# TABLE OF CONTENTS

**PREFACE** .......................................................................................................................................................... XV

Using the Blueprint .................................................................................................................................................. xvi

Changing the Blueprint ........................................................................................................................................... xvi

**SECTION 1. INTRODUCTION** .......................................................................................................................... 1

1 CMS Quality Measure Development .................................................................................................................. 3
   1.1 Meaningful Measures Initiative ............................................................................................................ 4
   1.2 Successes to Date .................................................................................................................................... 4
   1.3 Critical Challenges ............................................................................................................................... 5
   1.4 General Principles for Measure Development .................................................................................... 5
   1.5 Technical Principles for Measure Development ................................................................................ 6

2 The Measures Management System .................................................................................................................. 8
   2.1 Role of the Measure Developer ........................................................................................................... 8
   2.2 Role of the Contracting Officer’s Representative for the Measure Development Contract .................. 9
   2.3 Role of the Measures Manager .......................................................................................................... 9

**SECTION 2. THE MEASURE LIFECYCLE** ......................................................................................................... 11

1 Introduction to the Measure Lifecycle ............................................................................................................... 13

2 Measure Conceptualization .............................................................................................................................. 15
   2.1 Information Gathering ......................................................................................................................... 16
   2.2 Business Case Development .............................................................................................................. 17
   2.3 Stakeholder Engagement ..................................................................................................................... 20
      2.3.1 Technical Expert Panel .............................................................................................................. 20
      2.3.2 Person and Family Engagement ............................................................................................... 21
      2.3.3 Public Comment ....................................................................................................................... 21

3 Measure Specification ........................................................................................................................................ 22
   3.1 Technical Specification ....................................................................................................................... 23
   3.2 Harmonization ..................................................................................................................................... 24
   3.3 Stakeholder Engagement ..................................................................................................................... 26

4 Measure Testing .................................................................................................................................................. 27
   4.1 The Measure Testing Process ............................................................................................................. 27
   4.2 Stakeholder Engagement ..................................................................................................................... 29

5 Measure Implementation .................................................................................................................................... 30
   5.1 Pre-Rulemaking ................................................................................................................................. 30
   5.2 Other Submissions .............................................................................................................................. 30
   5.3 Stakeholder Engagement ..................................................................................................................... 31
   5.4 The Implementation Process ............................................................................................................... 31
6 Measure Use, Continuing Evaluation, and Maintenance ........................................................................... 33
   6.1 Continuing Evaluation ...................................................................................................................... 33
   6.2 Measure Maintenance ....................................................................................................................... 33
       6.2.1 Measure Production and Monitoring ...................................................................................... 35
       6.2.2 Measure Maintenance Reviews ............................................................................................ 35
       6.2.3 Stakeholder Engagement ...................................................................................................... 35
7 Tools And Resources for Measure Developers ............................................................................................. 36
   7.1 Bonnie ............................................................................................................................................. 36
   7.2 CMS Data Element Library (DEL) .................................................................................................. 36
   7.3 CMS Measures Inventory Tool ........................................................................................................ 36
   7.4 Cypress .......................................................................................................................................... 36
   7.5 Electronic Clinical Quality Improvement Resource Center (eCQI Resource Center) ....................... 36
   7.6 Environmental Scanning Support Tool (ESST) .............................................................................. 36
   7.7 Measure Authoring Tool (MAT) .................................................................................................... 37
   7.8 National Quality Forum Incubator (NQF Incubator) ......................................................................... 37
   7.9 National Quality Forum Quality Positioning System (QPS) .......................................................... 37
   7.10 Value Set Authority Center (VSAC) ............................................................................................ 37

SECTION 3. IN-DEPTH TOPICS ...................................................................................................................... 39
1 Healthcare Quality Priorities .................................................................................................................... 41
   1.1 CMS Goals and Priorities .............................................................................................................. 41
   1.2 Meaningful Measures Initiative .................................................................................................... 42
   1.3 Quality Across CMS ....................................................................................................................... 44
2 Measure Prioritization and Planning ......................................................................................................... 45
   2.1 Measure Prioritization ..................................................................................................................... 45
   2.2 CMS Measure Planning Inputs ..................................................................................................... 46
       2.2.1 Meaningful Measures Framework ......................................................................................... 46
       2.2.2 Legislative Mandates ............................................................................................................. 46
       2.2.3 Other Legislation Influencing CMS Quality Programs ......................................................... 48
       2.2.4 Quality Measure Development Plan (MDP) ......................................................................... 49
       2.2.5 Patients, Public, and Other Stakeholders ............................................................................. 50
       2.2.6 Impact Assessment and Other Reports ................................................................................. 50
   2.3 Role of the Measure Developer in Priorities Planning ............................................................ 51
   2.4 Role of the Measures Manager in Priorities Planning ................................................................. 52
3 Roles in Measure Development ............................................................................................................... 54
   3.1 Measure Developers ....................................................................................................................... 54
   3.2 Measure Stewards ............................................................................................................................ 54
   3.3 Guiding Committees ....................................................................................................................... 54
       3.3.1 Quality Measures Technical Forum ....................................................................................... 54
       3.3.2 eCQM Governance Group .................................................................................................. 55
4 Measure Classification ............................................................................................................................ 57
5 Selected Measure Types ....................................................................................................................... 66
  5.1 Cost and Resource Use Measures .................................................................................................. 66
  5.2 Composite Performance Measures .............................................................................................. 67
    5.2.1 Purpose of Composite Measures ......................................................................................... 67
    5.2.2 Component Performance Measures .................................................................................... 68
  5.3 Patient-Reported Outcome Measures .......................................................................................... 68
    5.3.1 Patient-Reported Outcomes ............................................................................................ 69
    5.3.2 Patient-Reported Outcome Measurement Tools .............................................................. 69
    5.3.3 Patient-Reported Outcome-based Performance Measures .............................................. 70
    5.3.4 Approaches to Developing Patient-Reported Outcome-based Performance Measures ........................................................................................................... 70
  5.4 Multiple Chronic Conditions Measures ....................................................................................... 71
    5.4.1 Multiple Chronic Condition Definition ............................................................................... 72
    5.4.2 Need for Measure Development ......................................................................................... 72
    5.4.3 Considerations for Measure Development Targeting Persons with Multiple Chronic Conditions .............................................................................................................................. 72
  5.5 Other Measure Types .................................................................................................................. 74

6 Introduction to Electronic Clinical Quality Measures ........................................................................ 75
  6.1 eCQM Components ....................................................................................................................... 76
  6.2 Collaborative Measure Development Workspace and Data Element Repository ....................... 77
  6.3 Special Considerations for eCQMs ............................................................................................... 78
  6.4 eCQM Annual Update and Change Review Process (CRP) .......................................................... 79
    6.4.1 Change Review Process ...................................................................................................... 79
    6.4.2 eCQM Specifications .................................................................................................. 79
    6.4.3 Supplemental Documents .......................................................................................... 80
    6.4.4 eCQM Standards and Tools ........................................................................................ 80

7 eCQM Standards-Based Guidance and Tools ............................................................................... 81
  7.1 Introduction ............................................................................................................................... 81
  7.2 Standards ................................................................................................................................. 81
    7.2.1 Health Quality Measure Format ......................................................................................... 81
    7.2.2 Quality Data Model ..................................................................................................... 82
    7.2.3 Clinical Quality Language ............................................................................................ 83
    7.2.4 Quality Reporting Document Architecture ................................................................ 84
  7.3 Tools ........................................................................................................................................ 86
    7.3.1 Development Tools .................................................................................................... 86
    7.3.2 Certification Tools .................................................................................................. 87

8 Master List of Deliverables and Document Requirements ................................................................ 89

9 Information Gathering ..................................................................................................................... 93
  9.1 Conduct an Environmental Scan .............................................................................................. 93
  9.2 Conduct an Empirical Data Analysis, as Appropriate .................................................................. 94
  9.3 Evaluate Information Collected during Environmental Scan and Empirical Data Analysis .......... 94
  9.4 Conduct a Measurement Gap Analysis to Identify Areas for New Measure Development .............. 97
  9.5 Determine the Appropriate Basis for Creation of New Measures ............................................ 97
  9.6 Apply Measure Evaluation Criteria ............................................................................................ 98
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.7</td>
<td>Submit the Information Gathering Report</td>
<td>98</td>
</tr>
<tr>
<td>9.7.1</td>
<td>Summary of Literature Review (Annotated Bibliography)</td>
<td>98</td>
</tr>
<tr>
<td>9.7.2</td>
<td>Summary of Clinical Practice Guidelines Review</td>
<td>99</td>
</tr>
<tr>
<td>9.7.3</td>
<td>Review of Existing Measures, Related Measures, and Gap Analysis Summary</td>
<td>99</td>
</tr>
<tr>
<td>9.7.4</td>
<td>Empirical Data Analysis Summary</td>
<td>99</td>
</tr>
<tr>
<td>9.7.5</td>
<td>Summary of Solicited and Structured Interviews (If Applicable)</td>
<td>100</td>
</tr>
<tr>
<td>9.8</td>
<td>Prepare an Initial List of Measures or Measure Topics</td>
<td>100</td>
</tr>
<tr>
<td>9.9</td>
<td>Tools and Resources for Information Gathering</td>
<td>100</td>
</tr>
<tr>
<td>9.9.1</td>
<td>CMS Data Element Library</td>
<td>100</td>
</tr>
<tr>
<td>9.9.2</td>
<td>CMS Measures Inventory Tool</td>
<td>100</td>
</tr>
<tr>
<td>9.9.3</td>
<td>Electronic Clinical Quality Improvement Resource Center (eCQI Resource Center)</td>
<td>100</td>
</tr>
<tr>
<td>9.9.4</td>
<td>Environmental Scanning Support Tool (ESST)</td>
<td>101</td>
</tr>
<tr>
<td>9.9.5</td>
<td>National Quality Forum Quality Positioning System</td>
<td>101</td>
</tr>
<tr>
<td>9.9.6</td>
<td>National Quality Forum Incubator</td>
<td>101</td>
</tr>
<tr>
<td>10</td>
<td>Environmental Scan</td>
<td>102</td>
</tr>
<tr>
<td>10.1</td>
<td>Literature Review</td>
<td>102</td>
</tr>
<tr>
<td>10.2</td>
<td>Quality of the Body of Evidence</td>
<td>103</td>
</tr>
<tr>
<td>10.3</td>
<td>Clinical Practice Guidelines</td>
<td>104</td>
</tr>
<tr>
<td>10.4</td>
<td>Existing and Related Measures</td>
<td>104</td>
</tr>
<tr>
<td>10.5</td>
<td>Stakeholder Input to Identify Measures and Important Measure Topics</td>
<td>105</td>
</tr>
<tr>
<td>10.6</td>
<td>Call for Measures</td>
<td>106</td>
</tr>
<tr>
<td>11</td>
<td>Business Case</td>
<td>107</td>
</tr>
<tr>
<td>11.1</td>
<td>MMS Business Case Best Practices</td>
<td>111</td>
</tr>
<tr>
<td>11.2</td>
<td>Business Case Template</td>
<td>111</td>
</tr>
<tr>
<td>12</td>
<td>Technical Expert Panel</td>
<td>112</td>
</tr>
<tr>
<td>12.1</td>
<td>Timing of TEP Input</td>
<td>112</td>
</tr>
<tr>
<td>12.2</td>
<td>TEP Structures</td>
<td>113</td>
</tr>
<tr>
<td>12.2.1</td>
<td>Traditional TEP Structure</td>
<td>113</td>
</tr>
<tr>
<td>12.2.2</td>
<td>Standing TEP Structure</td>
<td>114</td>
</tr>
<tr>
<td>12.3</td>
<td>Steps for the TEP</td>
<td>114</td>
</tr>
<tr>
<td>12.3.1</td>
<td>Draft the TEP Charter and Consider Potential TEP Members for Recruitment</td>
<td>115</td>
</tr>
<tr>
<td>12.3.2</td>
<td>Complete the Call for TEP Web Page Posting Form</td>
<td>115</td>
</tr>
<tr>
<td>12.3.3</td>
<td>Notify Relevant Stakeholder Organizations</td>
<td>116</td>
</tr>
<tr>
<td>12.3.4</td>
<td>Post the Call for Nominations Following COR Review</td>
<td>116</td>
</tr>
<tr>
<td>12.3.5</td>
<td>Select TEP and Notify the COR of the Membership List</td>
<td>117</td>
</tr>
<tr>
<td>12.3.6</td>
<td>Select Chair or Meeting Facilitator</td>
<td>118</td>
</tr>
<tr>
<td>12.3.7</td>
<td>Post the TEP Composition Documentation (Membership) List and Projected Meeting Dates</td>
<td>118</td>
</tr>
<tr>
<td>12.3.8</td>
<td>Arrange TEP Meetings</td>
<td>118</td>
</tr>
<tr>
<td>12.3.9</td>
<td>Send Materials to the TEP</td>
<td>118</td>
</tr>
<tr>
<td>12.3.10</td>
<td>Conduct the TEP Meetings</td>
<td>119</td>
</tr>
<tr>
<td>12.3.11</td>
<td>Prepare TEP Summary Report and Propose Recommended Set of Candidate</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>Measures</td>
<td></td>
</tr>
<tr>
<td>12.3.12</td>
<td>Post the TEP Summary Report</td>
<td>121</td>
</tr>
</tbody>
</table>
13 Person and Family Engagement .................................................................................................................. 122
  13.1 Background and Definition .................................................................................................................. 122
  13.2 Options for Engagement and Selected Best Practices .......................................................................... 122
  13.3 Engagement Activities: Virtual vs. In-Person .................................................................................. 126
  13.4 Recruitment .................................................................................................................................. 126
  13.5 Options for Engagement by Measure Lifecycle Stage and Selected Best Practices ...................... 127
    13.5.1 Measure Conceptualization ................................................................................................. 128
    13.5.2 Measure Specification ......................................................................................................... 128
    13.5.3 Measure Testing .................................................................................................................. 129
    13.5.4 Measure Implementation .................................................................................................... 130
    13.5.5 Measure Use, Continuing Evaluation, and Maintenance .................................................. 131
  13.6 Other Considerations ......................................................................................................................... 131

14 Public Comment .................................................................................................................................. 133
  14.1 Timing of Public Comment ............................................................................................................... 133
  14.2 Federal Rulemaking .......................................................................................................................... 133
  14.3 Steps for Public Comment ................................................................................................................ 134
    14.3.1 Prepare the Call for Public Comment .................................................................................. 134
    14.3.2 Notify Relevant Stakeholder Organizations ......................................................................... 135
    14.3.3 Post the Measures Following COR Approval ....................................................................... 135
    14.3.4 Collect Information .............................................................................................................. 136
    14.3.5 Summarize Comments and Produce Report ......................................................................... 136
    14.3.6 Send Comments to the TEP for Consideration .................................................................. 136
    14.3.7 Finalize the Public Comment Report, Including Verbatim Comments .......................... 136
    14.3.8 Arrange for the Final Public Comment Summary Report to be Posted on the Website .... 136

15 MMS Website Posting .......................................................................................................................... 138
  15.1 Posting Time Frame .......................................................................................................................... 139
  15.2 Posting Format .................................................................................................................................. 139
  15.3 Posting Template .............................................................................................................................. 139
  15.4 Documents to Include ...................................................................................................................... 139

16 Measure Technical Specification .......................................................................................................... 140
  16.1 Measure Specifications by Measure Category .................................................................................. 141
  16.2 Define the Data Source(s) ............................................................................................................... 141
    16.2.1 Administrative Data .............................................................................................................. 143
    16.2.2 Claims Data .......................................................................................................................... 143
    16.2.3 Patient Medical Records (Paper-based or Electronic) .......................................................... 143
    16.2.4 Electronic Clinical Data ....................................................................................................... 143
    16.2.5 Registries ............................................................................................................................. 143
    16.2.6 Standardized Patient Assessments ...................................................................................... 144
    16.2.7 Patient-Reported Data and Surveys ..................................................................................... 144
  16.3 Develop Specifications and Definitions ............................................................................................... 144
    16.3.1 Define the Initial Population ............................................................................................... 144
    16.3.2 Define the Denominator ...................................................................................................... 144
    16.3.3 Determine if a Denominator Exclusion is Needed ................................................................. 145
    16.3.4 Define the Numerator .......................................................................................................... 146
    16.3.5 Determine if a Denominator Exception is Needed ................................................................. 146
Table of Contents

16.3.6 Determine if a Numerator Exclusion is Needed................................................................. 147
16.3.7 Define Stratification Scheme................................................................................................ 148
16.3.8 Use Positive Evidence........................................................................................................ 149

16.4 Specify the Code and/or Code Systems .................................................................................. 149

16.5 Construct Data Protocol.......................................................................................................... 149
16.5.1 Define Key Terms, Data Elements, Codes, and Code Systems ........................................ 149
16.5.2 Describe the Level of Measurement/Analysis..................................................................... 150
16.5.3 Describe Sampling................................................................................................................ 151
16.5.4 Determine Risk Adjustment................................................................................................ 151
16.5.5 Clearly Define Any Time Intervals...................................................................................... 151
16.5.6 Describe How the Measure Results are Scored and Reported ........................................ 152
16.5.7 Develop the Calculation Algorithm..................................................................................... 152

16.6 Document the Measures and Obtain COR Approval............................................................. 152
16.6.1 Finalize Measure Name and Description.............................................................................. 152
16.6.2 Making Changes.................................................................................................................. 154
16.6.3 Submission Forms............................................................................................................. 154

17 Codes, Code Systems, and Datasets......................................................................................... 155
17.1 Code Systems/Vocabularies/Terminologies............................................................................ 155
17.1.1 Encoding Clinical Information............................................................................................ 155
17.1.2 Health Information Technology Standards Committee Recommendations ..................... 156
17.1.3 Quality Data Model Categories with Recommended Code Systems .................................. 157
17.1.4 Use of Specific Code Systems............................................................................................. 165

17.2 Value Sets................................................................................................................................ 171
17.2.1 Use of Value Sets................................................................................................................ 171
17.2.2 Constructing Value Sets....................................................................................................... 171
17.2.3 Representing the Codes to be Included............................................................................. 171
17.2.4 Determining the Value Set Code System........................................................................... 172
17.2.5 Value Set Mapping............................................................................................................. 174

17.3 Value Set Reuse .................................................................................................................... 175

17.4 Value Set Authority Center..................................................................................................... 175

17.5 VSAC Collaboration Tool......................................................................................................... 176

17.6 Unique Situations for Value Set Use in eCQMS...................................................................... 176
17.6.1 Use of Value Sets to Address Negation Rationale in eCQMs.............................................. 176
17.6.2 Value Set Review................................................................................................................ 177
17.6.3 Pre-MAT Value Set Review................................................................................................. 178
17.6.4 Post-MAT Value Set Review – eCQM Annual Update Only.............................................. 179
17.6.5 NLM Value Set Review – eCQM Annual Update Only...................................................... 179

18 Measure Harmonization........................................................................................................... 180
18.1 Respecified Measures............................................................................................................. 181
18.2 Adopted Measures.................................................................................................................. 182
18.3 New Measures....................................................................................................................... 183
18.4 Harmonization during Measure Maintenance......................................................................... 183
18.4.1 Decide Whether Harmonization is Indicated.................................................................... 183
18.4.2 Implement Harmonization Decisions................................................................................. 184
18.4.3 Test Scientific Acceptability of Measure Properties.......................................................... 184
18.4.4 NQF Evaluates for Harmonization during Measure Maintenance..................................... 184
## Table of Contents

19 Risk Adjustment .................................................................................................................................. 185
  19.1 Risk Adjustment Strategies – When to use Risk Adjustment with Stratification .................. 188
  19.2 Features of Risk Adjustment Models ....................................................................................... 189
    19.2.1 Defining the Appropriate Measure Development Sample ........................................... 189
    19.2.2 Appropriate Time Intervals ............................................................................................... 189
    19.2.3 High Data Quality ............................................................................................................ 190
    19.2.4 Appropriate Variable Selection ......................................................................................... 190
    19.2.5 Appropriate Analytic Approach ......................................................................................... 191
  19.3 Risk Adjustment Procedure ...................................................................................................... 192
    19.3.1 Choose and Define an Outcome ....................................................................................... 192
    19.3.2 Define the Conceptual Model ........................................................................................... 192
    19.3.3 Identify the Risk Factors and Timing ............................................................................... 193
    19.3.4 Acquire Data (Sample, if Necessary) ............................................................................... 193
    19.3.5 Analyze the Data .............................................................................................................. 193
    19.3.6 Assess the Model ............................................................................................................. 196
    19.3.7 Document the Model ....................................................................................................... 198
  19.4 Measure Reporting ....................................................................................................................... 201
    19.4.1 Attributing Resource Use Measures ............................................................................... 201
    19.4.2 Peer Group Identification and Assignment ..................................................................... 201
    19.4.3 Calculating Comparisons ............................................................................................... 202
    19.4.4 Setting Thresholds ........................................................................................................... 202
    19.4.5 Providing Detailed Feedback ......................................................................................... 202
    19.4.6 Reporting with Descriptive Statistics .......................................................................... 202

20 Cost and Resource Use Measure Specification ............................................................................... 200
  20.1 Measure Clinical Logic .............................................................................................................. 200
  20.2 Measure Construction Logic ..................................................................................................... 200
    20.2.1 Time Intervals .................................................................................................................. 200
    20.2.2 Assigning and Triaging Claims ....................................................................................... 200
  20.3 Adjusting for Comparability ..................................................................................................... 201
    20.3.1 Define Risk Adjustment Approach ................................................................................... 201
    20.3.2 Define Stratification Approach ........................................................................................ 201
    20.3.3 Define Costing Methodology .......................................................................................... 201
  20.4 Measure Reporting ....................................................................................................................... 201
    20.4.1 Attributing Resource Use Measures ............................................................................... 201
    20.4.2 Peer Group Identification and Assignment ..................................................................... 201
    20.4.3 Calculating Comparisons ............................................................................................... 202
    20.4.4 Setting Thresholds ........................................................................................................... 202
    20.4.5 Providing Detailed Feedback ......................................................................................... 202
    20.4.6 Reporting with Descriptive Statistics .......................................................................... 202

21 Composite Measure Technical Specifications ............................................................................... 203

22 Measure Testing ............................................................................................................................... 207
  22.1 Alpha and Beta Testing .............................................................................................................. 207
    22.1.1 Alpha Testing ................................................................................................................... 207
    22.1.2 Beta Testing ..................................................................................................................... 207
    22.1.3 Comparing Alpha and Beta Testing ............................................................................... 207
  22.2 The Measure Testing Process .................................................................................................... 209
    22.2.1 Develop the Testing Work Plan ....................................................................................... 209
    22.2.2 Perform Sampling ............................................................................................................. 210
    22.2.3 Submit the Plan and Obtain COR Approval ..................................................................... 211
    22.2.4 Implement the Plan ......................................................................................................... 211
    22.2.5 Analyze the Test Results ................................................................................................. 211
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>26.3 Measure Developer’s Role during NQF Evaluation</td>
<td>238</td>
</tr>
<tr>
<td>26.4 Trial Use Approved Measures</td>
<td>239</td>
</tr>
<tr>
<td>26.5 Measure Maintenance for NQF</td>
<td>239</td>
</tr>
<tr>
<td>27 Measure Selection</td>
<td>241</td>
</tr>
<tr>
<td>27.1 Pre-Rulemaking Process</td>
<td>241</td>
</tr>
<tr>
<td>27.1.1 Measures Under Consideration</td>
<td>241</td>
</tr>
<tr>
<td>27.1.2 CMS Measure Selection Criteria</td>
<td>242</td>
</tr>
<tr>
<td>27.1.3 MAP Recommendations</td>
<td>242</td>
</tr>
<tr>
<td>27.1.4 CMS Considers MAP Input for Final Selection</td>
<td>242</td>
</tr>
<tr>
<td>27.2 CMS Rulemaking Processes</td>
<td>242</td>
</tr>
<tr>
<td>27.3 Rollout, Production, and Monitoring of Measures</td>
<td>243</td>
</tr>
<tr>
<td>27.4 Measure Maintenance</td>
<td>243</td>
</tr>
<tr>
<td>27.5 Impact Assessment of Medicare Quality Measures</td>
<td>245</td>
</tr>
<tr>
<td>28 Measure Rollout</td>
<td>246</td>
</tr>
<tr>
<td>28.1 Measures are Selected by CMS</td>
<td>246</td>
</tr>
<tr>
<td>28.2 Implement the Rollout Plan</td>
<td>248</td>
</tr>
<tr>
<td>28.3 Implement the Data Management Process</td>
<td>248</td>
</tr>
<tr>
<td>28.4 Develop the Auditing and Validation Plan</td>
<td>248</td>
</tr>
<tr>
<td>28.5 Develop an Appeals Process</td>
<td>249</td>
</tr>
<tr>
<td>28.6 Implement Education Processes</td>
<td>249</td>
</tr>
<tr>
<td>28.7 Conduct the Dry Run</td>
<td>250</td>
</tr>
<tr>
<td>28.8 Submit Reports</td>
<td>250</td>
</tr>
<tr>
<td>29 Measure Production and Monitoring</td>
<td>251</td>
</tr>
<tr>
<td>29.1 Conduct Data Collection and Ongoing Surveillance</td>
<td>252</td>
</tr>
<tr>
<td>29.2 Respond to Questions about the Measure</td>
<td>253</td>
</tr>
<tr>
<td>29.3 Produce Preliminary Reports</td>
<td>253</td>
</tr>
<tr>
<td>29.4 Report Measure Results</td>
<td>253</td>
</tr>
<tr>
<td>29.5 Monitor and Analyze the Measure Rates and Audit Findings</td>
<td>254</td>
</tr>
<tr>
<td>29.6 Perform Measure Maintenance or Ad Hoc Review, When Appropriate</td>
<td>254</td>
</tr>
<tr>
<td>29.7 Provide Information that CMS can Use in Measure Priorities Planning</td>
<td>255</td>
</tr>
<tr>
<td>30 Measure Maintenance Reviews</td>
<td>256</td>
</tr>
<tr>
<td>30.1 Annual Update</td>
<td>256</td>
</tr>
<tr>
<td>30.1.1 Potential for Harmonization</td>
<td>257</td>
</tr>
<tr>
<td>30.1.2 Procedure</td>
<td>257</td>
</tr>
<tr>
<td>30.2 Comprehensive Reevaluation</td>
<td>260</td>
</tr>
<tr>
<td>30.2.1 Measures in CMS Programs for Which CMS is Not the Steward or Providing Maintenance Through a CMS Measure Developer/Measure Maintenance Contractor</td>
<td>261</td>
</tr>
<tr>
<td>30.2.2 Harmonization</td>
<td>261</td>
</tr>
<tr>
<td>30.2.3 Procedure</td>
<td>261</td>
</tr>
<tr>
<td>30.3 CMS Ad Hoc Review</td>
<td>265</td>
</tr>
<tr>
<td>30.3.1 Trigger for an Ad Hoc Review</td>
<td>266</td>
</tr>
<tr>
<td>30.3.2 Deferring an Ad Hoc Review</td>
<td>266</td>
</tr>
<tr>
<td>30.3.3 Procedure</td>
<td>267</td>
</tr>
<tr>
<td>30.4 NQF Ad Hoc Reviews</td>
<td>269</td>
</tr>
<tr>
<td>30.5 Possible Outcomes of Maintenance Reviews</td>
<td>270</td>
</tr>
</tbody>
</table>
# Table of Contents

**SECTION 4. FORMS AND TEMPLATES**

1. Environmental Scan Outline ................................................................. 273
2. Business Case Form Instructions ............................................................. 276
3. Business Case Template ............................................................................ 280
4. Call for Measures Web Posting ................................................................. 284
5. Measure Information Form Instructions .................................................... 285
6. Blank Measure Information Form Template .............................................. 291
7. Measure Justification Form Instructions .................................................... 293
8. Blank Measure Justification Form Template .............................................. 311
9. Measure Evaluation Criteria and Instructions ............................................ 317
11. CMS Measures Management System Web Page Posting Instructions ....... 336
12. Public Comment Call Web Posting ........................................................... 337
13. Public Comment Summary Web Posting .................................................. 338
15. Technical Expert Panel Call for TEP Web Posting .................................... 342
16. Technical Expert Panel Nomination Form Template ................................ 344
17. Technical Expert Panel Composition (Membership) List Template .......... 348
18. Technical Expert Panel Composition (Membership) List Web Posting .... 349

**SECTION 5. GLOSSARY AND ACRONYMS** ................................................. 355

Glossary ......................................................................................................... 357
Acronyms ...................................................................................................... 375

**SECTION 6. APPENDICES** ........................................................................... 383

Appendix A: Summary of Changes to Blueprint ........................................... 385
Appendix B: ONC Project Tracking System (Jira) .......................................... 387
Appendix C: eCQM Metadata ...................................................................... 389
Appendix D: eCQM Logic Quality Assurance Checklist .................................. 397
Appendix E: Registries .................................................................................. 399
Appendix F: Measure Calculations ............................................................... 403
Appendix G: References ............................................................................... 413
<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Timeline of Measure Lifecycle</td>
<td>13</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Interaction Among Measure Lifecycle Phases</td>
<td>14</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Flow of the Measure Lifecycle—Measure Conceptualization</td>
<td>15</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Information Gathering Deliverables and Requirements</td>
<td>16</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Business Case Deliverables</td>
<td>19</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Technical Expert Panel Deliverables and Requirements</td>
<td>20</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Public Comment Deliverables and Requirements</td>
<td>21</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Flow of the Measure Lifecycle—Measure Specification</td>
<td>22</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Measure Specification Deliverables and Requirements</td>
<td>23</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Flow of the Measure Lifecycle—Measure Testing</td>
<td>27</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Measure Testing Deliverables and Requirements</td>
<td>28</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Overview of Measure Testing</td>
<td>29</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Flow of the Measure Lifecycle—Measure Implementation</td>
<td>31</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Measure Implementation Deliverables and Requirements</td>
<td>32</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Measure Maintenance Deliverables and Requirements</td>
<td>33</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Flow of the Measure Lifecycle—Measure Use, Continuing Evaluation, and Maintenance</td>
<td>34</td>
</tr>
<tr>
<td>Figure 17</td>
<td>CMS Strategic Priorities</td>
<td>42</td>
</tr>
<tr>
<td>Figure 18</td>
<td>Meaningful Measures Framework</td>
<td>43</td>
</tr>
<tr>
<td>Figure 19</td>
<td>CMS Priorities Planning and Measure Selection</td>
<td>46</td>
</tr>
<tr>
<td>Figure 20</td>
<td>Relationship between Resource Use, Value, and Efficiency</td>
<td>66</td>
</tr>
<tr>
<td>Figure 21</td>
<td>Relationship between PROs, PROMs, and PRO-PMs</td>
<td>69</td>
</tr>
<tr>
<td>Figure 22</td>
<td>Changes in eCQM Information Structure</td>
<td>81</td>
</tr>
<tr>
<td>Figure 23</td>
<td>QDM Structure</td>
<td>83</td>
</tr>
<tr>
<td>Figure 24</td>
<td>Connections between Standards</td>
<td>85</td>
</tr>
<tr>
<td>Figure 25</td>
<td>Connections between Tools and Standards</td>
<td>86</td>
</tr>
<tr>
<td>Figure 26</td>
<td>Environmental Scan Data Sources</td>
<td>94</td>
</tr>
<tr>
<td>Figure 27</td>
<td>Potential Costs, Benefits, and Savings</td>
<td>107</td>
</tr>
<tr>
<td>Figure 28</td>
<td>Inputs and Uses for the Business Case</td>
<td>109</td>
</tr>
<tr>
<td>Figure 29</td>
<td>Flow of Inputs into Business Case Development</td>
<td>110</td>
</tr>
<tr>
<td>Figure 30</td>
<td>Person and Family Engagement Concepts</td>
<td>122</td>
</tr>
<tr>
<td>Figure 31</td>
<td>TEP and Working Group Best Practices</td>
<td>125</td>
</tr>
<tr>
<td>Figure 32</td>
<td>Recruitment Practices</td>
<td>127</td>
</tr>
<tr>
<td>Figure 33</td>
<td>Measure Conceptualization Best Practice</td>
<td>128</td>
</tr>
<tr>
<td>Figure 34</td>
<td>Measure Conceptualization and Specification Best Practice</td>
<td>129</td>
</tr>
<tr>
<td>Figure 35</td>
<td>The Web Posting Process</td>
<td>138</td>
</tr>
<tr>
<td>Figure 36</td>
<td>VSAC Value Set Life Cycle</td>
<td>173</td>
</tr>
<tr>
<td>Figure 37</td>
<td>Risk Adjustment Requirements</td>
<td>192</td>
</tr>
<tr>
<td>Figure 38</td>
<td>Example of ROC Curves</td>
<td>197</td>
</tr>
<tr>
<td>Figure 39</td>
<td>Applying Measure Evaluation Criteria</td>
<td>215</td>
</tr>
<tr>
<td>Figure 40</td>
<td>Measure Review Cycle Timeline</td>
<td>240</td>
</tr>
<tr>
<td>Figure 41</td>
<td>NQF Consensus Development Process</td>
<td>240</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>42</td>
<td>Sample Pre-rulemaking Process Timeline</td>
<td>241</td>
</tr>
<tr>
<td>43</td>
<td>CMS Criteria for Measure Disposition: Revise or Retain</td>
<td>244</td>
</tr>
<tr>
<td>44</td>
<td>CMS Criteria for Measure Disposition: Retire</td>
<td>244</td>
</tr>
<tr>
<td>45</td>
<td>CMS Criteria for Measure Disposition: Remove</td>
<td>245</td>
</tr>
<tr>
<td>46</td>
<td>Overview of the Measure Rollout Process</td>
<td>246</td>
</tr>
<tr>
<td>47</td>
<td>Production and Monitoring Requirements</td>
<td>251</td>
</tr>
<tr>
<td>48</td>
<td>Overview of the Measure Monitoring Process</td>
<td>252</td>
</tr>
<tr>
<td>49</td>
<td>Annual Update Requirements</td>
<td>256</td>
</tr>
<tr>
<td>50</td>
<td>Comprehensive Reevaluation Deliverable</td>
<td>261</td>
</tr>
<tr>
<td>51</td>
<td>Ad Hoc Review Requirements</td>
<td>266</td>
</tr>
<tr>
<td>52</td>
<td>QCDR Measure Development, Review, and Posting Process</td>
<td>400</td>
</tr>
<tr>
<td>53</td>
<td>Proportion Measure Populations</td>
<td>403</td>
</tr>
<tr>
<td>54</td>
<td>CV Measure Populations</td>
<td>407</td>
</tr>
<tr>
<td>55</td>
<td>Ratio Measure Populations</td>
<td>410</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Framework for Progression of Payment to Clinicians and Organizations in Payment Reform (adapted from Rajkumar et al., 2014)</td>
<td>3</td>
</tr>
<tr>
<td>Table 2</td>
<td>Legislation Initiating CMS Quality Programs for Acute Care</td>
<td>47</td>
</tr>
<tr>
<td>Table 3</td>
<td>Legislation Initiating CMS Quality Programs for Ambulatory Care</td>
<td>47</td>
</tr>
<tr>
<td>Table 4</td>
<td>Legislation Initiating CMS Quality Programs for Post-acute Care</td>
<td>48</td>
</tr>
<tr>
<td>Table 5</td>
<td>Legislation Initiating CMS Quality Programs Across Settings</td>
<td>48</td>
</tr>
<tr>
<td>Table 6</td>
<td>Crosswalk of Measure Types</td>
<td>57</td>
</tr>
<tr>
<td>Table 7</td>
<td>CMIT</td>
<td>62</td>
</tr>
<tr>
<td>Table 8</td>
<td>Examples of Measures Addressing Each of the Meaningful Measurement Areas</td>
<td>63</td>
</tr>
<tr>
<td>Table 9</td>
<td>Measure Types for Pre-Rulemaking</td>
<td>64</td>
</tr>
<tr>
<td>Table 10</td>
<td>NQF’s Measure Categories and Types as Listed in the 2018 Evaluation Criteria and Guidance</td>
<td>64</td>
</tr>
<tr>
<td>Table 11</td>
<td>The Business Case is an Ongoing Process that Occurs throughout the Measure Lifecycle</td>
<td>108</td>
</tr>
<tr>
<td>Table 12</td>
<td>Best Practices for Implementing Person/Family Engagement Activities, by Phase of Engagement</td>
<td>123</td>
</tr>
<tr>
<td>Table 13</td>
<td>Measure Specifications by Measure Category</td>
<td>141</td>
</tr>
<tr>
<td>Table 14</td>
<td>Pros and Cons of Different Data Sources</td>
<td>142</td>
</tr>
<tr>
<td>Table 15</td>
<td>ONC HITSC Recommended Vocabulary Summary</td>
<td>156</td>
</tr>
<tr>
<td>Table 16</td>
<td>ONC HITSC Transition Vocabulary Standards Summary and Plan</td>
<td>157</td>
</tr>
<tr>
<td>Table 17</td>
<td>QDM Categories with ONC HITSC Recommended Vocabularies</td>
<td>158</td>
</tr>
<tr>
<td>Table 18</td>
<td>Value Set Review Areas and Remedial Actions</td>
<td>178</td>
</tr>
<tr>
<td>Table 19</td>
<td>Harmonization Decisions during Measure Development</td>
<td>181</td>
</tr>
<tr>
<td>Table 20</td>
<td>Harmonization Decisions during Measure Maintenance</td>
<td>184</td>
</tr>
<tr>
<td>Table 21</td>
<td>Framework for Risk Adjustment Strategies</td>
<td>188</td>
</tr>
<tr>
<td>Table 22</td>
<td>Features of Risk Adjustment Models</td>
<td>189</td>
</tr>
<tr>
<td>Table 23</td>
<td>Types of Composite Measure Scoring</td>
<td>204</td>
</tr>
<tr>
<td>Table 24</td>
<td>Features of Alpha and Beta Testing</td>
<td>208</td>
</tr>
<tr>
<td>Table 25</td>
<td>Extent of Measure Evaluation as a Function of Prior Comprehensive Evaluation and Measure Use</td>
<td>264</td>
</tr>
</tbody>
</table>
Preface
USING THE BLUEPRINT

This Blueprint is composed of six sections.

Section 1, Introduction, presents an overview of the guiding principles of Centers for Medicare & Medicaid Services (CMS) measure development, background information on the Measures Management System (MMS), and administrative details about interfacing with the Measures Manager.

Section 2, The Measure Lifecycle, covers the basics of the measure development process as defined by the Measure Lifecycle. Five of its chapters correspond to the five phases of the lifecycle: (1) Measure Conceptualization; (2) Measure Specification; (3) Measure Testing; (4) Measure Implementation; and (5) Measure Use, Continuing Evaluation, and Maintenance. Each of these chapters touches on the fundamental steps that measure developers undertake in each phase, as well as the CMS contract deliverables that they develop in the process. Each chapter then refers the reader to Section 3, In-Depth Topics, for more detailed information on specific measure development topics.

Section 3, In-Depth Topics, contains a suite of standalone, detailed articles on each aspect of measure development. These article topics range from CMS priorities planning to the details of risk adjustment. Although these articles chronologically follow the Measure Lifecycle, they are not meant to be read as a single entity from beginning to end. Rather, they are individual reference articles with a high degree of granularity for a more detailed understanding of each aspect of the measure development process.

Section 4, Forms and Templates, contains all the forms and templates required for completion of the measure development process and the delivery of CMS contract deliverables.

Section 5 contains the Glossary and Section 6 contains the Appendices.

The Blueprint does not address measures at the portfolio/measure set level. It addresses single measures or a small set of closely related measures. The deliverables and requirements identified in the Blueprint are from the CMS Measure and Instrument Development and Support (MIDS) Umbrella Statement of Work (USOW) and may not be included in all MIDS contracts or all measure development contracts.

Throughout the document, a modified American Psychological Association (APA) format has been adopted. APA format inline citations are used for journal articles and books with hyperlinks to the article or book, if available. For grey literature, document names with hyperlinks are used. APA format is used for punctuation, abbreviations, bullets, and the reference list.

Note: eCQM-specific information is identified throughout this Blueprint by a computer icon.

CHANGING THE BLUEPRINT

From Version 1 through the present, this Blueprint has been updated to incorporate changes in the regulatory environment and in healthcare quality measurement science and to meet the evolving needs of measure developers. Each year, input has been systematically gathered, formally tracked, and considered for implementation in subsequent Blueprint updates. For a high-level list of changes in this latest version, refer to Appendix A.

Recommendations for changes to the content, structure, or organization of the Blueprint are welcome. Submit all suggestions to the MMS support mailbox. Include specifics about the recommended change, including:
Recommended changes will be considered year-round and incorporated into the next review cycle of the document.
Section 1. Introduction
1 CMS Quality Measure Development

A transformation is under way in the United States (U.S.) healthcare system. It is a transformation fundamentally driven by performance measurement. In nearly every setting of care, CMS is moving from paying for services to paying for value. The CMS goal is to foster value by promoting the highest quality, safety, and care experience with the most affordable, cost-efficient service possible for Americans. Table 1 highlights four payment categories that represent the progression of payment reform for clinicians and facilities for their services. Initiated with the passage and implementation of the Patient Protection and Affordable Care Act (ACA) and more recently driven by the Medicare Access and Children’s Health Insurance Program (CHIP) Reauthorization Act (MACRA), CMS is well on its way to transitioning from a fee for service (FFS) system to a payment system based on quality and value. In the near term, few payments in the Medicare program will continue to be based on Category 1 and there will be a rapid transition to the majority of payments falling under Categories 3 and 4.

Table 1. Framework for Progression of Payment to Clinicians and Organizations in Payment Reform (adapted from Rajkumar et al., 2014)

<table>
<thead>
<tr>
<th>Category 1: FFS—No Link to Quality</th>
<th>Category 2: FFS—Link to Quality</th>
<th>Category 3: Alternative Payment Models (APMs) on FFS Architecture</th>
<th>Category 4: Population-Based Payment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td></td>
<td>Description</td>
<td>Description</td>
</tr>
<tr>
<td>Payments are based on volume of services and not linked to quality or efficiency</td>
<td>At least a portion of payments vary based on the quality or efficiency of healthcare delivery</td>
<td>Some payment is linked to the effective management of a population or an episode of care</td>
<td>Payment is not directly triggered by service delivery, so volume is not linked to payment</td>
</tr>
<tr>
<td>Medicare</td>
<td></td>
<td>Hospital value-based purchasing</td>
<td>Clinicians and organizations are paid and responsible for the care of a beneficiary for a long period (e.g., &gt; one year)</td>
</tr>
<tr>
<td>• Limited in Medicare FFS</td>
<td>• Hospital readmissions/hospital-acquired condition reduction programs</td>
<td>Accountable care organizations</td>
<td>Medicare Advantage plan payments to clinicians and organizations</td>
</tr>
<tr>
<td>• Majority of Medicare payments now are linked to quality</td>
<td></td>
<td>Medical homes</td>
<td></td>
</tr>
<tr>
<td>Medicaid</td>
<td></td>
<td>Bundled payments</td>
<td>Medicaid managed care plan payments to clinicians and organizations</td>
</tr>
<tr>
<td>Varies by state</td>
<td></td>
<td></td>
<td>Medicaid–Medicaid (duals) plan payments to clinicians and organizations</td>
</tr>
<tr>
<td>• Primary care case management</td>
<td></td>
<td></td>
<td>Some Medicare–Medicaid beneficiaries</td>
</tr>
<tr>
<td>• Some managed care models</td>
<td></td>
<td></td>
<td>Medicaid health homes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Medicaid shared savings models</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Medicaid waivers for delivery reform incentive payments</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Episodic-based payments</td>
</tr>
</tbody>
</table>
The stakes are higher than ever for patients and providers, as measurement is used in payment programs, and in transparent public reporting, which may impact not only patient choice, but also accountability for healthcare providers. As such, clinical quality measures (CQMs) must be meaningful, robust, valid, feasible, based in scientific evidence, and well tested to ensure that the measures do not lead to unintended negative consequences or burden for patients or providers. In striving to achieve the goals of the Meaningful Measures Initiative, developed measures must be meaningful to patients and the providers who serve them, represent opportunities for improvement in care quality and hence, actionable, and differentiate quality in a meaningful and valid way. To accomplish these goals, the strategies for measure development outlined in this document must be kept at the forefront.

1.1 **Meaningful Measures Initiative**

CMS’s comprehensive initiative “Meaningful Measures” was launched in 2017 and identifies high priority areas for quality measurement and improvement. Its purpose is to improve outcomes for patients, their families, and providers while also reducing burden and moving payment toward value through focusing everyone’s efforts on the same quality areas. The Meaningful Measures Initiative also helps to identify and close important gap areas of measures, align measures across the continuum of care and across payers, and spur innovation in new types of measures such as patient-reported measures and electronic measures. High-value, Meaningful Measures meet these principles:

- Address high-impact measure areas that safeguard public health
- Are patient-centered and meaningful to patients
- Are outcome-based, where possible
- Fulfill requirements in programs’ statutes
- Minimize level of burden
- Provide significant opportunity for improvement
- Address measure needs for population-based payment through alternative payment models (APMs)
- Align across programs and/or with other payers (Medicaid, commercial payers).

For in-depth information on healthcare quality strategies, refer to Section 3, Chapter 1.2, Meaningful Measures Initiative.

1.2 **Successes to Date**

For the first time in many years, a number of critically important metrics at the national level have significantly improved, such as hospital readmission rates, central line-associated blood stream infections (CLABSI), surgical site infections, early elective deliveries, and ventilator-associated pneumonia. There has also been a sustained decrease in total Medicare per capita costs. In the Medicare Advantage programs, plans are rated by stars to reflect the quality of the services they offer, and beneficiaries are increasingly choosing plans that have higher star ratings. These improvements are real and measurable and are increasing the length and quality of beneficiaries’ lives.

Many measures that CMS has developed are National Quality Forum (NQF)-endorsed and/or recommended by the NQF-convened Measure Applications Partnership (MAP). However, as performance on quality metrics is increasingly tied to provider payment, there is an increasing need for measure development to be a more flexible and efficient process with a shorter development time.

---

1 Throughout the document, “robust” refers to measures with the most vigorous quality action or guidance or as a descriptor to describe strong, vigorous, or thoroughly vetted components of a measure.
frame. CMS continually seeks innovations and process improvement to meet these challenges. CMS has also started to remove measures from programs that have minimal room for improvement, are no longer supported by evidence, are duplicative of other measures, or are of low value from the patient or clinical workflow perspective.

CMS is rebalancing the portfolio of measures to contain more outcome measures and fewer process measures, with the goal of better addressing performance gaps in the Meaningful Measures Initiative. As part of this effort, the Measures Manager maintains the CMS Measures Inventory. The Inventory is inclusive of measures throughout the measure development lifecycle for dozens of CMS programs and initiatives including measures under development (MUD), measures under consideration (MUC), measures that have entered the rulemaking process, measures actively in a program, and measures that have been removed from programs. The Inventory includes measure title, description, NQF endorsement status, measure type, and NQS domain with the goal of providing users with a complete picture of how the measure quantifies performance quality within the various CMS programs.

1.3 **Critical Challenges**

The challenges to developing measures that are meaningful and appropriate for payment programs are described in detail in the CMS Quality Measurement Development Plan (MDP) and cannot all be enumerated here. However, some of the key opportunities include:

- Partnering with patients in the measure development process
- Partnering with frontline clinicians and professional societies
- Aligning measures across programs, payers, and payment systems
- Reducing clinician burden of data collection for measure reporting
- Shortening the time frame for measure development
- Streamlining data acquisition for measure testing
- Identifying and developing meaningful outcome measures
- Developing patient-reported outcome measurement (PROM) tools and appropriate use measures
- Developing measures that promote shared accountability across settings and providers.

CMS is committed to addressing these challenges head-on using process improvement techniques, such as Lean, in all phases of measure development. CMS wants developers to identify ways to most meaningfully engage patients in the measure development process and to share best practices with CMS and its contractors.

1.4 **General Principles for Measure Development**

These principles are to be used throughout the measure development process, especially when identifying concepts for new measures. These principles serve as overarching guidelines for measure development that meet the standards and rigor expected of a meaningful, valid, and useful measure. Measures should be developed to:

- Focus on what is best for patients and most meaningful to patients, caregivers, and providers.
- Explicitly align with Meaningful Measures and its goals and objectives.
- Align across payers, including Medicare, Medicaid, the Exchanges, other federal partners, and private payers.
- Address a performance gap where there is known variation in performance, not just a measure gap.
• Be developed in a rapid-cycle fashion, in accordance with process improvement techniques, such as Lean and human-centered design.
• Encourage collaboration among measure developers and share best practices/new learnings freely.
• Reorient and align around patient-centered outcomes that span across clinical settings, which may require different “versions” of the same measure (i.e., different cohorts, but same numerator); it is important to test each of these setting-specific versions for reliability and validity.
• Value-based care that produces quality outcomes.
• Be focused on outcomes (including patient-reported outcomes [PROs]), safety, patient experience, care coordination, appropriate use/efficiency, and cost.
• Identify and eliminate disparities in the delivery of care.
• Guard against unintended consequences of measure implementation, including overuse and underuse of care.
• Engage stakeholders early and often in the measure development process.
• Strive to reduce clinician burden in reporting measures.  

1.5 TECHNICAL PRINCIPLES FOR MEASURE DEVELOPMENT

These principles should be applied when developing measures for consideration for quality reporting and value-based purchasing programs:

• Develop a rigorous business case for an evidence-based measure concept—a critical first step in the development process.
• Prioritize electronic clinical data sources (e.g., electronic health records [EHRs] and registries), where appropriate, and reduce dependency on data from claims and chart abstraction whenever possible.
• Maintain a focus on iterative testing using both real and synthetic data.
• Consider approaches to aggregate multiple data sources (e.g., hybrid measures) to achieve the most accurate assessment of quality until universal interoperability can be achieved.
• Define outcomes, risk factors, cohorts, and inclusion/exclusion criteria based on clinical and empirical evidence.
• Judiciously select exclusions to capture as broad a patient population as possible and appropriate; consider developing a paired measure to capture and measure the care received for the excluded patients if a significant number of patients are excluded (e.g., for all patients seen in the emergency department, if those patients who are transferred directly to another acute care facility for tertiary treatment are excluded, a paired measure would address those patients who were transferred out of the original facility).
• Develop risk adjustment models to distinguish performance between providers rather than predict patient outcomes.
• Include measure stratification and risk adjustment approaches to patient demographic characteristics that promote equitable quality comparisons.
• Harmonize measure methodologies, data elements, and specifications, when applicable and feasible.
• Develop each measure with sufficient statistical power to detect and report statistically significant differences in provider performance.

2 Adapted from the Measure Development Plan (MDP).
• Consider strategies to enable clinicians that have smaller practices and low-volume facilities to reliably report a measure.
• Strive to develop measures that can progress to multi-payer applicability using all-payer databases where available.³
• Consider the clinical workflow needed in the electronic record for electronic clinical quality measures (eCQMs).

³ Adapted from the MDP.
2 THE MEASURES MANAGEMENT SYSTEM

The MMS is a standardized system for developing and maintaining the quality measures used in CMS’s various quality initiatives and programs. The primary goal of the MMS is to provide guidance to measure developers to help them produce high-caliber healthcare quality measures. CMS-funded measure developers (or contractors) should follow this Blueprint, which documents the core set of business processes and decisions criteria when developing, implementing, and maintaining measures.

Measure developers who do not currently hold CMS contracts are encouraged to use the Blueprint as a guide in their measure development process, especially if they have a future interest in working within CMS programs. The Blueprint process produces high-caliber measures that stand up to review for reliability, validity, and importance.

Within the MMS, the measure developer, the measure developer’s Contracting Officer’s Representative (COR), and the Measures Manager all have distinct roles and responsibilities. Refer to Section 3, Chapter 3, Roles in Measure Development for more information.

2.1 ROLE OF THE MEASURE DEVELOPER

Measure developers are responsible for the development, implementation, and maintenance of measures, as required by individual contracts with CMS. The Blueprint guides entities holding contracts for quality measure-related activities including contracts for measure development, public reporting, analytics, and measure maintenance, and will most often address the user as the measure developer. However, other terms with similar meanings are used in various situations; the entities may also be called measure contractors. For the most part, the term “measure developer” is synonymous with measure contractor, but in some situations, the primary contractor may subcontract with other entities as measure developers to work on various tasks of the contract.

Another term used for entities involved with measures is measure steward. The NQF Phrasebook defines measure steward as “an individual or organization that owns a measure is responsible for maintaining the measure. Measure stewards are often the same as measure developers, but not always. Measure stewards are also an ongoing point of contact for people interested in a given measure.” CMS will be the steward for most measures developed under contract for CMS. However, for NQF-endorsed measures, the contracted measure developer will be responsible for carrying out the tasks required by the Measure Steward Agreement on behalf of CMS.

Measure developers fulfill CMS measure development, implementation, and maintenance requirements by:

- Using the processes and forms detailed in this Blueprint.
- Giving attention to Blueprint updates as provided by the Measures Manager (Section 1, Chapter 2.3, Role of the Measures Manager).
- Reviewing Blueprint requirements in context of their measure contract and good business practice; if the context requires flexible interpretation of the activities specified in the Blueprint, discuss options with the measure developer’s COR (henceforth referred to as COR) and the Measures Manager.
- Consulting with the Measures Manager with any questions about Blueprint processes.
- Attending forums and webinars related to measure development and the MMS.
- Providing feedback on the Blueprint to the COR and the Measures Manager.
• Ensuring that all deliverables are provided to the COR and relevant deliverables are also sent to their point of contact on the Measures Management team, or as directed by the contract and the COR.
• Copying the COR on all communications with the Measures Manager.

2.2 ROLE OF THE CONTRACTING OFFICER’S REPRESENTATIVE FOR THE MEASURE DEVELOPMENT CONTRACT

Although the measure developers are responsible to develop, implement, and maintain measures as specified in their contracts, CMS remains the measure steward, which means that CMS holds and retains ultimate responsibility for measures developed by its measure developers. Within the context of this Blueprint, the COR must ensure that tasks in the measure development, implementation, and maintenance contracts are completed successfully. The COR achieves this mission by:

• Notifying the Measures Manager COR when a new measure development, maintenance, or implementation contract is awarded.
• Ensuring that the relevant chapters of the Blueprint and required deliverables are appropriately incorporated into the requests for proposal, task orders, or other contracting vehicles and the ensuing contract.
• Requiring the measure developer’s compliance with the Blueprint, supporting basic training, and providing first-line technical assistance to the measure developer for the Blueprint.
• Ensuring that the measure developer is submitting copies of appropriate deliverables (as specified in the Schedule of Deliverables) to the Measures Manager.
• Determining when flexible application of Blueprint processes is appropriate and providing or obtaining CMS authorization for this variation.
• Providing or obtaining CMS approval of the measure developer’s deliverables at the specified processes defined in the Blueprint.
• Notifying the Measures Manager COR when a contract has ended.

2.3 ROLE OF THE MEASURES MANAGER

The Measures Manager supports CMS and its measure developers as they use the Blueprint to develop, implement, and maintain the healthcare quality measures. The Measures Manager achieves this mission by:

• Supporting CMS in its work of prioritizing and planning measurement activities and quality initiatives.
• Collecting a library of deliverables submitted as part of measure development contracts.
• Supporting CMS measure development communication, coordination, and collaboration meetings.
• Offering technical assistance to measure developers and CMS during measure development and monitoring processes, which includes soliciting feedback and implementing process improvements.
• Providing expertise and cross-cutting perspectives to CMS and measure developers regarding measures and measurement methods and strategies.
• Scanning the measurement environment to inform CMS of issues related to quality measures.
• Leading efforts to identify opportunities for harmonization of measures and measure activities across settings of care, programs, and initiatives.
- Reviewing draft documents and lists of potential measures to identify opportunities for measure harmonization and alignment.
- Facilitating measure harmonization work between measure developers as approved by the COR.
- Helping CMS coordinate across CMS, between multiple internal Department of Health and Human Services (HHS), and external key organizations: NQF, quality alliances, and major measure developers. This assistance is critical in establishing consensus on measurement policies, coordinating measure inventories, and promoting alignment across programs and settings of care.
- Ensuring, to the extent possible, that the Blueprint processes are aligned with NQF requirements.
- Refining the Blueprint continuously based on the evolving needs of CMS, customer feedback, and ongoing changes in the science of quality measurement.
- Conducting informational sessions on updates to the Blueprint and other key measurement-related activities.
- Facilitating posting of Calls for Measures, Calls for Public Comment, and Calls for Technical Expert Panel (TEP) on the CMS MMS website.
- Copying CORs on all Measures Manager to measure developer communications.
- Maintaining electronic mailing lists of measure development stakeholders and communicating with measure development public about topics of interest.
- Updating and maximizing the utility of the CMS Measures Inventory.
Section 2. The Measure Lifecycle
INTRODUCTION TO THE MEASURE LIFECYCLE

The end product of measure development is a precisely specified, valid, reliable, and clinically significant measure that is directly linked to CMS quality goals. Figure 1 shows a high-level view of the major tasks and timeline involved in developing measures from the time of the initial measure development contract award through measure implementation and maintenance. Although the figure depicts the five phases of the Measure Lifecycle in a linear, sequential fashion, measure developers have some flexibility to adjust the sequence or carry out steps concurrently and iteratively. Given this flexibility, the timeline in Figure 1 is only an estimate of the possible timeline of the Measure Lifecycle. CMS aims to shorten the current measure timeline so that new measures can be developed and implemented in a more timely fashion. Figure 2 depicts the way in which the Measure Lifecycle phases may overlap and interact in a nonlinear fashion.

Figure 1. Timeline of Measure Lifecycle
Note that the discussions in this document reflect both traditional measures and eCQMs. Information pertaining uniquely to eCQMs is designated with a computer icon.

For more background on eCQMs, refer to **Section 3, Chapter 6, Introduction to Electronic Clinical Quality Measures**.
2 **MEASURE CONCEPTUALIZATION**

In the first phase of the Measure Lifecycle, measure conceptualization, measure developers identify an opportunity for quality improvement in an area that would have a high impact on improving healthcare and quantify the potential impact on patients by compiling the evidence base for the concept and the basic elements of the measures. Determining the need for the measure is based on identification of the healthcare priority area, conceptual framework, environmental scan and gap analysis, and stakeholder input.

Potential measures should be reviewed for whether existing measures may be adopted or respecified to fit the desired purpose. If no measures are identified that match the desired purpose for the measure, the measure developer works with a TEP to develop new measures. Depending on the information gathering findings, including application of the measure evaluation criteria, the TEP will consider potential measures. These measures can be either newly proposed or derived from existing measures. The measure developer then submits the list of candidate measures, selected with TEP input, to the COR for approval. Upon approval from the COR, the measure developer proceeds with the development of draft specifications for the measures. Figure 3 depicts measure conceptualization in the context of the entire Measure Lifecycle.

![Figure 3. Flow of the Measure Lifecycle—Measure Conceptualization](image)

It is important to consider early in measure conceptualization what has been used before (e.g., what other developers have learned regarding feasibility or scientific evidence from existing measures or previous measure concepts that had been considered) to express concepts under consideration. Starting
very early in the measure concept stage will encourage selection of more feasible measure elements at the outset of measure development and identify possible unintended negative consequences with opportunity to mitigate them to avoid rework later in the process.

The main components of measure conceptualization include:

- Information gathering
- Business case development, including specific benefit when measure successfully meets its goal
- Stakeholder input:
  - TEP
  - Person/family engagement
  - Public comment.

### 2.1 INFORMATION GATHERING

Information gathering is a term that includes an environmental scan (i.e., literature review, clinical practice guidelines search, interviews, search for existing measures,⁴ and other related activities) and empirical data analysis. These activities are conducted to obtain information that will guide the prioritization of topics or conditions, gap analysis, business case building, and compilation of existing and related measures. This section describes the various sources of information that can be gathered as well as instructions for documenting and analyzing the collected information. Deliverables are outlined in Figure 4.⁵ Templates and instructions for the Measure Information Form (MIF) and Measure Justification Form (MJF) and a template for the Business Case are provided in Section 4. Forms and Templates. Note that some of the deliverables will evolve over the course of the measure lifecycle, such as the MIF and MJF.

Comprehensive information gathering will provide a significant knowledge base that includes the quality goals, strength of scientific evidence (or lack thereof) pertinent to the topics or conditions of interest, and information with which to build a business case for the measure. It will also produce evidence of general agreement on the quality issues pertinent to the topics/conditions of interest along with diverse or conflicting views. Underlying costs associated with the condition, procedure, or healthcare issue can

---

⁴ One resource to identifying existing measures is the CMS Measures Inventory Tool.
⁵ These deliverables and requirements are associated with the CMS MIDS USOW. A master list of all deliverables associated with the MIDS measure development process from the USOW is included in Section 3, Chapter 8.
also be explored in the early stages of information gathering to start the development of a return on investment calculation as a part of the business case.

At a minimum, the five measure evaluation criteria—importance, scientific acceptability of measure properties, feasibility, usability and use, and related and competing measures—will serve as a guide for conducting information gathering activities and for identifying priority topics/conditions or measurement areas. The fifth criterion—consideration of related and competing measures—refers to measure harmonization and should be considered from the very beginning of measure development. Both the measure specifications and measure evaluation are documented during this process in the MIF and MJF.6

Information gathering is conducted via 10 steps, which may or may not occur sequentially:

- Identify the healthcare quality issue to be addressed and determine if it is in a priority area
- Conduct an environmental scan (refer to Section 3, Chapter 10 Environmental Scan)
- Conduct an empirical data analysis, as appropriate
- Evaluate information collected during the environmental scan and empirical data analysis
- Conduct a measurement gap analysis to identify areas for new measure development
- Determine the appropriate basis for creation of new measures
- Apply measure evaluation criteria and propose a list of potential measures
- Submit the Information Gathering Report
- Prepare an initial list of measures or measure topics
- Explore possible data sources while considering feasibility (e.g., understanding the data captured in EHRs).

Complete details about these information gathering steps are found in Section 3, Chapter 9, Information Gathering.

2.2 BUSINESS CASE DEVELOPMENT

The business case provides CMS with the information needed to assess the anticipated benefits of a measure against the resources and costs required to develop and implement a measure. It should include enough information to demonstrate the strategic fit of the measure in CMS’s measure portfolio, addressing the strategic goals and objectives of the Meaningful Measures Initiative, its value to the public, the capacity of the healthcare system to respond to the quality action defined by the measure, and the affordability and achievability of the measure in terms of quality improvement and performance measurement. The initial business case information is gathered during the initial information gathering process. Vital information can be obtained during later stages of measure development and should be added to the business case to produce a final business case.

Noting that the first three of these categories listed align with the Institute for Healthcare Improvement’s (IHI) Triple Aim (Berwick et al., 2008), these types of information should be systematically evaluated to build the business case as factors considered when evaluating a measure (Leatherman et al., 2003):

---

6 Completed NQF measure submission forms may be used for contract deliverables in lieu of the MIF and MJF; permissible only if the contract allows for it.
Better Care:
- Improve overall quality ensuring healthcare is safe, timely, effective, efficient, equitable, and patient-centered (Institute of Medicine [IOM], 2001), as well as affordable.
- Improve processes of care delivery where every patient receives the appropriate care every time in the most appropriate setting (e.g., preventive screening services, immunizations, pharmacotherapy, and counseling).
- Improve intermediate outcomes (e.g., lower blood pressure, lower HbA1c values among diabetics) and outcomes of care (e.g., fewer heart attacks and decubitus ulcers).
- Decrease rates of untoward effects or complications of care and the likelihood of their occurrence (e.g., bleeding from anticoagulation, death from low blood glucose levels).
- Include the patient as a member of the care team and implement PRO measures.
- Promote coordinated care across settings (e.g., promoting interoperability).

Healthy People/Healthy Communities:
- Improve the health of the U.S. population by supporting proven interventions to decrease incidence and prevalence of disease in the population.
- Focus on population health and not an individual patient, including but not limited to an emphasis on cross-continuum and coordination of care.
- Decrease variation in care across disadvantaged subgroups related to race, ethnicity, and other social risk factors.

Affordable Care:
- Decrease cost of implementing the measure.
- Increase efficiency of implementing clinical processes.
- Save money by preventing complications, errors, and adverse events.
- Promote value through right care in right place at right time for right patient.
- Reduce cost by promoting wellness and preventing illness.

Continual Improvement:
- Monitor the magnitude and time frame of the expected benefits improvement with significant room for improvement.
- Require a projected measure performance trajectory, including estimation of when performance may top out.
- Demonstrate how the improvements from measure development and implementation have the potential for far-reaching, long-term benefits.

Joy in Work:
- Consider how data collection will impact the clinical workflow.
- Work to achieve health equity (Perlo et al., 2017).
Burden Reduction:

- Minimize data collection through the use of existing data elements and electronic sources, where possible.
- Standardize specifications for related measure when they have the same measure focus, same target population, or apply to many measures.
- Align and harmonize across CMS programs and care settings.

While other models may be used, a cost savings model is the most prevalent for evaluating the potential quality measures’ business case (i.e., the aggregate effect of cash inflows and outflows accruing to an organization as a result of implementing a specific process or treatment). This model presents a result that, given its quantitative method, can be more easily interpreted, and it can be reliably compared to rank multiple events. If anticipated savings are not expected to be realized until future years, the savings should be adjusted to a net present value. This model also applies to many outcome measures. For example, if increased physician follow-up visits are required to reduce hospital readmissions, the savings equals the cost saved by not being readmitted minus the cost of the additional physician visits.

The cost savings model is not the only way to quantify benefits of implementing a specific measure or measure set. Better health and better care should be measured with quantifiable anticipated benefits assigned to a model that can then be tested. These assertions should be supported by high-quality, consistent evidence. Using the example mentioned above, improved care coordination could not only reduce expenses associated with unnecessary readmissions, but it could also reduce mortality in selected populations and improve patient satisfaction.

Regardless of the model used, a hypothesis that can be used for later testing should be stated in explicit terms and, at a minimum, predict how the measure will have an effect over time (the trajectory). It is essential that these details are presented in the business case so comparisons can be made during measure use, continuing evaluation, and maintenance. When possible, historical data and baseline data should be included. Historical and baseline data, in this context, refer to data collected from the measure (if completing for maintenance) or similar measures (identified during the environmental scan), established and valid findings from the literature or other reliable resources, or information collected from practice guidelines and similar guidance documents. These historical and baseline data provide reviewers a resource from which to determine the existence and extent of performance gaps as well as changes in those gaps, as possible.

After measures have been implemented and are in use, the measure developer should reevaluate the business case with the other measure evaluation criteria and report, to CMS and in NQF’s reevaluation process, whether the projected improvements were achieved. This consideration will impact continued use (or modification) of the measures.

Complete details about the business case are found in Section 3, Chapter 11, Business Case. The Business Case Template is provided in Section 4, Forms and Templates. Deliverables for this step are outlined in Figure 5.7

---

7 These deliverables and requirements are associated with the CMS MIDS USOW. A master list of all deliverables associated with the MIDS measure development process from the USOW is included in Section 3, Chapter 8.
2.3 **STAKEHOLDER ENGAGEMENT**

2.3.1 **Technical Expert Panel**

A TEP is a group of experts and stakeholders who contribute direction and thoughtful input to the measure developer in every phase of the measure development process, from conceptualization through maintenance. While panel members may be involved at any time during the development process, CMS requires the panel to be asked for input at specific times, including but not limited to: when developing the **business case** (i.e., why the measure makes sense and is important), when reviewing testing results, and when deciding which measures should be recommended to CMS. Because an important use of quality measures is to provide information to patients and their caregivers on the quality of care provided, their perspective on what is important and useful to measure and evaluate is vital and cannot be overlooked. One way that this goal may be accomplished is by having a patient or caregiver on the panel.

Although TEP input is critical to the measure developer in advising their process, TEP input cannot be used to advise CMS. The **Federal Advisory Committee Act (FACA)** has specific rules about advising the government directly, so it is important to be familiar with them. Measure developers should be clear in all materials and references that the TEP is advising the measure developer and not CMS directly. In addition, federal representatives to a TEP should serve only as non-voting members.

The TEP process involves three postings to the dedicated [MMS page](#) on the CMS website. These three postings include:

- TEP (Call for TEP) Nominations
- TEP Composition Documentation with Meeting Dates
- TEP Summary Report.

Measure developers will communicate and collaborate with the Measures Manager for these postings. The website posting process is detailed in [Section 3, Chapter 15, MMS Website Posting](#), and may take up to 5 working days.

The steps for TEP are detailed in [Section 3, Chapter 12, Technical Expert Panel](#), and should be performed when convening the TEP and conducting the TEP meetings. TEP deliverables and requirements are outlined in Figure 6.8 Instructions and/or templates for most deliverables and requirements are found in [Section 4, Forms and Templates](#).

---

**Technical Expert Panel Deliverables and Requirements**

1. Call for TEP member nominations
2. TEP nomination forms
3. TEP charter
4. TEP composition documentation (TEP membership list) (Deliverable 4-1)
5. TEP meeting materials and minutes
6. Potential measures presented to the TEP
7. Measure Evaluation Report(s) (Deliverable 4-2)
8. Updated MIF and MJF
9. TEP Summary Report (Deliverable 4-3)

---

8 A master list of all deliverables associated with the measure development process is included in [Section 3, Chapter 8](#).
2.3.2 Person and Family Engagement

Involving persons and family representatives in the measure development process (e.g., on TEPs, in focus groups, during testing) is among the many ways that CMS strives to accomplish its goal of strengthening person and family engagement as partners in their care. In this context, a person is a non-healthcare professional representing those who receive healthcare. Family representatives are other non-healthcare professionals, such as caregivers, supporting those who receive healthcare. Guidance for obtaining input from persons and family member stakeholders is provided in Section 3, Chapter 13, Person and Family Engagement.

2.3.3 Public Comment

Public comment ensures that measures are developed using a transparent process with balanced input from relevant stakeholders and other interested parties. During a public comment period, measure developers may receive critical suggestions that were not previously considered by the measure developer and the TEP. The procedures described next will apply whenever public comment is obtained.

These eight steps are essential to successfully soliciting public comment. Deviation from the procedure requires COR approval. Deliverables and requirements associated with soliciting public comments are outlined in Figure 7.10 Instructions and/or templates for several of the deliverables and requirements are found in Section 4, Forms and Templates.

1. Prepare the Call for Public Comment
2. Post the measures following COR approval
3. Notify relevant stakeholder organizations
4. Collect information
5. Summarize comments and produce report
6. Send comments to the TEP for consideration
7. Finalize the Public Comment Report, including verbatim comments
8. Arrange for the final Public Comment Summary Report to be posted on the website.

More detail on these steps is included in Section 3, Chapter 14, Public Comment.

---

9 The term “persons” is used in lieu of “patients” throughout the Blueprint to reflect an individual’s identity as more than a patient and to recognize his or her participation in prevention and wellness. The term “patients” is not used here because the term “patient” refers to individuals receiving inpatient care in hospitals.

10 These deliverables and requirements are associated with the CMS MIDS USOW. A master list of all deliverables associated with the measure development process is included in Section 3, Chapter 8.
3 **Measure Specification**

The process of developing measure specifications occurs throughout the measure development process. The measure specification process is defined by both technical specification and [harmonization](#), along with stakeholder engagement through public comment on the process. Final technical specifications provide the comprehensive details that allow the measure to be collected and implemented consistently, reliably, and effectively. Measure developers complete and update the fields within the MIF as the measure progresses from Measure Conceptualization to [Measure Testing](#), and finally, when/if submitted to NQF for endorsement consideration. Figure 8 depicts the measure specification portion of the Measure Lifecycle that is discussed here.

![Figure 8. Flow of the Measure Lifecycle—Measure Specification](image)
3.1 TECHNICAL SPECIFICATION

The MIF is used to document the technical specifications of the measures. At this stage, the technical specifications are likely to include high-level numerator and denominator statements and initial information on potential exclusions, if applicable, and will continue to be completed throughout the development process as more information is obtained. Deliverables and requirements associated with the measure specification process are outlined in Figure 9. Instructions and/or templates for many of the deliverables and requirements are found in Section 4, Forms and Templates.

Developing technical specifications is an iterative process. Prior to drafting initial specifications, the measure developer should consider the data elements necessary for the proposed measure and conduct preliminary feasibility assessments. The measure developer then drafts the initial specifications and the TEP will review and may suggest changes.

During the development process, alpha (formative) testing of the measure occurs. For measures based on electronic, administrative, or claims-based data, the draft technical specifications may be provided to the programming staff responsible for data retrieval and for developing the programming logic necessary to produce the measure. The programmers will assess the feasibility of the technical specifications as written and may provide feedback. For measures based on chart abstraction, data collection tools are developed and tested. When the specifications are more fully developed, beta (field) testing occurs. Section 2, Chapter 4, Measure Testing, provides details of the procedures for beta testing. As a result of testing, technical specifications will continue to evolve, becoming more detailed and precise.

---

11 Completed NQF measure submission forms may be used for contract deliverables in lieu of the MIF and MJF.
12 These deliverables and requirements are associated with the CMS MIDS USOW. A master list of all deliverables associated with the MIDS measure development process from the USOW is included in Section 3, Chapter 8.
The key components of technical specifications are:

- Measure name/title
- Measure description
- **Initial population**
- **Population** descriptions (denominator statement and numerator statement for ratio and proportion measures)
- Exclusions and exceptions
- **Data sources**
- Key terms, data elements, codes, and code systems
- Unit of measurement or analysis
- Sampling and stratification, as needed
- Risk adjustment, as needed (refer to Section 3, Chapter 19, Risk Adjustment)
- Timing and time intervals
- Calculation algorithm.

*eCQMs* must also have a machine-readable eXtensible Markup Language (XML) document in Health Quality Measure Format (HQMF), CQL, Expression Logical Model (ELM) XML, ELM JSON XML, an HTML document, and link to the value sets in the Value Set Authority Center (VSAC). Most of the key components listed are in the eCQM files.

These steps are performed to develop the full measure technical specifications:

- Develop the candidate measure list
- Develop precise technical specifications and update the MIF
- Define the data source
- Specify the codes, code systems, code sets, code lists, and/or value sets
- Construct the data protocol
- Document the measures and obtain COR approval.

Details on the execution of each of these steps is included in Section 3, Chapter 16, Measure Technical Specification.

In certain cases, risk adjustment of the measure is also a component of the specification process, specifically, for outcome measures. Refer to Section 3, Chapter 19, Risk Adjustment, for more detail on determining when risk adjustment is necessary and steps on performing the risk adjustment.

Technical specifications are also slightly different in execution for cost and resource use measures and composite measures. Refer to Section 3, Chapter 20, Cost and Resource Use Measure Specification and Chapter 21, Composite Measure Technical Specification, for details on developing technical specifications for those measure types.

### 3.2 Harmonization

When specifying measures, measure developers should consider whether a similar measure exists for the same condition, process of care, outcome, or care setting. Measure developers should consider harmonization for every measure under development or maintenance throughout the Measure Lifecycle. Measures should be harmonized unless there is a compelling reason for not doing so (e.g., significant risk variation by age, comorbidity, race) that would justify a separate measure. Harmonization standardizes similar measures when their differences do not make them scientifically stronger or more valuable. Harmonization should not result in inferior measures, but in measures that are scientifically strong, clinically valuable, and important to persons/families/caregivers. Quality measures should be based on the best way to calculate whether and how often the healthcare system does what it should. It should not be assumed that an endorsed measure is better than a new measure.
When developing specifications, measure developers should consider various aspects of the measure for potential harmonization. Harmonization often requires close inspection of specification details of the related measures. Harmonizing measure specifications during measure development is more efficient than harmonizing after a measure has been fully developed and specified. The earlier in the process that related or competing measures are identified, the sooner problematic issues may be resolved.

Harmonization may include, but is not limited to:

- Age ranges
- Measurement period
- Allowable values for medical conditions or procedures (e.g., codes, code systems, code lists, descriptions)
- Allowable conditions for inclusion in the denominator (e.g., codes, code systems, code lists, descriptions)
- Exclusion categories, whether the exclusion is from the denominator or numerator, and whether optional or required
- Calculation algorithm
- Risk adjustment methods.

Examples:

- NQF 0417: Diabetic Foot & Ankle Care, Peripheral Neuropathy – Neurological Evaluation (Steward: American Podiatric Medical Association) is a process measure reporting the frequency of those evaluations by providers; the proposed measure addresses peripheral neuropathy outcomes.
- Influenza immunization measures exist for many care settings, but the new measure is for a new care setting.
- Readmission rates exist for several conditions, but the new measure is for a different condition.
- A set of new hospital measures may be able to use data elements already in use for existing hospital measures.

If the measure can be harmonized with any characteristics of existing measures, then use the existing definitions for those attributes. Consult with the Measures Manager and other resources (e.g., CMS Measures Inventory Tool [CMIT] and the electronic Clinical Quality Improvement [eCQI] Resource Center) to review specifications to identify opportunities for further harmonization. If measures should not be harmonized, then document those reasons and include any literature used to support this decision. Some reasons not to harmonize include:

- The science, such as clinical practice guidelines, behind the new measure does not support using the same variable(s) found in the existing measure.
- Measure’s intention varies across programs/payers and requires the difference.
- The measures have differing denominator populations at significantly different risk (i.e., the denominators are stratified by risk).
Examples:

- An existing diabetes measure includes individuals aged 18 to 75; a new process of care measure is based on new clinical practice guidelines that recommend a specific treatment only for individuals aged 65 years and older.
- An existing diabetes measure includes individuals aged 18 to 75; CMS has requested measures for beneficiaries aged 75 years and older.

For more detail on measure harmonization, refer to Section 3, Chapter 18, Measure Harmonization.

### 3.3 Stakeholder Engagement

Although it is advisable to obtain public comments at several points during measure development, a key time to obtain additional public comments is the measure specification drafting phase. Comments received during the public comment period are reviewed and taken into consideration by the measure developer, CMS, and the TEP and may result in revisions to the measure specifications. For more detail on stakeholder engagement, refer to Section 3, Chapter 14, Public Comment.
4 **Measure Testing**

Measure testing enables a measure developer to assess the suitability of the quality measure’s technical specifications and acquire empirical evidence to help assess the strengths and weaknesses of a measure with respect to the NQF Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement. Information gathered through measure testing is part of full measure development, and this information can be used in conjunction with expert judgment to evaluate a measure. For Blueprint purposes, measure testing refers to testing quality measures, including the components of the quality measures, such as the data elements, the instruments, and the performance score.

Figure 10 describes how testing fits into the flow of the Measure Lifecycle.

![Figure 10. Flow of the Measure Lifecycle—Measure Testing](image)

4.1 **The Measure Testing Process**

Properly conducting measure testing and analysis is critical to approval of a measure by CMS and endorsement by the NQF. Section 3, Chapter 22, Measure Testing, describes the types of testing that may be conducted during measure development (alpha and beta testing), the procedure for planning and testing under the direction of the COR, and key considerations when analyzing and documenting results of testing and analysis, including incorporation of stakeholder inputs after testing is complete.
When testing a measure (or set of measures) for CMS, a measure developer is required to submit specific reports. Although reports are always required after completion of beta testing, measure developers should discuss the need for reporting upon more formative alpha testing with the COR, especially if the alpha testing is intended to precede beta testing under the same measure development contract. Figure 11 lists the deliverables and requirements associated with measure testing. Instructions and templates for the MIF, MJF, and Measure Evaluation Report are found in Section 4, Forms and Templates.

Figure 12 shows the relationships between the eight steps of measure testing. The first few steps address planning and execution of testing and are identical for alpha and beta testing; the last steps address reporting and follow-up after the conclusion of testing:

1. Develop the testing work plan
2. Submit the plan and obtain CMS approval
3. Implement the plan
4. Analyze the test results
5. Refine the measure, including incorporation of stakeholder inputs
6. Retest the refined measure
7. Compile and submit deliverables to CMS
8. Support CMS during NQF endorsement process.

---

13 These deliverables and requirements are associated with the CMS MIDS USOW. A master list of all deliverables associated with the MIDS measure development process from the USOW is included in Section 3, Chapter 8.
4.2 Stakeholder Engagement

It is often appropriate to obtain stakeholder inputs at several points during the testing process, which includes obtaining face validity inputs at alpha testing, feasibility and burden inputs at beta testing, and other inputs based on a review of overall results. These inputs can take many forms, including but not limited to, formal TEPs, consultation with subject matter experts (SMEs), outreach to professional associations or patient advocacy groups, and public comments. Once obtained at a given step, it is important to follow up on those communications by providing additional opportunities for stakeholders to comment on the results of their inputs at future stages. Such follow up maximizes the likelihood that the developer operationalizes the inputs in a way consistent with the stakeholders’ needs. It also improves the likelihood stakeholders will remain engaged for ongoing support on current or future measures.
5  **MEASURE IMPLEMENTATION**

**Quality measure** implementation includes all activities associated with taking a measure from a development state to an active, in-use state, which includes but is not limited to, consensus endorsement processes, measure selection processes, and measure rollout. CMS identifies and selects measures through a transparent process that is open to stakeholders and public comment. Depending on the CMS program, there are different paths that a measure can take for implementation. When considering a measure for a topic already measured in another program, CMS prefers to use the same measure or a harmonized measure. However, the measure must be tested for **reliability** and **validity** with the new **population**.

5.1  **PRE-RULEMAKING**

The CMS programs involved in CMS pre-rulemaking and rulemaking process include those identified under **ACA Section 3014**. Measures for these programs are submitted to the **MUC** list, which makes publicly available a list of measures HHS is considering adopting through the federal rulemaking process for use in a select number of Medicare payment program(s).

Through the NQF-convened Measure Applications Partnership (**MAP**) process, multi-stakeholder groups provide input to HHS on the selection of quality and **efficiency measures**. The MAP also considers program and measure **alignment** when deciding which measures to recommend. After considering the **MAP recommendations**, CMS proposes which measures they intend to implement. They also publish the rationale for the use of any quality and efficiency measures that are not endorsed by the consensus-based entity (CBE) as well as the MAP recommendation for the measures.

Measures may also be submitted from organizations other than CMS. For example, a specialty society may submit a set of measures to be considered for programs covered under this process.

5.2  **OTHER SUBMISSIONS**

Some measures or measure programs do not use the pre-rulemaking or rulemaking process; however, CMS still requires the same level of rigor. To maintain rigor, the steps differ only slightly from those used in measures that require rulemaking; measures still undergo the identification and finalization steps through a public process:

- CMS issues a call letter to solicit measures and/or identify measures considered for removal.
- Submitted measures follow the HHS clearance process.
- Cleared measures go through a consensus development process, which might include the MAP process; this step is not required for all programs.
- Developers solicit public comments on all measures.
- Once satisfied with the measures, CMS issues a final letter of implementation for the selected measures.

These measure programs have their own submission processes. Measure developers should check the relevant program’s requirements for additional guidance.
5.3 **STAKEHOLDER ENGAGEMENT**

MUCs for implementation are publicly submitted for comment either through the formal federal rulemaking process or through an ad hoc public comment process for measures that are not subject to rulemaking. Measure developers convene stakeholder meetings regarding the implementation of considered measures, and their questions about the measures are resolved iteratively as the measure remains under consideration. The measure implementation process is completely transparent and open to the public for comments and questions.

5.4 **THE IMPLEMENTATION PROCESS**

Figure 13 depicts the process of measure implementation, which encompasses three phases:

- NQF endorsement, if applicable
- Measure selection
- Measure rollout.

The process of implementing measures varies significantly from one measure to another depending on many factors, which may include, but is not limited to:

- Scope of measure implementation:
  - Implemented in a new program
  - Added to an existing program
- Healthcare provider being measured
- Data collection processes
- Ultimate use of the measure (e.g., quality improvement, public reporting, pay-for-reporting, or value-based purchasing)
- Program into which the measure is being added.
Figure 14 outlines the deliverables associated with measure implementation.\textsuperscript{14} For detailed information on measure implementation phases, refer to Section 3, Chapter 26, NQF Endorsement and Maintenance; Section 3, Chapter 27, Measure Selection; and Section 3, Chapter 28, Measure Rollout.

---

**Measure Implementation Deliverables**

1. Public description of quality measures (Deliverable 4-12)
2. Timeline for data item and/or quality measure implementation (Deliverable 8-1)
3. Implementation stakeholder meetings (Deliverable 8-2)
4. Question and answer support (Deliverable 8-3)
5. Implementation process roadmap (Deliverable 8-4)
6. Measure calculations/results (Deliverable 8-5)
7. Pre-posting preview results (Deliverable 9-2)
8. Compare site files and measures (as applicable) (Deliverable 9-1)
9. Implementation algorithm (also called calculation algorithm/measure logic) (Deliverable 9-3)
10. Data Use Agreement (Deliverable 10-1)

---

\textsuperscript{14} These deliverables and requirements are associated with the CMS MIDS USOW. A master list of all deliverables associated with the MIDS measure development process from the USOW is included in Section 3, Chapter 8.
6 **Measure Use, Continuing Evaluation, and Maintenance**

To help CMS ensure the continued soundness of the measures, the measure developer must provide strong evidence that a measure currently in use continues to add value to quality reporting and incentive programs and that its construction continues to be sound throughout its lifecycle. This work also helps CMS ensure that its measures obtain or maintain NQF endorsement.

### 6.1 Continuing Evaluation

The measure developer uses the continuing evaluation process to update the MJF and any changes to the technical specifications to demonstrate that:

- The aspects of care included in the specifications continue to be highly important to measure and report because the measurement results can supply meaningful information to consumers and healthcare providers.
- The measurement results continue to drive significant improvements in healthcare quality and health outcomes where there is variation in and/or overall less-than-optimal performance.
- The data elements, codes, and parameters included in the specifications are the best ones to use to quantify the specific measure because they most accurately and clearly target the aspects of the measure that are important to collect and report, and they do not place undue burden on resources in order to collect the data.
- The calculation methods included in the specifications remain valid because they reflect a clear and accurate representation of the variation in the quality or efficiency of the care delivered or the variation in the health outcome of interest.
- The measure continues to be either unique for its topic or it is the “best in class” when compared to competing measures.
- The measure is comparable to other measures in its clinical significance or difficulty.

### 6.2 Measure Maintenance

As depicted in Figure 16, there are multiple steps to measure maintenance. These steps are reported via three basic types of measure maintenance reviews: annual updates, comprehensive reevaluations, and ad hoc reviews, with stakeholder inputs being a critical component of this review process.

---

**Measure Maintenance Deliverables and Requirements**

1. Maintenance Reevaluation (Deliverable 4-9)
2. Comprehensive Reevaluation (Deliverable 4-10)
3. Ad Hoc Reevaluation (Deliverable 4-11)
4. Program and Initiative Assessment Report (Deliverable 8-6)
5. Audit and Validation Reports
6. Audit and Validation Appeals Reports
7. Preview Reports, if required by the CMS program using the measure
8. Periodic Measure Rate Trend Reports
9. Analysis of the measure results
10. Ad hoc analyses, as requested by CMS
11. Questions and answers support
12. Periodic environmental scans
13. Data files of the measure rates and/or demographic information suitable for posting on CMS website
14. Updated MJF (Deliverable 3-4), as necessary
15. Submit documentation to NQF for reendorsement

*Figure 15. Measure Maintenance Deliverables and Requirements*
Figure 16. Flow of the Measure Lifecycle—Measure Use, Continuing Evaluation, and Maintenance
6.2.1 Measure Production and Monitoring

Seven (7) steps are involved in the continuous production and monitoring of implemented measures:

1. Conduct data collection and ongoing surveillance
2. Respond to questions about the measure
3. Produce preliminary reports
4. Report measure results
5. Monitor and analyze the measure rates and audit findings
6. Perform measure maintenance or ad hoc review, when appropriate
7. Provide information that CMS can use in measure priorities planning.

Details on these steps can be found in Section 3, Chapter 29, Measure Production and Monitoring.

6.2.2 Measure Maintenance Reviews

Three (3) types of maintenance reviews are described in Section 3, Chapter 30, Measure Maintenance Reviews, including deliverables and the steps required for each:

- Annual update
- Comprehensive reevaluation
- Ad hoc review.

For more information on harmonization and evaluation during measure maintenance, refer to Section 3, Chapter 18, Measure Harmonization and Section 3, Chapter 23, Measure Evaluation.

6.2.3 Stakeholder Engagement

This Blueprint describes the annual update, comprehensive reevaluation, and ad hoc review as distinct and separate activities; however, in practice, these activities sometimes overlap and are conducted concurrently. All steps require solicitation of inputs via public comment. Results from, and progress on, each of these review processes are reported publicly. Stakeholders are engaged for comment and, in some cases, participation in formal panel review. Ideally, the measure maintenance schedule is aligned with the NQF endorsement maintenance cycle, which also includes requirements for public review and comment. However, in practice, these schedules may not align completely.
7  TOOLS AND RESOURCES FOR MEASURE DEVELOPERS

Numerous tools and resources are available to assist measure developers in the different phases of the measure lifecycle. New tools and resources are always under development. This list is not exhaustive and provides examples.

7.1  BONNIE

Bonnie is a software tool that allows eCQM developers to test and verify the behavior of their eCQM logic. The Bonnie application allows measure developers to independently load measures that they have constructed using the Measure Authoring Tool (MAT) and helps measure developers execute the measure logic against the constructed patient test deck and evaluate whether the logic aligns with the intent of the measure.

7.2  CMS DATA ELEMENT LIBRARY (DEL)

The DEL is the centralized resource for CMS assessment instrument data elements, questions and responses, and their associated information technology (IT) standards. In support of the Improving Medicare Post-Acute Care Transformation Act of 2014 (IMPACT Act), the goals of the DEL are to:

- Serve as a centralized resource for CMS assessment data elements (questions and response options)
- Promote the sharing of electronic CMS assessment data sets and health IT standards
- Influence and support industry efforts to promote EHR and other health IT interoperability.

7.3  CMS MEASURES INVENTORY TOOL

CMIT is a repository for information about CMS measures. CMS and its partners use the inventory to inform stakeholders, manage its measure portfolio, and guide measure development. The functions allow users to find measures quickly; compile and refine sets of related measures; identify measures across the continuum of care; and coordinate measurement efforts across all conditions, settings, and populations.

7.4  CYPRESS

Cypress is an open source testing tool used by vendors to certify their EHRs and health IT modules for calculating eCQMs. Cypress is an official testing tool for the Office of the National Coordinator for Health Information Technology (ONC) Health IT Certification Program. Testing involves Cypress generating synthetic patient records for the subset of published eCQMs selected for certification and testing the ability of the EHR systems and health IT modules to accurately record, import, calculate, filter, and report eCQMs.

7.5  ELECTRONIC CLINICAL QUALITY IMPROVEMENT RESOURCE CENTER (eCQI RESOURCE CENTER)

The eCQI Resource Center is a website that provides eCQI resources and connections with the community of professionals who are dedicated to electronic clinical quality improvement for better health. It serves as “the one-stop shop for the most current resources to support electronic clinical quality improvement.” It is the source of truth for specifications of eCQMs in CMS programs and is the
home to the Collaborative Measure Development Workspace (CMD) and eCQM Data Element Repository (DERep).

The eCQI Resource Center has an eCQI Tools and Resources Library providing information and links to tools and resources used in electronic clinical quality improvement.

7.6 **ENVIRONMENTAL SCANNING SUPPORT TOOL (ESST)**

The ESST makes it easier to complete the environmental scans required in the information gathering process to develop and maintain quality measures. The ESST reduces the time needed to scan literature from months to hours, saving immense amounts of resources. Conventional literature reviews use keyword searches to find the relevant research that has been published on a topic, a collection that may contain millions of documents. The ESST uses an automated natural language processing (NLP) approach that rapidly scans literature in PubMed and PubMed Central (and in the future, CINAHL) to:

- Identify relevant documents (abstracts and full-text articles)
- Identify and extract the specific knowledge within each relevant document that applies to the measure’s potential opportunity for improvement.

7.7 **MEASURE AUTHORING TOOL (MAT)**

The MAT is a web-based tool that allows measure developers to author eCQMs using CQL and the Quality Data Model (QDM). The MAT provides the capability to express complex measure logic and export measures in several formats.

7.8 **NATIONAL QUALITY FORUM INCUBATOR (NQF INCUBATOR®)**

The NQF Incubator facilitates measure development and testing through collaboration and partnership. Example goals of the NQF Incubator are to fill measure gaps with more Meaningful Measures, encourage development of eCQMs, and advance measurement science.

7.9 **NATIONAL QUALITY FORUM QUALITY POSITIONING SYSTEM (QPS)**

The QPS is a web-based inventory tool developed by the NQF to help people select and use NQF-endorsed measures. It allows a user to search for NQF-endorsed measures in a number of ways, for example by type of measure, and then export search results. Using QPS, a user can find NQF-endorsed measures on particular topics, track and receive reminders about measures that are important to them, provide feedback on measures, and discover which measures others are using.

7.10 **VALUE SET AUTHORITY CENTER (VSAC)**

The VSAC is a repository and authoring tool for public value sets created by external programs. Value sets are lists of codes and corresponding terms, from National Library of Medicine (NLM)-hosted standard clinical vocabularies (such as Systematized Nomenclature of Medicine—Clinical Terms [SNOMED CT®], RxNorm, Logical Observation Identifiers Names and Codes [LOINC®] and others) and billing terminologies (such as International Classification of Diseases-10th Revision-Clinical Modification [ICD-10-CM]), that define clinical concepts to support effective and interoperable health information exchange. The VSAC does not create value set content. With a free Unified Medical Language System (UMLS®) Metathesaurus license, the VSAC also provides downloadable access to all official versions of value sets specified by CMS eCQMs.
Section 3. In-Depth Topics
1 HEALTHCARE QUALITY PRIORITIES

1.1 CMS GOALS AND PRIORITIES

At CMS, the top priority is putting patients first, with the patient always being at the center of CMS’s work. CMS’s strategic goals support the patient and overall patient experience by:

- Improving the CMS customer experience
- Ushering in an era of state flexibility and local leadership
- Supporting innovative approaches to improve quality, accessibility, and affordability
- Empowering patients and clinicians to make decisions about their healthcare.

In order to put patients first across all programs – Medicaid, Medicare, and the Health Insurance Exchanges – CMS must empower patients to work with their physicians and make healthcare decisions that are best for them.

This empowerment means giving patients meaningful information about quality and costs to be active healthcare consumers. It also includes supporting innovative approaches to improving quality, accessibility, and affordability, while finding the best ways to use innovative technology to support patient-centered care.

**Empowering Patients:** CMS puts patients at the center of our healthcare system by ensuring they have the resources they need to make the best decisions for themselves and their families.

**Focusing on Results:** CMS uses new flexibilities and incentives, working to make sure that patients receive the right care, at the right time, in the right place while protecting taxpayers by paying for care based on results.

**Unleashing Innovation:** CMS continues to remove the barriers that too often limit innovation. Innovations are needed to make a healthcare system where providers and health plans compete to deliver better care at lower costs.

These goals are framed into a strategic wheel (Figure 17) reflecting the strategic initiatives across the agency.
Much of the quality measurement work across CMS aligns with the strategic initiative, “Ensuring Safety & Quality.” With a focus on better patient health outcomes, CMS holds providers accountable for providing safe and effective care, while minimizing administrative burden to ensure clinicians can spend more time with patients. We are focused on ensuring beneficiaries are empowered to make decisions about their healthcare based on quality and cost information by moving our quality programs to measure value and to provide consumers access to information in an understandable and actionable way.

1.2 MEANINGFUL MEASURES INITIATIVE

Regulatory reform and reducing regulatory burden are high priorities. CMS continuously works to find ways to reduce burden on providers, while empowering patients. By identifying the highest priorities for quality measurement and improvement, the Meaningful Measures Initiative provides a framework for core issues that are most vital to improving patient outcomes. The Meaningful Measures Initiative represents a new approach to quality measures that will reduce the collection and reporting burden, while producing quality measurement that is more focused on meaningful outcomes.
The objectives of the Meaningful Measures Framework include:

- Addressing high-impact measure areas that safeguard public health
- Focusing on areas that are patient-centered and meaningful to patients
- Developing measures that are outcome-based, where possible
- Fulfilling each program’s statutory requirements
- Minimizing the level of burden for healthcare providers
- Identifying areas for significant opportunity for improvement
- Addressing measure needs for population-based payment through APMs
- Aligning across programs and/or with other payers.

In order to achieve these objectives, CMS has identified 19 Meaningful Measures areas and mapped them to 6 overarching quality priorities as shown in the Figure 18.15

Figure 18. Meaningful Measures Framework

15 The Meaningful Measures Framework will be regularly evaluated and updated to reflect stakeholder feedback given to CMS and any shift in CMS priorities. For the most up to date information, visit the Meaningful Measures website.
1.3 Quality Across CMS

CMS supports the healthcare priorities by developing quality measures that address these priorities and goals, and implements them through provider feedback, public reporting, and links to payment incentives. CMS has long played a leadership role in quality measurement and public reporting. CMS started by measuring quality in hospitals and dialysis facilities and now measures and publicly reports the quality of care across settings of care, including nursing homes, home health agencies, physician offices, and drug and health plans. Beginning in 2012, CMS efforts expanded the quality reporting programs to include physician offices, inpatient rehabilitation facilities, inpatient psychiatric facilities, cancer hospitals, and hospices. CMS is also transforming from a passive payer to an active value purchaser by implementing payment mechanisms that reward providers who achieve better quality or improve the quality of care they provide. CMS has been seeking “to transition from setting-specific, narrow snapshots...to assessments that are broad-based, meaningful, and patient-centered in the continuum of time [and delivery modalities] in which care is delivered” (Conway et al., 2013, p. 2215).

In addition, CMS is committed to supporting states’ efforts to measure and improve the quality of healthcare for children and adults enrolled in Medicaid and CHIP. CMS is building on its experiences in provider quality measurement and reporting to support similar state Medicaid programs and CHIP. CMS is mindful that state Medicaid agencies, health plans, and providers will want to use measures that are aligned, reflect beneficiary priorities, provide value, have impact, and are not administratively burdensome.

CMS contracts with external organizations to develop and implement quality measurement programs. These organizations include Quality Innovation Network-Quality Improvement Organizations (QIN-QIOs), university researchers, health services research organizations, and consulting groups. The Measures Manager supports the CORs and their various measure developers in their work implementing the MMS.
2 MEASURE PRIORITIZATION AND PLANNING

CMS responds to a variety of inputs to develop and implement its quality measurement agenda for the next 5 to 10 years. CMS develops and implements measures with the primary purpose of improving care in a spectrum of healthcare service delivery settings such as hospitals, outpatient facilities, physician offices, nursing homes, home health agencies, hospices, inpatient rehabilitation facilities, and dialysis facilities. CMS selects measures based on the priorities articulated in the Meaningful Measures Framework. CMS places emphasis on electronically specified measures for implementation in quality initiatives. These measures include public reporting, value-based purchasing, and other payment incentive and accountability programs.

In broad terms and in context of recent legislative mandates, CMS continues to pursue measure development and maintenance work based on the Meaningful Measures Framework and national healthcare priorities, with an emphasis on outcome- and patient-centered measures. These focus areas drive measure development, selection, and implementation activities. CMS also sets priorities based on inputs from the National Impact Assessment reports. Although the current CMS measurement programs are setting-specific, there is an increasing need to move toward a more patient-centric approach that spans the continuum of care. The Meaningful Measures Initiative helps to identify and close important gap areas of measures, align measures across the continuum of care and across payers, and spur innovation in new types of measures such as patient-reported measures and electronic measures. PROMs and measures using patient-generated data are needed and highly prioritized in CMS programs.

With the implementation of many quality initiatives, quality measures are proliferating. While measurement gaps still exist, significant progress has been made. With the NQF comprehensive evaluation process, there has been substantial work done to identify “best in class measures” and to harmonize related and competing measures. The pre-rulemaking process required under Section 3014 of the ACA has instituted the MAP discussion and review process in areas such as safety, care coordination, cardiovascular conditions, diabetes, and dual-eligible beneficiaries. The 2015 IOM Vital Signs report and the 2018 Impact Report, “Findings and Actions to Consider” will further the momentum toward “measures that matter.” Future editions of the Blueprint will incorporate these findings and actions into the topics and processes documented.

2.1 MEASURE PRIORITIZATION

Figure 19 outlines how CMS priority planning informs quality measurement through measure selection, implementation, and maintenance activities. Section 3014 of the ACA, which created sections 1890A and 1890(b)(7)(B) of the Social Security Act (SSA), requires HHS to establish a federal pre-rulemaking process for the selection of quality and efficiency measures for use in certain Medicare programs. To comply with the statutory requirement, HHS annually posts the list of measures to be considered for inclusion in Medicare programs. The MUC list is made available to the public no later than December 1 of every year.
Figure 19. CMS Priorities Planning and Measure Selection

Around the second quarter of each fiscal year, through a call for quality and efficiency measures, CMS begins the annual pre-rulemaking cycle of collecting and compiling the MUC list. Early in the year, stakeholders are invited to submit proposed quality and efficiency measures. Stakeholders submitting measures include other federal HHS agencies, organizations contracted with these federal agencies, and healthcare advocacy groups.

Following submission, the pre-rulemaking process includes providing the opportunity for multi-stakeholder groups to offer input to HHS on the selection of quality and efficiency measures. NQF, the CBE with a current contract under Section 1890 of the SSA, convenes the MAP in December of each year to review and comment on the measures proposed on the annual MUC list. The MAP consists of multiple working groups such as clinicians, post-acute care/long-term care, and hospitals. Annually, the MAP working groups and the Coordinating Committee meet to provide program-specific recommendations to HHS by February 1. Measure developers are strongly encouraged to attend the MAP meetings. Participation by measure developers leads to the MAP making decisions using complete and accurate information on the measures on the MUC list. United States Code provides general requirements for informal rulemaking.

2.2 CMS Measure Planning Inputs

2.2.1 Meaningful Measures Framework

The Meaningful Measures Framework sets a course for improving the quality of health and healthcare for all Americans. It serves as a framework for healthcare stakeholders across the country—patients; providers; employers; health insurance companies; academic researchers; and local, state, and federal governments—that help prioritize quality measurement efforts.

2.2.2 Legislative Mandates

Most CMS quality reporting and incentive programs are born out of legislation, which in turn amend the SSA. MACRA, ACA, and the American Recovery and Reinvestment Act of 2009 (ARRA), including the Health Information Technology for Economic and Clinical Health (HITECH) Act, have the largest influence on CMS’s quality measurement priorities, which have led to the broad payment reform and quality-based payment models.
Table 2 provides the legislation and SSA section for CMS quality programs for acute care. Table 3 provides the legislation and SSA section for CMS quality programs for ambulatory care. Table 4 provides the legislation and SSA section for CMS quality programs for post-acute care. Table 5 provides the legislation and SSA section for CMS quality programs across programs.

**Table 2. Legislation Initiating CMS Quality Programs for Acute Care**

<table>
<thead>
<tr>
<th>Quality Program</th>
<th>Legislation Initiating the Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambulatory Surgical Center Quality Reporting (ASCQR)</td>
<td>Section 109(B) of the <a href="http://www.gpo.gov/fdsys/pkg/PLAW-109publ2985/pdf/PLAW-109publ2985.pdf">Tax Relief and Health Care Act of 2006</a> (p. 2985) amended <a href="http://www.gpo.gov/fdsys/pkg/SSA-2013/pdf/SSA-2013.pdf">Section 1833(i)</a> of the SSA to authorize, but does not require, the Secretary to implement the revised Ambulatory Surgical Center (ASC) payment system</td>
</tr>
<tr>
<td>Hospital Inpatient Quality Reporting (Hospital IQR)</td>
<td>Mandated in Section 501(b) of <a href="http://www.gpo.gov/fdsys/pkg/PLAW-109publ2289-2290/pdf/PLAW-109publ2289-2290.pdf">Medicare Prescription Drug, Improvement, and Modernization Act of 2003</a> (pp. 2289-2290)</td>
</tr>
<tr>
<td>Hospital Readmissions Reduction Program (HRRP)</td>
<td>Section 3025 of the <a href="http://www.gpo.gov/fdsys/pkg/PLAW-109publ353/pdf/PLAW-109publ353.pdf">ACA</a> (pp. 408-412), which amended Section 1886(q) of the SSA establishing the program</td>
</tr>
<tr>
<td>Hospital Value-Based Purchasing (HVBP)</td>
<td>Section 3001(a) of the <a href="http://www.gpo.gov/fdsys/pkg/PLAW-109publ353/pdf/PLAW-109publ353.pdf">ACA</a> (pp. 408-412) and amended <a href="http://www.gpo.gov/fdsys/pkg/SSA-2013/pdf/SSA-2013.pdf">Section 1886(o)</a> of the SSA establishing the program</td>
</tr>
<tr>
<td>Inpatient Psychiatric Facility Quality Reporting (IPFQR)</td>
<td>Section 3401(f) and Section 10322(a) of the <a href="http://www.gpo.gov/fdsys/pkg/PLAW-109publ952/pdf/PLAW-109publ952.pdf">ACA</a> (pp. 483 and 952) and amended <a href="http://www.gpo.gov/fdsys/pkg/SSA-2013/pdf/SSA-2013.pdf">Section 1886(s)</a> of the SSA establishing the program</td>
</tr>
<tr>
<td>Promoting Interoperability (Formerly the Medicaid and Medicaid EHR Incentive Program for Eligible Hospitals and Critical Access Hospitals)</td>
<td>Title XIII (aka HITECH Act) of the <a href="http://www.gpo.gov/fdsys/pkg/PLAW-109publ226-279/pdf/PLAW-109publ226-279.pdf">ARRA of 2009</a> (pp. 226-279) amended Titles XVIII and XIX of the SSA establishing the program</td>
</tr>
<tr>
<td>Prospective Payment System-Exempt Cancer Hospital Quality Reporting</td>
<td>Mandated by Section 3005 of the <a href="http://www.gpo.gov/fdsys/pkg/PLAW-109publ371-372/pdf/PLAW-109publ371-372.pdf">ACA</a> (pp. 371-372) and added <a href="http://www.gpo.gov/fdsys/pkg/SSA-2013/pdf/SSA-2013.pdf">Section 1866(k)</a> of the SSA developing and mandating the program</td>
</tr>
</tbody>
</table>

**Table 3. Legislation Initiating CMS Quality Programs for Ambulatory Care**

<table>
<thead>
<tr>
<th>Quality Program</th>
<th>Legislation Initiating the Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicaid Quality Reporting</td>
<td>Section 401 [Children’s Health Insurance Program Reauthorization Act of 2009 (CHIPRA)] (pp. 72-82) added <a href="http://www.gpo.gov/fdsys/pkg/SSA-2013/pdf/SSA-2013.pdf">Section 1139A</a> to the SSA requiring identification and publishing a core measure set of children’s healthcare quality measures for voluntary use by Medicaid and CHIP programs</td>
</tr>
<tr>
<td>Promoting Interoperability (Formerly the Medicaid EHR Incentive Program for Eligible Professionals)</td>
<td>Title XIII (aka HITECH Act) of the <a href="http://www.gpo.gov/fdsys/pkg/PLAW-109publ226-279/pdf/PLAW-109publ226-279.pdf">ARRA of 2009</a> (pp. 226-279) amended Titles XVIII and XIX of the SSA</td>
</tr>
<tr>
<td>Medicare Part C Star Rating</td>
<td>Section 3201 of the <a href="http://www.gpo.gov/fdsys/pkg/PLAW-109publ567-568/pdf/PLAW-109publ567-568.pdf">ACA</a> (pp. 567-568) establishes the star ratings as basis of Quality Bonus Payments</td>
</tr>
<tr>
<td>Medicare Part D Star Rating</td>
<td>No legislative requirement; Medicare Part D Star rating began in 2008</td>
</tr>
<tr>
<td>Merit-based Incentive Payment System (MIPS)</td>
<td>Section 101(c) of the <a href="http://www.gpo.gov/fdsys/pkg/PLAW-114publ1848/pdf/PLAW-114publ1848.pdf">MACRA of 2015</a> (p. 92) added <a href="http://www.gpo.gov/fdsys/pkg/SSA-2013/pdf/SSA-2013.pdf">Section 1848(q)</a> of the SSA establishing the program</td>
</tr>
<tr>
<td>Quality Payment Program (QPP)</td>
<td><a href="http://www.gpo.gov/fdsys/pkg/PLAW-114publ1848/pdf/PLAW-114publ1848.pdf">MACRA of 2015</a></td>
</tr>
</tbody>
</table>

**Note:** The legislation and SSA section numbers are cited for reference purposes only and may not be the most current or complete. For the most up-to-date information, please consult the official documents and regulations provided by the Centers for Medicare & Medicaid Services (CMS) and the Social Security Administration (SSA).
Table 4. Legislation Initiating CMS Quality Programs for Post-acute Care

<table>
<thead>
<tr>
<th>Quality Program</th>
<th>Legislation Initiating the Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>End-Stage Renal Disease Quality Incentive Program (ESRD QIP)</td>
<td>Section 153(c) of the Medicare Improvements for Patients and Providers Act of 2008 (p. 2556) to amend Section 1881(h) of the SSA to establish the program</td>
</tr>
<tr>
<td>Home Health Quality Reporting (HHQR)</td>
<td>Section 5201(c)(2) of the Deficit Reduction Act of 2005 (pp. 46-47) added Section 1895(b)(3)(B)(v)(II) of the SSA mandating the reporting of home health quality data</td>
</tr>
<tr>
<td>Hospice Quality Reporting (HQR)</td>
<td>Section 3004 of the ACA (p. 368) amended Section 1814(ii)(5) of the SSA directing the Secretary to establish quality reporting requirements for hospice programs</td>
</tr>
<tr>
<td>Inpatient Rehabilitation Facility (IRF) Quality Reporting (QR)</td>
<td>Section 3004(b) of the ACA (pp. 369-340) amended Section 1886(i)(7) of the SSA directing the Secretary to establish quality reporting requirements for IRFs</td>
</tr>
<tr>
<td>Long-Term Care Hospital Quality Reporting (LTCH QR)</td>
<td>Section 3004(a) of the ACA (pp. 368-369) amended Section 1886(m)(5) of the SSA directing the Secretary to establish quality reporting requirements for long-term care hospitals</td>
</tr>
<tr>
<td>Nursing Home Quality Initiative</td>
<td>No legislative requirement; pilot and program initiation in 2002</td>
</tr>
<tr>
<td>Post-Acute Care Quality Initiatives</td>
<td>Improving Medicare Post-Acute Care Transformation Act of 2014 (the IMPACT Act)</td>
</tr>
<tr>
<td>Program of All-Inclusive Care for the Elderly (PACE)</td>
<td>Section 4801 of the Balanced Budget Act of 1997 (BBA) (pp. 528-538) added Section 1894 to the SSA establishing the PACE program</td>
</tr>
<tr>
<td></td>
<td>Section 4802 of the Balanced Budget Act of 1997 (BBA) (pp. 538-549) added Section 1934 to the SSA, which established PACE as a state option</td>
</tr>
<tr>
<td>Skilled Nursing Facility Quality Reporting Program (SNF QRP)</td>
<td>Improving Medicare Post-Acute Care Transformation Act of 2014 to amend Section 1899 of the SSA developing and requiring submission of quality data (the IMPACT Act)</td>
</tr>
<tr>
<td>Skilled Nursing Facility Value-Based Purchasing Program (SNF VBP)</td>
<td>Section 215 of the Protecting Access to Medicare Act of 2014 (pp. 1048-1053) added Sections 1888(g) and (h) to SSA establishing the program</td>
</tr>
</tbody>
</table>

Table 5. Legislation Initiating CMS Quality Programs Across Settings

<table>
<thead>
<tr>
<th>Quality Program</th>
<th>Legislation Initiating the Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicare Shared Savings (MSS)</td>
<td>Section 3022 of the ACA (p. 395) amended Section 1899 of SSA establishing the program</td>
</tr>
<tr>
<td>Health Insurance Exchange Quality Rating System (QRS)</td>
<td>Section 1311(c)(3) of the ACA (p.175)</td>
</tr>
</tbody>
</table>

2.2.3 Other Legislation Influencing CMS Quality Programs

The **21st Century Cures Act** (Public Law [P.L.] 114-255), commonly referred to as the Cures Act, was enacted in 2016 and aligns with many of CMS’s quality measurement priorities. The Cures Act serves largely to increase choice, access, and quality of care for patients. The Cures Act mandates efforts to reduce administrative burden on healthcare providers and calls for the alignment and simplification of quality measures across federal programs and other payers. The Cures Act specifically mandates the reduction of regulatory or administrative burden related to the use of EHRs and calls for the identification of priority uses for the data arising from the implementation of value-based payment programs. The Cures Act aims to increase data-sharing and interoperability via the expanded use of health IT and the creation of partnerships between health information exchange organizations and healthcare providers. The increase in data-sharing is also meant to increase patient access to EHRs and thus improve patient care.

**MACRA of 2015 (P.L. 114-10)** defined five quality domains, including: (i) clinical care; (ii) safety; (iii) care coordination; (iv) patient and caregiver experience; and (v) population health and prevention. In response to this Act and the laws it amends, CMS conducts measure priorities planning across these domains and emphasizes: (a) outcome measures, including PROM and functional status measures;
(b) patient experience measures; (c) care coordination measures; and (d) measures of appropriate use of services, including measures of overuse.

With MACRA (P.L. 114-10) in 2015, Congress mandated that several quality reporting and incentive programs phase out in 2018, while the Merit-based Incentive Payment System (MIPS) would continue well beyond 2019. Under MACRA, CMS has developed performance assessment methods using composite scoring for the determination of MIPS adjustment factors for all MIPS eligible clinicians. This effort is supported by the funding provided under the ACA for the creation of a wide array of quality measures, including outcome measures and measures for settings that are new to quality reporting such as inpatient rehabilitation facilities, hospices, long-term care hospitals, inpatient psychiatric facilities, and Prospective Payment System-exempt cancer hospitals. In addition, under MACRA and ACA, Medicaid and other HHS programs will continue to develop and implement quality measures. MACRA also supports the gains made under ARRA. ARRA launched a period of significant funding for the development of standards for EHRs and the widespread adoption of certified EHR technology (CEHRT) across providers. MACRA continues this support with a mandate for widespread interoperability among these systems with requirements for CMS to develop metrics for successful interoperability, as well as incentives and payment penalties to encourage rapid achievement of that goal.

The Improving Medicare Post-Acute Care Transformation Act of 2014, commonly referred to as the IMPACT Act, required the submission of standardized data by long-term care hospitals (LTCHs), skilled nursing facilities (SNFs), home health agencies (HHAs), and inpatient rehabilitation facilities (IRFs), in addition to initiating the SNF Quality Reporting Program. The IMPACT Act required implementation of specified clinical assessment domains using standardized (uniform) data elements to be nested within the assessment instruments currently required for submission by LTCH, IRF, SNF, and HHA providers. The Act further required that CMS develop and implement quality measures from five quality measure domains using standardized assessment data. Also required was the development and reporting of measures pertaining to resource use, hospitalization, and discharge to the community. Through the use of standardized quality measures and standardized data, the goal is to enable interoperability and access to longitudinal information for such providers to facilitate coordinated care, improved outcomes, and overall quality comparisons. The DEL is an output of the IMPACT Act and is the centralized resource for CMS assessment instrument data elements (e.g., questions and responses) and their associated health IT standards.

ARRA was an economic stimulus package that affected many sectors (e.g., federal tax relief, expansion of unemployment benefits, education, infrastructure, and healthcare). Title XIII of ARRA is the HITECH Act, which initiated the EHR Incentive Programs, now called Promoting Interoperability. The primary goal of the HITECH Act was to promote and expand the meaningful use of health IT, but it also included funding for other things such as workforce education and health information exchanges. The EHR Incentive Programs (Promoting Interoperability) provide payments to eligible professionals and eligible hospitals if they demonstrate meaningful use of CEHRT and penalize those who do not. The belief is that EHR use will improve the quality, safety, and efficiency of healthcare. The Medicare EHR Incentive Program for Eligible Professionals was rolled into MIPS.

2.2.4 Quality Measure Development Plan (MDP)

On May 2, 2016, CMS finalized the Quality MDP, mandated under the MACRA, to support the new MIPS and advanced APMs. MACRA supports a transition to value-based payment incentives for physicians and other clinicians to be based on quality, rather than quantity, of care.
The CMS Quality MDP is an essential resource in this transition, as it provides the foundation and a strategic framework for building and implementing a measure portfolio to support the quality payment programs under MACRA. The CMS Quality MDP highlights known clinical and specialty measurement and performance gaps and recommends prioritized approaches to close these gaps through the development, adoption, and refinement of quality measures.

Through the application of the principles included in the MDP and the quality measure development funded by MACRA, CMS is committed to increased transparency and partnerships with persons and families, clinicians, and professional societies to develop measures that are meaningful, applicable, and useful across payers and healthcare settings. These quality measures are essential to address critical performance gaps, facilitate alignment across settings and payers, and promote efficient data collection. CMS intends for the MDP and related quality measures to be key levers of delivery system reform, promoting movement toward paying for value rather than volume and improved national healthcare delivery.

For a copy of the MDP, MDP Annual Reports, and MDP Environmental Scan and Gap Analysis Reports and more information, view the CMS Quality MDP and Annual Report website.

2.2.5 Patients, Public, and Other Stakeholders

CMS conducts its measurement activities in a transparent manner. The information gathered through various methods described in Section 2, Chapter 2, Measure Conceptualization informs HHS and CMS about future measurement needs. Additionally, Section 101(f) of MACRA requires that CMS solicit, accept, and respond to input from stakeholders, including physician specialty societies, applicable practitioner organizations, and other stakeholders for episode groups (i.e., care episode groups and patient-condition groups). Care episode groups include those patients whose care included similar treatments and procedures, taking into consideration patients’ clinical diagnoses and problems during the care episode, care setting, and level of acuity, and principal procedures or services furnished. Patient-condition groups include those patients with similar conditions, taking into consideration patients’ medical and surgical histories, comorbid conditions, overall health status, and eligibility or dual-eligibility status. Patients and families are extremely important stakeholders in the quality measurement enterprise, and CMS is committed to gathering their input during priorities planning. More detail about ways in which the patient’s voice can be heard is found in Section 2, Chapter 2, Measure Conceptualization.

2.2.6 Impact Assessment and Other Reports

Once a measure is in use, it requires ongoing monitoring and maintenance in addition to formal periodic reevaluations to determine whether it remains appropriate for continued use. The measure developer will conduct measure trend analyses, evaluate barriers, and identify unintended consequences associated with specific measures in their purview.

Measure maintenance reports yield information that CMS leadership may find valuable for setting priorities. This information may include barriers to implementation of measures, unintended consequences, lessons learned, measure impact on providers, care disparities, and gaps in care. Measure maintenance includes assessment of the performance of the measure, including trend analyses, and comparison to the initial projected performance, found in the Business Case. CMS uses this input to decide whether to remove, retire, modify, or retain measures in use.

In addition to measure maintenance, CMS conducts various evaluations and assessments of its measures and programs to determine the effectiveness of its various programs. Many of these programs use
quality measures, and these analyses evaluate the usefulness of the measures as they are used in the programs.

The triennial National Impact Assessment of the CMS Quality Measures Reports required by Section 3014 of the ACA aims to contribute to the overall, cross-cutting evaluation of CMS quality measures. The analyses in these reports are not intended to replace or duplicate program-specific assessments, nor are they intended to replace the analyses individual measures must undergo as part of ongoing measure maintenance. Rather, they are intended to help the federal government and the public understand the overall impact of its investments in quality measurement and reflect on future needs.

Several organizations analyze the performance of CMS-implemented quality measures, and these studies provide valuable input into CMS measure priority planning. These reports and studies may provide information on disparities, gaps in care, and other findings related to measurement policies. Some of these entities and their associated reports include:

- Medicare Payment Advisory Committee (MedPAC) and Medicaid and CHIP Payment and Access Commission (MACPAC) quality reports
- Agency for Healthcare Research and Quality (AHRQ) National Healthcare Quality and Disparities Reports
- CMS Office of Minority Health
- CMS Office of Enterprise Data and Analytics – Chronic Conditions among Medicare Beneficiaries
- Universities, researchers, and healthcare facilities, including their journal articles and conference presentations.

Together, these inputs influence CMS planning for future measure development, implementation, and maintenance activities.

### 2.3 Role of the Measure Developer in Priorities Planning

The measure developer plays a key role in supporting CMS’s priorities planning. It is important for measure developers to be knowledgeable about how CMS plans its measure development and maintenance activities so that appropriate measures are developed and maintained based on the priorities established by CMS, and measure harmonization and alignment are achieved to the greatest degree possible.

Measure developers are expected to be knowledgeable of inputs into the measurement priority-setting activities. At a minimum, measure developers should follow the Blueprint processes for soliciting public and stakeholders’ input into the MUD. Section 2, Chapter 2, Measure Conceptualization provides further details. Measure developers are responsible for monitoring all feedback and input provided on their measures. It is their responsibility to report this information to their COR, who will ensure that CMS staff members working on measure priorities planning receive this information.

During measure development, it is important that measure developers conduct a thorough environmental scan and are knowledgeable about measures that may be like those they are seeking to develop. To the extent possible, measure developers are to avoid developing competing measures—those that essentially address the same concepts for the target process, condition, event, or outcome, and the same target patient population. Competing measures are conceptually similar, but their technical specifications may differ.

Measure developers should consider HHS and CMS goals and priorities when identifying a list of potential measures for pre-rulemaking, rulemaking, and eventual program adoption. Measure
developers may be required to help the COR develop the MUC list, which may include providing CMS with the justification and assessment of the potential impact of the new measure developed, providing the performance trends and evaluation of an implemented measure, and helping CMS evaluate how the measure developer’s measures address the quality priorities and Meaningful Measures goals. This information can be useful to the MAP in evaluating the MUC.

CMS often contracts with organizations to support the rulemaking process. While this work may be performed under a support contract separate from the measure development contract, the contractor that developed or is maintaining the measure may also be asked to provide information. During the proposed phase of rulemaking, the measure developers may be asked to monitor the public comments that are submitted on the measures and begin drafting responses for CMS. For the final rule, measure developers may also be asked to provide additional information about their measures. Measure developers are also strongly encouraged to attend the MAP meetings.

Measure developers must convey to their COR the lessons learned from the measure rollout, implementation, and ongoing monitoring of the measures. During measure maintenance, it is important that the measure developers analyze the measure performance trends, including feedback through helpdesks and trainings, to determine if the measure undergoing reevaluation is still the best or most relevant measure and if there are unintended consequences that need to be addressed.

2.4 ROLE OF THE MEASURES MANAGER IN PRIORITIES PLANNING

The Measures Manager’s role supporting CMS with setting measure priorities is to research and consider a wide variety of measure-related information and materials to help CMS prioritize and coordinate measure development activities, which may include:

- Review HHS and CMS strategic plans, goals, and initiatives.
- Monitor the progress of CMS measure development and maintenance projects against the quality priorities and identify Meaningful Measurement areas in need of measure development.
- Produce harmonization and alignment tools and reports.
- Develop white papers to help CMS formulate measurement policies.
- Research legislative mandates, proposed and final rules, and priorities of key external stakeholders.
- Support various HHS, CMS, and interagency working groups that focus on coordination of measure development, measure alignment, and harmonization.
- Support CMS’s collection of measures for and management of the MUC list for pre-rulemaking.
- Maintain a CMS inventory of measures for policy and program use. The CMS Measures Inventory is updated quarterly and includes a wide array of measures.
For priorities planning, based on status or year of anticipated use, the measures are separated into 11 categories:

- **Considered** – A measure that has been submitted to the pre-rulemaking process and has been accepted for consideration by a CMS program, has been cleared through the HHS clearance process, and published on an annual MUC list.
- **Declined** – A measure that was submitted but not accepted by a CMS program through the pre-rulemaking process and, therefore, will not be published on the MUC list.
- **Development** – A measure that is currently being developed for eventual use in a CMS program or initiative.
- **Discontinued** – A measure that is no longer being developed for use in a CMS program or initiative.
- **Finalized** – A measure that was proposed for use in a CMS program that has been finalized per Federal Rule. The measure will be implemented within a designated time frame noted within the Final Rule.
- **Implemented** – A measure that is currently used within a CMS program to impact incentive or reimbursement payments, or is used to improve performance.
- **Proposed** – A measure that is introduced in a published Proposed Rule for possible use within a CMS Program if the measure is eventually finalized in the federal rulemaking process.
- **Reference** – A placeholder for additional information about the use of the measure in the program.
- **Removed** – A measure that has been removed from a CMS program via Federal Rule. The measure is no longer implemented, meaning it no longer impacts payments or performance scoring.
- **Rescinded** – A measure that was proposed in a published Proposed Rule, but was not finalized for use within a CMS Program.
- **Suspended** – A finalized or implemented measure, which has been suspended from current use within a program. The measure is no longer implemented, meaning it will not impact payment or performance scoring.
3 ROLES IN MEASURE DEVELOPMENT

Multiple stakeholders are involved in measure development. Throughout the Blueprint there is an emphasis on stakeholder engagement. These stakeholders include Technical Expert Panels, persons and families, clinicians, and the public. There are committees (e.g., the Quality Measures Technical Forum (QMTF) and the eCQM Governance Group, that provide guidance as to how measures are developed along with the processes and procedures for different phases of the measure lifecycle and/or types of measures. Measure developers have distinct roles and responsibilities throughout the lifecycle of a measure. As the measure steward, so does CMS and its CORs. Developers perform editing functions, whereas stewards approve the work of the developers and submit measures for publication. CMS functions as a unique steward within measure development because they are contracting the development of the measures.

3.1 MEASURE DEVELOPERS

Measure developers create, edit, and submit measures to a designated steward, in this case, CMS. Developers submit measures to their assigned stewards for approval. It is also the responsibility of the developer to circulate their measure content for feedback and to collaborate on potential measure changes suggested by other authors or other entities. Refer to Section 3, Chapter 2.3 Role of the Measure Developer in Priorities Planning.

3.2 MEASURE STEWARDS

Stewards have permission to approve, reject, and publish measures that their assigned developer groups create and submit. Stewards provide overall coordination and management of the measures created by developers under a specific program or for a specific purpose. Stewards are responsible for approving measure content. Stewards may withdraw measures from approval.

3.3 GUIDING COMMITTEES

These committees guide how measures are developed and the processes and procedures used throughout the measure lifecycle. They work to coordinate, align, and harmonize measures across CMS programs, while ensuring best practices in measure development and measurement policies.

3.3.1 Quality Measures Technical Forum

The mission of the QMTF is to align quality measures across all CMS programs. The QMTF accomplishes this mission by coordinating and advising CMS components on policy and research used to inform recommendations on the prioritization of measures, development of core sets of measures, direction for measure development, and measurement policies.

The QMTF Guiding Principles include:

- Alignment with the Meaningful Measures Initiative
- Promoting transparency
- Continuous improvement and coordination of measures.
The QMTF fosters alignment of measurement activities across CMS by:

- Gathering pertinent data or information from across CMS.
- Analyzing issues pertaining to the development and/or implementation of quality measures and making formal recommendations to CMS components.
- Reviewing and approving measure concepts, measures, and measure sets for implementation into quality reporting, public reporting, and value-based purchasing programs.
- Providing a forum for CMS components to routinely share information about their plans for the development of new quality measures and/or re-engineering of existing measures.
- Sharing information on new and changing implementation of quality measures.
- Identifying opportunities for continuous quality improvement of CMS coordination efforts for the development and deployment of measures across all CMS programs.
- Providing feedback and guidance to CMS components regarding plans for measure implementation, measure changes, and projects such as measure development.

The QMTF continues to align and coordinate the development, maintenance, and implementation of quality measures across CMS programs.

### 3.3.2 eCQM Governance Group

The mission of the eCQM Governance Group is to ensure broad stakeholder collaboration, coordination, and communication of key eCQM development, implementation, and reporting decisions for stakeholders involved in eCQM development for HHS quality reporting programs. The group achieves this mission through a consensus-based, decision-making approach and through dissemination of information to stakeholders across HHS divisions and within the healthcare and health IT industry.

The eCQM Governance Group consists of representatives from federal health agencies and their respective contractors in eCQM development, implementation, reporting, and/or testing. It is a decision-making group that proposes, reviews, and approves decisions regarding all technical aspects of eCQMs, including but not limited to, harmonization, testing, certification, publication, implementation, reporting standards, specifications, and maintenance. The group gathers and disseminates information, fosters coordination and alignment across programs and functions, and provides a mechanism to generate efficiencies and lessons learned to continuously improve the measure and quality program process.

The eCQM Governance Group:

- Provides a forum for HHS divisions, federal partners, and their respective contractors to share information needed to make key decisions.
- Requests additional information and/or participants necessary to make decisions on eCQM topics, either proactively or in response to feedback.
- Communicates final decisions to a broad range of stakeholders.
- Identifies opportunities for continued improvement, coordination, and automation in the development process surrounding eCQMs.
- Works collectively with all stakeholders to facilitate revisions and continued improvements in eCQM tools and requirements, such as the MAT, VSAC, Bonnie eCQM testing tool, Cypress certification tool, eCQI Resource Center, and United States Healthcare Information Knowledgebase (USHIK).
• Coordinates significant activities across various stakeholders regarding measure development, testing, certification, communication, and implementation/feedback processes for eCQMs, both during initial development and the eCQM Annual Updates that affect the greater eCQM community.
• Achieves consensus on the use of standard code sets and versions across eCQMs, naming conventions (e.g., identifiers and filenames for eCQMs), related artifacts, and publications.
4 MEASURE CLASSIFICATION

Measures may be classified according to a variety of schemes, including measurement domain,16 by the Meaningful Measurement area addressed, CMS pre-rulemaking, types, or NQF submission types. Measure classification types and names may be dictated by legislation, consensus, or other methodology, and the types and names can and do change over time. As such, complete alignment in types and names would be difficult, if not impossible, to attain. Table 6 provides a crosswalk of measure types and definitions from the CMIT User Guide, the 2019 MUC User Guide, and the 2018 NQF Measure Evaluation Guidance. Elements of different classification schemes with examples are provided in Tables 7, 8, 9, and 10.

Table 6. Crosswalk of Measure Types

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>(No definition in CMIT User Guide, listed as a measure type in CMIT, but no measures assigned to the category.)</td>
<td>(No definition in the 2019 MUC User Guide.)</td>
<td>Assess the ability to obtain needed healthcare services in a timely manner, including the perceptions and experiences of people regarding their ease of reaching health services or health facilities in terms of proximity, location, time, and ease of approach. Examples may include, but are not limited to, measures that address the timeliness of response or services, time until next available appointment, and availability of services within a community.</td>
</tr>
<tr>
<td>Communication and Care Coordination</td>
<td>(No definition in CMIT User Guide, listed as a measure type in CMIT, but no measures assigned to the category.)</td>
<td>(No definition in the 2019 MUC User Guide.)</td>
<td>(No definition in the 2018 NQF Measure Evaluation and Guidance).</td>
</tr>
</tbody>
</table>
| Composite                    | A measure that contains two or more individual measures, resulting in a single measure and a single score. Composite measures may be composed of one or more process measures and/or one or more outcome measures. | A combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score. | Combine two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score. For the purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composite performance measures:
  - Measures with two or more individual performance measure scores combined into one score for an accountable entity.
  - Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity, including all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient). |

---

16 Note that CMS and other HHS agencies define and use the term “domain” and classify measures differently from one another; within this Blueprint, the term “domain” is defined differently in different contexts, depending on the relevant agency within the discussion and different measure classification types.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost/Resource</td>
<td>Broadly applicable and comparable measures of health service counts. A resource</td>
<td>(No definition in the 2019 MUC User Guide.)</td>
<td>Broadly applicable and comparable measures of health services counts (in terms of units or dollars) that</td>
</tr>
<tr>
<td></td>
<td>measure counts the frequency of defined health system resources; some may further</td>
<td></td>
<td>are applied to a population or event (broadly defined to include diagnoses, procedures, or encounters). A</td>
</tr>
<tr>
<td></td>
<td>apply a dollar amount to each unit of resource. (No measures assigned to this</td>
<td></td>
<td>resource use measure counts the frequency of use of defined health system resources; some may further</td>
</tr>
<tr>
<td></td>
<td>category in CMIT.)</td>
<td></td>
<td>apply a dollar amount (e.g., allowable charges, paid amounts, or standardized prices) to each unit of</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>resource use.</td>
</tr>
<tr>
<td>Cost/Resource Use</td>
<td>(No definition in CMIT User Guide, however 97 measures are assigned to this</td>
<td>A count of the frequency of units of defined health system services or resources; some</td>
<td>(NQF previously used this category name for the current Cost/Resource).</td>
</tr>
<tr>
<td></td>
<td>category.)</td>
<td>may further apply a dollar amount (e.g., allowable charges, paid amounts, or standardized prices) to</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>each unit of resource use (i.e., monetize the health service or resource use units).</td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td>Refers to a measure concerning the cost of care associated with a specified level</td>
<td>The cost of care associated with a specified level of health outcomes.</td>
<td>Combine the concepts of resource use and quality. NQF has defined efficiency broadly as the resource use</td>
</tr>
<tr>
<td></td>
<td>of health outcome.</td>
<td></td>
<td>(or cost) associated with a specific level of performance with respect to the other five Institute of</td>
</tr>
<tr>
<td>Instrument-based Performance</td>
<td>(No definition in the CMIT User Guide nor is it a CMIT Measure Type.)</td>
<td></td>
<td>Medicine (IOM) aims of quality: safety, timeliness, effectiveness, equity, and patient centeredness.</td>
</tr>
<tr>
<td>Measure</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Cost/Resource**
  - Broadly applicable and comparable measures of health service counts. A resource measure counts the frequency of defined health system resources; some may further apply a dollar amount to each unit of resource. (No measures assigned to this category in CMIT.)

- **Cost/Resource Use**
  - (No definition in CMIT User Guide, however 97 measures are assigned to this category.)

- **Efficiency**
  - Refers to a measure concerning the cost of care associated with a specified level of health outcome.

- **Instrument-based Performance Measure**
  - (No definition in the CMIT User Guide nor is it a CMIT Measure Type.)

- **2019 MUC User Guide**
  - (No definition in the 2019 MUC User Guide.)

- **2018 NQF Measure Evaluation and Guidance**
  - Use data derived from instruments. "Instruments" is a generic term that researchers use for a measurement device (e.g., survey, test, questionnaire, scale). Instruments are used for consistently obtaining (or presenting) data from respondents. The data derived from an instrument may include ratings or ranking output that is included in the calculation of a performance measure. Instruments may be used to collect information from a variety of individuals; examples include patients, observers (e.g., family, or other caregivers takers), or clinicians. Data from instruments can be used in the calculation of structure, process, or outcome performance measures. Instruments specific to PROs may be referenced as PROMs.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate Outcome</td>
<td>(No definition in the CMIT User Guide. However, 71 CMIT measures are assigned this measure type.)</td>
<td>Refers to a change produced by a healthcare intervention that leads to a long-term outcome (e.g., a reduction in blood pressure is an intermediate outcome that leads to a reduction in the risk of long-term outcomes such as cardiac infarction or stroke).</td>
<td>Intermediate clinical outcome is a change in physiologic state that leads to a long-term health outcome.</td>
</tr>
<tr>
<td>Not Specified</td>
<td>(No definition in the CMIT User Guide. However, 18 CMIT measures are assigned this measure type.)</td>
<td>(No definition in the 2019 MUC User Guide.)</td>
<td>(No definition in the 2018 NQF Measure Evaluation and Guidance).</td>
</tr>
<tr>
<td>Outcome</td>
<td>A measure that assesses the results of healthcare that are experienced by patients: clinical events, recovery and health status, experiences in the health system, and efficiency/cost.</td>
<td>Assess the results of healthcare that are experienced by patients. They include endpoints like well-being, ability to perform daily activities, or even death.</td>
<td>An outcome of care is the health status of a patient (or change in health status) resulting from healthcare—desirable or adverse.</td>
</tr>
<tr>
<td>Patient Engagement/Experience</td>
<td>(No definition in the CMIT User Guide. However, 31 measures are assigned this measure type.)</td>
<td>These measures use direct feedback from patients and their caregivers about the experience of receiving care. The information is usually collected through surveys.</td>
<td>(No definition in the 2018 NQF Measure Evaluation and Guidance).</td>
</tr>
<tr>
<td>Patient Perspective</td>
<td>(No definition in the CMIT User Guide. However, it is a measure type listed in CMIT, but no measures are assigned this measure type.)</td>
<td>(No definition in the 2019 MUC User Guide.)</td>
<td>(No definition in the 2018 NQF Measure Evaluation and Guidance).</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Patient-Reported Outcome (PRO)</td>
<td>(No definition in the CMIT User Guide. However, 86 measures are assigned this measure type in CMIT.)</td>
<td>Refers to a measure of a patient’s feelings or what they are able to do as they are dealing with diseases or conditions. These types may include PROM, which is an instrument, scale, or single-item measure that gathers the information directly from the patient and PRO-Based Performance Measure (PRO-PM), which is a way to aggregate the information that has been shared by the patient and collected into a reliable, valid measure of health system performance.</td>
<td>(No definition in the 2018 NQF Measure Evaluation and Guidance).</td>
</tr>
<tr>
<td>Patient-Reported Outcome Measure (PROM)</td>
<td>(No definition in the CMIT User Guide and it is not a CMIT measure type.)</td>
<td>An instrument, scale, or single-item measure that gathers the information directly from the patient.</td>
<td>(No definition in the 2018 NQF Measure Evaluation and Guidance).</td>
</tr>
<tr>
<td>Population Health Quality Measure</td>
<td>(No definition in the CMIT User Guide and it is not a CMIT measure type.)</td>
<td>A mechanism to assess the degree to which public health providers or the health system serving a population effectively and safely delivers health services that are appropriate for the population in the optimal time period.</td>
<td>(No definition in the 2018 NQF Measure Evaluation and Guidance).</td>
</tr>
<tr>
<td>PRO-Based Performance Measure</td>
<td>(No definition in the CMIT User Guide and it is not a CMIT measure type.)</td>
<td>A way to aggregate the information that has been shared by the patient and collected into a reliable, valid measure of health system performance.</td>
<td>(No definition in the 2018 NQF Measure Evaluation and Guidance).</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Process</td>
<td>A measure that focuses on steps that should be followed to provide good care. There should be a scientific basis for believing that the process, when executed well, will increase the probability of achieving a desired outcome.</td>
<td>A healthcare service provided to, or on behalf of, a patient – may include, but is not limited to, measures that may address adherence to recommendations for clinical practice based on evidence or consensus. Separate definition – Process measures: Assess steps that should be followed to provide good care.</td>
<td>Process of care is a healthcare-related activity performed for, on behalf of, or by a patient.</td>
</tr>
<tr>
<td>Process:</td>
<td>(No definition in the CMIT User Guide and it is not a CMIT measure type.)</td>
<td>(No definition in the 2019 MUC User Guide.)</td>
<td>Appropriate Use is a type of process measure that has been used to evaluate procedures and medical technologies. Appropriate use measures are neither cost/resource use measures nor efficiency measures.</td>
</tr>
<tr>
<td>Appropriate Use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structure/Structural</td>
<td>A structural measure is one that assesses features of a healthcare organization or clinician relevant to its capacity to provide healthcare.</td>
<td>Assess healthcare infrastructure.</td>
<td>Structure of care is a feature of a healthcare organization or clinician related to its capacity to provide high-quality healthcare.</td>
</tr>
</tbody>
</table>
Table 7. CMIT

<table>
<thead>
<tr>
<th>Measurement Type</th>
<th>Definition (From the CMIT User Guide)</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite</td>
<td>A measure that contains two or more individual measures, resulting in a single measure and a single score. Composite measures may be composed of one or more process measures and/or one or more outcome measures.</td>
<td>Patient Safety for Selected Indicators (PSI90 - Composite)</td>
</tr>
<tr>
<td>Cost/Resource</td>
<td>Broadly applicable and comparable measures of health service counts. A resource measure counts the frequency of defined health system resources; some may further apply a dollar amount to each unit of resource.</td>
<td>Hospital-level, risk-standardized payment associated with a 90-day episode of care for elective primary total hip and/or total knee arthroplasty (THA/TKA)</td>
</tr>
<tr>
<td>Efficiency</td>
<td>A measure concerning the cost of care associated with a specified level of health outcome.</td>
<td>Adult Sinusitis: Computerized Tomography (CT) for Acute Sinusitis (Overuse)</td>
</tr>
<tr>
<td>Outcome</td>
<td>A measure that assesses the results of healthcare that are experienced by patients: clinical events, recovery and health status, experiences in the health system, and efficiency/cost.</td>
<td>Acute Care Hospitalization During the First 60 Days of Home Health</td>
</tr>
<tr>
<td>Patient-Reported Outcome (PRO)</td>
<td>A measure that focuses on a patient’s report concerning observations of and participation in healthcare.</td>
<td>Average Change in Leg Pain Following Lumbar Discectomy and/or Laminotomy</td>
</tr>
<tr>
<td>Process</td>
<td>A measure that focuses on steps that should be followed to provide good care. There should be a scientific basis for believing that the process, when executed well, will increase the probability of achieving a desired outcome.</td>
<td>Acute Otitis Externa (AOE): Systemic Antimicrobial Therapy - Avoidance of Inappropriate Use</td>
</tr>
<tr>
<td>Structural</td>
<td>A measure that assesses features of a healthcare organization or clinician relevant to its capacity to provide healthcare.</td>
<td>Melanoma: Continuity of Care - Recall System</td>
</tr>
</tbody>
</table>
### Table 8. Examples of Measures Addressing Each of the Meaningful Measurement Areas

<table>
<thead>
<tr>
<th>CMS Priorities</th>
<th>Meaningful Measurement Areas</th>
<th>Examples of Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strengthen Person &amp; Family Engagement as Partners in their Care</td>
<td>Care is Personalized and Aligned with Patient’s Goals</td>
<td>Care Plan</td>
</tr>
<tr>
<td>Strengthen Person &amp; Family Engagement as Partners in their Care</td>
<td>End of Life Care According to Preferences</td>
<td>Hospice Visits while Death is Imminent</td>
</tr>
<tr>
<td>Strengthen Person &amp; Family Engagement as Partners in their Care</td>
<td>Patient’s Experience of Care</td>
<td>Consumer Assessment of Healthcare Providers and Systems (CAHPS) In-center Hemodialysis Survey</td>
</tr>
<tr>
<td>Strengthen Person &amp; Family Engagement as Partners in their Care</td>
<td>Functional Outcomes</td>
<td>Functional Status Assessment for Total Hip Replacement</td>
</tr>
<tr>
<td>Promote Effective Communication &amp; Coordination of Care</td>
<td>Medication Management</td>
<td>Use of High-Risk Medications in the Elderly</td>
</tr>
<tr>
<td>Promote Effective Communication &amp; Coordination of Care</td>
<td>Admissions and Readmissions to Hospitals</td>
<td>Plan All-cause Readmissions</td>
</tr>
<tr>
<td>Promote Effective Communication &amp; Coordination of Care</td>
<td>Transfer of Health Information Technology and Interoperability</td>
<td>Use of an EHR</td>
</tr>
<tr>
<td>Promote Effective Prevention &amp; Treatment of Chronic Disease</td>
<td>Preventive Care</td>
<td>Timeliness of Prenatal Care</td>
</tr>
<tr>
<td>Promote Effective Prevention &amp; Treatment of Chronic Disease</td>
<td>Management of Chronic Conditions</td>
<td>Osteoporosis Management in Women Who Had a Fracture</td>
</tr>
<tr>
<td>Promote Effective Prevention &amp; Treatment of Chronic Disease</td>
<td>Prevention, Treatment, and Management of Mental Health</td>
<td>Follow-up After Hospitalization for Mental Illness</td>
</tr>
<tr>
<td>Promote Effective Prevention &amp; Treatment of Chronic Disease</td>
<td>Prevention and Treatment of Opioid and Substance Use Disorders</td>
<td>Alcohol Use Screening</td>
</tr>
<tr>
<td>Promote Effective Prevention &amp; Treatment of Chronic Disease</td>
<td>Risk-Adjusted Mortality</td>
<td>Hospital 30-day All-Cause, Risk-standardized Mortality Rate Following Heart Failure Hospitalization</td>
</tr>
<tr>
<td>Work with Communities to Promote Best Practices of Healthy Living</td>
<td>Equity of Care</td>
<td>(No measure yet developed)</td>
</tr>
<tr>
<td>Work with Communities to Promote Best Practices of Healthy Living</td>
<td>Community Engagement</td>
<td>Discharge to Community—Post-Acute Care</td>
</tr>
<tr>
<td>Make Care Safer by Reducing Harm Caused in the Delivery of Care</td>
<td>HAI</td>
<td>Catheter-associated Urinary Tract Infection</td>
</tr>
<tr>
<td>Make Care Safer by Reducing Harm Caused in the Delivery of Care</td>
<td>Preventable Healthcare Harm</td>
<td>Early Elective Delivery</td>
</tr>
<tr>
<td>Make Care Affordable</td>
<td>Appropriate Use of Healthcare</td>
<td>Appropriate Treatment for Children with Upper Respiratory Infection</td>
</tr>
<tr>
<td>Make Care Affordable</td>
<td>Patient-Focused Episode of Care</td>
<td>Medicare Spending Per Beneficiary</td>
</tr>
<tr>
<td>Make Care Affordable</td>
<td>Risk-Adjusted Total Cost of Care</td>
<td>Oncology Care Model</td>
</tr>
</tbody>
</table>
Each measure submitted to CMS for consideration in pre-rulemaking must be assigned a single measure type to each measure submitted.

Table 9. Measure Types for Pre-Rulemaking

<table>
<thead>
<tr>
<th>Measure Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite</td>
<td>A combination of two or more component measures, each of which individually reflects quality of care, into a single quality measure with a single score.</td>
</tr>
<tr>
<td>Cost/Resource Use</td>
<td>A count of the frequency of units of defined health system services or resources; some may further apply a dollar amount (e.g., allowable charges, paid amounts, or standardized prices) to each unit of resource use (i.e., monetize the health service of resource use units).</td>
</tr>
<tr>
<td>Efficiency</td>
<td>The cost of care associated with a specified level of health outcomes.</td>
</tr>
<tr>
<td>Intermediate Outcome</td>
<td>Refers to a change produced by a healthcare intervention that leads to a long-term outcome (e.g., a reduction in blood pressure is an intermediate outcome that leads to a reduction in the risk of long-term outcomes such as cardiac infarction or stroke).</td>
</tr>
<tr>
<td>Outcome</td>
<td>The health state of a patient (or change in health status) resulting from healthcare, which can be desirable or adverse.</td>
</tr>
<tr>
<td>Patient-Reported Outcome (PRO)</td>
<td>A measure of a patient’s feelings or what they are able to do as they are dealing with diseases or conditions. These types may include PROM, which is an instrument, scale, or single-item measure that gathers the information directly from the patient, and Patient-Reported Outcome-based Performance Measure (PRO-PM), which is a way to aggregate the information that has been shared by the patient and collected into a reliable, valid measure of health system performance.</td>
</tr>
<tr>
<td>Process</td>
<td>A healthcare service provided to, or on behalf of, a patient, which may include, but is not limited to, measures that may address adherence to recommendations for clinical practice based on evidence or consensus.</td>
</tr>
<tr>
<td>Structure</td>
<td>Features of a healthcare organization or clinician relevant to the capacity to provide healthcare, which may include, but is not limited to, measures that address health IT infrastructure, provider capacity, systems, and other healthcare infrastructure supports.</td>
</tr>
<tr>
<td>Other</td>
<td>An explanation is required.</td>
</tr>
</tbody>
</table>

The NQF’s measure categories and types are very similar to CMS’s pre-rulemaking measure types.

Table 10. NQF’s Measure Categories and Types as Listed in the 2018 Evaluation Criteria and Guidance

<table>
<thead>
<tr>
<th>Measure Type</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>Assess the ability to obtain needed healthcare services in a timely manner, including the perceptions and experiences of people regarding their ease of reaching health services or health facilities in terms of proximity, location, time, and ease of approach. Examples may include, but are not limited to, measures that address the timeliness of response or services, time until next available appointment, and availability of services within a community.</td>
<td>No examples</td>
</tr>
</tbody>
</table>
| Composite          | Combine two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score. For the purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composite performance measures:  
  • Measures with two or more individual performance measure scores combined into one score for an accountable entity.  
  • Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity, including all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient). | Hospital Risk-Standardized Complication Rate following Implantation of Implantable Cardioverter-Defibrillator (ICD) |
<table>
<thead>
<tr>
<th>Measure Type</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost/Resource</td>
<td>Broadly applicable and comparable measures of health services counts (in terms of units or dollars) that are applied to a population or event (broadly defined to include diagnoses, procedures, or encounters). A resource use measure counts the frequency of use of defined health system resources; some may further apply a dollar amount (e.g., allowable charges, paid amounts, or standardized prices) to each unit of resource use.</td>
<td>Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Combine the concepts of resource use and quality. NQF has defined efficiency broadly as the resource use (or cost) associated with a specific level of performance with respect to the other five IOM aims of quality: safety, timeliness, effectiveness, equity, and patient centeredness.</td>
<td>Cardiac Imaging for Preoperative Risk Assessment for Non-Cardiac: Low Risk Surgery</td>
</tr>
<tr>
<td>Instrument-based</td>
<td>Use data derived from instruments. “Instrument” is a generic term that researchers use for a measurement device (e.g., survey, test, questionnaire, scale). Instruments are used for consistently obtaining (or presenting) data from respondents. The data derived from an instrument may include ratings or ranking output that is included in the calculation of a performance measure. Instruments may be used to collect information from a variety of individuals; examples include patients, observers (e.g., family, or other caregivers takers), or clinicians. Data from instruments can be used in the calculation of structure, process, or outcome performance measures. Instruments specific to PROs may be referenced as PROMs.</td>
<td>Depression Remission at 12 Months</td>
</tr>
<tr>
<td>Intermediate Clinical</td>
<td>A change in physiologic state that leads to a long-term health outcome.</td>
<td>Hemodialysis Vascular Access: Long-term Catheter Rate</td>
</tr>
<tr>
<td>Outcome</td>
<td>The health status of a patient (or change in health status) resulting from healthcare—desirable or adverse.</td>
<td>Acute Care Hospitalization During the First 60 Days of Home Health</td>
</tr>
<tr>
<td>Patient-Reported Outcome</td>
<td>Any report of the status of a patient’s (or person’s) health condition, health behavior, or experience with healthcare that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else. Key PRO domains include health-related quality of life/functional status, symptom/symptom burden, experience with care, and health-related behavior.</td>
<td>No examples</td>
</tr>
<tr>
<td>Process</td>
<td>A healthcare-related activity performed for, or on behalf of, a patient.</td>
<td>Angiotensin converting enzyme (ACE) inhibitors or angiotensin II receptor blockers (ARB) for left ventricular systolic dysfunction – Acute Myocardial Infarction (AMI) Patients</td>
</tr>
<tr>
<td>Process: Appropriate</td>
<td>A type of process measure that has been used to evaluate procedures and medical technologies. Appropriate use measures are neither cost/resource use measures nor efficiency measures.</td>
<td>No examples</td>
</tr>
<tr>
<td>Structure</td>
<td>Structure of care is a feature of a healthcare organization or clinician related to its capacity to provide high-quality healthcare.</td>
<td>Nursing Hours per Patient Day</td>
</tr>
</tbody>
</table>
5  SELECTED MEASURE TYPES

5.1  COST AND RESOURCE USE MEASURES

The CMS strategic goals include providing better care and lower cost of care for all Americans. That strategy addresses affordable care by aiming to reduce the cost of quality healthcare for individuals, families, employers, and government. Measures of cost and resource use can be used to assess the variability of the cost of healthcare and to direct efforts to make healthcare more affordable. Some terms related to measures addressing affordable care include:

- **Resource use**—Broadly applicable and comparable measures of health services counts (in terms of units or dollars) applied to a population or event (broadly defined to include diagnoses, procedures, or encounters). A resource use measure counts the frequency of defined health system resources; some may further apply a dollar amount (e.g., allowable charges, paid amounts, standardized prices) to each unit of resource use—that is, monetize the health service or resource use units.
- **Cost of care**—Total healthcare spending, including total resource use and unit price(s), by payer or consumer, for a healthcare service or group of healthcare services, associated with a specified patient population, time period, and unit(s) of clinical accountability.
- **Quality of care**—Quality measures assess performance on the six healthcare aims specified by the IOM in Crossing the Quality Chasm: A New Health System for the 21st Century: safety, timeliness, effectiveness, efficiency, equity, and patient centeredness.
- **Efficiency**—Measuring cost of care associated with a specified level of quality of care.
- **Value of care**—Includes a specified stakeholder’s preference-weighted assessment of a combination of quality and cost of care performance. The stakeholder could be an individual patient, consumer organization, payer, provider, government, or society. The value of care would be the combination of quality and cost, weighted by the stakeholder’s preference.

The Organization for Economic Co-operation and Development (OECD) noted in their Health at a Glance 2017, the United States is a country with high healthcare costs, but poorer than expected health outcomes relative to many parts of the world. The challenge for CMS, the largest payer for healthcare in the country, is to identify the best, most efficient means by which to improve care, while ensuring care remains patient-centered and of equal quality for all populations. Resource use measures can be valuable building blocks to understanding efficiency and value. NQF Measurement Framework: Evaluating Efficiency Across Patient-Focused Episodes of Care has broadly defined efficiency as “the resource use (or cost) associated with a specific level of performance with respect to the other five IOM aims of quality: safety, timeliness, effectiveness, equity, and patient centeredness” (p. 16). NQF uses Figure 20 (adapted) to illustrate the relationship between resource use, efficiency, and value.

Cost and resource use measures must be linked to quality outcomes as well as to the processes that are required to achieve those outcomes. Measure developers should consider how cost and resource use measures can be paired. Methodologies
for adding the stakeholder preference factors necessary to measure value are still being defined. There also remain challenges to truly identify benchmarking cohorts for accountability comparisons.

5.2 **COMPOSITE PERFORMANCE MEASURES**

The NQF, in their Composite Performance Measure Evaluation Guidance, defines a composite performance measure as “a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score” (p. 27). These measures are useful for a variety of purposes. Composite performance measures can group measures into a common construct that can provide a broader assessment of quality care. There are two primary types of composite performance measures:

- Measures of two or more individual performance areas scored using an algorithm that produces as single score as its only output. With this type of composite, the individual components cannot produce individual scores.
- Measures with two or more individual component measures assessed separately and then aggregated into one score. Component elements of this type of composite stand alone, but their combination produces a richer representation of the target construct (e.g., optimal diabetes care).

Composite performance measures consist of two or more measures, possibly already specified and endorsed. Measure development is unique for composites because the intended use of the composite and relationships between the component measures should be examined and understood.

Composite performance measures can be useful in situations such as public reporting websites and pay-for-performance programs. They take several components and combine them into a single metric summarizing overall performance. Composite performance measures can also be referred to as a composite index, composite indicator, summary score, summary index, or scale. Composite performance measures can evaluate various levels of the healthcare system such as individual patient data, individual practitioners, practice groups, hospitals, or healthcare plans.

This section discusses development of composite measures intended for quality measurement in accountability programs. Quality indicator aggregations such as the Nursing Home Compare star rating and other similar collections of measures are not covered in the Blueprint.

5.2.1 **Purpose of Composite Measures**

For measures to be grouped as a composite, they must have a purpose for which the composite will be used (e.g., comprehensive assessment of adult cardiac surgery quality of care). There also needs to be a delineated quality construct to be measured (e.g., the four domains of cardiac surgery quality, which include perioperative medical care, operative care, operative mortality, and postoperative morbidity).

Composite performance measure development should be conducted using the principles (American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures, 2010):

- The purpose, intended audience, and scope of a composite performance measure should be explicitly stated.
- The individual measures used to create a composite performance measure should be evidence-based, valid, feasible, and reliable.
• The methods used for weighting and combining individual measures into a composite performance measure should be transparent and empirically tested.
• The scientific properties of these measures, including reliability, accuracy, and predictive validity, should be demonstrated.
• Composites should be useful for clinicians and/or payers to identify areas for quality improvement.

5.2.2 Component Performance Measures

The NQF Composite Performance Measure Evaluation Guidance suggests some considerations for selecting measures to be included in a composite:

• Components should be justified based on clinical evidence.
• NQF endorsement is not required; however, measures need to be justified in terms of feasibility, reliability, and validity.
• Individual components generally should demonstrate a gap in care; however, if included, a clinical or analytic justification needs to be made for including components that do not demonstrate a gap in care.
• Individual components may not be sufficiently reliable independently, but they can be included if they contribute to the reliability of the composite.

Components of the composite should be assessed for internal consistency. Internal consistency is the extent to which several measures of a given construct provide similar information about that construct. For instance, in NQF 0729 Optimal Diabetes Care (Composite Measure), the consensus endorsement entity agreed with the steward that the optimal management of hemoglobin A1c, blood pressure, statin use, tobacco non-use, and daily aspirin or anti-platelet use for patients with diagnosis of ischemic vascular disease adequately represented excellent management of diabetes mellitus by preventing or reducing future complications associated with poorly managed diabetes. Each of these measures individually represent good care of diabetes symptoms, and as a group are internally consistent with the construct of comprehensive diabetes management. Consistency may be less relevant if the goal of the composite is to combine multiple distinct dimensions of quality rather than a single dimension. Standard psychometric criteria would not apply to that scenario; therefore, it may be difficult to evaluate internal consistency for composites with multiple distinct dimensions.

The Measure Authoring Tool (MAT) currently supports eCQM composite measures within the metadata section. Users who are defining a composite measure can indicate the measure type as composite and can then identify measures in the metadata that are component measures.

5.3 Patient-Reported Outcome Measures

PROMs are quality measures that are derived from outcomes reported by patients and are a high priority for CMS. These measures present some design challenges that are described in this section, with some approaches to those challenges.

Ensuring that patients and families are engaged as partners in their care—one of the CMS priorities—can also be an effective way to measure the quality of patient care. Although patient reports of their health and experience with care are not the only outcomes that should be measured, they certainly are an important component. Historically CMS used surveys to collect patient experience and satisfaction with care data, but the infrastructure to collect these data more timely and in other ways for PROMs in quality and reporting programs continues to be developed. Tools to collect these data (e.g., Patient-
Reported Outcomes Measurement Information System [PROMIS tools](#) have mostly been used in academic settings and are being tested for clinical application. Figure 21 depicts the relationship between PROs, PROMs, and PRO-PMs.

**Figure 21. Relationship between PROs, PROMs, and PRO-PMs**

### 5.3.1 Patient-Reported Outcomes

The United States Food and Drug Administration’s (FDA) [Guidance for Industry, Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims](#) defines a PRO as “any report of the status of a patient’s (or person’s) health condition, health behavior, or experience with healthcare that comes directly from the patient (i.e., outcome data)” (p. 2). Self-reported patient data provide a rich data source for outcomes. This definition reflects the key domains listed in the NQF report on PRO-PM:

- Health-related quality of life (including functional status)
- Symptoms and symptom burden (e.g., pain, fatigue)
- Experience with care
- Health behaviors (e.g., smoking, diet, exercise).

### 5.3.2 Patient-Reported Outcome Measurement Tools

PROMs are tools that are used to collect patient-reported outcomes. Some examples of patient self-reported data collection tools include:

- [Patient-Reported Outcomes Measurement Information System](#) (PROMIS)—Funded by the National Institutes of Health (NIH), these tools measure patient self-reported health status.
- [Health Outcomes Survey](#) (HOS)—The HOS was the first outcome measure used in Medicare Advantage plans. The goals of the Medicare HOS program are to gather valid and reliable health
status data in Medicare managed care for use in quality improvement activities, plan accountability, public reporting, and health improvement. All managed care plans with Medicare Advantage contracts must participate.

- **Focus On Therapeutic Outcomes (FOTO)**—This tool is used to measure the functional status of patients who received outpatient rehabilitation through the use of self-reported health status questionnaires. Because the measures are assessed at intake, during, and at discharge from rehabilitation, the change in functional status can be calculated.

However, the outcomes collected by the tools are insufficient individually for measuring performance and cannot be used directly as part of accountability programs. A performance measure must be constructed that applies the outcome data collected by the tools to measure the quality of care.

### 5.3.3 Patient-Reported Outcome-based Performance Measures

A PRO-PM is a way to aggregate the information from patients into a reliable, valid (tested) measure of performance. NQF only endorses PRO-PMs that can be used in performance improvement and accountability. The same measure evaluation and justification principles that apply to other outcome measures also apply to PRO-PMs.

Several PRO-PMs are available, for example:

- **Average Change in Back Pain Following Lumbar Discectomy/Laminotomy**
- **Average Change in Functional Status Following Total Knee Replacement Surgery**

### 5.3.4 Approaches to Developing Patient-Reported Outcome-based Performance Measures

Although PROs are a special type of outcome measure, the principles for development are the same. PRO-based measure development will be used as an example of the steps involved in developing all outcome measures. **Section 3, Chapter 19, Risk Adjustment** details the procedure for risk-adjusting outcome measures. NQF outlined a pathway for PROs to move from simple patient-reported data to measurement, to performance measurement, and finally to endorsed measures in use for reporting and accountability.

#### 5.3.4.1 Choose and Define a Patient-Reported Outcome

Many kinds of data are reported by patients or are collected directly from patients without clinician interpretation. To choose outcomes that will become performance measures, measure developers must first identify quality issues for a target population.

An appropriate outcome has clinical or policy relevance. For example, whether the patient did or did not develop a surgical site infection after cataract surgery would not be a good PRO. A patient could report redness, swelling, and drainage, but not actually whether he/she has an infection. A better outcome measure in this instance might be a clinically meaningful measure of improvement in vision.

Outcome performance measures must also be meaningful to the target population and usable by the providers being held accountable. Whenever possible, clinical experts should be consulted to more relevantly define appropriate and meaningful outcomes.

#### 5.3.4.2 Determine the Appropriate Way to Collect the PRO Using a PROM (Tool)

Measure development always begins with an environmental scan and literature review to identify whether there are existing tools to collect the outcome in the target population. Many tools in this area have been developed for research and have existing psychometric data establishing reliability and
validity. While the tools are not themselves measures, with further testing in clinical settings, the information from these tools may be used to develop and test the construct of a PROM. Feasibility must also be tested for the relevant clinical applications.

It is important that these tools have been tested with the population on which the measure focuses. It should also be noted that there may be differences between the reliability and validity of a PRO tool in more controlled settings (e.g., clinical trials, academic research projects) compared to use in real-world practice settings, but most PRO tools have only been tested in the former.

5.3.4.3 Determine the Appropriate Performance Measure: the PRO-PM

The outcomes for target populations can be reported as average change or percentage improvement determined by the topic of interest. All must be tested for reliability, usability, feasibility, validity, and threats to validity, including how missing data are handled and appropriate risk adjustments. To appropriately distinguish variations in performance between providers, the outcome must capture the results of the care given and not the influence of comorbidities or other extraneous variables. However, as in any other outcome measurement, risk adjustment should not be allowed to mask disparities. Section 3, Chapter 19, Risk Adjustment contains a discussion on determining the need for risk adjustment and development, and evaluation of risk adjustment models.

5.3.4.4 Evaluate the Outcome Measure

Outcome measures, including those based on PROs, must be evaluated against standard criteria in the same way that all MUD must be evaluated. Detailed specifications must be submitted, using the MIF.17 Some of the unique considerations (in addition to the others in each category) that apply to evaluating PRO-PMs include:

- **Importance**—The measures must be patient-centered. Patients must be involved in identifying the PROs to be used for performance measurement.
- **Scientific Acceptability**—Specifications must include methods of administration, how proxy responses are handled, response rate calculations, and how the responses affect results. Reliability and validity must be established not only for the data measurement instrument (i.e., PROM) but also for the derived performance measurement (i.e., PRO-PM).
- **Feasibility**—Burden to respondents must be minimized. Illness may complicate accessibility issues. Language, literacy, and cultural issues must also be considered.
- **Usability** and Use—Not only must patients find the results of PRO-PMs useful, but providers must also be able to use the information to improve quality of care.

The NQF endorsement criteria for PRO-PMs are enumerated in NQF’s final report, PROs in Performance Measurement. Documentation of these items should be submitted to the COR as specified in the contract.

5.4 Multiple Chronic Conditions Measures

CMS noted in the Chronic Conditions Charts: 2015 that the prevalence of two or more chronic conditions among FFS Medicare beneficiaries in 2015 was 65%. These individuals constitute a challenge to the healthcare system because their conditions complicate each other, are ongoing, and are extremely costly to both the persons involved and the nation overall. The effects of their comorbidities are more than simply additive; they multiply both morbidity and mortality (Tinetti et al., 2011). CMS

---

17 The NQF submission may be acceptable for this deliverable.
found in 2015 that Medicare beneficiaries with multiple chronic conditions (MCC) were the heaviest users of healthcare services. Those with six or more chronic conditions accounted for 51% of Medicare spending on hospitalizations. However, Giovannetti et al. (2013) identified few measures that are specifically designed to evaluate the quality of care provided to people with MCCs.

5.4.1 Multiple Chronic Condition Definition

HHS contracted with NQF to develop a measurement framework for persons with MCC. The NQF MCC Measurement Framework defined MCC as “having two or more concurrent chronic conditions that collectively have an adverse effect on health status, function, or quality of life and that require complex healthcare management, decision-making, or coordination” (pp. 7-8).

Assessment of the quality of care provided to the MCC population should consider persons with two or more concurrent chronic conditions that require ongoing clinical, behavioral, or developmental care from members of the healthcare team and whose conditions act together to significantly increase the complexity of management and coordination of care—including, but not limited to, potential interactions between conditions and treatments.

NQF MCC Measurement Framework (p. 8) stated, from an individual’s perspective, the presence of MCC would:

- Affect functional roles and health outcomes across the lifespan
- Compromise life expectancy
- Hinder a person’s ability to self-manage or a caregiver’s capacity to assist in that individual’s care.

5.4.2 Need for Measure Development

Although persons with MCC represent a growing proportion of society who use an increasingly large amount of healthcare services, existing quality measures do not adequately address comorbidity. Current quality measures are largely based on performance standards derived from clinical practice guidelines for management of a single disease (Tinetti et al., 2004). Patients with MCC have often been excluded from the evidence-generating clinical trials that form the basis of many clinical practice guidelines. The randomized clinical trials used in clinical practice guideline development focus mainly on single diseases to produce robust guidance for specific disease treatments. Rigid adherence to these disease-specific guidelines could potentially harm those with MCC. For example, medications prescribed in adherence to guidelines for several diseases individually may result in a patient suffering adverse effects of polypharmacy (American Geriatrics Society Expert Panel on the Care of Older Adults with Multimorbidity, 2012). Few, if any, measures exist to evaluate inappropriate care in these situations.

5.4.3 Considerations for Measure Development Targeting Persons with Multiple Chronic Conditions

5.4.3.1 What to Consider when Choosing Appropriate Measure Concepts

Without evidence-based guidelines specifically directed to care of persons with MCC, best practices may remain up to the clinical judgment of the providers. However, measurable quality topics exist that are especially pertinent to people with MCC. The NQF MCC Measurement Framework (p. 9) identified these measurement concepts as having potential for high-leverage in quality improvement for patients with MCC:

- Optimizing function, maintaining function, or preventing further decline in function
- Seamless transitions between multiple providers and sites of care
• Patient-important outcomes (includes PROs and relevant disease-specific outcomes)
• Avoiding inappropriate, non-beneficial care, including at the end of life
• Access to a usual source of care
• Transparency of cost (total cost)
• Shared accountability across patients, families, and providers
• Shared decision-making.

These measure concepts represent cross-cutting areas with the greatest potential for reducing factors of cost, disease burden, and improving well-being that are highly valued by providers, patients, and families.

5.4.3.2 When Determining How to Address Key Issues

5.4.3.2.1 Guiding Principles

NQF MCC Measurement Framework identified that quality measures for persons with MCC should be guided by several principles. Quality measures should:

• Promote collaborative care among providers
• Consider various types of measures that address appropriateness of care
• Prioritize optimum outcomes that are jointly established by considering patient preferences
• Address shared decision-making
• Assess care longitudinally
• Be as inclusive as possible
• Illuminate and track disparities through stratification and other approaches
• Use risk adjustment for comparability (of outcome measures only) with caution, as it may obscure serious gaps in quality of care
• Standardize inputs from multiple sources, particularly patient-reported data.

5.4.3.2.2 Time Frame Issues to Consider

Measurement time frame is particularly important with chronic conditions because the nature of chronic conditions requires observation over time. Especially in the case of outcome measures for patients with MCCs, it is extremely difficult to know where to attribute responsibility unless the measurement time frame is carefully considered and specified. Measures for this population should assess care across episodes and across providers and staffing using a longitudinal approach. Delta measures of improvement (or maintenance rather than decline) over extended periods are particularly relevant in this population.

5.4.3.2.3 Attribution Issues to Consider

Issues of attribution are compounded when adding the factor of MCC. Since multiple conditions usually means multiple providers, it becomes difficult to choose who should be credited for good outcomes and which provider gave inadequate care when the treatment for one condition might exacerbate the other. These issues may require a more aggregated level of analysis such as at a provider group level or population rather than individual level. Since beneficiaries with MCC have multiple providers, it would be more appropriate to measure and attribute the outcomes for the population to the care provided by the team of providers.
5.4.3.2.4 Methodological Issues to Consider

Quality measures for this population should be designed to be as inclusive as possible. Methodological approaches should be designed to reveal and track variances in care and outcomes.

The empirical link between quality processes with the outcomes of those healthcare processes is even more difficult to establish when dealing with MCCs. Risk adjustment should be used with caution in the situation of MCCs. Stratification may allow quality comparison across populations without masking important distinctions of access, care coordination, and other issues. Section 3, Chapter 19, Risk Adjustment provides an in-depth discussion on how to determine when risk adjustment is appropriate and how to evaluate risk adjustment models when they are applied.

Quality measures for this population should address quality across multiple domains. Measures should be harmonized across levels of the healthcare system to provide a comprehensive picture of care.

5.4.3.2.5 Data Gathering Issues to Consider

There may be difficulties gathering data systematically, especially for this population. Particularly, patient-reported data may be problematic to collect because of the interacting conditions. For example, it might be difficult to collect fatigue data from a person with both chronic lung disease and history of stroke, because each condition may contribute to a patient’s fatigue, and it may be hard to assess the contribution of each disease to that fatigue. Interpretation of different types of data is needed, as the data may come from multiple providers, multiple sources, in multiple formats, and over extended periods of time. It is important for measure developers to standardize data collection methods.

5.4.3.3 When Testing and Evaluating Measures for Persons with Multiple Chronic Conditions

Evaluation methods described elsewhere in this Blueprint also apply to measures of quality care for persons with MCC. In addition, MCC measures should successfully carry out the guiding principles from the NQF Framework. Functional status and other outcomes should be examined using delta measures of change over time. If new tools and methods of data collection are developed, those tools must also be carefully assessed. Formative, or alpha, testing may be particularly important early during development, not only for new tools designed for these types of measures, but also to test the feasibility of linking data from a variety of sources.

5.5 Other Measure Types

Other measure types that exist may not be covered in this chapter, but the standard measure development and maintenance processes should apply to them. Hybrid measures that use more than one data type or method of data collection are one example. Some hybrid measures use both claims and EHR data or survey and chart-abstracted data. A measure developer may require additional guidance for other situations that are not covered here. For those situations, contact the Measures Manager and the appropriate COR.
INTRODUCTION TO ELECTRONIC CLINICAL QUALITY MEASURES

Collecting and reporting accurate healthcare performance data has historically been a highly structured and time-consuming manual process. To limit the need for extensive record reviews required by chart-abstracted measures, early performance measures used routinely available claims data. Subsequently, clinically enhanced measures provided increased relevance by supplementing claims information with electronically available laboratory results and pharmaceutical usage data. Increasing use of EHRs and other electronic clinical systems has the potential to provide access to a significantly greater set of clinical information. By addressing such electronic data captured during the routine process of clinical care, the eCQM has become a critical component of the quality reporting framework. When measures are unambiguously represented as eCQMs, they can be used to guide collection of EHR data, which are then assembled into quality reports and submitted to organizations such as CMS. CMS considers using the data routinely collected through EHRs an essential tool for reducing burden.

eCQMs, formerly known as eMeasures, can promote greater consistency and improved uniformity in defining clinical concepts and measure logic across measures and increased comparability of performance results. Through standardization of a measure’s structure, metadata, definitions, and logic, the HQMF strives for quality measure consistency and unambiguous interpretation. HQMF is a component of a larger quality end-to-end framework, which has evolved to a normative HL7 standard. eCQMs intend to significantly reduce measurement errors due to manual abstraction and highlight encoding issues.

eCQMs are designed to be turned into queries that automatically retrieve the necessary information from the EHR’s data repositories and generate quality data reports. From there, individual and/or aggregate patient quality data can be transmitted to the appropriate agency using Quality Reporting Document Architecture (QRDA) Category I (individual patient data) or Category III (aggregate patient data) reports. As the nation makes progress toward health IT adoption, much of the successes will rely on solid electronic representation of measurement and clinical decision support.

eCQM developers need to be knowledgeable of:

- The Blueprint – This guide for the CMS MMS.
- The types of clinical data that are typically encoded using standardized terminology (i.e., a code system) in EHR systems and consider the impact on workflow and data fidelity for organizations that will need to map local codes to standard terminologies used in an eCQM.
- Quality Data Model (QDM) – An information model used to define clinical concepts in a standardized format to enable electronic quality performance measurement. More information on QDM is found in Section 3, Chapter 7, eCQM Standards-Based Guidance and Tools.
- MAT – A web-based tool that enables measure developers to author eCQMs in HQMF using the QDM elements, Clinical Quality Language (CQL), and healthcare industry standard terminologies. The representation of an eCQM is simplified and standardized when measure developers author their eCQMs in the MAT. Measure developers should consult with their COR if they have questions about using the MAT. More information on the MAT is found in Section 3, Chapter 7, eCQM Standards-Based Guidance and Tools.
- CQL – A Health Level Seven International (HL7) standard that provides the ability to express logic that is human-readable yet structured enough for processing a query electronically.

New and evolving standards have been adopted (e.g., Clinical Quality Language, and others are under consideration, like Fast Healthcare Interoperability Resources (FHIR)) to help harmonize with clinical decision support standards and increase interoperability.
information on CQL is found in Section 3, Chapter 7, eCQM Standards-Based Guidance and Tools.

- **Value Set Authority Center (VSAC)** – Provided by the NLM, in collaboration with the ONC and CMS. Requiring a free UMLS license for access, the VSAC provides searchable and downloadable access to all official versions of value sets used by each of the eCQM releases used in CMS and other quality reporting programs (e.g., The Joint Commission). More information on the VSAC is found in Section 3, Chapter 17, Codes, Code Systems, and Datasets.

- **Bonnie** – A software tool that allows eCQM developers to test and verify the behavior of their eCQM logic. More information on Bonnie is found in Section 3, Chapter 7.3.1 Development Tools.

- **ONC Project Tracking System (Jira)** – An issue tracking system licensed by ONC. It is a collaboration platform that supports the implementation of health IT by providing a space in which internal and external users can transparently log, prioritize, and discuss issues with appropriate SMEs on a host of topics. More information about Jira is found in Appendix B.

- **eCQI Resource Center** – A website that provides eCQI resources and connections. It is the source of truth for specifications of eCQMs in CMS programs. It serves as “the one-stop shop for the most current resources to support electronic clinical quality improvement.”

### 6.1 eCQM Components

eCQMs are written to conform to the HL7 CQL-based HQMF standard for representing a health quality measure as an electronic XML document. eCQMs are specified using patient-level information coded in a format intended for extraction from EHRs and other electronic clinical systems.

An eCQM has three components:

- **Computable representations** of the eCQM, which contain important details about the measure, how the data elements are defined, and the underlying logic of the measure calculation. The files include:
  - HQMF XML syntax (.xml). The HQMF includes a header and a body. The header identifies and classifies the document and provides important metadata about the measure. Appendix C lists the metadata along with definitions, measure developer guidance, and whether the element is required or not applicable. The HQMF body contains eCQM sections (e.g., definitions, population criteria, supplemental data elements). Refer to Section 3, Chapter 7, eCQM Standards-Based Guidance and Tools for a full list of the eCQM sections.
  - CQL file (.cql). Provides the expression logic for data criteria, population criteria, and supplemental data elements. It provides a formal description of the computable content in the measure and is organized into libraries that can be reused or shared between measures and other artifacts. Refer to Section 3, Chapter 7.2.3 Clinical Quality Language.
  - Shared CQL libraries (.cql, .xml, and .json). The shared libraries are the basic units of sharing CQL. They consist of a foundation of CQL statements used within a measure. Every measure has at least one main CQL library and is referenced from HQMF.
  - ELM XML document (.xml). Provides a machine-readable representation of the measure’s logic in XML. The ELM file is intended for machine processing and provides the information needed to automatically retrieve data from an EHR.
  - ELM JavaScript Object Notation (JSON) file (.json). The JSON file is the ELM file in JavaScript Notation, as opposed to XML.
• **Human-readable representation** of the eCQM, which displays the eCQM content in a human-readable format directly in a web browser, HTML file (.html). This file does not include the underlying HQMF syntax, but the narrative content at the top of the HTML is extracted from the HQMF header.

• **Value sets and direct reference codes (DRCs),** which convey specific coded value(s) that are allowed for the data elements within the eCQM. Value sets are identified by an object identifier (OID) and include several metadata elements that describe the inclusion and exclusion criteria for the codes in the set. The value set includes a list of codes (i.e., the value set “expansion code set”) acceptable or valid for a specific data element in the measure, descriptors of those codes, the code system from which the codes are derived, and the version of that code system. DRCs are specific codes that are referenced directly in the eCQM logic to describe a data element or one of its attributes. Value sets and DRCs are found in the VSAC.

### 6.2 Collaborative Measure Development Workspace and Data Element Repository

The [CMD](#) and [Data Element Repository (DERep)](#) are a result of the outreach CMS conducted as part of the eCQM Strategy project. The goal of the CMD Workspace is to bring together a set of interconnected resources, tool, and processes to promote transparency and better interaction across stakeholder communities that develop, implement, and report eCQMs.

The CMD Workspace will be comprised of five modules to assist clinicians, eCQM developers, implementers, and submitters during the entire eCQM lifecycle, from initial measure concept, through development, implementation, and reporting to CMS. New content will be added over time and stakeholders are encouraged to review and participate. Goals of the CMD Workspace are to:

- Provide detailed data element definitions to support implementation
- Achieve harmonization across measures, data elements, and value sets
- Improve alignment of measure concepts with clinical need and newly published guidelines
- Demonstrate how new measures fill existing quality reporting gaps
- Increase involvement by clinical experts and EHR vendors during measure development
- Offer transparency of test results during measure development
- Provide notification of updates to MUD.

One of the five modules is available. The eCQM DERep provides all the data elements associated with published and tested eCQMs for use in CMS quality reporting programs as well as the definitions and clinical focus for each data element. An end user can sort information by data element, eCQM, union, QDM attribute, or QDM category and datatype data element. The DERep currently has the calendar year (CY)2019 reporting/performance period elements and will add more information as measures are updated or added to CMS programs.

The four future modules include:

- **eCQM Concepts** – The eCQM concept workspace will provide users the ability to submit new measure concepts, align new measure with Meaningful Measures criteria, and identify whether similar measures exist.
- **New eCQM Clinical Workflow** – Groups will be able access all the measure development tools in the CMD Workspace and work in an iterative manner to perform measure development activities. Stakeholders can provide early comments, clinical workflow concerns, and guidance during the measure development. Lessons learned from previous measure development
activities can help developers address implementation-specific issues that arise during development.
- **eCQM Test Results** – Draft measure test results will offer transparency into the feasibility, reliability, and validity of the eCQM, a testing scorecard, and additional characteristics of test sites including types of health IT used, number of test sites, and rating of each data element in the testing process for each measure.
- **Subscribe to CMD Workspace Updates** – Participants will be able to sign up for alerts on the progress of evolving measures.

### 6.3 Special Considerations for eCQMs

The measure development process for eCQMs does not significantly differ from that used for non-eCQMs. The measure conceptualization process is the same for eCQMs as for measures developed using other data sources. While the process of creating an eCQM is like the process of creating other types of measures with respect to defining measure metadata and measure components for each measure scoring type (e.g., proportion, continuous variable [CV], ratio), eCQMs require additional steps to map measure data elements to corresponding QDM components and standard terminologies to assemble the data criteria. eCQMs are based on information that should exist in a structured format in electronic clinical systems such as EHRs.\(^\text{19}\) In principle, all information should be available and accessed without impacting the normal workflow; hence, it is essential to carefully consider how, by whom, and in what context the desired information is being captured.

Evaluation of the scientific acceptability (i.e., validity and reliability) of eCQMs is based on some unique assumptions and special considerations, including:

- **eCQM evaluation** is based on use of only data elements that can be expressed using the QDM.
- Quality measures that are based on electronic clinical systems should significantly reduce measurement errors due to manual abstraction, coding issues, and inaccurate transcription errors.
- eCQMs are subject to some of the same potential implementation issues as non-eCQMs, which could result in low evaluation ratings for the reliability and validity of data elements and measure scores.
- Careful analysis, such as through systematic audits of patient data used in reporting (Provonost, Wu, & Austin, 2017), is required to avoid the potential, unintended consequences of selecting data elements that are infrequently or inconsistently captured. For example, problem lists may not be updated in a timely manner and may not be reconciled to remove or “resolve” health concerns that are no longer active. Therefore, using information from problem lists may not necessarily provide valid and reliable data.\(^\text{20}\) Given that eCQMs rely on accurately maintained, specifically encoded EHR data, increased attention to improved documentation will be important. Examples of potential sources of error that might be introduced as a result of implementation include:
  - EHR system structure or programming that does not comply with standards for data fields, coding, or exporting data.
  - Data fields used in different ways or multiple ways to enter the same data; for example, variation in clinical workflow resulting in entries made into the EHR fields other than those used to retrieve data to calculate the measure.

\(^{19}\) Data not in a structured field may be used in conjunction with NLP software or similar tools.

\(^{20}\) eCQM specifications, as defined by QDM data elements, do not designate where (e.g., Problem List) in the EHR the data should be extracted.
Inaccurate interpretation of data by NLP software used to analyze information from text fields.
- Variability in the mapping of data encoded using a non-standard (local) terminology to that of the standard terminology expected by the eCQM.
- Although data element reliability (repeatability) is assumed with computer programming of an eCQM, empirical evidence is required to evaluate the reliability of the measure score.
- To test data element validity:
  - Compare the electronic extract with the manual abstract.
  - If using NLP, need to assure NLP is correct.

### 6.4 eCQM Annual Update and Change Review Process (CRP)

The eCQM Annual Update includes updates to eCQM specifications, the CRP, supporting documentation, eCQM tools, and may include updates to eCQM standards. CMS updates eCQMs annually to align with current clinical guidelines, code systems, and eCQM standards so they remain relevant and actionable within the clinical care setting. Updates may also be made in response to end user questions or suggestions usually submitted via the ONC Project Tracking System (Jira) CQM project. Selected issues submitted via Jira and other means go through the CRP.

#### 6.4.1 Change Review Process

The goal of the CRP is to work with eCQM implementers to determine the impact of an update, while minimizing provider and vendor burden in the collection, capture, calculation, and reporting of eCQMs. eCQM users have the opportunity to review and comment on proposed changes to the eCQM specifications through the Jira website. To participate in the CRP, users must have a Jira account. As Jira CQM tickets are posted for public comment, weekly digest emails serve to inform members of new issues posted for review and those that will be closing soon. The CRP occurs during the fall.

#### 6.4.2 eCQM Specifications

There are different phases to the eCQM specifications update. As part of the pre-MAT time frame, measure developers propose changes to the specifications to CMS based on the CRP feedback, standards changes, etc. Once CMS approves, the measure developers share marked-up specifications with standards and logic SMEs via the Jira annual update project – CQM Annual Update (CAU), which is a restricted access project. Measure developers identify any deprecated codes that need to continue in use in measure specifications (legacy codes) for look-back periods and share with the Value Set Authority Center (VSAC). VSAC in turn, updates the value set expansion profiles to include the legacy codes. Measure developers update the value sets using the VSAC Collaboration Tool.

Measure developers then input CMS-approved changes to the specifications in the MAT and export the revised packages for review and Bonnie testing. The post-MAT phase begins when the updated packages are posted to the CAU project for a second review by standards and logic SMEs. Updated draft measure packets are posted to the CQM Jira project for public review and comment. Measure developers finalize measure specifications based on reviews and feedback by updating the MAT packages and retest in Bonnie. Measure developers then develop the technical release notes (TRNs), which provide an overview of technical changes in the eCQM specifications. The final MAT packages are sent to VSAC and all the final value sets are moved into VSAC’s production environment. To view the details of the value sets, a free UMLS license is required. The VSAC also posts the Binding Parameter Specification (BPS) document. The BPS is a record of the value set metadata information that defines the value set code lists specified by published CMS eCQMs. Measure implementers and vendors can use the BPS to track
versions and other parameters that define the value set code lists for each eCQM release. The measure packages, to include the TRNs, are posted to the [eCQI Resource Center](https://www.ecri.org).

### 6.4.3 Supplemental Documents

There are supplemental documents updated annually. Prior to the publication of the updated specifications, CMS releases a Pre-Publication Document that contains technical and program changes and the standards and code sets approved by CMS. The purpose of this document is to give implementers advance notice of upcoming changes. About the same time as the updated measures are posted to the [eCQI Resource Center](https://www.ecri.org), the [eCQM Logic and Implementation Guidance document](https://www.cms.gov) is released. It provides general implementation guidance such as how specific logic and data elements should be conceptualized and addressed during implementation and how to use Jira to provide feedback, track issues, and ask questions. An appendix in the eCQM Logic and Implementation Guidance document provides the standards and code systems in use for the particular reporting/performance period. The [Guide for Reading eCQMs](https://www.ecri.org) describes the eCQM package contents, file naming conventions, brief descriptions of the standards, and tools used with eCQMs. The different sections of the human-readable HTML document are explained with examples. The [CQL Style Guide](https://www.cms.gov) is also updated to reflect any changes. The CQL Style Guide provides examples to standardize expression of measure concepts across eCQMs and define a uniform look and feel to eCQM logic using CQL. The guide focuses on a set of common best practices implemented across CQL-based eCQMs in CMS reporting programs and also promotes the use of consistent language within the framework of CQL, including libraries, aliases, definitions, functions, and conventions.

eCQM flows are developed for each measure and posted to the eCQI Resource Center eCQM Materials tables, usually by the end of the summer after publication of the updated specifications. The eCQM flows are flowcharts designed to assist in interpretation of the eCQM logic and calculation methodology for reporting rates. These flows provide an overview of each of the population criteria components and associated data elements that lead to the inclusion or exclusions into the eCQM’s quality action (numerator).

### 6.4.4 eCQM Standards and Tools

Several of the standards used with eCQMs are HL7 standards (CQL, HQMF, QRDA). The standards are reviewed and any significant changes must undergo the HL7 ballot process, which can take a year or more. Smaller changes can be done outside the ballot cycle. Changes to CQL, both the base standard and the CQL-based HQMF Implementation Guide (IG), may affect the QDM and vice versa. The QDM, a non-HL7 standard, is updated as necessary and changes go to the MAT Change Control Board (MCCB) for approval. Updates to CQL may also require updates to the [CQL to Expression Logical Model (ELM) Translator](https://www.hl7.org), which translates high-level CQL syntax into the canonical ELM representation. The MAT and Bonnie tools are updated to align with the standards, including coordinating with the VSAC team. If needed, user acceptance testing of the MAT and Bonnie tools is completed.

The Cypress tool is also updated. Cypress is an open source testing tool used by vendors to certify their EHRs and health IT modules for calculating eCQMs, and is an official testing tool for the ONC EHR Certification Program.
7  **eCQM Standards-Based Guidance and Tools**

7.1  **Introduction**

eCQM specification development and maintenance has evolved into a highly structured process that requires input from multiple stakeholders (e.g., CMS, NLM, measure steward) as well as use of multiple standards-based guidance documents and tools. The tools used to implement the standards discussed in this chapter during eCQM development and maintenance include measure authoring and information gathering tools (e.g., MAT, VSAC), testing tools (e.g., Bonnie), as well as Jira (refer to Appendix B). The standards-based guidance and tools described here apply to de novo eCQMs, respecified eCQMs, and eCQM maintenance.

7.2  **Standards**

Initially, the information container was HQMF using the QDM for both the data model and the logic expressions. Now, the same HQMF container continues to use QDM as the data model and the logic expressions are written in CQL (Figure 22). The HL7 CQL-based HQMF IG, replaced QDM-based logic with CQL. This means that QDM data criteria specify only the data of interest (e.g., value sets, effective time, properties) for the eCQM, and the previous use of QDM expressions that captured interrelationships between data criteria, such as “starts after end of,” or identified subsets of data, such as min, max, last, and first, are now represented with CQL expressions. The standards used to develop eCQMs are HQMF, QDM, and CQL. The standard used to report eCQMs is QRDA. Each will be discussed briefly.

7.2.1  **Health Quality Measure Format**

HQMF is an HL7 standard for representing a health quality measure as an electronic XML document. Through standardization of a measure’s structure, metadata, definitions, and logic, the HQMF provides consistency and unambiguous interpretation. HQMF is a component of a larger quality end-to-end framework, which has evolved to a normative HL7 standard. HQMF-defined eCQMs ideally can be turned into queries that automatically gather data from the EHR data repositories and generate

---

Figure 22. Changes in eCQM Information Structure

Definitions:
HQMF – Health Quality Measure Format
CQL – Clinical Quality Language
QDM – Quality Data Model

---

21 Refer to eCQI Resource Center pages for eligible hospital/critical access hospital and eligible professional/eligible clinician measures for up-to-date examples of how eCQMs will appear as XML and HTML documents.
22 A CQM encoded in HQMF was initially referred to as an eMeasure. The terminology has evolved, and an electronically specified CQM is now referred to as an eCQM.
The HQMF Release 1 (R1) was published as a standard for trial use (STU) in 2009 and is the underlying structured representation used by the CMS MAT for eCQMs developed through June 2014. HQMF was updated to STU Release 2.1 (R2.1) in 2015 and updated to a normative standard in January 2017. The normative standard is the version currently in use.

The components of an HQMF document include a header and a body. The header identifies and classifies the document and provides important metadata about the measure such as general descriptions; numerator and denominator statements; the measure steward, measure type, and measure scoring; guidance; and definitions, as well as information about whether the measure is NQF-endorsed. The body contains the eCQM chapters. The chapters are:

- Population Criteria
- Definitions
- Functions
- Terminology
- Data Criteria (QDM Data Elements)
- Supplemental Data Elements
- Risk Adjustment Variables.

Population criteria can contain narrative descriptions, and all chapters contain formally encoded HQMF entries. For more information on the HQMF eCQM Chapters, refer to the HL7 HQMF IGs.

Any eCQM intended to be submitted for CMS consideration for review by the MAP and NQF endorsement must be submitted in HQMF. This process is assured when measure developers author their eCQMs in the MAT. The MAT, refer to Section 3, Chapter 7.3.1.1, Measure Authoring Tool, was developed under contract with CMS to aid in the creation of eCQMs.

7.2.2 Quality Data Model

The QDM is a standard information model adopted by CMS that describes the data needed to represent information necessary for electronic quality assessment. The QDM was initially established by the Health Information Technology Expert Panel (HITEP) convened by NQF in 2009. CMS is now the sponsor.

The QDM allows definition of a data element, which is the smallest possible unit of information that has precise meaning to communicate the data required within a quality measure. Each data element is comprised of:

- Category—a particular group of information that can be addressed in a quality measure.
- QDM datatype—the information category and its context of expected use for any given data element.
- Values—a single code or list of codes used to define the specific data element. For values that require coded data, there are code system recommendations used for a data element’s category (e.g., SNOMED CT, LOINC, RxNorm). Refer to Section 3, Chapter 17, Codes, Code Systems, and Datasets.
- Attribute—specific detail about a data element that further constrains the concept.

A data element is specified by selecting a category, the QDM datatype in which the category is expected to be found with respect to electronic clinical data, a value or value set drawn from an appropriate code
system, and all necessary attributes. Refer to Section 3, Chapter 6.2 Collaborative Measure Development Workspace and Data Element Repository for more information on the Data Element Repository. An example in Figure 23 shows:

- QDM Category – Laboratory Test
- QDM Datatype – Laboratory Test, Performed
- Value Set – High Density Lipoprotein (HDL)
- QDM Data Element – Laboratory Test, Performed: HDL.

Information about current and prior QDM versions can be found at the eCQI Resource Center.

The QDM continues to evolve through input from the QDM User Group that reviews measure development and implementation needs and evaluates resolution of QDM project Jira tickets. If a change to QDM is needed, there is a process to update the QDM and subsequently, the MAT. The process for changes to the QDM are outlined in the QDM User Group Charter.

7.2.3 Clinical Quality Language

CQL is an HL7 standard that is a clinically focused, author-friendly language enabling more precise measure specifications. CQL provides the ability to express logic that is human-readable, yet structured enough for processing a query electronically. CQL allows for a more modular, flexible, and robust expression of the logic and enables logic to be shared between measures and clinical decision support.

The ELM provides a more streamlined format for automated sharing of executable measure logic. The ELM file is “the machine-readable representation of the CQL that has been designed for sharing and implementation applications. The ELM file provides the semantics necessary to retrieve the correct data from the EHR” (Measure Authoring Tool User Guide, p. 56).

CQL has replaced the logic expressions previously defined in the QDM, and the QDM (beginning with v5.0) includes only the conceptual model for defining the data elements (i.e., the data model).

The CQL Formatting and Usage Wiki serves as a collaborative workspace for the development of CQL formatting conventions and usage patterns for the representation of logic within quality measures. All users have edit rights to be able to submit edits and add comments and concerns. The CQL Style Guide provides examples to standardize expression of measure concepts across eCQMs and define a uniform “look and feel” to eCQM logic using CQL. The guide focuses on a set of common best practices implemented across CQL-based eCQMs in CMS reporting programs and also promotes the use of
consistent language within the framework of CQL, including libraries, aliases, definitions, functions, and conventions. Refer to Appendix D to help with review of eCQM logic.

7.2.4 Quality Reporting Document Architecture

Once eCQMs are specified, tested, and implemented, the EHR system vendors turn the eCQM into queries that retrieve the necessary information from the EHR’s data repositories and generate quality data reports. eCQM reporting (i.e., the transmission format) is another important component of the quality reporting end-to-end framework. Individual and aggregate patient quality data can be transmitted to the appropriate agency using QRDA Category I (i.e., individual patient data) and Category III (i.e., aggregate patient data) reports, respectively. Both QRDA Category I and Category III are HL7 standards for reporting quality measures.

QRDA is an HL7 Clinical Document Architecture (CDA)-based standard. As such, the QRDA conforms to the HL7 CDA standard. The HL7 QRDA IGs describe the constraints on the CDA. CMS further constrains the base QRDA and publishes Implementation Guides and Schematrons for CMS reporting.

Each QRDA Category I report contains quality data for one patient for one or more quality measures. For each QDM datatype, there is a one-to-one mapping of each QRDA Category I template to its corresponding QDM-based HQMF template. This tight coupling helps to streamline the end-to-end process from eCQM specification to eCQM reporting.

Like a QRDA Category I report, a QRDA Category III report also contains a Measure Section that lists the eCQM(s) being reported and a Reporting Parameters Section that provides information about the reporting period. However, instead of reporting raw individual patient data, the report includes an aggregated summary for all patient populations from a measure (i.e., a total count of patients who meet the denominator population criteria of a measure within a health system over a specific period of time).

As depicted in Figure 24, healthcare organizations receive the measure specifications, expressed in HQMF using QDM and CQL and then report results to CMS using QRDA.

Figure 24. Connections between Standards
7.3 Tools

Figure 25 shows the connections between eCQM Standards and the tools used to help develop and test eCQMs.

7.3.1 Development Tools

7.3.1.1 Measure Authoring Tool

The MAT is a publicly available, web-based tool that is used by measure developers to create eCQMs. Measure developers use the MAT to express measure criteria using QDM and CQL. The tool enables measure developers to create their eCQMs in HQMF, a structured document format, using CQL without extensive knowledge of the HQMF or CQL standards. The MAT is maintained, supported, and updated as needed by modifications to the QDM, HQMF, CQL, and other standards to meet future measure authoring requirements. All changes to the MAT are evaluated and prioritized by the MCCB, which is coordinated by CMS and other federal agencies.

The MAT provides the capability to express complex measure logic and export measures in several formats, including a human-readable document that can be viewed in a web browser, the CQL-based HQMF eCQM is composed of a CQL file containing the CQL library in its entirety, an eCQM HQMF XML
document, an ELM XML document, and a corresponding ELM JSON file. Measure developers use both Bonnie and MAT to promote test driven development.

As shown in Appendix C, eCQM metadata are summarized in the order in which they are conventionally displayed as generated from the MAT.

A MAT account is free and is available for anyone completing the application process. The application process requires notarized paperwork and can take up to 1 week to be processed.

7.3.1.2 Bonnie

Bonnie is a software tool that enables eCQM developers to test and verify the behavior of their eCQM logic. The main goal of the Bonnie application is to reduce the number of defects in eCQMs by providing a robust and automated testing framework. The Bonnie application allows measure developers to load measures that they have constructed using the MAT. Developers can then use the measure metadata to build a synthetic patient test deck for each measure from the clinical elements defined during the measure construction process. By using measure metadata as a basis for building synthetic patients, developers can quickly and efficiently create a test deck for a measure. The Bonnie application helps measure developers execute the measure logic against the constructed patient test deck and evaluate whether the logic aligns with the intent of the measure. Bonnie also shows which sections of the eCQM are currently tested by the test deck, allowing measure developers to ensure that all logic in the measure is covered by tests. Additionally, because the Bonnie tool implements the evolving eCQM standards early in the standards update process, it helps to verify that the standards can be implemented more broadly. Bonnie testing is required for NQF submissions. Refer to the Bonnie User Guide for more information.

7.3.1.3 CQL-to-ELM Translator

The CQL-to-ELM Translator is a specification that describes a formal mechanism for translating the high-level CQL syntax into the canonical called ELM representation. The reference implementation is intended to be used in support of clinical quality framework implementations as a tool to enable CQL output to be uniformly and automatically translated into ELM XML or JSON documents for sharing and distribution to support implementation, integration, translation, and execution of CQL-based artifacts. The MAT uses the CQL-to-ELM Translator for validation of syntactically correct CQL content.

The Translator is an artifact of the HL7 CQL specification and is maintained by the CMS Standards contractor. The Translator is open source and available on GitHub.

7.3.1.4 Value Set Authority Center

VSAC is not specific to eCQMs, but is required for eCQMs. Information about the VSAC is found in Section 3, Chapter 17, Codes, Code Systems, and Datasets.

7.3.2 Certification Tools

7.3.2.1 Cypress

Cypress is an open source testing tool used by vendors to certify their EHRs and health IT modules for calculating eCQMs. Cypress is an official testing tool for the ONC Health IT Certification Program. Testing involves Cypress generating synthetic patient records for the subset of published eCQMs selected for certification and testing the ability of the EHR systems and health IT modules to accurately record, import, calculate, filter, and report eCQMs. The Cypress test data are available in QRDA Category I for
import into an EHR system. Cypress tests an EHR system’s ability to generate accurate QRDA Category I and Category III documents for reporting to CMS.

7.3.2.2 National Committee for Quality Assurance (NCQA) Testing Resource

NCQA’s eCQM testing method was approved by the ONC for use in the ONC Health IT Certification Program in June 2017. NCQA’s program tests and validates the integrity of the software code that produces the eCQM results. NCQA creates unique sets of sample data or “test decks” for each eCQM, developed from randomly generated patient-level test data. Learn more about the NCQA program at their website.
8 MASTER LIST OF DELIVERABLES AND DOCUMENT REQUIREMENTS

Each phase of the lifecycle has specific deliverables and requirements associated with that phase, which is outlined in each chapter. The lists are consolidated in this table. Measure development/maintenance contractors should confer with their COR and contractual documents as to which deliverables are required by the contract, as not all deliverables are required in every contract.

<table>
<thead>
<tr>
<th>Phase/Topic</th>
<th>Deliverable Number (MIDS USOW)</th>
<th>Deliverable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Gathering</td>
<td>3-1</td>
<td>Information Gathering Report (Summary Report of Environmental Scan and Empirical Analysis)</td>
</tr>
<tr>
<td>Information Gathering</td>
<td>3-2</td>
<td>List of Potential Measures</td>
</tr>
<tr>
<td>Information Gathering</td>
<td>3-3</td>
<td>Measure Information Form (MIF) or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>for eCQMs, Machine-readable Files (HQMF, ELM, and JSON files), HTML, and Link to the VSAC.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MIF is Updated in Several Phases of the Measure Lifecycle:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Requirement Associated with Convening a Technical Expert Panel (TEP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Requirement Associated with Soliciting Public Comments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Deliverable 4-4 Risk Adjustment Section of MIF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Deliverable 4-5 Draft Documentation Set</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Deliverable 4-6 Final Documentation Set</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Requirement Associated with All Three Types of Reevaluation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Requirement Associated with Measure Testing</td>
</tr>
<tr>
<td>Information Gathering</td>
<td>3-4</td>
<td>Measure Justification Form (MJF)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MJF is Updated in Several Phases of the Measure Lifecycle:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Requirement Associated with Convening a Technical Expert Panel (TEP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Requirement Associated with Soliciting Public Comments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Deliverable 4-4 Risk Adjustment Section of MIF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Deliverable 4-5 Draft Documentation Set</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Deliverable 4-6 Final Documentation Set</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Requirement Associated with All Three Types of Reevaluation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Requirement Associated with Measure Testing</td>
</tr>
<tr>
<td>Information Gathering</td>
<td>3-5</td>
<td>Business Case:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Initial Business Case</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Final Business Case</td>
</tr>
<tr>
<td>Information Gathering</td>
<td>3-6</td>
<td>Expert Input Report</td>
</tr>
<tr>
<td>Quality Measure Development and Reevaluation</td>
<td>4-1</td>
<td>TEP Composition Documentation Form</td>
</tr>
<tr>
<td>Quality Measure Development and Reevaluation</td>
<td>4-2</td>
<td>Measure Evaluation Report for Each Measure</td>
</tr>
<tr>
<td>Quality Measure Development and Reevaluation</td>
<td>4-3</td>
<td>Summary of TEP Evaluation of Measures</td>
</tr>
<tr>
<td>Phase/Topic</td>
<td>Deliverable Number (MIDS USOW)</td>
<td>Deliverable</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------</td>
<td>-------------</td>
</tr>
</tbody>
</table>
| Requirements associated with convening a TEP | | • Call for TEP Member Nominations  
• TEP Nomination Forms  
• TEP Charter  
• TEP Composition Documentation ([TEP Membership List])  
• TEP Meeting Materials and Minutes  
• Potential Measures Presented to the TEP  
• Measure Evaluation Report(s)  
• Updated MIF and MJF  
• TEP Summary Report |
| Requirements associated with soliciting public comments | | • Public Comment Call Web Posting Template Form  
• List of Stakeholders for Notification  
• MIF and MJF (for candidate measures)  
• Verbatim Public Comments  
• Responses to Public Comment  
• Public Display of Outcome Measures Related to the Call for Public Comment  
• Public Comment Summary Report |
| Quality Measure Development and Reevaluation | 4-4 | Risk Adjustment Section of the MIF |
| Quality Measure Development and Reevaluation | 4-5 | Draft Documentation Set (contains draft MIFs and MJFs) (Deliverables 3-3 and 3-4) |
| Quality Measure Development and Reevaluation | 4-6 | Final Documentation Set (contains final MIFs and MJFs) (Deliverables 3-3 and 3-4) |
| Quality Measure Development and Reevaluation | 4-7 | Draft CBE Endorsement Submission Materials |
| Quality Measure Development and Reevaluation | 4-8 | Final CBE Endorsement Submission Materials |
| Quality Measure Development and Reevaluation | 4-9 | Maintenance Reevaluation |
| Quality Measure Development and Reevaluation | 4-10 | Comprehensive Reevaluation |
### Deliverables

<table>
<thead>
<tr>
<th>Phase/Topic</th>
<th>Deliverable Number (MIDS USOW)</th>
<th>Deliverable</th>
</tr>
</thead>
</table>
| Requirements associated with all three types of reevaluation |  |  - An updated [MIF](#) (Deliverable 3-3) detailing all recommended changes to the measure.  
  - A document summarizing changes made, such as release notes, if not included with the updated [MIF](#) (Deliverable 3-3).  
  - An updated [MIF](#) (Deliverable 3-4) documenting the environmental scan results, any new controversies about the measure, and any new data supporting the measure’s justification.  
  - An updated [Maintenance Reevaluation Report](#) (Deliverable 4-9) reflecting the experience of the measure compared to the measure evaluation criteria.  
  - CBE endorsement maintenance online submission documentation:  
    - Annual update (regardless whether any change were made to the measure).  
    - Submission documentation for any material changes to the measure.  
    - For Comprehensive Measure Reevaluations, CBE endorsement maintenance documentation.  
  Completed NQF measure submission forms may be used for contract deliverables in lieu of the [MIF](#) and [MJF](#); permissible only if the contract allows for it. |
| Quality Measure Development and Reevaluation | 4-11 | Ad Hoc Reevaluation |
| Quality Measure Development and Reevaluation | 4-12 | Public Description of Quality Measures |
| Instrument/Item Development | 5-1 | Data Collection Tools/Assessment Instruments, Manuals or Instructions, and Updates |
| Instrument/Item Development | 5-2 | Summary Report of New/Revised Data Collection Instruments or Data Items |
| Testing/Validation | 6-1 | Measure Testing Plan |
| Testing/Validation | 6-2 | Measure Testing Summary Report |
| Testing/Validation | 6-3 | Risk Adjustment Methodology Report |
| Requirements associated with measure testing |  |  - Updated [MIF](#) (Deliverable 3-3)  
  - Updated [MJF](#) (Deliverable 3-4)  
  - Updated [Measure Evaluation Report](#) (Deliverable 4-4), if risk-adjusted. For eCQMs, the HTML file includes instructions where the complete risk adjustment methodology may be obtained |
<p>| Reports/Approval Packages | 7-7 | Office of Management and Budget (OMB)/PRA Materials |
| Implementation, Production, and Assessment | 8-1 | Timeline for Data Item and/or Quality Measure Implementation |
| Implementation, Production, and Assessment | 8-2 | Implementation Stakeholder Meetings |
| Implementation, Production, and Assessment | 8-3 | Question and Answer Support |
| Implementation, Production, and Assessment | 8-4 | Implementation Process Roadmap |</p>
<table>
<thead>
<tr>
<th>Phase/Topic</th>
<th>Deliverable Number (MIDS USOW)</th>
<th>Deliverable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementation, Production, and Assessment</td>
<td>8-5</td>
<td>Measure Calculation/Results</td>
</tr>
<tr>
<td>Implementation, Production, and Assessment</td>
<td>8-6</td>
<td>Program and Initiative Assessment Report</td>
</tr>
<tr>
<td>Requirements associated with production and monitoring</td>
<td></td>
<td>• Audit and Validation Reports</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Audit and Validation Appeals Reports</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Preview Reports (if required by the CMS program using the measure)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Periodic Measure Reports</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Analysis of the Measure Results</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Ad hoc Analyses (as requested by CMS)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Periodic environmental scans of journal literature, guidelines, and stakeholder feedback, as directed by CMS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Program and Initiative Assessment</td>
</tr>
<tr>
<td>Public Reporting/Compare Sites</td>
<td>9-1</td>
<td>Compare Site Files and Measures</td>
</tr>
<tr>
<td>Public Reporting/Compare Sites</td>
<td>9-2</td>
<td>Pre-Posting Preview Report</td>
</tr>
<tr>
<td>Public Reporting/Compare Sites</td>
<td>9-3</td>
<td>Implementation Algorithm</td>
</tr>
<tr>
<td>Access to Systems/Data</td>
<td>10-1</td>
<td>Data Use Agreement (DUA)</td>
</tr>
</tbody>
</table>
9 INFORMATION GATHERING

Information gathering is conducted via 10 steps, which may or may not occur sequentially:

- Identify the healthcare quality issue to be addressed and determine if it is in a priority area
- Conduct an environmental scan (Refer to Section 3, Chapter 10, Environmental Scan)
- Conduct an empirical data analysis, as appropriate
- Evaluate information collected during the environmental scan and empirical data analysis
- Conduct a measurement gap analysis to identify areas for new measure development
- Determine the appropriate basis for creation of new measures
- Apply measure evaluation criteria and propose a list of potential measures
- Submit the Information Gathering Report
- Prepare an initial list of measures or measure topics
- Explore possible data sources while considering feasibility (e.g., understanding the data captured in EHRs).

9.1 CONDUCT AN ENVIRONMENTAL SCAN

The environmental scan is an essential part of building the case for quality measures. It serves as the foundation for the measurement plan. Developing a broad-based environmental scan that includes a strong review of the literature, regulatory environment, economic environment, and stakeholder needs and capabilities will guide thinking and decision-making. A strong, comprehensive environmental scan will improve the likelihood of project success.

According to the MIDS USOW, measure developers can conduct the environmental scan through various methods, including literature review, clinical performance guideline search, interviews, or other activities. In the case of new measures, the measure developer must identify any applicable measures in current use that might be appropriate for the specific task order, which would occur through analysis of resources, including, but not limited to, employers, commercial plans, managed care plans, TRICARE, NQF, MedPAC, National Academy of Medicine, IHI, Veterans Health Administration (VHA), and the Department of Defense (DoD). Depending on the nature of the contract, and if deemed necessary, the measure developer may also conduct interviews or post a Call for Measures as part of the environmental scan.

The scan should consider quality priorities, as well as Medicare, Medicaid, and other payer top volume and top cost conditions, as appropriate. Under a given task order, the government might require the measure developer to conduct a literature review and scan web-based sources for relevant sites, papers, clinical practice guidelines, competing measures, and other reliable sources of information relating to the topic. The government might also require the measure developer to evaluate existing quality measures to support development of outcome and process measures that have established histories of quality or process improvement. Measure development contracts may be based on developing measures of a specific type (e.g., process, outcome), specific setting (e.g., hospital inpatient, ambulatory), specific condition (e.g., kidney disease, heart disease), specific domain (e.g., care coordination, population health, patient safety), or specific population (e.g., Medicare, Medicaid).

Among the many important areas to scan, measure developers must consider the IOM’s Six Aims of Care, outlined in Crossing the Quality Chasm, which include safety, timeliness, efficiency, effectiveness, equitability, and patient centeredness. Measure developers must explore the various dimensions of
quality to develop informative quality measures. The resulting report of the environmental scan will include several findings:

- Identification of related, similar, or competing measures, including opportunities for consolidation, harmonization, and alignment.
- Listing of clinical guidelines pertinent to the clinical domain or topic specified in the task order.
- Review of studies that document the success of measures in the same or similar healthcare setting or domain covered by the task order.
- Discussion of scientific evidence supporting clinical leverage points that might serve as a basis for the measure (e.g., importance).

The environmental scan includes a literature review (white and grey), clinical practice guidelines review, review of legislation and regulations and their implications on measurement (e.g., MACRA), evaluation of existing related measures, empirical data analysis, expert input (including the TEP and other experts), and stakeholder input—inclusive of all relevant stakeholders, including patients (Figure 26).

![Diagram of Environmental Scan Data Sources](image)

**Figure 26. Environmental Scan Data Sources**

Refer to Section 3, Chapter 10, Environmental Scan, and Section 4, Forms and Templates, for detailed instructions and an example outline for conducting an environmental scan.

### 9.2 Conduct an Empirical Data Analysis, as Appropriate

If data are available, conduct an empirical data analysis to provide statistical information to support the importance of the measure, identify gaps or variations in care, and provide incidence/prevalence information and other data necessary for the development of the business case. This empirical data analysis may also provide quantitative evidence for inclusion or exclusion of a set of populations or
geographic regions or other considerations for the development of the measure. Data analysis is documented in the Importance section of the MJF, and in the business case.

Empirical analysis can be used to test the feasibility of data elements required for a measure. Feasibility considerations that can be assessed empirically include data availability (including standardization) and accuracy of data information.

Feasibility concerns should be identified early in the development of the measure, which will enable measure developers sufficient time to replace or revise data elements, consider an alternative measure type, assess implementation burden versus value of measure, or recommend halting further development of the measure concept.

If risk-adjusted measures are being developed, a preliminary feasibility assessment should also be applied to the risk variables.

9.3 Evaluate Information Collected during Environmental Scan and Empirical Data Analysis

If there are related measures, evaluate the measures to assess whether they meet the needs of the measure development contract. A detailed description of harmonization concepts is covered in Section 3, Chapter 18, Measure Harmonization.

An adopted measure has the same numerator, denominator, data source, and care setting as its parent measure, and the only additional information to be provided pertains to the measure’s implementation (e.g., data submission instructions).

Examples of an adopted measure include:

- Measures developed and endorsed for physician- or group-level use are specified for submission to a physician group practice demonstration project and are proposed for a new physician incentive program.
- An existing Joint Commission hospital measure not developed by CMS is now added to the CMS measure set.

If a related measure is found with a measure focus appropriate to the needs of the contract, but the measure is specified for a different population, setting, or data source, it may be possible for the measure developer to respecify the measure for the new use and test for reliability and validity specific to the new population.

A respecified measure is an existing measure that a measure developer changes to fit the current purpose or use, which may mean changing the measure to meet the needs of a different care setting, data source, or population. Or, it may mean changing the numerator, denominator, or adding additional specifications to fit the current use.

An example of a respecified measure:

- A measure for screening adult patients for depression is found. The current contract requires mental health screening measures for adolescents. It would then be appropriate for the steward of the adult depression screening measure to expand the population in the measure to the

24 The NQF submission may be acceptable for this deliverable.
adolescent population, if the evidence base supports that expansion AND the measure is tested for reliability and validity within the new population.

Begin evaluating whether to respecify a measure by assessing the applicability of the measure focus to the measure topic or setting of interest. Is the measure focus of the existing measure applicable to the quality goal of the new measure topic or setting? Does it meet the importance criterion for the new setting or population? If the population changes or if the types of data are different, new measure specifications would have to be developed and properly evaluated for scientific acceptability and feasibility before a determination regarding use in a different setting can be made. Section 3, Chapter 16, Measure Technical Specification describes the standardized process.

For measures that are being respecified for use in a different setting, the unit of measurement usually does not need to undergo the same level of development as for a new measure. However, aspects of the measure need to be evaluated and possibly adjusted for the new setting to show the importance of the measure to each setting for which the measures may be used. Additional testing of the measure in the new setting is also required. Section 3, Chapter 22, Measure Testing and Chapter 24, Testing for Special Types of Measures provide further details of the process.

Empirical analysis may be needed to evaluate whether it is appropriate to respecify the measure for the new purpose. The analysis may include, but is not limited to, evaluation of answers to the questions:

- Are there changes in the relative frequency of critical conditions used in the original measure specifications when applied to a new setting/population (e.g., when the exclusionary conditions have increased dramatically)?
- Is there a change in the importance of the original measure in a new setting (e.g., an original measure addressing a highly prevalent condition may not show the same prevalence in a new setting or evidence that large disparities or suboptimal care found using the original measure do not exist in the new setting/population)?
- Are there changes in the applicability of the original measure (i.e., the original measure composite contains preventive care components that are not appropriate in a new setting such as hospice care)?
- If the existing measure is NQF-endorsed, are the changes to the measure significant enough to require resubmission to NQF for endorsement?
- Will the measure steward be agreeable to the changes in the measure specifications to meet the needs of the current project?
- If a measure is copyright protected, are there issues relating to the measure’s copyright that need to be considered?

These considerations must be discussed with the COR and the measure steward. NQF endorsement status may need to be discussed with NQF.

If a measure is copyright protected, there may be issues relating to its stewardship or to proper referencing of the parent measure. In either case, contact the measure steward for permission or clarification. Upon receiving approval from the original developer to use the existing measure, include the original detailed specifications for the measure in the Information Gathering Report.
9.4 **CONDUCT A MEASUREMENT GAP ANALYSIS TO IDENTIFY AREAS FOR NEW MEASURE DEVELOPMENT**

The purpose of the gap analysis is to identify measure types or concepts that may be missing for the measure topic or focus. Use the information collected from the environmental scan, measure gap analysis, and other information gathering activities to determine whether there are existing or related measures before deciding to develop new measures. If there are no existing or related measures that can be respecified or adopted, then it is appropriate to develop a new measure. Measure developers should:

- Develop a framework to organize the measures gathered. Refer to the NQF website for an example of a framework for evaluating needed measures and measure concepts. Through this analysis, the measure developer may identify existing measures that can be adopted or respecified, or identify new measures that need to be developed.

- Provide recommendations based on the results of the environmental scan, measure gap analysis, initial feasibility assessment, and other information collected during the information gathering process. After the COR has approved the recommendations, develop a set of candidate measures (i.e., newly developed measures, respecified measures, or measures adopted from an existing set).

9.5 **DETERMINE THE APPROPRIATE BASIS FOR CREATION OF NEW MEASURES**

If no existing measures are suitable for adoption or respecification, then new measures must be developed, and the measure developer will determine the appropriate basis for the new measures by gathering supporting information. The appropriate basis will vary by type of measure. This information will also contribute to the business case.

It is important to note that the goal is to develop measures most proximal to the outcome desired. Measure developers should avoid selecting or constructing measures that can be met primarily through documentation without evaluating the quality of the activity—often satisfied with a checkbox, date, or code—for example, a completed assessment, care plan, or delivered instruction. Measure developers should consider these guidelines in their determination of the appropriate basis for new measures:

- If applicable to the contract, and as directed by the COR, the measure developer may choose to solicit TEP input to identify the appropriate basis for new measures.
- For **outcome measures**, there should be a rationale supporting the relationship of the health outcome to processes or structure of care.
- For **intermediate outcomes**, there should be a body of evidence that the measured intermediate outcome leads to a desired health outcome.
- For **process measures**, there should be a body of evidence that links the measured process to a desired health outcome.
- For **structure measures**, the appropriate basis is the evidence that the specific structural elements are linked to improved care and improved health outcomes.
- For cost and **resource use**, the measures should be linked with measures of quality care for the same topic. Ways to link cost and resource use measures to quality of care are discussed in **Section 3, Chapter 20, Cost and Resource Use Measure Specification**.
- For all measures, it is important to assess the relationship between the unit of analysis and the decision-maker involved. Consider the extent to which processes are under the control of the...
entity being measured. The measure topic should be attributed to an appropriate provider or setting, which is not an absolute criterion. In some cases, there is “shared accountability.” For example, for measures of health functioning and care coordination, no one provider controls the performance results.

9.6 **APPLY MEASURE EVALUATION CRITERIA**

If many measures or concepts are identified, narrow down the list of potential measures by applying the measure evaluation criteria—especially, the **importance** and **feasibility criteria** to determine which measures should move forward. At a minimum, consider the measure’s relevance to the population; effects on healthcare costs; gaps in care; the availability of well-established, evidence-based clinical guidelines; and/or supporting **empirical evidence** that can be translated into meaningful quality measures. Other criteria may be included depending on the specific circumstances of the measure set. If applicable to the contract, and as directed by the COR, the measure developer may choose to solicit TEP input to help narrow the list.

In the early stages of measure development, while narrowing the initial list of potential measures to candidate measures, the measure developers may find it appropriate to use a spreadsheet to present information for multiple measures in one document.

Completing an **MIF** and **MJF** for each measure should begin as early as possible during the development process. Before presenting measures to the TEP, the measure developer may choose to use a modified MIF and MJF to display partial information as it becomes available. At the end of the project, fully document each potential measure on the MIF and MJF. The MIF and MJF are aligned with the NQF **measure submission**. By the end of measure development, these forms should be completed in their entirety for new measures or measures that are significantly changed from the original.

Analyze the literature review results and the guidelines found and organize the evidence to support as many of the measure evaluation criteria as possible. Document this information in the MJF. Measures that are adopted and NQF-endorsed do not require further documentation in the MJF.

The MJF should be completed for respecified measures. These measures will require evidence of the importance of the topic for a new setting or population. The measures may also need to be assessed for **reliability**, **validity**, **feasibility**, and **usability**.

9.7 **SUBMIT THE INFORMATION GATHERING REPORT**

Prepare a report for the COR that summarizes the information obtained from the previous steps. This report should include, but not be limited to, the items outlined in Sections 9.7.1 through 9.7.5.

9.7.1 **Summary of Literature Review (Annotated Bibliography)**

Provide per citation, by individual measure or, if directed by CMS, provide the information by measure sets:

- Search methods, including a complete explanation of all research tools used (i.e., online publication directories, keyword combinations, and Boolean logic used to find studies and clinical practice guidelines).
- Complete literature citations.
- Level of evidence and rating scheme used.

---

25 The NQF submission may be acceptable for this deliverable.
- Characteristics of the study (i.e., population, study size, data sources, study type, and method).
- Measure evaluation criteria (i.e., importance, scientific acceptability, usability, and feasibility) that the study addresses. Sorting the literature review by these criteria will facilitate the development of the measure justification in the later phases of measure development or reevaluation.
- Information gathered to build the business case for the measure:
  - Incidence/prevalence of condition in the population.
  - Major benefits of the process or intermediate outcome under consideration for the measure.
  - Unintended consequences of process or intermediate outcome and likelihood of their occurrence.
  - Cost statistics relating to cost of implementing the process to be measured, savings that result from implementing the process, and cost of treating complications that may arise.
  - Current performance of process or intermediate outcome and identifying gaps in performance.
  - Size of improvement that is reasonable to anticipate.
- Summary of findings.
- Other pertinent information, if applicable.

9.7.2 Summary of Clinical Practice Guidelines Review

Provide by measure set or, if needed, provide for individual measures in the set:
- Guideline name
- Developer
- Year published
- Summary of major recommendations
- Level of evidence
- If multiple guidelines exist, note inconsistencies and rationale for using one guideline over another.

9.7.3 Review of Existing Measures, Related Measures, and Gap Analysis Summary

Provide a summary of findings and measurement gaps, including:
- Existing related measures
- Gap analysis.

9.7.4 Empirical Data Analysis Summary

For new measures, provide:
- Data source(s) used (if available)
- Time period
- Methodology
- Findings.
For a measure reevaluation contract, use the Measure Evaluation Report to:

- Obtain current performance data on each measure
- Analyze measure performance to identify opportunities to improve the measure
- Provide a summary of empirical data analysis findings.

9.7.5 Summary of Solicited and Structured Interviews (If Applicable)

For solicited and structured interviews, include, at a minimum:

- Summary of overall findings from the input received
- Name of the person(s) interviewed, type of organization(s) represented, date(s) of interview, and area of quality measurement expertise if the input was from patients or other consumers
- List of interview questions used.

9.8 Prepare an Initial List of Measures or Measure Topics

Develop an initial list of measures based on the results of the previous steps. This list may consist of adopted, respecified, new measures, or measure concepts. This list of initial measures should be included in the Information Gathering Report. The measure developer may document this list of measures or concepts in an appropriate format. One option is to present the measures in a grid or table. This table may include, but is not limited to, the measure name, description, rationale/justification, numerator, denominator, exclusion or exception, and measure steward. The initial measure list is then reviewed and narrowed to create the list of potential measures. Work closely with the Measures Manager to ensure that no duplication of measure development occurs. Provide measure development deliverables (e.g., candidate lists) to the Measures Manager, who will help the measure developer identify potential harmonization opportunities.

9.9 Tools and Resources for Information Gathering

There are numerous tools and resources available to assist with and inform the information gathering process. For example:

9.9.1 CMS Data Element Library

The CMS DEL’s use in information gathering is similar to that of the eCQM DERep, but for post-acute care measures.

9.9.2 CMS Measures Inventory Tool

CMIT can be used in information gathering to identify related, similar, and competing measures that are currently in, proposed for, have been considered, were removed from, and are in development for CMS programs.

9.9.3 Electronic Clinical Quality Improvement Resource Center (eCQI Resource Center)

The eCQI Resource Center’s use in information gathering is to identify current eCQMs and those retired from CMS programs. As the CMD and DERep grow, they will be essential tools for identifying eCQMs in development and data elements and their associated value sets and DRCs to help avoid duplication and increase alignment and harmonization. It can help identify gaps in eCQM topics and data elements.
9.9.4 Environmental Scanning Support Tool (ESST)

The ESST makes it easier to complete the environmental scans required in the information gathering process to develop and maintain quality measures. The ESST reduces the time needed to scan literature from months to hours, saving immense amounts of resources.

9.9.5 National Quality Forum Quality Positioning System

NQF QPS can be used in the information gathering process similar to that of CMIT. QPS includes measures that are NQF-endorsed that are in CMS programs, but also those NQF-endorsed measures that are not in CMS programs. The QPS also includes measures that have lost endorsement and eCQMs that are approved for trial use. The QPS provides contact information for the measure steward.

9.9.6 National Quality Forum Incubator

The NQF Incubator can be used in the information gathering process to identify potential partners for measure development and testing.
10 **ENVIRONMENTAL SCAN**

These six steps are fundamental to creating an environmental scan:

- Frame a series of unambiguous, structured questions to limit the search to a specific problem set and prevent distraction by other interesting, but unrelated topics.
- Determine the framework for relevant work, including literature databases and search engines, keywords and phrases, inclusion and exclusion criteria, and domain experts.
- Assess the literature using qualitative techniques and quantitative metrics such as impact (e.g., number of times a paper is cited, number of page views), innovativeness, consistency with other works on the topic, recency of citations used in the work, seminal/originality, and quality of writing.
- Qualitatively evaluate and summarize the evidence. Evaluate the effectiveness and value of the data sources used, sample sizes, data collection methods, statistical methods, periods, and research findings.
- Interpret findings by evaluating the similarities and differences among the findings through expansion of the techniques cited above. From this, draw conclusions to inform data collection and analyses.
- Refine research questions and develop hypotheses. Generate a general analysis plan, including data sources and estimation procedures.

In addition, measure developers will want to (Choo, 1999):

- Be strategic in planning and managing the scan
- Formalize their scanning process
- Design the scan in collaboration with domain experts
- Manage the information obtained.

10.1 **LITERATURE REVIEW**

Conduct a literature review to determine the quality issues associated with the topic or setting of interest, and to identify significant areas of controversy if they exist. Document the tools used (e.g., search engines, online publication catalogs) and the criteria (i.e., keywords and Boolean logic) used to conduct the search in the search methods section of the Information Gathering Report. Whenever possible, include the electronic versions of articles or publications when submitting the report.

Use the measure evaluation criteria described in the Section 3, Chapter 23, Measure Evaluation and Chapter 25 Evaluation for Special Types of Measures, to guide the literature search and organize the literature obtained.

Evidence should support that there is a gap in achievement of Better Care, Healthy People/Healthy Communities, and/or Affordable Care associated with the measure topic, which is especially true if:

- Clinical practice guidelines are unavailable
- Guidelines about the topic are inconsistent
- Recent studies have not been incorporated into the guidelines.

If recent studies contribute new information that may affect the clinical practice guidelines, the measure developer must document these studies, even if the measure developer chooses not to base a measure
on the relatively new evidence. Emerging studies or evidence may be an indication that the guideline may change, and if it does, this may affect the stability of the measure.

Evidence should directly apply to the specified measure, if possible. State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population.

10.2 QUALITY OF THE BODY OF EVIDENCE

Across studies in the body of evidence, summarize the certainty or confidence in the estimates of benefits and harms to patients resulting from study factors (i.e., study design/flaws, directness/indirectness of the evidence to the measure, imprecision/wide confidence intervals due to few patients/events). In general, randomized controlled trials (RCTs), studies in which subjects are randomly assigned to various interventions, are preferred. However, this type of study is not always available because of the strict eligibility criteria; further, RCTs may not be appropriate in some cases. In these cases, non-RCT studies may be relied upon, including quasi-experimental studies, observational studies (e.g., cohort, case-control, cross-sectional, epidemiological), and qualitative studies. Review the:

- Quantity—Per NQF’s Review and Update of Guidance for Evaluating Evidence and Measure Testing, five or more RCT studies are preferred, but this is a general guideline. This count refers to actual studies, not papers or journal articles written about the same study.
- Consistency of results across studies—Summarize the consistency of the direction and magnitude of clinically/practically meaningful benefits over harms to patients across the studies.
- Grading of strength/quality of the body of evidence—If the body of evidence has been graded, identify the entity that graded the evidence, including the balance of representation and any disclosures regarding bias. Measure developers are not required to grade the evidence; rather, they assess whether the evidence was graded and what the grading process entailed.
- Summary of controversy and contradictory evidence, if applicable.
- Information related to healthcare disparities—Review these demographics across patient, clinical care, and outcomes, which may include referenced statistics and citations that demonstrate potential disparities (i.e., race, ethnicity, age, social risk factors, income, region, gender, primary language, disability, or other classifications) in clinical care areas/outcomes across patient demographics related to the measure focus. If a disparity has been documented, a discussion of referenced causes and potential interventions should be provided, if available.

Literature that is reviewed should include, but not be limited to:

- Published in peer-reviewed journals.
- Written recently (i.e., within the past five years).
- Based on data collected within the past 10 years or and end date within 10 years
- Unpublished studies or reports such as those described as grey literature. Government agencies such as the AHRQ, CMS, and Centers for Disease Control and Prevention (CDC) produce studies and reports that are publicly available but not peer-reviewed.
- If available, systematic literature reviews to assess the overall strength of the body of evidence for the measure topic; evaluate each study to report the grade of the body of evidence for the topic.
Additional resources may include, but not be limited to:

- IOM report: Finding What Works in Health Care Standards for Systematic Reviews

10.3 **CLINICAL PRACTICE GUIDELINES**

Measure developers should search for the most recent clinical practice guidelines applicable to the measure topic (i.e., written within the past 5 years). Clinical practice guidelines vary in how they are developed. Guidelines developed by American national healthcare professional organizations or federal agencies are preferred. However, guidelines and other evidence documents developed by non-American organizations may also be acceptable and should be assessed to determine whether they are a sufficient basis for measure development.

They should also document the criteria used for assessing the quality of the guidelines. When guideline developers use evidence rating schemes, which assign a grade to the quality of the evidence based on the type and design of the research, it is easier for measure developers to identify the strongest evidence on which to base their measures. If the guidelines were graded, indicate which system was used: United States Preventive Services Task Force (USPSTF) or Grading of Recommendation, Assessment, Development, and Evaluation (GRADE).

It is important to note that not all guideline developers use such evidence rating schemes. If no strength of evidence is noted, document whether the guideline recommendations are valid, useful, and applicable, using examples from peer-reviewed literature.

If multiple guidelines exist for a topic, review the guidelines for consistency of recommendation. If inconsistencies among guidelines exist, evaluate the inconsistencies to determine which guideline will be used as a basis for the measure and document the rationale for selecting the guideline.

Sources for clinical practice guidelines review include the U.S. Preventive Services Task Force, ECRI Guidelines Trust™, and the IOM report Clinical Practice Guidelines We Can Trust.

10.4 **EXISTING AND RELATED MEASURES**

Search for similar or related measures that will help achieve the quality goals. Keep the search parameters broad to obtain an overall understanding of the measures in existence, including measures that closely meet the contract requirements. Look for measures endorsed and recommended by multi-stakeholder organizations whenever applicable. Include a search for measures developed and/or implemented by the private sector. Determine what types of measures are needed to promote the quality goals for a topic/condition or setting. Identify measurement gaps for the topic area, as well as existing measures that may be adopted or respecified for the project. For example, if a contract objective is the development of immunization measures for use in the home health setting, it will be necessary to identify and review existing home health measures and immunization measures used in other settings such as nursing homes and hospitals.

The COR and Measures Manager can help measure developers identify measures in development to reduce duplication of efforts and to ensure related measures are developed with harmonization in mind. Search parameters include:
• Measures used in the same setting, but for a different topic
• Measures used in a different setting, but for the same topic
• Measures constructed in a similar manner
• Quality indicators
• Accreditation standards
• NQF-preferred practices for the same topic.

Use a variety of databases and sources to search for existing and related measures such as:

• CMIT
• NQF’s QPS
• Physician Consortium for Performance Improvement (PCPI).

Search for other sources of information such as performance indicators, accreditation standards, or preferred practices that may pertain to the contract topic. Although they may not be as fully developed as quality measures, quality indicators could be further developed to create a quality measure by providing detailed and precise specifications. Providers seeking accreditation must comply with accreditation standards such as those developed by the Joint Commission or the NCQA. Measures aligned with those standards may be easier to implement and be more readily accepted by the providers. These standards are linked to specific desired outcomes, and quality measures may be partially derived from the preferred practices reflected in the standards.

10.5 Stakeholder Input to Identify Measures and Important Measure Topics

There are multiple ways to obtain information from patients early in the process, including conducting informal conversations with patients, conducting focus groups, or by including patients or their caregivers on the TEP. Measure developers should prepare a plan for how patient input will be solicited, gathered, and meaningfully incorporated into measure development and maintenance processes and discuss a plan with their COR. Section 3, Chapter 13, Person and Family Engagement includes information on best practices and sources for patient recruitment.

If patient input is to be obtained, the engagement activities could be convened in phases—early during the information gathering process and later when measure concepts are more fully developed and the focus can be more technical. Patient input may be obtained during the earlier, less technical, phases of TEP discussions. A COR may ask to discuss a plan for obtaining the patient perspective.

If applicable to the contract, and as directed by the COR, the measure developer may also contact and interview measure experts, SMEs (including clinicians and EHR system implementers), relevant stakeholders, and other measure developers to identify any measures in use or in development that are relevant to the topic of interest or to offer suggestions regarding appropriate topics for measure development. These or other experts may also be used to provide information about feasibility, importance, usability, and face validity early on before actual measure development begins. Details of how to conduct a TEP and other stakeholder meetings are covered in Section 3, Chapter 12, Technical Expert Panel.
10.6 **Call for Measures**

While conducting the environmental scan, if insufficient numbers or types of measures have been identified, discuss the situation with the COR to determine whether a Call for Measures is needed. If CMS approves, the measure developer may issue a Call for Measures to the public. Work with the COR to develop a list of relevant stakeholder organizations to notify that a Call for Measures is being issued.

Measure developers can notify relevant organizations or individuals about the Call for Measures before the posting goes live on the website. Electronic means can be used to notify the stakeholder community about upcoming Calls for Measures. Other, more targeted communication can be used to notify relevant stakeholder organizations that can, in turn, notify their members. Relevant stakeholder groups may include, but are not limited to, quality alliances, medical societies, scientific organizations, and other CMS measure developers. In the Call for Measures, a measure developer may request stakeholders to submit candidate measures or measure concepts that meet requirements of the measure contract. The measure developer then determines whether the steward of those measures or measure concepts is willing to expand the measures for use by CMS. A call period should be a minimum of 14 days, but 30 days is recommended. Communicate and coordinate with the point of contact from the Measures Management team to post the call at the Call for Measures website. Use the [Call for Measures Web Posting form](#).

It is important to note that this Call for Measures is for information gathering and should be distinguished from other calls during measure implementation. A Call for Measures during the implementation phase of development seeks fully developed measures that will be considered for implementation in CMS programs. [Section 2, Chapter 5, Measure Implementation](#), covers these types of calls.

Compile a list of the initial measures received during the Call for Measures and evaluate these measures using the measure evaluation criteria. If an existing measure is found with a measure focus appropriate to the needs of the contract, but the population is not identical, it may be possible for CMS to collaborate with the steward of the original measure to discuss issues related to stewardship, maintenance, and testing.
11 **BUSINESS CASE**

The business case documents all anticipated impacts of a quality measure, including but not limited to, financial outcomes and the resources required for measure development and implementation. Despite what the name suggests, the business case is not limited to a description of economic benefits. Impacts and outcomes resulting from quality improvement through measure implementation may include lives saved, costs reduced, complications prevented, clinical practice improved, and patient experience enhanced.

The anticipated benefits made explicit in the business case should outweigh the costs and burden of collection and implementation for the specific quality measure. All potential positive and negative impacts should be evaluated and reported (Figure 27). For example, to reduce mortality through early detection and treatment, there may be increased costs and potential complications of screening tests. The business case should demonstrate:

- Why the measure is needed and how it will further the aims and objectives of CMS.
- The value of the measure and why it is the best balance of cost, benefits, and risks.
- The viability of the measure as it relates to the healthcare sector’s ability to respond.
- Realistic and affordable costs.
- Sufficient capacity within the system to implement the measure.

Benefits from the quality improvement efforts associated with measures described in the business case include:

- Better care through improvement in the quality of care provided and positive influence on patients’ perception of their care.
- Better health through reduction in mortality and morbidity and improvements in quality of life.
- More affordable care through cost savings.

By documenting the potential improvement anticipated from implementing a specific measure, the measure developer can make a strong case explaining why CMS should invest resources in the development (or continued use) of the specific measure in its quality initiatives. At a minimum, the business case of a measure should state explicitly, in economic and societal terms, the expected costs and benefits of the measure.

The business case for a measure applies information gathered, as well as supports the measure importance evaluation criterion, by providing supplementary information to create a model that predicts performance of the measure and the impact it will have on health and financial outcomes.
The formal business case for a measure supports measure evaluation during its initial development and facilitates reevaluation during measure maintenance. Development of the business case starts early during measure conceptualization, is enhanced throughout measure development, and should be used to compare actual results during measure reevaluation and maintenance. Therefore, communication with the COR regarding the business case should be ongoing throughout the Measure Lifecycle, as delineated in Table 11.

Table 11. The Business Case is an Ongoing Process that Occurs throughout the Measure Lifecycle

<table>
<thead>
<tr>
<th>Measure Conceptualization</th>
<th>Measure Specification</th>
<th>Measure Testing</th>
<th>Measure Implementation</th>
<th>Measure Use, Continuing Evaluation, and Maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data gathered to assess pros and cons of measure implementation and to provide information as to whether continued development is warranted</td>
<td>Initial business case prepared and submitted to CMS</td>
<td>Final business case submitted and reviewed</td>
<td>Update business case based on implementation</td>
<td>Assess measure performance as it pertains to the business case Determine whether business case adequately captures benefits, outcomes, and costs</td>
</tr>
</tbody>
</table>

The importance criteria in the Measure Evaluation Report and in the Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement contain requirements for information that will be used to begin developing a business case. The guidance provided here and in the Business Case Form Instructions will help measure developers to identify additional information to collect and to construct a case to meet CMS requirements. To the greatest extent possible, CMS has aligned its Business Case Form with NQF’s Measure Submission Form. In some cases, a measure developer may be able to use text from their NQF Submission Forms to complete their CMS Business Case Form and vice versa. This practice is accepted and encouraged by CMS as it aligns with Lean quality improvement strategies.

Figure 28 diagrams business case inputs and the impact of the business case throughout the Measure Lifecycle, and Figure 29 shows the overall flow of inputs to business case development.
Figure 28. Inputs and Uses for the Business Case
The business case should be evaluated during measure development and maintenance. Evaluation of the strength of the business case is ongoing during measure development and is used to justify continued development of the measure. This business case will provide CMS with information when considering implementation of the measure in a program. This information can be provided to the MAP to inform their deliberations.

Strategies for measure evaluation in the context of its business case might include a maturity model for how a measure concept is intended to change over time, from structure, process, and the phases of an outcome. Another strategy might include a “portfolio” of measures with composites, where the portfolio is directed toward achieving some stated goal(s) for a patient-condition population.

The business case and the predictions about measure performance used to inform decision-making during measure development and selection for use should be compared against actual performance after the measure is implemented. If anticipated improvements in health, provider care performance, and increased cost savings are demonstrated as predicted, then the measure is succeeding with regard to the business case. If the anticipated improvements are not being realized, then the measure developer should reexamine the data, reevaluate the justification for the measure, and analyze why the improvements are not occurring. The business case should be adjusted for any changes in the environment or if the initial assumptions need to be revised. For annual updates of measures in use and continuing evaluation, simply reporting performance relative to predictions may be sufficient. For the
comprehensive reevaluation, a full analysis should be conducted, and the report should include recommendations for improvement.

Measure developers should submit an initial business case during the measure conceptualization process and present a final business case before measure implementation begins. At the time of the initial submission, some of the data and details may be limited and, as a result, some fields in the template may not be completed. However, the measure developer is expected to update the initial business case with complete, detailed responses to all items prior to measure implementation and throughout the measure development process. Measure implementation should not begin until the business case has been reviewed and approved by CMS.

11.1 MMS BUSINESS CASE BEST PRACTICES

There are five key elements in a well-constructed business case. The executive summary should focus on what is available and provide a concise, high-level overview (maximum 500 words). Key elements that may be included in the executive summary include:

- Precise Statement of Need
- Business Impact
- Proposed Solution/Alternatives
- Benefits Estimation
- Cost Estimation.

11.2 BUSINESS CASE TEMPLATE

The Business Case Template provides the elements required to construct a full business case. It includes prompts to direct the measure developer to consider the quality gaps that exist, the benefits that can be expected to accrue, the costs of implementing the measures, and a time trajectory when CMS can expect to realize the benefits. The Business Case Template also includes prompts that inform measure developers when fields request information that is also required for an NQF business case submission so that measure developers know when it may be possible to use existing materials to complete the form. The business case will be used during measure maintenance as a comparison to the actual data and to the performance of the measure. Additional elements may be required based on the types of MUD and maintenance. Consult with the COR on the types of information and final format of the business case required.
12 TECHNICAL EXPERT PANEL

When developing measures, it is important to obtain input from experts. TEPs should include stakeholders such as persons/family members and providers as well as recognized experts in relevant fields such as clinicians, statisticians, quality improvement experts, methodologists, and other SMEs. TEP members are chosen based on their expertise, personal experience, diversity of perspectives, background, and training. The membership should also reflect geographic and organizational diversity as well as the variety of organization types that may have an interest in the topic.

A TEP for an eCQM should include recognized SMEs in relevant fields such as:

- Implementers of EHR systems—clinicians with personal knowledge of EHR workflow
- Clinical informaticists
- EHR/health IT vendors—preferably at least two vendors
- Programmers
- Coding experts
- Other measure developers—the collaborative process encouraged by the MIDS
- Current EHR users (e.g., staff from measure testing sites).

12.1 TIMING OF TEP INPUT

TEP timing will depend on the type and focus of the measure or concept under development. If the developer holds the TEP early during the contract period, then the measure developer should post a call for the panel immediately upon contract award. Best practices from developers suggest posting the TEP call concurrent with the environmental scan, literature review, and other tasks that require TEP review. This timing makes findings available for review in advance of and during the TEP meetings. Occasionally, developers may find it necessary to convene a smaller, more focused group of SMEs, instead of the entire TEP, to provide specific expertise (e.g., on technical aspects of coding measure specifications or EHR clinical workflow). These smaller groups can inform the larger TEP on measure feasibility.

Consider obtaining TEP input at these points during the Measure Lifecycle:

Measure Conceptualization:

- Gathering information to give input on topics and importance
- Refining the candidate measure list
- Applying the measure evaluation criteria to the candidate measures
- Conducting feasibility assessment (i.e., the TEP should assess the feasibility of alternative methods to address the measurement opportunity, such as when a measure originally intended to be an eCQM was determined to not be feasible as an eCQM, but is feasible as a chart-abstracted measure); refer to the NQF website for the Feasibility Assessment Scorecard.

Measure Specification:

- Constructing technical specifications
- Risk-adjusting outcome measures.
Measure Testing:

- Analyzing test results
- Reviewing updated measure evaluation and updated specifications.

Measure Implementation:

- Responding to questions or suggestions from the NQF Steering Committee, public comment, and stakeholder input.

Measure Use, Continuing Evaluation, and Maintenance:

- Reviewing measure performance during comprehensive reevaluations
- Meeting as needed to review other information, specifications, and evaluation.

For most measure development contracts, measure developers will convene several TEP meetings, either virtual and/or face-to-face. During early TEP meetings, the members will review the results of the environmental scan and clarify measure concepts. They will also evaluate the list of potential measures and narrow them down to candidate measures. During subsequent meetings, the TEP will review and comment on the draft measure specifications, review the public comments received on the measures, and evaluate the measure testing results.

After implementation, measure maintenance plans should include TEP review of measure performance. The measure developer should continue conducting environmental scans of the literature about the measure; watch the general media for articles and commentaries about the measure; and scan the data that are being collected, calculated, and publicly reported. Results of these scans will provide information about measure performance, unintended consequences, and other issues for TEP review. During maintenance, TEPs should also compare measure performance to the business case of impact on quality. Refer to Section 2, Chapter 6, Measure Use, Continuing Evaluation, and Maintenance, for details of the procedures for TEP involvement in comprehensive reevaluation, annual updates, and ad hoc reviews.

In addition to developing measures that address measurement gaps, the measure developer should keep an overall vision for discerning the breadth of quality concerns and related goals for improvement. The developer should direct and encourage the TEP to think broadly about principal areas of concern regarding quality as they relate to the topic or contract at hand. Finally, at the end of the measure development process, the developer should be able to show how the recommended measures relate to quality priorities and Meaningful Measures goals.

CMS strongly recommends that developers include a patient or caregiver representative on the TEP roster as an effective way to ensure input on the quality issues that are important to patients. Although consumer and patient advocacy organizations participation may be desirable, their participation is not a substitute for actual patients.

12.2 TEP STRUCTURES

12.2.1 Traditional TEP Structure

Measure developers may follow a traditional TEP structure model in which a new TEP is selected and convened each time a new measure is to be developed. Under this model, TEPs are convened at the beginning of each measure development process, with measure developers undertaking a lengthy and resource-intensive nomination and review process as they are in the information gathering stage.
One challenge with this structure is the lack of opportunity to solicit input from stakeholders or eventual TEP members in the early stages of measure development, since a formal TEP is convened concurrent with the early measure development activities. Once a TEP is formed, expertise is solicited from the whole TEP on all aspects of the measure, which can often lead to confusion or feelings of exclusion by patient and caregiver members who lack detailed statistical knowledge to actively participate in the more intensive technical reviews. A standing TEP structure can help address some of these issues and concerns.

### 12.2.2 Standing TEP Structure

Some measure developers have migrated toward the creation of a standing TEP structure. Under this model, the measure developer nominates and gathers a standing TEP with a 2- or 3-year term of membership. This TEP has a diversity of membership with broad-based expertise (e.g., policy and program, measure development, clinician, patient/advocate, technical) that enables review of all general aspects of measures that the developer is producing across a multiple-year measure cycle. The standing TEP meets approximately once a quarter for several hours to consider the policy surrounding each of the MUD or considered for future development. This cross-cutting focus enables the standing TEP to view and help problem solve across the portfolio of MUD.

In concert with the standing TEP, the measure developer also convenes a series of expert working groups through targeted outreach. These working groups are condition- or measure-specific and are populated by SMEs (e.g., statisticians, specialty clinicians) with targeted expertise and a narrow focus to view and solve problems on a specific measure. They may also include standing TEP members with expertise in the specific topic. These experts meet in smaller groups more frequently than the standing TEP, and for shorter periods of time, to dive deep into the technical aspects of a measure. The expert working groups give guidance on their specific measure to be considered by the standing TEP, which will take their recommendation(s) into account in the broader context of the program.

Advantages of the standing TEP structure include:

- Time and resource efficiency through avoidance of a full TEP nomination process for every measure
- Continuity, perspective, and programmatic knowledge within the standing TEP membership
- Trust building among TEP members who meet regularly and become acquainted with each other
- Less alienation and confusion for patient and caregiver representatives, since technicalities are tackled separately in the expert working group
- Combination of broad and narrow feedback results from differing perspectives.

Disadvantages of the standing TEP structure include:

- Potential disagreement between expert working group and standing TEP
- More frequent meetings.

### 12.3 Steps for the TEP

The exact order and level of detail required for the steps in convening a TEP may vary depending on the phase of the Measure Lifecycle, but the same general process should be followed. The steps for convening a TEP are:
• Draft TEP Charter and consider potential TEP members for recruitment
• Complete Call for TEP Web Page Posting form
• Notify relevant stakeholder organizations
• Post Call for Nominations following COR review
• Select TEP and notify the COR of the membership list
• Select chair or meeting facilitator
• Post TEP Composition documentation (membership) list and projected meeting dates
• Arrange TEP meetings
• Send materials to the TEP members
• Conduct TEP meetings and take minutes
• Prepare TEP Summary Report and propose recommended set of candidate measures
• Post TEP Summary Report.

12.3.1 Draft the TEP Charter and Consider Potential TEP Members for Recruitment

Draft the charter using the TEP Charter Template. This draft will be ratified at the first TEP meeting. The draft is important so that prospective TEP members may know the purpose and level of commitment required. The primary items to consider are:

• TEP goals and objectives
• TEP scope of responsibilities and how its input will be used by the measure developer
• TEP use of the Measure Evaluation criteria
• Estimated number and frequency of meetings
• Interest in participating in future maintenance activities.

The TEP’s role may include activities such as working with the measure developer to develop the technical specifications and business case, review testing results, and identify potential measures for further development or refinement. Specify how the TEP input will be used by the measure developer. Describe clearly how issues of confidentiality, particularly for patients/family/caregivers representatives, will be handled in the TEP reports. The measure developer should also consider the expertise of the individual members needed for the TEP and include balanced representation.

Additionally, since the voice of the patient is required in the TEP process, the measure developer is strongly encouraged to recruit a patient, family member of a patient, or a caregiver who can adequately provide input based on patient experiences. Section 3, Chapter 10.5, Stakeholder Input to Identify Measures and Important Measure Topics, provides more details about patient and caregiver input into TEP deliberations.

12.3.2 Complete the Call for TEP Web Page Posting Form

TEP recruitment begins with the Call for TEP members. Use the Technical Expert Panel (Call for TEP) Web Page Posting form to document information. The Call for TEP documents should be written in plain language that non-expert participants can understand. These items should be included in the Call for TEP:

• Overview of the measure development project
• Overall vision for discerning the breadth of quality concerns and related goals for improvement identified for the setting of care
• Project objectives
• Measure development processes
Types of expertise needed
- Information from the draft charter that explains the objectives, scope of responsibilities, etc.
- Expected time commitment and anticipated meeting dates and locations, including any ongoing involvement that is expected to occur throughout the development process
- Instructions for required information (e.g., TEP Nomination form, letter of intent)
- Information on confidentiality of TEP proceedings and how the TEP summary will be used
- Measure developer’s email address for submitting TEP nominations and questions.

12.3.3 Notify Relevant Stakeholder Organizations

It is important to publicize the Call for TEP nomination. Notify stakeholder organizations regarding the Call for TEP nominations before the posting goes live or simultaneously with the posting. The purpose of notifying the stakeholder organizations is to seek potential nominations for the TEP. Contacts at the organizations may choose to nominate specific individuals who may fill a need, or they may help disseminate information about the Call for TEP nominations. Share the list of relevant stakeholder organizations for notification with the COR for review and input.

Relevant stakeholder groups to notify of the Call for TEP may include, but are not limited to:
- Organizations that might help with recruiting appropriate patients or their caregivers
- Quality alliances
- Medical and other professional societies
- Setting-specific associations (e.g., American Hospital Association, American College of Emergency Physicians)
- Scientific organizations related to the measure topic
- Provider groups that may be affected by the measures
- NQF measure developer groups
- EHR and interoperability standards development organizations and industry organizations involved with clinical data collection and exchange
- Clinical data registries
- Other measure developers.

Individuals and organizations should be aware that the persons selected for the TEP represent themselves and not their organization. TEP members will use their experience, training, and perspectives to provide input on the proposed measures.

12.3.4 Post the Call for Nominations Following COR Review

Work with the Measures Manager to post the approved Technical Expert Panel (Call for TEP) Web Page Posting form and TEP Nomination forms on the dedicated CMS MMS page. Information required for the Call for TEP and TEP nomination is included in the template forms. The posting process for the Call for TEP is the same as described earlier in this chapter.

Developers submit their TEP nomination packages to MMS Support for posting on the CMS TEP web page. The CMS Events Calendar also lists dates when TEP nomination periods open and close. The calendar can be found on the CMS Resource Materials web page.

If an insufficient pool of candidates is received during the Call for TEP nomination period, the measure developer should alert the COR, who will decide to either approach relevant organizations or individuals to solicit candidates, or to extend the Call for TEP nomination period. If patient recruitment efforts are
not successful, alternative ways to find patients or caregivers should be considered and documented in the TEP Summary Report.

12.3.5 Select TEP and Notify the COR of the Membership List

The average TEP ranges from 8 to 15 members. This number may be larger or smaller depending on the nature of the contract and level of expertise required. Contracts for multiple measure sets or measures for multiple topics may require multiple TEPs to function simultaneously or within a larger TEP. Individual members of the TEP may represent multiple areas of expertise.

Select a balanced panel that includes nationally recognized experts in the relevant fields, including clinicians (i.e., physicians, pharmacists, and registered nurses), statisticians, quality improvement experts, methodologists, consumers, experienced measure developers, and EHR vendors to communicate and collaborate with the measure developer to develop the technical specifications and business case for measure development. Each TEP “shall include explicit incorporation of the patient perspectives and preferences in measure development through patient and/or caregiver participation on TEPs. Each TEP shall have at minimum one patient or caregiver on its roster in order to provide input into quality issues that are important to patients.”

The measure developer proposes the list of TEP members to CMS.

Consider these factors when choosing the final list of TEP members:

- Geography—Include representatives from multiple areas of the country who show a diversity of geographic characteristics, such as from rural and urban settings.
- Expertise in the subject matter of the measure.
- Diversity of experience—Consider individuals with diverse backgrounds (e.g., different types of clinicians and information technology professionals) and experience in different types of organizations and organizational structures.
- Include patients, patient advocates, caregivers, caregiver advocates.
- Affiliation—Include members not predominately from any one organization.
- Fair balance—Make a reasonable effort to have differing points of view represented.
- Availability—Select individuals who can commit to attending meetings, whether they are face-to-face or via telephone, and who can be accessible throughout the performance period of the measure developer’s contract.
- **Conflict of Interest**—Review of TEP members—Following nominations by a measure developer, CMS may review TEP nominees for potential conflicts of interest to prevent, to the extent possible, such conflicts, or the appearance thereof, in the TEP’s performance of its responsibilities. Nominees may be periodically monitored throughout their membership on the TEP for emergence of new conflicts of interest and to ensure they remain in good standing with the Medicare and Medicaid Programs.

TEP participants, including patients, should understand that their input will be recorded in the meeting minutes. TEP proceedings will be summarized in a report that is disclosed to the public. If a participant has disclosed personal data by his or her own choice, then that material and those communications are not deemed to be subject to confidentiality laws. In general, project reports should not include personally identifiable medical information. Answer any questions that participants may have about confidentiality and how their input will be used.

---

26 Quotes from MIDS 2018 Umbrella IDIQ Statement of Work.
Prepare a TEP membership list to document the proposed TEP member’s name, credentials, organizational affiliation, city, state, and area of expertise and experience. Include brief points to clearly indicate why a TEP member was selected. Additional information, such as TEP member biographies, may also be sent to the COR. Notify the COR about the TEP membership list within 1 week after the close of the posting. Confirm each member’s participation on the TEP.

12.3.6 Select Chair or Meeting Facilitator

Prior to the first TEP meeting, select a TEP chair (and co-chair, if indicated) who have either content or measure development expertise. It is important that the meeting is guided by a person with strong facilitation skills to:

- Convene and conduct the meeting in a professional and timely manner
- Conduct the meeting according to the agenda
- Recognize speakers
- Build consensus.

The TEP chair should be available to represent the TEP at the NQF Steering Committee meetings and follow-up conference calls. Additionally, all TEP members need to be available for potential conference calls with the measure developer to discuss NQF recommendations.

Some measure developers may choose to add a meeting facilitator to help with some of these tasks. In this case, a TEP chair must still be identified.

12.3.7 Post the TEP Composition Documentation (Membership) List and Projected Meeting Dates

Finalize the membership list (with COR approval) and complete the TEP Composition (Membership) List Template. Use the TEP Composition (Membership) List Web Page Posting form that includes the meeting schedule. Patients, family members, and caregivers included on the TEP who indicated that they wanted their name to remain confidential on the TEP nomination will be identified as “Patient,” “Family Member,” and “Caregiver” on the posted membership list. Include the dates of the TEP meetings in the document. The information should be available until the TEP Summary Report is removed from the website, within 21 calendar days or as directed by the COR.

12.3.8 Arrange TEP Meetings

Organize and arrange all TEP meetings and conference calls. TEP meetings may occur face-to-face, virtual, or a combination of the two. If an in-person meeting is required, the measure developer should plan the meeting date, time, and venue, and help participants with travel and hotel arrangements, as needed.

The measure developer may decide that additional SMEs and staff are needed to support the TEP, including data management and coding representatives, EHR experts, health informatics personnel, and statisticians/health services researchers. These SMEs can contribute summarized technical information to the TEP for consideration.

12.3.9 Send Materials to the TEP

Send the meeting agenda, meeting materials, and supporting documentation to the COR and TEP members at least 1 week before the meeting. For TEP lay members (i.e., patients and caregivers), consideration must be given to present the materials in a manner that they will be able to understand. Patients should not be burdened with detailed technical documents.
At a minimum, prepare and disseminate:

- Instructions on the measure evaluation criteria and how they should be applied by the TEP. Materials should also indicate how the measure developer plans to use the TEP’s evaluation and recommendations.
- The list of initial or potential measures identified by the measure developer. Depending on the number of measures that the TEP will review, the measure developer may modify or shorten both the MIF and MJF.27
- Measure developers may modify the MIF to suit their contract needs. For example, the contract may not require the measure developer to develop detailed specifications, so a much shorter summary of the measure information could be used. Alternatively, measure developers who have identified many potential measures may present the information in a grid or table. This table may include, but is not limited to, the measure name, description, rationale, numerator, denominator, and exclusion.
- The TEP Charter, for ratification at the first meeting, and to orient members to their roles and responsibilities.
- Other documents, as applicable.

Remind TEP members that they must disclose any current and past activities that may cause a conflict of interest. If a member’s status changes and a potential conflict of interest arises at any time while a member is serving on the TEP, the TEP member is required to notify the measure developer and the TEP chair.

12.3.10 Conduct the TEP Meetings

It is recommended that TEP discussions be held in two phases. Measure developers can determine the timing of these phases during measure development in consultation with their COR. For example, Phase 1 TEP discussions may be held concurrently during the information gathering phase when collecting information on useful and important measure concepts.

The goal of Phase 1 is to develop an initial list of measure concepts. Patient/caregiver participation is mandatory in these Phase 1 TEP discussions. Phase 2 is focused on evaluating the measures for further development.

12.3.10.1 Phase 1 TEP Meetings

Phase 1 should focus primarily on discussions about the importance and usability of measure concepts and potential measures to the identified patient population. Given that, patient/caregiver input into the Phase 1 TEP discussions is crucial. Measure developers should pay attention to patient/caregiver ideas, comments, and points of view about the potential measures and concepts during these Phase 1 discussions.

During the initial Phase 1 TEP meeting, review and ratify the TEP Charter to ensure participants understand the TEP’s role and scope of responsibilities. Summarize the findings of the literature review and the environmental scan. Discuss any overall quality concerns such as measurement gaps, alignment across programs and settings, and overarching goals for improvement. Emphasize presenting materials in a manner that lay members of the TEP will be able to understand to highlight the important role of the patient/caregiver voice in measure development.

---

27 The NQF submission may be acceptable for these deliverables.
By the end of Phase 1 TEP discussions, the measure developer should be able to identify measures/measure concepts that are deemed important, usable, and valuable by the patient(s) on the TEP, which will be discussed further in Phase 2 TEP meetings.

12.3.10.2 Phase 2 TEP Meetings
Phase 2 TEP meetings may involve details about the feasibility of the measures and in-depth technical discussions about acceptability of the evidence base for the measures, face validity, and adequacy of measure specifications. Phase 2 technical discussions may be overwhelming or burdensome for some patient TEP members. Patients and caregivers may be excused at this point, if they wish. However, they may stay, if appropriate, and they wish to remain in the meeting.

For the Phase 2 TEP meetings, measure developers should compile a list of measures finalized after Phase 1. Depending on the specifics of the measure contract, the measure developer may focus TEP guidance on one or more measure evaluation criteria based on the TEP’s expertise. However, the TEP should be allowed to provide input on any or all measure evaluation criteria as part of its deliberations. Section 3, Chapter 23, Measure Evaluation and Chapter 25 Evaluation for Special Types of Measures provide descriptions of the evaluation criteria.

Measure developers can use the TEP discussions as input to complete the Measure Evaluation Report for each measure after the meeting. Alternatively, the measure developer may conduct a preliminary evaluation of the measures and complete a draft Measure Evaluation Report before the TEP meeting. These drafts can be presented to the TEP for discussion. Either way, maintain transparency by notifying the TEP regarding the way its evaluations are used.

The Measures Manager is available to work closely with measure developers throughout the TEP process. The Measures Manager can provide feedback on TEP process deliverables such as candidate measure lists, charters, and other meeting materials.

12.3.11 Prepare TEP Summary Report and Propose Recommended Set of Candidate Measures
Keep detailed minutes of all TEP meetings whether they are conducted face-to-face or via teleconference. TEP conference calls may be recorded to document the discussion. Announce to the participants if the session is being recorded. At a minimum, include in the minutes:

- A record of attendance
- Key points of discussion and input
- Decisions about topics presented to the TEP
- Copies of the meeting materials.

It is the responsibility of the measure developer to consider the input received by the TEP; however, any recommendations made to CMS are made by the measure developer. If the measure developer makes recommendations to CMS that are not consistent with the recommendations from the TEP, these differences should be noted and explained in the report. At a minimum, the summary will include:

- Name of the TEP
- Purpose and objectives of the TEP
- Description of how the measures meet the overall quality concerns and goals for improvement
- Key points of TEP deliberations
- Meeting dates
- TEP composition
- Recommendations on the candidate measures.
Measure evaluation reports for each of the measures considered are then delivered to CMS by the measure developer. The Measure Evaluation Report includes information on how each measure met or did not meet each subcriterion and provides CMS with information regarding the feasibility of strengthening the rating of any subcriterion that was rated “low.” At this point in the process, it may not be possible to evaluate all subcriteria. For example, reliability and validity may require further testing before the measure can be evaluated.

12.3.12 Post the TEP Summary Report

Communicate and coordinate with the Measures Manager to post the approved TEP Summary Report at the discretion of the COR using the TEP Summary Web Page Posting form, and the same process as the other postings. The report should remain on the website for at least 21 calendar days or as directed by the COR. After the public comment period, the measure developer and the TEP review the comments received and recommend appropriate action, particularly if the public comment period was asking for feedback on the measure specifications.

It is important to note that the TEP may be consulted during any stage of measure development, including when the measure is undergoing the NQF endorsement process. If the TEP has met several times on one topic, it may (at the COR’s discretion) be appropriate to summarize discussions held during multiple meetings.
13 PERSON AND FAMILY ENGAGEMENT

13.1 BACKGROUND AND DEFINITION

Person and family engagement in the measure development process is the process of involving persons and/or family representatives in a meaningful way throughout the measure lifecycle. Per the CMS Person & Family Engagement Strategy, “the term ‘person’ is used to reflect an individual’s identity as more than a patient, to recognize his or her participation in prevention and wellness.” In this context, family “is used broadly to include participants in a person’s healthcare, including informal caregivers, along with the primary caregivers of persons who are in need of the support of their caregivers to make informed healthcare decisions.” Advocates and advocacy groups can also be involved to provide the person and family perspective.

Strengthening persons and families as partners in care is important at CMS. Involving persons and family representatives in the measure development process is among the many ways that CMS is striving to achieve this goal. Engaging persons and family representatives benefits consumers by helping to identify issues that are important and meaningful from their perspective. It also supports identification of information that consumers need to make informed healthcare decisions. Person/family engagement helps developers and CMS produce high-quality measures that are easily understood, relevant, and useful to consumers. Their involvement helps CMS develop messaging that resonates with and reflects healthcare quality issues important to the public.

13.2 OPTIONS FOR ENGAGEMENT AND SELECTED BEST PRACTICES

Best practices for engaging persons and family members in measure development activities are discussed throughout this chapter and are summarized in Table 12. Regardless of the engagement methods used, it is critical that individuals involved with measure development efforts are provided with clear expectations about what their participation will entail. Developers may also consider the principles in the Patient-Centered Outcomes Research Institute (PCORI) Engagement Rubric when engaging consumers (Figure 30) and observe best practices for conducting qualitative research, survey and interview construction, and testing, as applicable and with approval from CMS.

<table>
<thead>
<tr>
<th>Concepts highlighted by PCORI that are applicable to person/family member engagement in the measure development process include:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reciprocal Relationships:</strong> Roles and decision-making authority of all involved are defined collaboratively and clearly stated.</td>
</tr>
<tr>
<td><strong>Co-Learning:</strong> It is important to ensure that all participants understand the measure development process, person and family engagement, and person-centeredness.</td>
</tr>
<tr>
<td><strong>Partnership:</strong> The time and contributions of person partners are valued. Time commitment and attendance requests for persons need to be thoughtful and reasonable. The research time is committed to diversity and demonstrates cultural competency, including disability accommodations, as appropriate.</td>
</tr>
<tr>
<td><strong>Trust, Transparency, Honesty:</strong> Measure developers are encouraged to express commitment to open and honest communication with person stakeholders, in a meaningful and usable way, and ensure that major decisions are made inclusively.</td>
</tr>
</tbody>
</table>

Figure 30. Person and Family Engagement Concepts
### Table 12. Best Practices for Implementing Person/Family Engagement Activities, by Phase of Engagement

<table>
<thead>
<tr>
<th>Phase</th>
<th>Best Practices for Implementing Person/Family Engagement Activities</th>
</tr>
</thead>
</table>
| Preparing for Person/Family Engagement Activities | • Set clear expectations. Inform potential person/family member participants during recruitment about the time commitment requirements and the nature of the input being sought from them. Be transparent about what stage of development the measure is in, the timeline for this phase of work, and the overall timeline for completing measure development.  
  • Ensure that individuals understand the nature of their participation, particularly around issues of confidentiality, and explain that their participation in measure development activities is voluntary. Confidentiality language is included in the TEP Nomination Form Template and in the TEP Charter Template.  
  • Prior to the session, provide participants with person-centered, read-ahead materials that are easy to understand. Provide individuals with ample time to review materials and ask questions. For individuals without email or Internet access, mail the printed materials to them.  
  • Conduct preparatory calls with participants.  
  • Remind participants of the date and time of the meeting 1 to 2 days before the meeting.  
  • For in-person meetings, when applicable, consider using a facility that allows the development team to observe the discussion and enables the moderator to check in with the team during the session.                                                                                                                                                                                                 |
| During Person/Family Engagement Activities | • Adhere to best practices for qualitative research. Cognitive and plain language testing are essentially semi-structured, in-depth qualitative interviews. Be sure to have a trained facilitator who knows how to develop and follow a protocol and work with a respondent in a neutral, engaged setting. If possible, use a facilitator who has experience working with the relevant patient population.  
  • Ensure that introductions clarify the purpose of the meeting and the role that each participant will play. Ensure persons and families have a clear understanding of what parts of the measure they can impact and which things are out of scope.  
  • Take time to clearly explain technical measure concepts and answer questions to ensure persons and families can participate effectively. Minimize the use of technical jargon.  
  • Ensure participants feel comfortable participating in the discussion and emphasize that everyone’s input is important. For TEPs, remind persons and families of the expertise they bring to measure development.  
  • Convey the expectation that the group should hear and respect each participant’s perspective.  
  • Foster freedom of thought. Encourage participants to be free with their ideas even if they feel it may not be pertinent to the discussion at hand. Communicate the plan for tracking suggested ideas that do not directly fit into the current discussion but may be relevant for future work.  
  • Assist person or family member participants who become stuck in a personal story or situation, acknowledging the power of their experience and linking it to the objectives of the meeting.  
  • Continue assisting with technology needs for virtual or teleconference meetings, as needed.                                                                                                                                                                                                                                           |
| Following Person/Family Engagement Activities | • Hold one-on-one calls to encourage ongoing participation and answer questions.  
  • Keep persons and families updated on future decisions and the next stages of measure development after the working group, TEP, or other engagement activity has ended so they can understand the impact of their participation.  
  • Debrief participants and emphasize that their input was valued.  
  • Listen to participants’ suggestions to improve their experience and the experience of others.                                                                                                                                                                                                                           |

Prior to measure conceptualization, developers should compile a comprehensive plan outlining how person and/or family representative input can be incorporated at each stage of the Measure Lifecycle. As described below, many techniques are available to measure developers for engaging persons and family representatives in the development process. To capture the person/family perspective adequately, developers should involve persons/family representatives as early as possible in the measure development process and should consider incorporating two or more techniques in their development work. Options for person/family engagement in the measure development process include, but are not limited to, the types discussed.
Member of Standard TEP. A TEP is a group of stakeholders and experts that contributes direction and thoughtful input to developers during the measure development and maintenance processes. The TEP may work with the measure developer to develop the technical specifications and business case for measure development, review testing results, and identify potential measures for further development or refinement. The steps for convening a TEP are further described Section 3, Chapter 12, Technical Expert Panel.

Including one or more persons or family representative(s) on a TEP has been used widely for engaging persons and family representatives in the measure development process. As members of the TEP, consumers serve alongside professionals and may be asked to share aspects of their experience as healthcare consumers. An advantage of including persons/family members on the TEP is that it ensures that clinical and research concerns are balanced against consumer perspectives in the process. Involving consumers in the TEP requires few additional resources to implement. However, the measure developer must recognize that the views expressed by these one or two individuals may not be representative of the larger consumer population.

Best Practices:

- Ensure participants are well prepared by providing read-ahead materials that describe in plain language terms what each proposed measure is intended to communicate.
- Assign an advocate. Link representatives with a peer or professional who is familiar with the measure development process and relevant terminology and can support them before, during, and after serving on the TEP by providing background information and answering questions.
- Include at least two individuals representing the person/family perspective on the TEP so they do not feel isolated being on a TEP by themselves. In some instances, developers have found appointing a patient as the leader of the TEP an effective strategy:
  - Ask persons or caregivers to share their journey or story at the outset of the TEP (e.g., their own or a family member’s experience with cancer treatment or with being hospitalized for heart failure). This process often engages and energizes the TEP.
  - Any time information is gathered outside of the formal TEP (e.g., during one-on-one interviews), ensure information is relayed back to the full TEP.

The Person- or Family-Representative-Only TEP. A variant of the standard TEP, where the TEP is composed solely of persons or family representatives. An advantage of this approach over the standard TEP is that representatives may feel more comfortable sharing their own experiences with others like them.

Focus Groups. In a focus group, a skilled facilitator guides a group of persons or family representatives through a discussion by posing specific questions to the group about their own (or a family member’s) experiences with health and healthcare-related issues. Condition-based groups involve guided discussions among persons who have the health condition relevant to the measure under development. Seasoned measure developers have found that a group of five or six persons and family representatives is the ideal size for discussion, as the group is small enough to promote informal conversation yet large enough that the developer hears multiple views. Recruiting widely is a good strategy for recruiting a diverse group representing a variety of perspectives.

Working Groups. Working groups are composed of a leader and five or six individuals such as patients, family members, consumers, and advocates. In the context of a working group, developers seek group input on a topic related to the measure(s) under development. Seasoned measure developers have
found that working groups often promote close partnerships among developers and person/family representatives. When forming a working group or a focus group, developers should consider issues related to group composition (e.g., whether it is acceptable to have both persons and family members in the same group), as persons and family members may have very different perspectives on some topics. Figure 31 contains a list of best practices for TEPs and working groups.

One-On-One Interviews. In the context of an interview, the measure developer converses with one individual at a time. This technique can be used as a one-time information gathering exercise, but also can be useful for touching base with individuals and keeping them engaged between TEP meetings or multiple working group meetings. An advantage of this technique is that it enables the developer to obtain in-depth information, encourages ongoing participation in the measure development effort, and provides developers with the opportunity to answer participants’ questions.

Testing. Three types of testing relevant to measure development are concept testing, cognitive testing, and plain language testing. Additional information about Measure Testing is provided in Section 2, Chapter 4, Measure Testing:

- **Concept testing** is the process of evaluating consumer interest in and response to measurement-related topics.
- **Cognitive testing** involves presenting consumers with measure-related definitions and concepts and asking them to interpret the terms in their own words. This technique is particularly useful for appraising measures that are designed to be patient-reported because it enables the developer to evaluate whether consumers’ interpretations are accurate.
- **Plain language testing** investigates whether individuals are accurately translating the technical measure specifications into a description of what is being measured and why. This technique is particularly useful for evaluating measures planned for public reporting.28

Surveys. Surveys can be effective for obtaining input when the developer has specific questions about the measure(s) under construction that can be asked with multiple choice questions or brief answers (e.g., “Would this measure help you decide whether to have cardiac surgery at Hospital X?). Depending on the project, surveys can be conducted using paper instruments, via telephone, or online. Surveys can be an efficient way to gather information from a broad group of individuals in a short time frame. While surveys enable consumers to provide responses at their convenience, a drawback is that they do not allow respondents to ask questions or exchange ideas with the developer.

Virtual Community. A virtual community is a network of individuals who interact through social media such as message boards, chat rooms, and social networking sites. Virtual communities can be used to promote discussion and commentary among persons/family representatives about measure development through use of focused questions and topic threads (e.g., “Describe your experience

---

28 Additional information about plain language testing can be found through resources such as [http://www.plainlanguage.gov/](http://www.plainlanguage.gov/) and [http://centerforplainlanguage.org/](http://centerforplainlanguage.org/).
selecting a nursing home for your family member.”). This technique may provide valuable insight into a person’s or family representative’s viewpoints. At all points in the measure development lifecycle, representatives can be engaged in the online panel to review and comment on information related to the measure and its development. A caveat is that text-based, virtual community discussions may not yield responses that are representative of the consumer population at large.

### 13.3 Engagement Activities: Virtual vs. In-Person

Except for the text-based virtual community, which is, by definition, conducted online, all techniques described above have the flexibility to be conducted in-person or virtually using web meetings, web cameras, telephones, and other technology. A primary advantage of using a virtual approach is that it presents low burden to participants and measure developers and typically costs less to convene than in-person meetings. When deciding whether virtual or in-person interaction is preferable, developers should consider the population of interest and the role that the person/family members will play in measure development. Virtual approaches should be used only when individuals can reasonably be expected to participate, given their potential literacy, socioeconomic, or technology-related constraints (e.g., some at-risk populations may not have reliable access to the Internet). Developers should confer with their COR as virtual and in-person activities require different levels of resources.

**Best Practices.** When using virtual technology, developers should work with all participants in advance of each meeting to ensure they know how to use the technology and that technical support is available to all participants prior to and during the meeting.

### 13.4 Recruitment

There are diverse options for reaching persons and family members; however, it can still be a challenge to find individuals who are willing and able to participate in measure development. Recruitment strategies such as posting the TEP (Call for TEP) Web Page Posting form may be used, but other sources and methods may also be required. This list includes some possible recruitment approaches:

- Network with providers or clinicians currently active on TEPs who may be willing to place recruitment materials where persons or their family members may see them.
- Reach out to consumer advocacy organizations such as the American Association of Retired Persons (AARP) Inc. In addition to the advocates, they may have information on persons who are capable and willing to contribute.
- Contact condition-specific advocacy organizations such as the American Diabetes Association or the Michael J. Fox Foundation for Parkinson’s Research that may know of individuals who are active in support groups and knowledgeable about quality for those specific conditions.
- Some organizations such as the PCORI Patient Engagement Advisory Panel have person engagement representatives, who are experienced mentors and know of persons who are able to participate.
- For panel participation that will involve reviewing detailed information, it may be useful to contact people who have served on local community advisory groups such as Patient Family Advisory Councils (PFACs).
Examples of websites of advocacy organizations and support groups that may provide ways to reach out to persons and/or family members who would be interested in being involved in quality measure development include:

- AARP
- The Empowered Patient Coalition
- WebMD
- AgingCare.com
- Caring.com
- Connecticut Center for Patient Safety
- Daily Strength
- HealthWise
- MD Junction
- Med Help
- Patients Like Me
- CMS Quality Measures Public Comment
- People for Quality Care
- NQF.

Social media can also be used for recruitment. These websites and similar sites often include contact information, including social media sources. Social networking pages such as Twitter, Facebook, and other social media hosts are other potential options. These forms of recruitment are low cost and can be extremely effective. Because the use of social media for recruitment is still somewhat new, measure developers working on CMS contracts should work with the COR to verify that their recruitment approach and language adheres to CMS policies.

**Best Practices.** For focus group and interviews where the goal is to find participants who represent the typical target population, it works well to recruit people from a variety of sources. It can also be beneficial to seek persons from diverse geographical and sociodemographic backgrounds so that multiple perspectives are represented. Figure 32 contains an example of a featured best practice for recruitment.

### 13.5 Options for Engagement by Measure Lifecycle Stage and Selected Best Practices

As discussed in Section 2, The Measure Lifecycle, the Measure Lifecycle consists of five stages: measure conceptualization; measure specification, measure testing; measure implementation; and measure use, continuing evaluation, and maintenance. The most useful engagement techniques for each stage of the Measure Lifecycle are described in the next sections.
13.5.1 Measure Conceptualization

During the measure conceptualization stage, the developer’s primary task is to generate and prioritize a list of concepts to be developed. Often, the developer starts by developing a framework or logic model that captures important domains or topics. While it is critical for the framework to be grounded in the scientific literature, perspectives of patients and family members can be very helpful in framing the problems and prioritizing steps for quality evaluation. Refer to Figure 33 for a best practice on measure conceptualization.

**Techniques.** Qualitative methods that enable the measure development team to learn from patients and families about their care stories are particularly useful during measure conceptualization. From these stories, the team can map out typical encounters or episodes of care. Prompts that may be useful for eliciting this information include “Tell us your story,” “What went well?” and “What could have been done better?”

Examples of methods include:

- **One-on-one interviews** with a skilled interviewer using a planned interview guide may be convenient and particularly useful when the care event under study is complex or highly personalized.
- **Focus groups** may also be useful because they allow persons or family members to compare notes and help the team identify common responses and priorities.
- **Concept testing** (performed in the context of either an interview or focus group) can also be advantageous at this stage. Developers can test the extent to which persons or family members find the concepts interesting or relevant to their own situation to determine the measures that are the best candidates for further development.

13.5.2 Measure Specification

During the measure specification stage, the measure developer drafts the measure specifications and conducts an initial feasibility assessment. Person and family representatives can provide input on a variety of measure specification decisions such as the clinical outcome of the measure, PRO performance measure instrument selection, defining the target population, risk adjustment approaches, and measure methodology. By including person and family perspectives during the measure specification stage, developers can optimize measure usability/interpretability to patients, and maximize how meaningful the measure can be. Persons can help measure developers prioritize areas for future analyses or research while there is still time to modify the measure development approach, if necessary. Refer to Figure 34 for a best practice on measure conceptualization and specification.
Techniques. Mechanisms that enable discussion and ongoing exchange of ideas work best during new measure development and specification:

- **Working groups** are an excellent way for developers and person/family collaborators to discuss technical concepts and provide persons/family members with the opportunity to ask questions.
- **TEPs** can be used to enable persons and families to weigh in on measure specifications and respond to other stakeholders in a multi-stakeholder environment.
- **One-on-one interviews** enable the developer to gather targeted information to inform specific aspects of the measure under development.

Best Practices. When conducting discussions about measure specifications, it is critical to ensure representatives have a clear understanding of which parts of the measure they can impact and which things are out of scope. This understanding will help focus the recommendations they provide to the developer.

13.5.3 Measure Testing

During the measure testing phase, the developer tests the measure to ensure it is working as intended. Engaging person and family representatives during this stage ensures that the measures make sense to the public and will be beneficial for public reporting. This is an opportunity for the measure developer to ensure the patient-centered measure they set out to develop is adequately translated. If there are gaps in understanding, the measure developer can determine whether adjustments are needed at the specification level or at the translation level.

During this stage, the developer should ensure that consumers understand and are able to answer each of these questions:

- Why is this measure important for the public to know and understand?
- How is this measure derived (i.e., what specifically is being measured)?
- What does the performance score mean (i.e., what influences whether a patient has a higher versus a lower score)?

Techniques

Mechanisms that enable individuals to evaluate what the measure means and explain how they interpret the measures work best at this stage. These one-on-one data collection methods are often useful:

### Featured Practice: Measure Conceptualization and Specification

The measure developer for the Hospital-Acquired Conditions (HAC) Reduction Program wants to identify new, potentially suitable measures to fill HAC performance gaps and examine the current scoring methodology to determine if modifications are needed. The developer utilizes a person or family advisory panel early on to obtain input on additional HACs that could be tracked and measured as part of the HAC Reduction Program, and which of these items persons/family members consider to be of the greatest importance. The measure developer uses this feedback to identify new suitable measures, and begins to work with statisticians to examine the current scoring methodology. The advisory panel is not involved in the meetings focused on scoring methodology. Later, as the measure developer has focused on two viable scoring methods, it re-engages with the person or family advisory panel to seek feedback on the revised scoring method concepts under consideration.

Figure 34. Measure Conceptualization and Specification Best Practice

---

**Figure 34. Measure Conceptualization and Specification Best Practice**
• **Cognitive testing** can be used to determine how person and family representatives are interpreting the measure, and whether they can accurately answer each of the key questions above.

• **Plain language testing** can be used to test whether consumers are accurately translating the measure specifications.

**Best Practices**

• *Test in a “realistic” environment.* Developers may consider testing using a webinar platform so the person or family representative can be in front of their computer and review the information as they would if they were using the Internet.

• *Write for the web and a web-based attention span.* Developers should consider that the average person will spend about 30 seconds evaluating the measure. Material should be presented in short, easy-to-understand paragraphs.

### 13.5.4 Measure Implementation

At the measure implementation stage, measure specifications are complete, and the focus of the work is the framing and presentation of the measure. Measure developers can partner with persons and families during measure implementation to obtain feedback on the way the measure will be presented to various stakeholders, including persons and families. Representatives can review language and displays that describe measure specifications and result interpretations, and measure importance for appropriate word choice, reading level, inclusion of concepts that are important to persons and families, and exclusion of concepts that may not be important. Including person/family input can ensure the language and displays used to describe the measure are both relevant to, and easily understood by individuals who may use the measure to inform their healthcare decision-making.

**Techniques**

Mechanisms that enable informal, interpretive, and reactive discussions or quick “knee-jerk” feedback are often effective at this stage of measure development:

• *Focus groups* can be used to observe individuals’ reactions to various language/display options and enable them to provide critical feedback and make suggestions for improvement. Focus groups can also be used to assess how proposed language/displays are interpreted and whether that interpretation is consistent with the developer’s intent.

• *Surveys* are an excellent tool to obtain knee-jerk reactions to descriptive text or display options, obtain quick preference ranking of several options, and assess interpretation of unguided wording/phrasing.

**Best Practices**

• *Set clear expectations.* Developers should explicitly state the goals of the implementation work (e.g., improving readability, testing the comprehension of various language or displays about the measure).

• *Provide appropriate framing or context.* Developers should explain why the descriptive language about the measure or measure display is in its current format and describe previously received feedback.
13.5.5 Measure Use, Continuing Evaluation, and Maintenance

During this stage, the measure developer will test the measure post-development and once the measure is in use (and potentially, being actively publicly reported). At this point in the measure development lifecycle, engaging person and family representatives ensures that the measure remains relevant. Clinical practices change over time, but so does the public’s understanding of concepts. It is important to ensure that over time, measures continue to resonate with person and family representatives and that they are still meaningful to them. Also, over the life of a measure, adjustments will be made (e.g., when specifications are updated to address changes in clinical guidelines). Measures will be refined to ensure more precise measurement. Any time a measure is updated, the language used to explain and describe that measure to the public needs to be updated, which requires retesting the measure with person and family representatives.

Techniques

As during the initial measure testing phase, mechanisms that enable individuals to evaluate what the measure means and explain how they interpret the measure work best at this stage. One-on-one data collection methods—in particular, cognitive testing and plain language testing—are beneficial at this stage. As during measure testing, the same types of questions need to be asked to ensure the measure is accurately understood and interpreted, and the measure can still help person and family representatives make informed healthcare decisions.

Best Practices

It is most important to remember to test measures at least every 2 to 3 years to ensure the concepts are fresh and relevant, and every time an edit is made to the measure. If the adjustment is small, testing with one or two individuals may be sufficient. Developers should verify the measure is still being accurately interpreted and understood, and never assume a small change will be intuitive or easy for the public to understand.

13.6 Other Considerations

Paperwork Reduction Act (PRA) Exemption for Measure Development Activities

The PRA mandates that all federal government agencies obtain approval from the OMB before collection of information that will impose a burden on the public. However, with the passage of the MACRA, data collection for many quality measure development projects is now exempt from PRA requirements. Measure developers working under contract with CMS should consult with the COR to determine if their project is eligible for an exemption. Developers working with CMS programs that are not PRA-exempt should factor time—6 to 8 months on average—into their project timeline for OMB to review their Information Collection Request.

Budgeting Considerations

During the budgeting/planning process, measure developers should include costs for activities related to engaging persons/family representatives at multiple time points during the measure development process in their project budgets. For work that is ongoing, developers should consider ways that person/family input can be gathered within the constraints of their existing project plan and budget. For both new and existing projects, lower cost options such as virtual/web-based meetings (as opposed to in-person meetings that may require significant travel-related expenses) may be worth considering.
Participant Compensation

In the past, compensation for person and family members contributing to measure development efforts has been provided on a case-by-case basis. Developers working on CMS-funded measure development contracts should consult with the COR about whether participant compensation should be considered for their project.
14 **PUBLIC COMMENT**

The public comment process is an essential way that CMS ensures its measures are developed using a transparent process with balanced input from relevant stakeholders. The public comment period provides an opportunity for the widest array of interested parties to provide input on the MUD and to provide critical suggestions not previously considered by the measure developer or the TEP. Public comments obtained during measure development (and maintenance) are separate from, and complement the public comment obtained during the NQF endorsement process.

14.1 **TIMING OF PUBLIC COMMENT**

Public comment can be obtained at several points during the Measure Lifecycle. The public comment periods that occur during the Measure Lifecycle are consistent with Lean principles because they enable potential issues to be identified early. Addressing issues raised in public comments can prevent errors and rework later. If issues are not addressed adequately, they might cause problems after the measures are proposed for use in specific programs. There is flexibility to determine the best time to obtain comments during measure development, depending on the needs of CMS and the measure developers related to specific measures and programs:

- During measure conceptualization and information gathering: Aggregate comments on the summary of the TEP meetings.
- During measure specification: Draft technical specifications can be posted with summaries of subsequent TEP meetings.
- During measure testing: If a TEP reviews testing results and updated specifications, those summaries can be posted for further public comment.
- During measure implementation:
  - The MUC list is posted for public comment as part of the pre-rulemaking process.
  - The MAP posts their reports for public comment.
  - Public comment opportunities are part of the NQF Consensus Development Process.
  - Proposed federal rules are posted for public comment.
  - Federal Register Notices are posted for public comment.
  - Feedback can be obtained during CMS listening sessions, Open Door Forums, Special Open Door Forums, and town hall meetings.
- During measure use, continuing evaluation, and maintenance:
  - NQF-endorsed measures are listed on the NQF Quality Positioning System website and have a mechanism for comment enabled.
  - Summaries of TEP meetings held during measure maintenance are posted for public comment.

14.2 **FEDERAL RULEMAKING**

The federal rulemaking process also includes a public comment period. The public comment period during rulemaking is a time when CMS receives feedback on its measures, because most CMS quality programs are included in rulemaking. However, the federal rulemaking process should not be the only time when public comments are sought and addressed. The federal rulemaking process usually occurs after the measure is developed and is being proposed for implementation; therefore, the measure developer could miss the opportunity to address issues earlier, during measure development. During measure use, continuing evaluation, and maintenance, public comments received as part of the federal
rulemaking process should be considered as part of ongoing surveillance. They should also be formally considered during the comprehensive reevaluation. Finally, comments received as part of federal rulemaking could also generate measure concept ideas for future development.

14.3 Steps for Public Comment

The Call for Public Comment involves several postings to the dedicated CMS MMS website. The measure developers will develop materials to send to the Measures Manager to post the call. Website postings involve two CMS divisions, and the process to post the materials will take approximately five working days. Measure developers must plan accordingly for the deadlines for submitting information to be posted for public comment and for the time needed for soliciting and receiving public comment. If these deadlines are not considered, then public feedback may not be able to be incorporated into the measure development process.

Public comment is gathered according to these eight steps:

1. Prepare the Call for Public Comment
2. Post the measures following COR approval
3. Notify relevant stakeholder organizations
4. Collect information
5. Summarize comments and produce report
6. Send comments to the TEP for consideration
7. Finalize the public comment report, including verbatim comments
8. Arrange for the final public comment summary report to be posted to the website.

These eight steps are essential to successfully soliciting public comment. Deviation from the procedure requires COR approval.

14.3.1 Prepare the Call for Public Comment

Measure developers may use the Call for Public Comment as a means of soliciting public comment on CMS measures. This document includes general information regarding the purpose of the call for comments and instructions on how to submit comments. Measure developers may also post an announcement on the CMS MMS Public Comment site that informs readers that a measure is up for comment on another website.

When organizing a Call for Public Comment, establish an email address to receive the comments. Alternatively, a web-based tool such as Survey Monkey or Jira, often used for eCQMs, can be used to receive comments. If so, set up the tool after contents have been approved by the COR.

For eCQMs, calls for public comment should also be sent to the eCQI Resource Center at ecqi-resource-center@hhs.gov. Additionally, the ONC Project Tracking System (Jira), a licensed software development platform is used for collecting and monitoring feedback on different stages of the measure development process, including collecting public comments. Jira is also used to collect measures for the MUC List. Refer to Appendix B for more information on Jira.

The public is encouraged to submit general comments on the entire measure set or comments specific to certain measures. When drafting the posting and questions for a web-based tool, consider PRA requirements.
14.3.2 Notify Relevant Stakeholder Organizations

Submit a list of relevant stakeholder organizations for notification about the public comment period to the COR for review and input prior to posting the call. Input can ensure that the list is complete and appropriately representative of all the types of experts that should be included. After approval by CMS, it may be appropriate to notify the stakeholder organizations before the posting goes live. Relevant stakeholder groups may include, but are not limited to:

- Organizations that might help with recruiting appropriate patients/their caregivers
- Quality alliances (e.g., Dental Quality Alliance, Pharmacy Quality Alliance)
- Medical and other professional societies
- Scientific organizations related to the measure topic
- Provider groups that may be affected by the measures
- NQF measure developer groups.

Notification methods may include, but are not limited to:

- Posting the notice on a related CMS website in addition to the Call for TEP page
- Announcing the notification during appropriate CMS working group or Open Door Forum calls and sending the notice to the related distribution list
- Sending the notice via email to the stakeholders’ email lists or having the stakeholder organizations post a notice on their websites
- Recruiting on patient support group sites and other consumer organizations
- Using social media (e.g., Twitter, Facebook, YouTube, LinkedIn); contact the COR for the process.

14.3.3 Post the Measures Following COR Approval

After obtaining COR approval, work with the Measures Manager to post the MIF and MJF on the dedicated CMS MMS website using the Call for Public Comment form and the posting process described in Section 3, Chapter 15, MMS Website Posting. When submitting the forms, prominently mark the MIF and MJF as “Draft.”

As a rule, the call should be posted on the website for at least two weeks to allow sufficient time for the public to provide comments. The COR makes the final decision as to how long the call should be posted.

The information to be posted may include, as directed by the COR:

- Objectives of the measure development contract
- Processes used to develop the measures:
  - Identifying important quality goals
  - Conducting literature reviews and grading evidence
  - Defining and developing specifications for each quality measure
  - Obtaining evaluation of the proposed measures by TEPs (as directed by the COR, the TEP Summary Report may be posted)
- Objectives of the solicitation (e.g., to help determine measure importance, to refine specifications, to comment on usability and feasibility)
- MIF, MJF, and the development stage of the measures
- Information about the measure developer and subcontractors developing this measure set.

---

The NQF submission may be acceptable for this deliverable.
14.3.4 Collect Information

Commenters submit their comments via email or other tools as directed on the CMS MMS website.

14.3.5 Summarize Comments and Produce Report

At the end of the public comment period, prepare a preliminary Public Comment Summary Report. The report should include verbatim comments as well as a summary and analysis of the public comments that were received. Preliminary recommendations may be stated in the report, pending discussion with the TEP. This report should be submitted to the COR and the TEP within two weeks following the end of the public comment period.

The report should include a:

- Summary of general comments posted and any other information that could apply to the set of measures and recommended action
- Summary of the comments for each measure and any preliminary recommendations for TEP consideration
- Listing of the verbatim public comments (if the submitter includes personal health information in relation to the measure, the measure developer should redact the sensitive portions).

When measure developers are asked to prepare responses to public comments on behalf of CMS (the measure steward), it is important to plan for close coordination and allow significant time for CMS deliberation and review. These discussions should start early, with roles, responsibilities, coordination protocols, and timeline clearly defined and agreed upon between CMS and the measure developer.

This level of coordination is critical to ensure that public comments are addressed efficiently, effectively, and in a timely manner by taking CMS policies and programs into consideration and to inform ongoing measure development. After the report has been reviewed by the COR, work with the Measures Manager point of contact to post the preliminary Public Comment Summary Report (including the verbatim comments) on the CMS MMS website.

14.3.6 Send Comments to the TEP for Consideration

Reconvene the TEP to discuss the submitted comments and preliminary recommended actions. After deliberations, the TEP may make recommendations to the measure developer concerning changes to the measures because of the public comments. This task may be performed via email, teleconference, or in-person meeting.

14.3.7 Finalize the Public Comment Report, Including Verbatim Comments

Document the TEP discussion and the recommended actions. The finalized report should include:

- Recommendations and actions taken in response to the comments received (e.g., candidate measures that are recommended to be eliminated from further consideration)
- Updated or revised measure specifications with notations about changes made.

Submit the report to the COR within one week after the TEP meeting to review the comments.

14.3.8 Arrange for the Final Public Comment Summary Report to be Posted on the Website

After obtaining COR approval, work with the Measures Manager to post the final Public Comment Summary Report, including verbatim comments as soon as possible (or as directed by the COR and Task Order Schedule of Deliverables) after the public comment period closes. Use the Public Comment
Summary Web Posting form to submit the report to the CMS website following the procedure described in Section 3, Chapter 15, MMS Website Posting.
15 MMS Website Posting

The procedures described below are used for all postings to the pages (i.e., Call for Measures, TEPs, and Call for Public Comment) linked through the CMS MMS Overview site.

Figure 35 demonstrates the steps in the process for posting to the CMS MMS website.

- The measure developer assembles Section 508 compliant materials to be posted and obtains final approval from their COR. Information about CMS Section 508 compliance is available on HHS website. All attachments to the submission template (not the template itself) must be in PDF and may not be submitted as zipped files.

    If Jira is to be used for comments, ensure that the link to the Jira ticket is activated before sending materials for posting.

    - Submissions must be sent to MMS Support inbox (MMSSupport@battelle.org) for posts on:
      - Call for Measures web page
      - Public Comment web page
      - TEP web page.
    - Upon receipt by the Measures Manager, the material to be posted is reviewed again for Section 508 compliance, completeness of information, confirmed COR approval, and compliance with formatting requirements. Non-compliance may cause delays.
    - The materials are then sent to the CMS Website Posting Coordinator.
    - The CMS Website Posting Coordinator creates the updated web page layout and submits it to the CMS Web group for posting. CMS Web and New Media Group, as part of the Office of Communications, are responsible for the entire CMS website. The group reviews the proposed web content to ensure it meets all CMS website requirements. The website is then moved to the production environment where the page “goes live.”
    - The CMS Website Posting Coordinator sends confirmation that the approved materials have been moved into the production environment.
    - The Measures Manager verifies that the materials are available on the site and notifies the measure developer and their COR.
15.1 **Posting Time Frame**

Allow at least five business days from submission to the Measures Manager for processing a post; however, posts may be posted prior to this time frame. If your post needs to be published on a specific date, note this in your email, and CMS will work to accomplish it by this date/time.

All posts will be removed from the website after 6 months unless otherwise specified with an open/close date and at the discretion of the COR.

15.2 **Posting Format**

A web posting document should be submitted in Word format *(every post must include a web posting document)*. All other documents/attachments to the post should be Section 508 compliant and submitted in PDF. Note: Tables must have repeated headers on every page.

15.3 **Posting Template**

All submissions must follow the content, format, and language of the relevant Blueprint template to be compliant. If they do not, CMS will ask you to revise them before submitting it as a final post. Templates for the web posting documents are in Section 4, Forms and Templates.

15.4 **Documents to Include**

Public Comment Documents to Include with Each Post:

- Call for Measures:
  - Call for Measures Web Posting document
- Call for Public Comments:
  - Public Comment Call Web Posting document (Word format)
  - Other files, if any, to be included with the Call for Public Comment (PDF format)
- Public Comment Summary report:
  - Public Comment Summary Web Posting document (Word format)
  - Public Comment Summary Report (PDF format)

TEP Documents to Include with Each Post:

- Call for TEP:
  - TEP Call for TEP Web Page Posting document (Word format)
  - TEP Nomination Form (PDF format)
  - TEP Charter (PDF format)
- TEP Composition (Membership List):
  - TEP Composition (Membership List) Web Page Posting document (Word format)
  - TEP Composition (Membership) List (PDF format)
- TEP Summary Report:
  - TEP Summary Web Page Posting document (Word format)
  - TEP Composition (Membership) List (PDF format)
  - TEP Summary report (PDF format)
16 **MEASURE TECHNICAL SPECIFICATION**

This chapter provides guidance for the measure developer to ensure measures developed for CMS have complete technical specifications that are detailed and precise. Measure specification follows the development of the initial intent of the measure, based on clinical practice guidelines and evidence identified in the environmental scan.

Measure technical specifications are the technical instructions for how to build and calculate a measure. The intent of measure specification is that each measure should reach its appropriate target population, but not over-reach or under-reach, for such errors in specification not only waste resources, but also may generate misleading conclusions about care quality.

The building blocks of a measure to be described in the technical specifications include:

- **Initial population**
- **Numerator** statement and definitions
- **Denominator statement** and definitions
- **Exclusions**
- Denominator exceptions
- **Target population**
- **Time interval**
- **Stratification** scheme, or how results might be split to show differences across groups
- **Risk adjustment** methodology
- **Calculation algorithm**, or how results are calculated
- Sampling methodology
- **Data source**
- **Level of analysis**
- **Attribution** model, or how data are attributed to providers and/or hospitals
- Care setting.

Different information sources influence the development of technical specifications for a measure:

- Literature review
- Clinical practice guidelines
- Existing measures
- TEP, SME, and stakeholder input
- Public comment
- Alpha testing
- Beta testing.

These inputs will improve the precision of the technical specifications and increase the **validity** and **reliability** of the measure. Measures must be specified with sufficient details to be distinguishable from other measures and to support consistent implementation. The NQF Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement notes measures should be specified with the broadest applicability (e.g., target population, setting, level of measurement/analysis) as supported by the evidence.
16.1 MEASURE SPECIFICATIONS BY MEASURE CATEGORY

Almost all measure technical specifications depend at least in part on the category of measure being specified. Measure categories, for the purposes of this chapter, are ratio, proportion, and CV measures.

A ratio is a score that is derived by dividing a count of one type of data by a count of another type of data (e.g., number of patients with central lines who develop infection divided by the number of central line days). The key to the definition of a ratio is that the numerator is not in the denominator.

A proportion is a score derived by dividing the number of cases that meet a criterion for quality (the numerator) by the number of eligible cases within a given time frame (the denominator) where the numerator cases are a subset of the denominator cases (e.g., percentage of eligible women with a mammogram performed in the last year).

CV is a score in which each individual value for the measure can fall anywhere along a continuous scale and can be aggregated using a variety of methods such as the calculation of a mean or median (e.g., mean number of minutes between presentation of chest pain to the time of administration of thrombolytics).

Table 13 defines the measure specifications relevant to each measure category.

<table>
<thead>
<tr>
<th>Measure Specifications by Measure Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
</tr>
<tr>
<td>Initial Population</td>
</tr>
<tr>
<td>Denominator</td>
</tr>
<tr>
<td>Denominator Exclusion</td>
</tr>
<tr>
<td>Denominator Exception</td>
</tr>
<tr>
<td>Numerator</td>
</tr>
<tr>
<td>Numerator Exclusion</td>
</tr>
</tbody>
</table>

The steps outlined and described in detail in Sections 16.2 through 16.6 are performed in coordination with the COR to develop the full measure technical specifications. As each step is completed, update the MIF accordingly.

- Define the data source
- Develop specifications and definitions
- Specify the codes and code systems
- Construct data protocol
- Document the measures.

16.2 DEFINE THE DATA SOURCE(S)

Measure specifications should include the data sources and the method of data collection that are acceptable. The data source used to calculate a measure can determine reliability, usability, validity, and feasibility of the measure. If the measure is calculated from more than one data source, develop detailed specifications (Deliverable 5-1) for each data source. Collect evidence that the results calculated from the different data sources are comparable.
Examples of data sources include:

- Administrative data
- Claims data
- Patient medical records – paper and electronic
- Electronic clinical data such as device data
- Registries
- Standardized patient assessments
- Patient-reported data and surveys.

When contemplating the source of data, the measure developer must consider the feasibility and methods of collecting data from that source. Examples of the pros and cons of using each data source are included in Table 14.

Table 14. Pros and Cons of Different Data Sources

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Pro</th>
<th>Con</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative data</td>
<td>Can provide information not usually found in a clinical database</td>
<td>These data are not collected primarily for the purpose of quality measurement, they are collected for other purposes</td>
</tr>
<tr>
<td></td>
<td>Less burdensome to providers for data collection</td>
<td></td>
</tr>
<tr>
<td>Claims data</td>
<td>Professionally coded; drawn from large populations (i.e., more representative of the populations of interest)</td>
<td>These data are not collected primarily for the purpose of quality measurement, they are collected for other purposes; varying degrees of clinical detail; often limited in content, completeness, timeliness, and accuracy</td>
</tr>
<tr>
<td></td>
<td>Less burdensome to providers for data collection</td>
<td></td>
</tr>
<tr>
<td>Paper patient medical records</td>
<td>Detailed clinical data with a rich description of care</td>
<td>Abstraction is time-intensive; requires expert staff (cost and time) to interpret each record and input data findings into a format suitable for analysis; abstraction can be open to subjectivity and interpretation or lack of consistency in how data are abstracted</td>
</tr>
<tr>
<td>Electronic patient medical records</td>
<td>Reduced cost of accessing clinical information from the patient medical record; detailed clinical data with a rich description of care</td>
<td>Identifying test sites that can serve as data sources can be difficult; extracting the data requires expertise, time, and money; hurdles related to continuing use of paper notes for point-of-care documentation; use of drop downs and structured fields can reduce the richness of the clinical data and descriptions of care</td>
</tr>
<tr>
<td>Electronic clinical data</td>
<td>Reduced cost of accessing clinical information from the patient medical record; or personal health device (e.g., home blood glucose monitor)</td>
<td>Identifying test sites that can serve as data sources can be difficult; extracting the data requires expertise, time, and money; hurdles related to continuing use of paper notes for point-of-care documentation; device data may be external to the patient medical record still only partially implemented in most hospitals</td>
</tr>
<tr>
<td>Registries</td>
<td>Data from multiple sources and across care settings; often available as an electronic upload</td>
<td>Unknown how registry requirements impact workflow; feasibility of data collection is determined by the data requirements imposed by the registry</td>
</tr>
<tr>
<td>Standardized patient assessments</td>
<td>Well validated and tested</td>
<td>Potential for bias as some have mixed use for determining reimbursement, meeting conditions of participation, and assessing quality; may be proprietary</td>
</tr>
</tbody>
</table>

30 Data are submitted electronically directly from the registry.
16.2.1 Administrative Data

Administrative data include multiple types of information, for example, covered services used by enrollees in a program, organizational staffing, organizational policies, birth registries, and tax records.

Similar data elements may exist in the provider’s billing system.

Other types of administrative data include patient demographics obtained from eligibility or enrollment information, organizational policies, crime reports, and census information. Payroll data and other databases containing information about providers can also be a source for some types of measures.

16.2.2 Claims Data

Claims data are derived from reimbursement information or the payment of bills. This information can come from claims that have been submitted and adjudicated or from the provider’s billing system. Claims are unique sources of data containing data elements that can be used (but were not intended) in the development of a quality measure. Claims include admission and discharge dates, diagnoses, procedures, and source of care.

16.2.3 Patient Medical Records (Paper-based or Electronic)

Patient medical records are a traditional source of clinical data for measures, and the data may be documented on paper or electronically and may include data from the clinical laboratory, imaging services, personal health records, and pharmacy.

16.2.4 Electronic Clinical Data

Electronic clinical data consist of patient-level information that can be extracted in a format that can be used in a measure, for example, data from personal health devices. Personal health device data may be uploaded to the EHR.

16.2.5 Registries

The term registry can apply to a variety of electronic sources of clinical information that can be used as a data source for quality measures. In general, a registry is a collection of clinical data for assessing clinical performance quality of care. The Quality Payment Program (QPP) accepts data from Qualified Clinical Data Registries (QCDRs) and Qualified Registries. For more information on QPP registries, refer to Appendix E.

Registries may be part of a larger regional or national system that may operate across multiple clinicians and institutions. Examples of national registries include the Chest Pain – MI Registry™ (from the American College of Cardiology and American Heart Association), the Society of Thoracic Surgeons (STS) National Database, and the Paul Coverdell National Acute Stroke Registry.

Registries have been used by public health departments for many years to record cases of diseases with importance to public health. This type of registry can provide epidemiological information that can be used to calculate incidence rates and risks, maintain surveillance, and monitor trends in incidence and
mortality. Immunization registries are used to collect, maintain, and update vaccination records to promote disease prevention and control.

16.2.6 Standardized Patient Assessments

CMS uses data items or elements from validated health assessment instruments and question sets to provide the requisite data properties to develop and calculate quality measures. Examples of these types of data include the Long-Term Care (LTC) Facility Resident Assessment Instrument (RAI), the Outcome and Assessment Information Set (OASIS), IRF Patient Assessment Instrument (PAI), and others.

16.2.7 Patient-Reported Data and Surveys

Data may be collected directly from a patient in the form of a survey, questionnaire, or assessment. Surveys (e.g., CAHPS surveys31 that collect information on beneficiaries' experiences of care) are advantageous because they ask about concepts such as satisfaction and patients’ feelings. PROs, such as pain assessments and quality of life indices, provide the patient’s perspective on their health, quality of life, or functional status.

16.3 DEVELOP SPECIFICATIONS AND DEFINITIONS

The basic construction of measure specifications begins with the outline of initial population, numerator, denominator, exclusions, exceptions, and measure logic. Then, the measure concept is given increasing amounts of detail, including precisely defined data elements and the appropriate values or value sets. Every part of the measure specification requires explicitly defined elements with accompanying analysis to identify constraints and criteria of the specification. Additional considerations for both numerator and denominator include alignment with other measures conceptually and technically. Refer to Section 3 Chapter 16.1, Measure Specifications by Measure Category for definitions of proportion, CV, and ratio measures and Appendix F for explicit calculation of proportion, CV, and ratio measures.

16.3.1 Define the Initial Population

The initial population refers to the cohort from which the denominator population is selected. Some measures (e.g., ratio measures) will require multiple initial populations, one for the numerator and one for the denominator.

Details often include information based on specific age groups, diagnoses, diagnostic and procedure codes, and enrollment periods.

If the measure is part of a measure set, the broadest group of population for inclusion in the set of measures is the initial population. The codes or other data necessary to identify this cohort, as well as any sequencing of steps that are needed to identify cases for inclusion, must also be specified.

16.3.2 Define the Denominator

The denominator statement describes the population evaluated by the individual measure. The target population defined by the denominator can be the same as the initial population or it can be a subset of the initial population to further constrain the population for the measure. The denominator statement should be sufficiently described so that the reader understands the eligible population or composition of the denominator. Codes should not be used in lieu of words to express concepts in written descriptions.

31 CAHPS® is a registered trademark of AHRQ.
However, codes may be useful references when writing measure algorithm in the MIF. The denominator statement should be precisely defined and include parameters such as:

- **Age ranges**
- **Setting**
- **Diagnosis**
- **Procedures**
- **Time interval**
- **Other qualifying events.**

Format—Patients, age [age or age range], with [condition] in [setting] during [time frame]

Examples:

- Patients 18-75 years of age by the end of the measurement year, who had a diagnosis of diabetes (type 1 or type 2) during the measurement year or the year prior to the measurement year (NQF 0062).
- All patients aged 18 and older with a diagnosis of Chronic Obstructive Pulmonary Disease (COPD) (NQF 0091).
- All patients at least 18 years old as of the first day of the reporting month who are determined to be maintenance hemodialysis patients (in-center and home HD [hemodialysis]) for the complete reporting month at the same facility (NQF 2978).
- All patients aged 65 years and older with a history of falls (history of falls is defined as two or more falls in the past year or any fall with injury in the past year). Documentation of patient-reported history with falls is sufficient (NQF 0101).

### 16.3.3 Determine if a Denominator Exclusion is Needed

Denominator exclusions refer to criteria that result in removal from the denominator before calculating the numerator. An exclusion means that the numerator event is not applicable. One example of an exclusion is a screening mammography for a woman who had a bilateral mastectomy.

The goal of denominator exclusion criteria is to have a population or sample with a similar profile in terms of meeting the numerator criteria.

Format of the exclusion statement—Denominator-eligible patients who [have some additional characteristic, condition, procedure]

Missing data should not be specified as an exclusion. The *NQF 2011 Consensus Standards Approval Committee (CSAC) Guidance on Quality Performance Measure Construction* notes systematic missing data (e.g., when poor performance is selectively not reported) reduces the validity of conclusions that can be made about quality.

An allowable exclusion must be supported by:

- Evidence of sufficient frequency of occurrence such that the measure results will be distorted without the exclusion, AND
- Evidence that the exclusion significantly improves the measure validity.
16.3.4 Define the Numerator

The numerator statement describes the process, condition, event, or outcome that satisfies the measure focus or intent. The numerator statement includes parameters such as:

- The event or events that will satisfy the numerator requirement
- The performance period or time interval in which the numerator event must occur, if it is different from that used for identifying the denominator.

Format—Patients who received/had [measure focus] {during [time frame] if different than for target population}

Examples:

- Patients receiving a nephropathy screening or monitoring test or having evidence of nephropathy during the measurement year (NQF 0062).
- Patients with documented spirometry results in the medical record (FEV1 and FEV1/FVC) (NQF 0091).
- The number of adult patient-months in the denominator who were on maintenance hemodialysis using a catheter continuously for 3 months or longer as of the last hemodialysis session of the reporting month (NQF 2978).

16.3.5 Determine if a Denominator Exception is Needed

An exception permits the exercise of clinical judgment and implies that the treatment was at least considered for, or offered to, each potentially eligible patient. Exceptions are most appropriate when contraindications to drugs or procedures being measured are relative (Spertus et al., 2010). A denominator exception is only used in proportion measures. It is not appropriate for ratio or CV measures.

An example of an exception allowing for clinical judgment in the case of two chronic conditions includes:

- Asthma is an allowable denominator exception for the performance measure of the use of beta blockers for patients with heart failure. Thus, physician judgment may determine there is greater benefit for the patient to receive this treatment for heart failure than the risk of a problem occurring due to the patient’s coexisting condition of asthma. Because the medication was given, the measure implementer does not search for exceptions, and the patient remains in the denominator. If the medication is not given, the implementer looks for exceptions and removes the patient, in this example a patient with asthma, from the denominator. If the medication was not given and the patient does not have any exceptions, the patient remains in the denominator and the provider fails the measure.

An exception should be specifically defined where capturing the information in a structured manner fits the clinical workflow. Allowable reasons fall into three general categories: medical, patient, and system:

- Medical reasons should be precisely defined and evidence-based. The events excepted should occur often enough to distort the measure results if they are not accounted for. A broadly defined medical reason, such as “any reason documented by physician,” may create an uneven comparison if some physicians have reasons that may not be evidence-based. Medical reasons resulting in an exception, if found to be in high enough volume and of universal applicability, should be considered for redefinition as an exclusion. Example: The medication specified in the
The numerator is shown to cause harm to fetuses and the patient’s pregnancy is documented as the reason for not prescribing an indicated medication.

- A patient’s reasons for not receiving the service specified may be an exception to allow for patient preferences. Example: The patient has a religious conviction that precludes the patient from receiving the specific treatment. The physician explained the benefits of the treatment and documented the patient’s refusal in the record.
- System reasons are generally rare. They should be limited to identifiable situations that are known to occur. Example: A vaccine shortage prevented administration of the vaccine.

The exception must be captured by explicitly defined data elements that allow analysis of the exception to identify patterns of inappropriate exception and gaming and to detect potential healthcare disparity issues. Analysis of rates without attention to exception information has the potential to mask disparities in healthcare and differences in provider performance.

Examples:

- Inappropriate exception: A notation in the medical record indicates a reason for not performing the specified care, and the reason is not supported by scientific evidence.
- Gaming: Patient refusal may be an exception; however, it has the potential to be overused. For example, a provider does not actively encourage the service, explain its advantages, or attempt to persuade the patient, and then uses patient refusal as the reason for nonperformance.
- Disparity issues: The use of a patient reason for exception for mammograms are noted to be high for a minority population, which may indicate a need for a more targeted, culturally appropriate patient education or closer examination of patient access such as lack of transportation or childcare.

To ensure transparency, an allowable exception must be captured in a way that it could be reported separately, in addition to the overall measure rate. An allowable exception must be supported by:

- Evidence of sufficient frequency of occurrence such that the measure results will be distorted without exception
- Evidence that the exception is clinically appropriate to the eligible population for the measure.

Although no single agreed-upon approach to exceptions exists, there seems to be consensus that exceptions provide valuable information for clinical decision-making. Measure developers that build exceptions into measure logic should be cautioned that—once implemented—exception rates may be subject to reporting, auditing, endorsement/maintenance review, and validation of appropriateness, and these factors need to be factored into the measure design and development. The difficulty in capturing an exception as part of clinical workflow makes the incorporation of an exclusion more desirable in an EHR environment.

16.3.6 Determine if a Numerator Exclusion is Needed

Numerator exclusions are used only in ratio measures to define elements that should not be included in the numerator data.

Example:

- If the number of central line blood stream infections per 1,000 catheter days were to exclude infections with a specific bacterium, that bacterium would be listed as a numerator exclusion.
16.3.7 Define Stratification Scheme

Measure developers may define a stratification scheme in lieu of risk adjustment, by stratifying the population based on their risk for an outcome or procedure. They may also stratify according to a reporting scheme, for example if data are reported in strata by age groups. Measure developers should also always consider stratifying by sociodemographic characteristics, as CMS has a continued interest in identifying and mitigating disparities in clinical care areas/outcomes across patient demographics. Stratification may effectively detect potential disparities in care/outcomes among populations related to the measure focus.

When the measure definition includes stratification, each population in the measure definition should be reported both without stratification and by each stratification criteria. For measures with multiple numerators and/or strata, each patient/episode must be scored for inclusion/exclusion to every population. For example, if a measure has two numerators, and the patient is included in the first numerator, the patient should be scored for inclusion/exclusion from the populations related to the other numerators as well (e.g., Antidepressant Medication Management (NQF 0105)). Measures may also be stratified by organizational characteristics, which is known as peer grouping stratification. Peer group stratification is appropriate in any circumstance where there is unmeasured systematic and persistent patient heterogeneity and the characteristics of the organizational setting is related to that unmeasured patient heterogeneity (e.g., location). There should be an explicit hypothesis or rationale as to why the characteristic is related to the unmeasured patient heterogeneity, but not the quality construct.

If multiple rates or stratifications are required for reporting, state this in the specifications. If the allowable exclusion is included in the numerator, specify the measure to report the overall rate as well as the rate of each exclusion. If results are to be stratified by population characteristics, describe the variables used.

Examples of possible stratification schemes:

- A vaccination measure numerator that includes: (1) the healthcare worker received the vaccine; (2) the healthcare worker was offered the vaccine and declined; or (3) the healthcare worker has an allergy or other medical contraindication/condition to vaccine (e.g., Influenza Vaccination Coverage Among Healthcare Personnel (NQF 0431)).
- Overall rate includes the median time for all patients arriving to an emergency department (ED) and admitted to the facility and two strata based on principal diagnosis (e.g., Median Time from ED Arrival to ED Departure for Admitted ED Patients (NQF 0495)).
- Overall rate is reported along with the percentage of the population in each of the two strata (e.g., Screening for Depression and Follow-up Plan: Age 18 and Older (CDF AD) NQF 0418).
- A measure is to be stratified by a population type (race, ethnicity, age, social risk factors, income, region, gender, primary language, disability) (e.g., Chlamydia Screening for Women (NQF 0033)).

For eCQMs, the Reporting Stratification section is included in the human-readable document. If a measure does not have reporting strata defined, “None” is displayed as the default. If a measure contains reporting stratification, each of the reporting strata are listed separately under the population criteria section.
16.3.8 Use Positive Evidence

All queries for measure specifications should be based on the principle of positive evidence. Positive evidence is defined as data that can be used to confirm that a given criterion was met. The principle is particularly relevant where there are no data or where there are conflicting data. Where, for instance, a numerator criterion is “low density lipoprotein (LDL) cholesterol is less than 100” and there is no LDL cholesterol result in the patient record, then there is no positive evidence, and the criterion is not met. Where, for instance, a denominator criterion is “ejection fraction is less than 40%” and there is both an ejection fraction of less than 40% and an ejection fraction of greater than 40% in the patient record, then because there is positive evidence of an ejection fraction less than 40%, the criterion is met.  

16.4 Specify the Code and/or Code Systems

Most CMS measures rely at least in part on the use of various standardized codes or code systems for classifying healthcare provided in the United States. All codes (plus the code system and version they came from) that are required for the measure should be listed along with their source and instructions pertaining to their use should be explicitly stated. Specifications may require certain codes to be accompanied by other codes, or to occur in specific locations in the record, or on claims from specific provider types. Some code sets, such as Current Procedural Terminology (CPT) Codes, may require copyright statements to accompany their use.

Some CMS claims-based measures use Quality Data Codes (QDCs) to report quality measure data to CMS. QDCs are CPT Category II or Level II G-codes (Healthcare Common Procedure Codes [HCPCS]). When appropriate, QDCs are added to the CMS form 1500 version 02/12. The codes are in the measure specifications so measure developers must identify the codes when developing measure specifications as they would other codes (e.g., ICD-10-CM codes) for determining numerators and denominators.

Refer to Section 3, Chapter 17, Codes, Code Systems, and Datasets for guidance.

16.5 Construct Data Protocol

Explicitly identify the types of data and how to aggregate or link these data so that the measure can be calculated reliably and validly. Merging data from different sources or systems must be performed carefully so that errors in assumptions are not made. Some potential areas where problems may occur include:

- Difficulty in determining which data represent duplicates
- Different units of measurement used by the different data sources (such as different age groups, time frames, etc.)
- Different quality controls used by data sources.

It may be necessary to clean the merged data. If inaccurate, incomplete, or unreasonable data are found, correct the data errors or omissions.

16.5.1 Define Key Terms, Data Elements, Codes, and Code Systems

Terms used in the numerator or denominator statement, or in allowable exclusions and exceptions, need to be precisely defined. Some measures are constructed by using precisely defined components or discrete pieces of data, often called data elements. Technical specifications include the “how” and

---

32 Many measures will be more specific with respect to which observation to use when comparing against a criterion, such as “MOST RECENT ejection fraction is less than 40%.”
“where” to collect the required data elements and measures should be fully specified including all applicable definitions and codes. Precise specifications are essential for successful implementation.

Example

Up-to-date vaccination status—the type of vaccinations to be assessed need to be clearly defined along with the definition of “up-to-date.”

Patient medical record data from EHRs (for eCQMs or measures specified for use in an EHR) consist of patient-level information coded in such a way that it can be extracted in a format that can be used in a measure. Information captured by an EHR, but not coded in a structured field, may require special processing by measure implementers.

Patient medical record data from paper charts and EHRs (if not specified for an EHR) will require instructions for abstraction. The level of detail may require specifying allowable terms, allowable places in the record, and the allowable values.

Examples:

- Allowable terms that can be used from the record: hypertension (HTN); high blood pressure, (↑BP).
- Allowable places within the record: problem list, history and physical, and progress notes.
- Allowable values: systolic blood pressure < 130, urine dipstick result +1 or greater.

Claims data will require information regarding type of claim, data fields, code types, and lists of codes.

Example

The AMI mortality measure includes admissions for Medicare FFS beneficiaries aged ≥ 65 years discharged from non-federal acute care hospitals having a principal discharge diagnosis of AMI and with a complete claims history for the 12 months prior to the date of admission. The codes are International Classification of Diseases, 10th Revision, Clinical Modification, (ICD-10-CM) code I21.xx, excluding those with I22.xx (AMI, subsequent episode of care).

Include enough detailed information in the denominator, numerator, exclusions, and exceptions so that each person collecting data for the measure will interpret the specifications in the same way. If multiple data collection methods are allowed, produce detailed specifications for each separate method.

16.5.2 Describe the Level of Measurement/Analysis

The unit of measurement/analysis is the primary entity upon which the measure is applied. The procedure for attributing the measure should be clearly stated and justified. Measures should be specified with the broadest applicability (e.g., target population, setting, level of measurement/analysis) as supported by the evidence. However, a measure developed for one level may not be valid for a different level and, therefore, should be respecified and tested for reliability and validity for each setting/population.

Examples:

- A measure created to measure performance by a facility such as a hospital may or may not be valid to measure performance by an individual physician.
If a claims-based measure is being developed for Medicare use and the literature and guidelines support the measure for all adults, consider not limiting the data source to “Medicare Parts A and B claims.”

**16.5.3 Describe Sampling**

If sampling is allowed, describe the sample size or provide guidance in determining the appropriate sample size. Any prescribed sampling methodologies need to be explicitly described.

Sampling is not applicable to eCQMs.

**16.5.4 Determine Risk Adjustment**

Risk adjustment is the modeling of health outcomes or costs as a function of various risk factors. Risk adjustment is important because health outcomes and costs are often a result of the interplay of demographic, clinical, and socioeconomic factors. In the context of quality measurement, the purpose of risk adjustment is to enhance meaningful comparison across healthcare providers, health plans, or individual clinicians. It facilitates a fairer comparison of providers with different patient populations and aims at leveling the playing field (Iezzoni, 2013).

All measure specifications, including the risk adjustment methodology, should be fully disclosed to ensure transparency and accountability. The risk adjustment statistical modeling methodology including risk adjustment model performance, data elements and risk factors, and algorithm are to be fully described in the MIF and the Risk Adjustment Methodology Report. The risk adjustment model should be updated or recalibrated as needed to adjust for changes in cohort/population change, coding changes (e.g., provider coding practice, change from ICD-9 to ICD-10 coding, new codes added), newly added clinically relevant data elements (e.g., NIH Stroke Scale for stroke measures), or social risk factors (e.g., patient dual status). Section 3, Chapter 19, Risk Adjustment provides details of the procedure.

**16.5.5 Clearly Define Any Time Intervals**

Time intervals must be explicitly stated in the measure specification whenever they are used to determine cases for inclusion in the denominator, numerator, or exclusion. The measure developer must clearly indicate the index event used to determine the time intervals. Also, identify how often the numerator should be reported for each patient as well as how often a patient is included in the denominator. For example, if the numerator is performed during an episode of community-acquired pneumonia, how is that episode of community-acquired pneumonia captured correctly if a patient has three episodes of pneumonia during the measurement period?

Measure developers must:

- Avoid the use of ambiguous semantics when specifying time intervals
- State the exact interval units required to achieve the sensitivity necessary for measurement
- State the exact interval units required to achieve the level of granularity necessary to ensure the validity and reliability of the measure calculation.

International Organization for Standardization (ISO) 8601:2004 defines data elements and interchange formats for the representation of dates and times, including time intervals. The HL7 CQL specification also provides conventions that are intended to standardize the time calculation units for durations (e.g.,

---

35 The NQF submission may be acceptable for this deliverable.
difference between two date/time elements). These standards can be used in time interval calculations for any type of CQM, not just eCQMs.

Example

Medication reconciliation must be performed within 30 days following hospital discharge. Thirty (30) days is the time interval and the hospital discharge date is the index event. If the minimum sensitivity and level of granularity desired was 1 month instead of 30 days, then the measure specification should state “month” instead of “day” as the unit of time. However, as the length of a month is variable by month, it is preferable to express time intervals in terms of days.

16.5.6 Describe How the Measure Results are Scored and Reported

Most quality measures produce rates; however, there are other scoring methods such as categorical value, CV, count, frequency distribution, non-weighted score/composite/scale, ratio, and weighted score/composite/scales. Measure information is required to include a description of the scoring type.

A description of the type of scoring should be accompanied by an explanation of how to interpret the score, such as:

- Higher score indicates better quality; improvement noted as an increase in rate.
- Lower score indicates better quality; improvement noted as a decrease in the median value.
- A score within a defined interval (or a “passing score” over or under a certain threshold) indicates better quality.

16.5.7 Develop the Calculation Algorithm

The calculation algorithm—sometimes referred to as the performance calculation—is an ordered sequence of data element retrieval and aggregation through which numerator and denominator events or CV values are identified by a measure. The developer must describe how to combine and use the data collected to produce measure results. The calculation algorithm can be a graphical representation (e.g., flowchart), text description, or combination of the two. A calculation algorithm is required for the MIF and is an item in the NQF Intent to Submit and measure submission.

The development of the calculation algorithm should be based on the written description of the measure. If the written description of the measure does not contain enough information to develop the algorithm, additional details should be added to the measure. The algorithm is to be checked for consistency with the measure text, as it will serve as the basis for the development of computer programming to produce the measure results.

16.6 DOCUMENT THE MEASURES AND OBTAIN COR APPROVAL

Complete the detailed technical specifications, including any additional documents required to produce the measure as it is intended. The complete specifications, including all attachments, are documented in the MIF and MJF.34

16.6.1 Finalize Measure Name and Description

The measure name should be a very brief description of the measure’s focus and target population. If the measure is NQF-endorsed, use the NQF-endorsed title.

---

34 The NQF Submission may be acceptable for this deliverable.
Format—[target population] who received/had [measure focus]

Examples:

- Diabetes: Medical Attention for Nephropathy (NQF 0062)
- Percent of Residents Whose Need for Help with Activities of Daily Living has Increased Long Stay (NQF 0688)
- COPD: Spirometry Evaluation (NQF 0091)

For measures based on appropriate use criteria addressing overuse of certain services, there are three standardized title lead-ins:

- Appropriate Use of ...
- Appropriate Non-Use of ...
- Inappropriate Use of ... (for inverse measures—the least desirable approach).

For the measure description, measure developers should briefly describe the type of score (e.g., percentage, percentage rate, proportion, number), the target population, and the focus of measurement.

Format—Patients in the target population who received/had [measure focus] {during [time frame] if different than for target population}

The measure description should consist of standardized phrases in a standard order:

- “The percentage of”
- [gender qualifier] if applicable (e.g., “female”)
- “patients or individuals”
- [environment qualifier] (e.g., admitted to a post-anesthesia care unit [PACU])
- [age qualifier] (e.g., aged 18 years and older)
- [denominator definition] (e.g., who are under the care of an anesthesia practitioner)
- [numerator criteria] (e.g., in which a formal post-anesthetic transfer of care protocol or checklist) is used that includes key transfer of care elements.

It is important that performance measures be worded positively, when possible (i.e., to demonstrate which clinical activity is being captured in the numerator).

Examples:

- The percentage of patients 18-75 years of age with diabetes who had a nephropathy screening test or evidence of nephropathy during the measurement period. (NQF 0062)
- Percentage of healthcare personnel who receive the influenza vaccination. (NQF 0431)
- Percentage of patients 18 years of age and older who were treated with antidepressant medication, had a diagnosis of major depression, and who remained on an antidepressant medication treatment. Two rates are reported: a) percentage of patients who remained on an antidepressant medication for at least 84 days (12 weeks); b) percentage of patients who remained on an antidepressant medication for at least 180 days (6 months). (NQF 0105)
- Percentage of adult hemodialysis patient-months using a catheter continuously for 3 months or longer for vascular access. (NQF 2978)
16.6.2 Making Changes

Information from measure testing, the public comment period, updated information gathering, or other stakeholder input may result in the need to make changes to the technical specifications. The measure developer will communicate and collaborate with the TEP to incorporate these changes before submitting the MIF and MJF to the COR for approval and the Intent to Submit to the NQF.

16.6.3 Submission Forms

The MIF and MJF have been aligned with the NQF measure submission, as appropriate, to guide the measure developer throughout the measure development process in gathering the information in a standardized manner. The forms also provide a crosswalk to the fields in the NQF measure submission to facilitate online information entry if CMS decides to submit the measure for endorsement. If approved by the COR, an equivalent document that contains the same information/elements may be used.
17  CODES, CODE SYSTEMS, AND DATASETS

17.1  CODE SYSTEMS/VOCABULARIES/TERMINOLOGIES

A code system is a managed collection of concepts wherein each concept is represented by at least one internally unique code that may include a language-dependent description. Some concepts are very specific and others can be quite general. Some code systems have complex ideas that include multiple, nuanced sub-elements such as the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM). Some have internal hierarchies built upon increasing specificity (IS-A) and may also include relationships among the concepts (e.g., caused-by or finding-site). Technically, “terminology,” “vocabulary,” and “code system” are not synonyms, but in measure development these phrases are often used interchangeably. This Blueprint preferentially uses the term code system to describe the managed concept collections from which value set content is drawn.

While some code systems are broad in scope (e.g., SNOMED CT), most are focused on a specific domain (e.g., laboratory tests for LOINC, medications for RxNorm) and, therefore, only represent concepts within the domain. Many code systems overlap in coverage (e.g., ICD-10-CM and SNOMED CT); when they do, the overlap may not result in simple one-to-one mapping between the members. Each code system has an area of focused use that tends to shape how the concepts are crafted and the relationships among these concepts. For example, ICD-10-CM is focused on categorizing disorders that cause mortality and morbidity into unique buckets such that any single disorder will always be associated with only one ICD code, and this categorization is useful for healthcare billing. Other code systems are multi-hierarchical such that the concepts capture multiple nuances and serve multiple purposes.

A code system should be managed by a code system authority, such as SNOMED International for SNOMED CT, which is responsible for ongoing maintenance such as updates and corrections, and for content coherence and consistency. Code systems are not simply a list of words. Code systems are a collection of concepts (ideas) with unique identifiers that exist in some sort of structure. The code system structure should provide each concept with a code-system-specific meaning, a concept identifier (a code), a string description (the name, and a definition of the concept meaning). Code systems should ensure meaning permanence for all the concepts in the code system, which means that if the meaning of the concept changes, the code system may need to retire the old concept and introduce one or more new ones to better characterize the meaning. This practice is performed to provide consistency in data analysis and retrieval over time. Some local environments define their own code systems, and these are not easily shared outside the local institution. Successful interoperability is dependent on either using common code systems for data capture or through mapping the local content to a code system that is defined as interoperable.

17.1.1  Encoding Clinical Information

Representing clinical information using code system concepts is called encoding the clinical information. Not all useful information in a clinical record is encoded and there is significant value in simply providing free text to support clinician-to-clinician information exchange, particularly if the free text is in specifically identified sections within the EHR. However, encoded content is critical to computable interoperability because it enables computer-based systems to find and operate upon data without human intervention. Encoding also benefits clinical interoperability by enabling clinicians from one organization to understand the meaning of transmissions from another organization. That is not to say that encoding results in perfect representation of clinical information such that no review or human
analysis is necessary; encoding of the nuances in clinical care is fraught with difficulty and almost always requires compromises in precision. The best approach is to reduce the number of mapping steps required by focusing on content that can be easily captured during clinical care where metrics that are useful in the care of the patient match those used in quality assessment and decision support systems.

In the past, healthcare organizations used billing codes and human review, chart abstraction, and communication between coding personnel and clinicians to clarify information used in clinical quality assessment. This practice helped to overcome differences in understanding based on coding alone. Currently, the ability to compute quality measures and to provide direct clinical decision support entirely from detailed encoded data increases documentation time and complexity during the care process. Tradeoffs among the alternatives still need to be considered as the industry learns how to best manage the demands of fully computable and interoperable information.

17.1.2 Health Information Technology Standards Committee Recommendations

To reduce the need to encode the same information in multiple code systems, the Health Information Technology Standards Committee (HITSC) provided guidance (Table 15) on what code system is best suited for encoding key clinical domains. HITSC recognized that moving from a focus on billing codes to encoded clinical detail that uses clinical vocabularies would be a challenge. Therefore, HITSC developed a transition plan that includes a list of acceptable transition vocabularies and associated time frames for their use, reconfirming the list in a 2015 review. A transmittal letter from HITSC to the ONC provides the latest Task Force’s transitional vocabulary recommendations.

It is always important to use the appropriate standardized terminology for the task. When a measure is validated based on the use of billing codes, which are considered “transitional code systems” for eCQMs, then measure authors should fully review the impact of changing to a clinically focused code system such as SNOMED CT.

<table>
<thead>
<tr>
<th>Clinical Terminology Standards</th>
<th>Primary Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNOMED CT</td>
<td>Clinical conditions, procedures, and general clinical information only when not available in LOINC</td>
</tr>
<tr>
<td>LOINC</td>
<td>Observables including laboratory test ordered or performed and clinical observations</td>
</tr>
<tr>
<td>RxNorm</td>
<td>Non-vaccine medications</td>
</tr>
<tr>
<td>CVX—Vaccines Administered</td>
<td>Vaccine medications</td>
</tr>
<tr>
<td>Centers for Disease Control and Prevention (CDC) Race and Ethnicity</td>
<td>Race and ethnicity</td>
</tr>
<tr>
<td>UCUM</td>
<td>Units of measure</td>
</tr>
<tr>
<td>ISO-639</td>
<td>Language</td>
</tr>
<tr>
<td>Public Health Data Standards Consortium (PHDSC) Source of Payment Typology</td>
<td>Source of payment</td>
</tr>
</tbody>
</table>
The transition vocabulary standards summary and plan are listed in Table 16.

Table 16. ONC HITSC Transition Vocabulary Standards Summary and Plan

<table>
<thead>
<tr>
<th>Transition Vocabulary</th>
<th>Transition Period: Acceptable for Reporting eCQM Results</th>
<th>Final Date for Reporting eCQM Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Classification of Diseases-9th Revision-Clinical Modification (ICD-9-CM)</td>
<td>With dates of service before the implementation of ICD-10</td>
<td>Not acceptable for reporting eCQM results for services provided after the implementation of ICD-10.</td>
</tr>
<tr>
<td>ICD-10-CM</td>
<td>With dates of service on or after the implementation of ICD-10</td>
<td>Final Date: Initially proposed 1 year after Stage 3 is effective; the 2015 HITSC proposal asks ONC to set a date in the future as the time for finalizing the transition.</td>
</tr>
<tr>
<td>ICD-10-Procedure Coding System (PCS)</td>
<td>With dates of service on or after the implementation of ICD-10</td>
<td>Final Date: Initially proposed 1 year after Stage 3 is effective; the 2015 HITSC proposal asks ONC to set a date in the future as the time for finalizing the transition.</td>
</tr>
<tr>
<td>CPT</td>
<td>During Stage 1, 2, 3 if unable to report using clinical vocabulary standards</td>
<td>Final Date: Initially proposed 1 year after Stage 3 is effective; the 2015 HITSC proposal asks ONC to set a date in the future as the time for finalizing the transition.</td>
</tr>
<tr>
<td>Healthcare Common Procedure Coding System (HCPCS)</td>
<td>During Stage 1, 2, 3 if unable to report using clinical vocabulary standards</td>
<td>Final Date: Initially proposed 1 year after Stage 3 is effective; the 2015 HITSC proposal asks ONC to set a date in the future as the time for finalizing the transition.</td>
</tr>
</tbody>
</table>

17.1.3 Quality Data Model Categories with Recommended Code Systems

Table 17 provides guidance on the recommended code system to be used when a noted clinical concept is required for an eCQM, which includes general guidance for clinical concepts in any CQM as well as transitional vocabularies, where specified. The code system guidance included indicates specific code system hierarchies, semantic types, or expected concept (code) attributes that characterize the concepts that should be used for that specific QDM category. These codes are strongly recommended constraints on concept choices for use in the value sets needed, but there can be situations where the recommended approach does not provide the needed concept (e.g., when the recommendation for a result of an observation is SNOMED CT). However, in some cases the expected result is numeric and not encoded. The result for a LOINC observation is restricted to a normative list of LOINC answer codes, or the answer concept is new and not yet available in SNOMED CT. In these situations, measure developers are encouraged to ask for guidance after considering the approach outlined.

---

35 These recommendations were made by HITSC in 2012 and 2015 using program information and language current at the time. The table includes some further clarifications from the eCQM Governance Group to increase harmonization among measure developers. With the adoption of the Quality Payment Program and other changes to quality reporting programs, these recommendations may be updated in the future.

36 ONC continues support for transitional vocabularies and has not identified a final date.

37 ICD-9-CM can be used for look-back and historical data.
### Table 17. QDM Categories with ONC HITSC Recommended Vocabularies

<table>
<thead>
<tr>
<th>General Clinical Concept</th>
<th>QDM Datatypes</th>
<th>QDM Attribute</th>
<th>Clinical Vocabulary Standards</th>
<th>Transition Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adverse Effect/Allergy/Intolerance</td>
<td>Adverse Event</td>
<td>Code (the causative agent of the adverse event)</td>
<td>Medication: RxNorm ingredient type (TTY)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SNOMED CT Substance for <strong>drug class only</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other causative agents: SNOMED CT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(product, substance if not a product)</td>
<td></td>
</tr>
<tr>
<td>Adverse Effect/Allergy/Intolerance</td>
<td>Adverse Event</td>
<td>Type (the reaction)</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Adverse Effect/Allergy/Intolerance</td>
<td>Allergy/Intolerance</td>
<td>Code (the causative agent of the allergy/intolerance)</td>
<td>Medication: RxNorm ingredient type (TTY)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SNOMED CT Substance for <strong>drug class only</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other causative agents: SNOMED CT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(substance)</td>
<td></td>
</tr>
<tr>
<td>Adverse Effect/Allergy/Intolerance</td>
<td>Intervention, Adverse Event Intervention, Intolerance</td>
<td>Type (the reaction)</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Care Experience</td>
<td>Patient Care Experience</td>
<td>Code</td>
<td>SNOMED CT (or LOINC if part of an Evaluation Tool)</td>
<td>N/A</td>
</tr>
<tr>
<td>Care Experience</td>
<td>Provider Care Experience</td>
<td>Code</td>
<td>SNOMED CT (or LOINC if part of an Evaluation Tool)</td>
<td>N/A</td>
</tr>
<tr>
<td>Substance</td>
<td>Substance, Administered Substance, Order Substance, Recommended</td>
<td>Code</td>
<td>SNOMED CT (substance if not a product)</td>
<td>N/A</td>
</tr>
<tr>
<td>Substance</td>
<td>Substance, Administered Substance, Order Substance, Recommended</td>
<td>Negation rationale</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Care Goal</td>
<td>Care Goal</td>
<td>Code</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Condition/Diagnosis/Problem</td>
<td>Diagnosis</td>
<td>Code</td>
<td>SNOMED CT (disorders, findings)</td>
<td>ICD-9-CM, ICD-10-CM</td>
</tr>
<tr>
<td>Condition/Diagnosis/Problem</td>
<td>Diagnosis</td>
<td>Anatomical location site</td>
<td>SNOMED CT (body structure)</td>
<td>N/A</td>
</tr>
<tr>
<td>Symptom</td>
<td>Diagnosis</td>
<td>Severity</td>
<td>SNOMED CT (qualifier)</td>
<td>N/A</td>
</tr>
<tr>
<td>Symptom</td>
<td>Diagnosis</td>
<td>Code</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

---

38 These recommendations were made by HITSC in 2012 and 2015 using program information and language current at the time and are consistent with the 2019 Interoperability Standards Advisory. With the adoption of the Quality Payment Program and other changes to quality reporting programs, these recommendations may be updated in the future.
<table>
<thead>
<tr>
<th>General Clinical Concept</th>
<th>QDM Datatypes</th>
<th>QDM Attribute</th>
<th>Clinical Vocabulary Standards</th>
<th>Transition Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encounter (any patient-provider interaction [e.g., telephone call, email] regardless of reimbursement status, status—includes traditional face-to-face encounters)</td>
<td>Encounter, Order, Encounter, Performed</td>
<td>Code</td>
<td>SNOMED CT (procedure)</td>
<td>CPT, HCPCS, ICD-9-CM Procedures, ICD-10-CM, ICD-10-PCS</td>
</tr>
<tr>
<td>Encounter (any patient-provider interaction [e.g., telephone call, email] regardless of reimbursement status, status—includes traditional face-to-face encounters)</td>
<td>Encounter, Recommended</td>
<td>Negation rationale</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Encounter (any patient-provider interaction [e.g., telephone call, email] regardless of reimbursement status, status—includes traditional face-to-face encounters)</td>
<td>Encounter, Recommended</td>
<td>Reason</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Encounter (any patient-provider interaction [e.g., telephone call, email] regardless of reimbursement status, status—includes traditional face-to-face encounters)</td>
<td>Encounter, Performed</td>
<td>Diagnoses</td>
<td>SNOMED CT (disorders, findings)</td>
<td>ICD-9-CM, ICD-10-CM</td>
</tr>
<tr>
<td>Encounter (any patient-provider interaction [e.g., telephone call, email] regardless of reimbursement status, status—includes traditional face-to-face encounters)</td>
<td>Encounter, Performed</td>
<td>Principal diagnosis</td>
<td>SNOMED CT (disorders, findings)</td>
<td>ICD-9-CM, ICD-10-CM</td>
</tr>
<tr>
<td>Encounter (any patient-provider interaction [e.g., telephone call, email] regardless of reimbursement status, status—includes traditional face-to-face encounters)</td>
<td>Encounter, Performed</td>
<td>Facility location</td>
<td>HL7 HealthcareServiceLocation codes (HSLOC)</td>
<td>N/A</td>
</tr>
<tr>
<td>General Clinical Concept</td>
<td>QDM Datatypes</td>
<td>QDM Attribute</td>
<td>Clinical Vocabulary Standards</td>
<td>Transition Vocabulary</td>
</tr>
<tr>
<td>--------------------------------------------------------------</td>
<td>----------------------------------------------------</td>
<td>--------------------</td>
<td>----------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Encounter (any patient-provider interaction [e.g., telephone call, email] regardless of reimbursement status, status—includes traditional face-to-face encounters)</td>
<td>Encounter, Performed</td>
<td>Admission source</td>
<td>SNOMED CT (environment)</td>
<td>N/A</td>
</tr>
<tr>
<td>Encounter (any patient-provider interaction [e.g., telephone call, email] regardless of reimbursement status, status—includes traditional face-to-face encounters)</td>
<td>Encounter, Performed</td>
<td>Discharge disposition</td>
<td>SNOMED CT (environment)</td>
<td>N/A</td>
</tr>
<tr>
<td>Family History</td>
<td>Family History</td>
<td>Code</td>
<td>SNOMED CT (person)</td>
<td>N/A</td>
</tr>
<tr>
<td>Family History</td>
<td>Family History</td>
<td>Relationships</td>
<td>SNOMED CT (person, situation)</td>
<td>N/A</td>
</tr>
<tr>
<td>Device</td>
<td>Device, Applied</td>
<td>Code</td>
<td>SNOMED CT (physical object)</td>
<td>N/A</td>
</tr>
<tr>
<td>Device</td>
<td>Device, Applied</td>
<td>Anatomical approach site</td>
<td>SNOMED CT (body structure)</td>
<td>N/A</td>
</tr>
<tr>
<td>Device</td>
<td>Device, Applied</td>
<td>Anatomical location site</td>
<td>SNOMED CT (body structure)</td>
<td>N/A</td>
</tr>
<tr>
<td>Device</td>
<td>Device, Applied</td>
<td>Reason</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Device</td>
<td>Device, Applied</td>
<td>Negation rationale</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Physical Exam (definition of the components of the physical exam performed)</td>
<td>Physical Exam, Order</td>
<td>N/A</td>
<td>LOINC</td>
<td>N/A</td>
</tr>
<tr>
<td>Physical Exam (expression of the answers/responses for the physical exam component)</td>
<td>Physical Exam, Performed</td>
<td>Result</td>
<td>SNOMED CT (disorders, findings) or LOINC Normative Responses</td>
<td>N/A</td>
</tr>
<tr>
<td>Laboratory Test (names)</td>
<td>Laboratory Test, Order</td>
<td>Code</td>
<td>LOINC</td>
<td>N/A</td>
</tr>
<tr>
<td>General Clinical Concept</td>
<td>QDM Datatypes</td>
<td>QDM Attribute</td>
<td>Clinical Vocabulary Standards</td>
<td>Transition Vocabulary</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>---------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Laboratory Test (names)</td>
<td>Laboratory Test, Order Laboratory Test, Performed Laboratory Test, Recommended</td>
<td>Reason</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Laboratory Test (names)</td>
<td>Laboratory Test, Order Laboratory Test, Performed Laboratory Test, Recommended</td>
<td>Negation rationale</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Laboratory Test (results)</td>
<td>Laboratory Test, Performed</td>
<td>Result</td>
<td>SNOMED CT (disorders, findings) or LOINC Normative Responses</td>
<td>N/A</td>
</tr>
<tr>
<td>Diagnostic Study Test Names</td>
<td>Diagnostic Study, Order Diagnostic Study, Performed Diagnostic Study, Recommended</td>
<td>Code</td>
<td>LOINC</td>
<td>HCPCS</td>
</tr>
<tr>
<td>Diagnostic Study Test Names</td>
<td>Diagnostic Study, Order Diagnostic Study, Performed Diagnostic Study, Recommended</td>
<td>Reason</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Diagnostic Study Test Names</td>
<td>Diagnostic Study, Order Diagnostic Study, Performed Diagnostic Study, Recommended</td>
<td>Negation rationale</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Diagnostic Study Test Results</td>
<td>Diagnostic Study, Performed</td>
<td>Result</td>
<td>SNOMED CT (disorders, findings) or LOINC Normative Responses</td>
<td>N/A</td>
</tr>
<tr>
<td>Units of Measure for Results</td>
<td>Laboratory test, Performed Diagnostic Study, Performed</td>
<td>Result (units)</td>
<td>UCUM – Unified Code for Units of Measure</td>
<td>N/A</td>
</tr>
<tr>
<td>Intervention</td>
<td>Intervention, Order Intervention, Performed Intervention, Recommended</td>
<td>Code</td>
<td>SNOMED CT (disorders, findings)</td>
<td>CPT, HCPCS, ICD-9-CM Procedures, ICD-10-PCS</td>
</tr>
<tr>
<td>Intervention</td>
<td>Intervention, Order Intervention, Performed Intervention, Recommended</td>
<td>Reason</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Intervention</td>
<td>Intervention, Order Intervention, Performed Intervention, Recommended</td>
<td>Negation rationale</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Intervention</td>
<td>Intervention, Performed</td>
<td>Result</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Procedure</td>
<td>Procedure, Order Procedure, Performed Procedure, Recommended</td>
<td>Code</td>
<td>SNOMED CT (procedures, regime/therapy)</td>
<td>CPT, HCPCS, ICD-9-CM Procedures, ICD-10-PCS</td>
</tr>
<tr>
<td>Procedure</td>
<td>Procedure, Order Procedure, Performed Procedure, Recommended</td>
<td>Reason</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>General Clinical Concept</td>
<td>QDM Datatypes</td>
<td>QDM Attribute</td>
<td>Clinical Vocabulary Standards</td>
<td>Transition Vocabulary</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>---------------------------------------------------</td>
<td>-----------------</td>
<td>-------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Procedure</td>
<td>Procedure, Order Procedure, Performed Procedure, Recommended</td>
<td>Negation rationale</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Procedure</td>
<td>Procedure, Performed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment Instrument Questions (e.g., questions for assessing patient status used as part of clinical workflow, clinical outcome evaluation, social functional and emotional status, patient preference, experience, characteristics)</td>
<td>Assessment, Performed Assessment, Recommended</td>
<td>Code</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Assessment Instrument Questions (e.g., questions for assessing patient status used as part of clinical workflow, clinical outcome evaluation, social functional and emotional status, patient preference, experience, characteristics)</td>
<td>Assessment, Performed Assessment, Recommended</td>
<td>Reason</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Assessment Instrument Questions (e.g., questions for assessing patient status used as part of clinical workflow, clinical outcome evaluation, social functional and emotional status, patient preference, experience, characteristics)</td>
<td>Assessment, Performed Assessment, Recommended</td>
<td>Negation rationale</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Assessment Instrument Questions (e.g., questions for assessing patient status used as part of clinical workflow, clinical outcome evaluation, social functional and emotional status, patient preference, experience, characteristics)</td>
<td>Assessment, Performed</td>
<td>Result</td>
<td>SNOMED CT (disorders, findings) or LOINC Normative Responses</td>
<td>N/A</td>
</tr>
<tr>
<td>Categories of Function</td>
<td>Assessment, Performed Assessment, Recommended</td>
<td>Code</td>
<td>ICF – International Classification of Functioning, Disability, and Health $^{39}$</td>
<td>N/A</td>
</tr>
</tbody>
</table>

$^{39}$ The HITSC Task Force recommended use of ICF; however, there is a fee associated with ICF. The 2019 Interoperability Standards Advisory recommends LOINC for observations and SNOMED CT for observation values.
<table>
<thead>
<tr>
<th>General Clinical Concept</th>
<th>QDM Datatypes</th>
<th>QDM Attribute</th>
<th>Clinical Vocabulary Standards</th>
<th>Transition Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>Communication: from Patient to Provider  Communication: from Provider to Patient Communication: from Provider to Provider</td>
<td>N/A</td>
<td>SNOMED CT (disorders, findings)</td>
<td>CPT, HCPCS</td>
</tr>
<tr>
<td>Communication</td>
<td>Communication: from Patient to Provider  Communication: from Provider to Patient Communication: from Provider to Provider</td>
<td>Negation rationale</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Medications (administered, excluding vaccines)</td>
<td>Medication, Active Medication, Administered Medication, Discharge Medication, Dispensed Medication, Order</td>
<td>Code</td>
<td>RxNorm</td>
<td>N/A</td>
</tr>
<tr>
<td>Medications (administered, excluding vaccines)</td>
<td>Medication, Active Medication, Administered Medication, Discharge Medication, Dispensed Medication, Order</td>
<td>Reason</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Vaccines</td>
<td>Immunization, Administered Immunization, Order</td>
<td>Code</td>
<td>CVX—Vaccines</td>
<td>N/A</td>
</tr>
<tr>
<td>Vaccines</td>
<td>Immunization, Administered Immunization, Order</td>
<td>Code (for procedure for administering a vaccine)</td>
<td>SNOMED CT (procedures)</td>
<td>CPT</td>
</tr>
<tr>
<td>Vaccines</td>
<td>Immunization, Administered Immunization, Order</td>
<td>Reason</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Vaccines</td>
<td>Immunization, Administered Immunization, Order</td>
<td>Negation rationale</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Patient Characteristic, Date of Birth</td>
<td>Patient Characteristic Birthdate</td>
<td>Code</td>
<td>Fixed to LOINC code 21112-8 (birth date); therefore, cannot be further qualified with a value set</td>
<td>N/A</td>
</tr>
<tr>
<td>Patient Characteristic, Expired</td>
<td>Patient Characteristic Expired</td>
<td>Code</td>
<td>Fixed to SNOMED CT code 419099009 (dead); therefore, cannot be further qualified with a value set</td>
<td>N/A</td>
</tr>
<tr>
<td>Patient Characteristic, Sex</td>
<td>Patient Characteristic Sex</td>
<td>Code</td>
<td>ONC Administrative Sex — VSAC OID 2.16.840.1.113762.1.4.1</td>
<td>N/A</td>
</tr>
<tr>
<td>General Clinical Concept</td>
<td>QDM Datatypes</td>
<td>QDM Attribute</td>
<td>Clinical Vocabulary Standards</td>
<td>Transition Vocabulary</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------</td>
<td>---------------</td>
<td>------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Patient Characteristic, Ethnicity</td>
<td>Patient Characteristic Ethnicity</td>
<td>Code</td>
<td>CDC National Center for Health Statistics – VSAC OID 2.16.840.1.114222.4.11.837 Detailed Ethnicity: HL7 Terminology – VSAC OID 2.16.840.1.114222.4.11.877</td>
<td>N/A</td>
</tr>
<tr>
<td>Patient Characteristic, Race</td>
<td>Patient Characteristic Race</td>
<td>Code</td>
<td>CDC National Center for Health Statistics – VSAC OID 2.16.840.1.114222.4.11.836 Race Value Set: HL7 Terminology – VSAC OID 2.16.840.1.113883.1.11.14914</td>
<td>N/A</td>
</tr>
<tr>
<td>Patient Characteristic, Preferred Language</td>
<td>Individual Characteristic</td>
<td>Code</td>
<td>RFC 5646</td>
<td>N/A</td>
</tr>
<tr>
<td>Patient Characteristic, Payer</td>
<td>Patient Characteristic Payer</td>
<td>Code</td>
<td>Payer Typology (PHDSC Payer Typology) (scope) VSAC OID 2.16.840.1.114222.4.11.3591 <a href="http://phdsc.org/standards/payer-typology.asp">http://phdsc.org/standards/payer-typology.asp</a></td>
<td>N/A</td>
</tr>
<tr>
<td>Patient Characteristic</td>
<td>Patient Characteristic, Clinical Trial Participant</td>
<td>Code</td>
<td>SNOMED CT (findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Patient Characteristic</td>
<td>Patient Characteristic, Clinical Trial Participant</td>
<td>Reason</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Patient Characteristic, Unspecified</td>
<td>Patient Characteristic (unspecified)</td>
<td>Code</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
<tr>
<td>Provider Characteristic, Unspecified</td>
<td>Provider Characteristic (unspecified)</td>
<td>Code</td>
<td>SNOMED CT (disorders, findings)</td>
<td>N/A</td>
</tr>
</tbody>
</table>
17.1.4 Use of Specific Code Systems

17.1.4.1 International Classification of Diseases

ICD is used to represent patient information on claims records, data collection for use in performance measurement, reimbursement for medical claims, and more. In the United States, data submitted to CMS transitioned from ICD-9-CM to ICD-10-CM/PCS beginning October 1, 2015. It is generally a good practice not to change originally captured patient information. There is not a simple method to crosswalk from ICD-9-CM to ICD-10-CM/PCS, so most legacy data using ICD-9-CM will remain archived in that form. The ICD-10 classification systems provide significant improvements through greater detailed information and the ability to capture additional advancements in clinical medicine, but the transition does create difficulties for monitoring trends when data are captured using both code systems.

ICD-10-CM/PCS consists of two parts:

- ICD-10-CM—Diagnosis classification system developed by the CDC National Center for Health Statistics (NCHS) for use in the United States healthcare treatment settings. Diagnosis coding under this system uses three to seven alpha and numeric digits and full code titles, but the format is the same as ICD-9-CM.
- ICD-10-PCS—Procedure classification system developed by CMS for use in the United States for inpatient hospital settings. The new procedure coding system uses seven alpha or numeric digits, whereas the ICD-9-CM coding system uses three or four numeric digits.

Codes that are not valid for clinical coding (e.g., ICD-10-CM Group Codes) should not be included in value sets. Specifically, codes that are associated with sections or groups of codes should not be used in value sets:

- Codes for dermal burns. Use the fourth digit (0–9) to identify the percentage of body surface and the fifth digit to indicate the percentage with third degree burns:
  - T31.0, Burns involving less than 10% of body surface
  - T31.10, Burns involving 10-19% of body surface with 0-9% third-degree burns
  - T31.11, Burns involving 10-19% of body surface with 10-19% third-degree burns
  - T31.20, Burns involving 20-29% of body surface with 0-9% third-degree burns
  - T31.21, Burns involving 20-29% of body surface with 10-19% third-degree burns
  - T31.22, Burns involving 20-29% of body surface with 20-29% third-degree burns.
- A09, Infectious Gastroenteritis and Colitis, unspecified—a standalone code and does not require additional digits to be valid.
- A08, Viral and Other Specified Intestinal Infections—a non-billable code that must have additional digits to be valid (e.g., A08.11 Acute Gastroenteropathy Due to Norwalk Agent).

When a developer submits ICD-10-CM/PCS codes for consideration by the NQF for measures that previously used ICD-9-CM codes, the NQF requires additional requirements be met. These requirements are outlined in the NQF document Inclusion of ICD-10 Codes in Measures Using ICD-9 Codes.

Processes for requesting changes to ICD-10-PCS are found on the CMS website. Processes for requesting changes to ICD-10-CM are found on the CDC NCHS website. Measure developers should consider contractual timelines when considering applying for new concepts.


CPT is a registered trademark of the American Medical Association (AMA) for the CPT, Fourth Edition (CPT4). The CPT Category I (CPT I) codes are a listing of descriptive terms and identifying codes for
reporting medical services and procedures performed by physicians. The purpose of the terminology is to provide a uniform language that will accurately describe medical, surgical, and diagnostic services, and thereby provides an effective means for reliable nationwide communication among physicians, patients, and third parties.

Each CPT record corresponds to a single observation or diagnosis. The CPT codes are not intended to transmit all possible information about an observation or diagnosis; they are only intended to identify the observation or diagnosis. The CPT code for a name is unique and permanent.

CPT Category II (CPT II) codes, developed through the CPT Editorial Panel for use in performance measurement, serve to encode the clinical actions described in a measure’s numerator. CPT II codes consist of five alphanumeric characters in a string ending with the letter “F.”

CPT Category II codes are not used in eCQMs.

The AMA requires users to include a set of notices and disclosures when publishing measures using CPT codes. Contact the COR to obtain the current full set of notices and disclaimers that includes:

- Copyright notice
- Trademark notice
- Government rights statement
- AMA disclaimer.

CPT codes are updated annually. For questions regarding the use of CPT codes, contact the AMA CPT Information and Education Services at 800-634-6922 or at http://www.ama-assn.org. Measure developers should consider contractual timelines when considering applying for new concepts.

17.1.4.3 SNOMED CT

SNOMED CT is owned and maintained by SNOMED International. SNOMED CT contains more than 357,000 healthcare concepts with unique meanings and formal logic-based definitions organized into hierarchies. Each hierarchy is identified with a unique “semantic type” that is included (in parenthesis) in the fully specified name of every concept in the hierarchy. These semantic type identifiers are included in Table 17. It should be noted that the recommended semantic types noted in the table are strong guidance but some “conditions” may only be found in the event, or the situation hierarchy. The fully populated code system list with unique descriptions for each concept contains more than 957,000 descriptions. Approximately 1.37 million semantic relationships exist to enable reliability and consistency of data retrieval.

SNOMED CT is a general clinical “reference terminology,” meaning that it is intended to represent clinical concepts across many domains, which includes conditions, diagnoses, symptoms, and signs, all of which are a type of “finding.” SNOMED CT also represents procedures, observations, and some laboratory tests, drugs, and devices. There are also concepts for ancillary aspects that can be used for documentation of the domains noted in the table. As a general reference terminology, SNOMED CT is expected to provide many of the concepts needed for clinical information encoding, and unless a specific terminology is otherwise noted, it should be the primary source for standardized terminology encoding.

SNOMED International maintains the SNOMED CT technical design, core content architecture, SNOMED CT Core content, and SNOMED CT documentation. SNOMED CT content includes the technical specification of SNOMED CT and fully integrated multi-specialty clinical content. The core content
includes the concepts table, description table, relationships table, history table, ICD-10-CM mapping, and Technical Implementation Guide.

In the United States, SNOMED CT is distributed by the NLM, which acts as the U.S. SNOMED CT Release Center. There are multiple SNOMED CT Release Centers across the globe and many, including the NLM, release a specific “edition” of SNOMED CT for use in their specific “realm.” In the United States, this release is called the U.S. Edition of SNOMED CT and contains the combination of the International Core (same in every edition of SNOMED CT) and a U.S.-specific extension that contains concepts only in use within the United States. Thus, the U.S. Edition of SNOMED CT will contain some SNOMED CT concepts that do not appear in other editions, such as that used in Canada or in the United Kingdom. Over time, realm-specific concepts will be promoted from the realm-specific edition to the International Core and then be available to all users. When this occurs, the concept identifier will not change. Only the U.S. Edition of SNOMED CT is available for use in the United States.

The SNOMED CT codes are not intended to transmit all possible information about a condition, observation, or procedure. They are only intended to identify the condition, observation, or procedure. The SNOMED CT code for a concept is unique and permanent.

At times, there may be a need to request new SNOMED CT concepts. The request should be submitted through the US SNOMED CT Content Request System (USCRS) of the NLM. The NLM will evaluate the request and determine if a useful addition should be included only in the U.S. Edition of SNOMED CT or also be promoted to the International Core. Measure developers must sign up for a UMLS Terminology Services account to log into the USCRS. Measure developers should consider contractual timelines when considering applying for new concepts. For information on obtaining the standard, contact SNOMED International or the NLM.

Allergy Value Sets

SNOMED CT may be used to represent allergen drug class concepts only when following the guidance below. Refer to the RxNorm section below for additional guidance.

Measure Developer Guidance

Beginning with the 2020 reporting year, in addition to the required approach for representing medication allergen (or ingredient) substances using RxNorm ingredient type concepts noted below, eCQMs can add SNOMED CT drug class concepts to represent medication allergens. If a drug class concept is deemed appropriate for use in the measure the following is expected to occur:

- Only include SNOMED CT drug class concepts when a general drug class concept is expected to be found in patient records as an indication that the patient is considered allergic to all drugs in the class.
- Measure authors should keep in mind that when a drug class concept is used, this means that no drug in the class can be an expected therapy for the patient.
- Measure authors should review all drugs that are defined to be included in the class when choosing to include the SNOMED CT drug class concept.
- Measure authors should also define an RxNorm Allergy value set with the specific ingredient (IN) and precise ingredient (PIN) TTY drug ingredients that represent the drug class.
- A grouping value set, while not required, can be created that groups both the RxNorm ingredient type allergy value set and the SNOMED CT drug class allergy value set into one grouping value set that is then referenced in the measure. Using a grouping value set eases development and testing burden on EHR vendors.
The SNOMED CT substance hierarchy is expected to be used for coding non-medication allergy-inducing entities.

17.1.4.4  RxNorm

The NLM produces RxNorm. As described by the NLM, RxNorm is a normalized naming system for generic and branded drugs and is a tool for supporting semantic interoperation between drug terminologies and pharmacy knowledgebase systems.

17.1.4.4.1  Purpose of RxNorm

Hospitals, pharmacies, and other organizations use computer systems to record and process drug information. Because these systems use many different sets of drug names, it can be difficult for one system to communicate with another. To address this challenge, RxNorm provides normalized names and unique identifiers for medicines and drugs. The goal of RxNorm is to allow computer systems to communicate drug-related information efficiently and unambiguously.

17.1.4.4.2  Scope of RxNorm

RxNorm contains the names of prescription and many over-the-counter drugs available in the United States. RxNorm includes generic and branded drugs including:

- Clinical drugs—pharmaceutical products given to (or taken by) a patient with therapeutic or diagnostic intent.
- Drug packs—packs that contain multiple drugs, or drugs designed to be administered in a specified sequence.

As noted on the RxNorm Overview page, radiopharmaceuticals, bulk powders, contrast media, food, dietary supplements, and medical devices such as bandages and crutches are generally out of scope for RxNorm.

17.1.4.4.3  RxNorm Term Types

RxNorm characterizes each concept in the code system as having a specific term type (TTY). Term types are a semantic tag that describe the type of information that concept conveys. A list of all RxNorm term types can be found in an appendix of the RxNorm Technical Documentation. In subsequent sections of this chapter, specific TTYs are recommended when building value sets for specific uses.

17.1.4.4.4  RxNorm Use in Quality Measures

RxNorm is the recommended national standard for medication vocabulary for clinical drugs and drug delivery devices. RxNorm is intended to cover all prescription medications approved for human use in the United States. RxNorm should be used to reference a medication for administration, order, and dispensing. RxNorm should also be used to represent the object (i.e., the causative agent) of an allergy, adverse reaction, or intolerance due to a drug.

Because every drug information system that is commercially available today follows somewhat different naming conventions, a standardized nomenclature is needed for the smooth exchange of information. The goal of RxNorm is to allow various systems using different drug nomenclatures to share data efficiently at the appropriate level of abstraction. Each RxNorm clinical drug name reflects the active ingredients, strengths, and dose form comprising that drug. When any of these elements vary, a new RxNorm drug name is created as a separate concept.
Note that Blueprint content is the broadest interpretation of the RxNorm TTYs with which a measure developer could align, but some eCQM releases include value sets that focus on the minimum needed RxNorm identifiers for all general representations of the necessary drugs. While the Blueprint includes branded TTYs in the guidance, authoring guidance has encouraged measure developers not to include branded term types because changes in branded identifiers for any single “general drug” (such as a Semantic Clinical Drug [SCD]) occur throughout the year and, even with the inclusion of value set addendum releases, there can be value sets that are out of synch with some implementer system content. Also, it provides for impartiality reducing the perception of branded drug favoritism. Given that RxNorm application content (and all drug information vendor products) can be used to map from the more stable general identifier to a branded identifier, and from other code systems such as National Drug Code (NDC) or proprietary code systems, the branded RxNorm TTYs were often not included under the assumption that if an implementer had a different identifier, they could map to the included SCD RXCUI or GPCK RXCUI or any other TTY and ID according to the intention.

More information can be found at the RxNorm website.

17.1.4.4.5 Allergy Value Sets

Allergy/intolerance value sets, when drawn from RxNorm, should include only the IN or PIN TTY. Authors should also review the guidance provided in the SNOMED CT section, 17.1.4.3.

Measure Developer Guidance:

- Always consider including a measure expression that appropriately removes a patient from a numerator or denominator population when there is an expectation that the patient should have received a substance, but the patient has an allergy/intolerance to the expected substance.
- Understand that if an “allergy/intolerance” value set is created and a patient has an allergy/intolerance to any one of the substances, that will likely remove the patient from consideration for any substance in that value set.
- When an allergy/intolerance value set is needed, it may be reasonable to identify the active ingredients for those medications included in the value sets used for expected therapies and then create the allergy/intolerance value set using that list of ingredients.
- The allergy/intolerance value sets only indicate the substance/agent considered as the cause of the reaction. RxNorm is not used to indicate the reaction.
- The naming convention for value sets used for allergy/intolerance is to end the value set name with the word “allergen,” (e.g., “Antithrombotic Therapy Allergen” and “Beta Blocker Therapy Allergen”).

17.1.4.4.6 Medication Value Sets

Medication value sets when used to represent administered, ordered, or recommended substances, must be drawn from RxNorm and must include RxNorm concepts having TTYs:

- Generic Pack (GPCK)
- SCD.

It is preferred that value sets used to represent acceptable medications would only include the RxNorm TTYs listed because any other drug representation can be mapped to one of these using the RxNorm content and applications, which means that value sets should have any formulation that is appropriate for use. It also means that implementers are expected to map content found in some EHR records to the most appropriate concept in the value set, preferably the concept with either a SCD or GPCK TTY.
17.1.4.5  Logical Observation Identifier Names and Codes

LOINC is a code system (i.e., set of identifiers, names, and codes) for clinical and laboratory observations, healthcare screening/survey instruments, and document type identifiers. Each LOINC record corresponds to a single observation of almost any type (i.e., observables) and is best known for concepts that represent laboratory tests. It is also used to represent document types and, thus, is frequently used to represent a document section in consolidated clinical document architecture (C-CDA) and other templated exchange standards. The LOINC codes are not intended to transmit all possible information about a test or observation; they are only intended to identify the observations. The LOINC code for a name is unique and permanent. LOINC codes must always be transmitted with a hyphen before the check digit (e.g., 10154-3). The numeric code is transmitted as a variable length number, without leading zeros. LOINC codes are available for commercial use without charge, subject to the terms of a license that assures the integrity and ownership of the codes.

17.1.4.5.1  Special Situations with LOINC Survey/Evaluation Tools

LOINC is used to represent survey instrument questions (observations). Because of the tight alignment in survey instruments of the question and the acceptable set of answers, LOINC often also includes specific answer sets for the survey questions, called LOINC Answers (LA codes). When the set of answers is defined specifically within the survey, the LA set is normative, which means only the specified LA codes are acceptable responses (values) for that LOINC observable. Defined normative LA sets can occur anywhere in LOINC, and because they are a requirement when defined, users of LOINC must look for and respect these restrictions.

Most LOINC survey/evaluation tools are assigned a LOINC code for the overall tool that is a type of “LOINC Panel” because that is the LOINC construct that collects other LOINC codes that are to be used together. In addition, each evaluation question/observation included in the survey or evaluation tool is also assigned a completely different LOINC code. In some cases, the LOINC survey will reuse observations that have been defined in a different (usually related) survey. In many cases, the LOINC survey will include information on the ownership of the survey and indicate whether the individual observations are copyright protected and whether they can be used as independent observations outside of the complete survey instrument as a complete questionnaire. Many survey instruments include summary final scores (i.e., a LOINC observation) that are based on a sum of the values associated with the specified LA codes allowed for all the component questions. All elements of a LOINC tool must be considered before use. Care must be taken when using LOINC observables with specified answer codes to determine whether the LA codes must be used and are the only set that can be used.

There may also be a need to request new LOINC concepts. Instructions and tools to request LOINC concepts can be found at the LOINC website. Measure developers should consider contractual timelines when considering applying for new concepts.

17.1.4.6  Other Important Code Systems

Some data elements are best encoded using code systems that represent a specific type of information, particularly when the code system is in widespread use. When considering inclusion of data in a measure that is not already identified in Section 3, Chapter 17.1.3, QDM Categories with Recommended Code Systems, determine whether a specific authoritative code system is in widespread use and consider including that code system into the measure.

---

LOINC codes are copyrighted by Regenstrief Institute and the Logical Observation Identifier Names and Codes Consortium.
Examples:

- NHSN Healthcare Facility Patient Care Location (HSLOC in VSAC)
- CVX (for vaccines)
- CDT – Code on Dental Procedures and Nomenclature
- UCUM – the Unified Code for Units of Measure
- ICF—International Classification of Functioning, Disability, and Health
- Source of Payment Typology (PHDSC Payer Typology Sub-Committee)
- HL7 (e.g., Administrative Sex, Discharge Disposition).

17.2 Value Sets

Value sets are a subset of concepts (represented by a code) drawn from one or more code systems, where the concepts included in the subset share a common scope of use. For a quality measure, value sets are used to identify a set of concepts whereby any one of the concepts included can be used to identify a patient of interest. Value sets are used in quality measures to collect all the coded concepts that can occur in the clinical record (or administrative data) and to represent patients that should be in the same population for analysis.

17.2.1 Use of Value Sets

*Coded* data elements in quality measures are bound to (i.e., may use) one of:

- A single specific code (drawn from a code system) that is directly referenced within the measure and, as such, is not in a value set; therefore, it is a DRC.
- A value set (i.e., a set of codes) where each code is considered equivalent for use in the context of that data element.

In quality measures, the patients identified using any of the codes in a value set are considered equivalent for the measure data element using the value set.

17.2.2 Constructing Value Sets

Value sets must be created with the thoughtful input of SMEs familiar with the clinical or administrative information needed, combined with the input of terminology experts familiar with the code systems to be used. This work should be strongly influenced by knowledge of how information is captured currently (both electronic encoding and traditional textual material) and the workflow necessary to accurately capture the expected information.

17.2.3 Representing the Codes to be Included

When constructing a value set, the author is actually constructing a value set definition (VSD) that may have multiple versions over time. A VSD describes the value set using metadata noted below and includes a Content Logical Definition (CLD) that identifies the specific concepts (i.e., codes) to be included in the value set expansion. A value set expansion is the actual list of codes, calculated using a specified expansion profile of code system versions and any predetermined retired (legacy) codes that are to be used when the value set is implemented. Many VSDs are constructed by enumerating each specific code that is desired, which has traditionally been called an enumerated or extensional definition. However, many value sets would be best defined logically or intensionally using the structure of the code system used (e.g., “All the codes that are descendants of the condition INSULIN DEPENDENT DIABETES MELLITUS”). VSAC provides tooling to support both extensional and intensional VSDs.
Value sets have a lifecycle similar to many persistent objects. The VSAC is a tool suite developed by NLM to support the creation, maintenance, and retrieval of value sets. In addition to the lifecycle noted in Figure 36, ongoing maintenance of value sets occur where value set authors modify the content to address improvements in clinical understanding, changes in available coded concepts that occur with updates to the code system, and errors.

17.2.4 Determining the Value Set Code System

Using the guidance noted in Section 3, Chapter 17.1.3, QDM Categories with Recommended Code Systems, a value set may need concepts from more than one code system, which is particularly important when the data element is associated with transitional code systems that may be in use currently, but the ongoing intent of the measure is to use a different code system. In this case, quality measures may reference one single value set that groups each of the value sets using different code systems (e.g., SNOMED CT value set, ICD-9-CM value set, and ICD-10-CM value set) to capture the same scope. VSAC has defined a grouping mechanism for this scenario to create a parent value set for use in the measure. Measure developers define separate value sets for different code systems, such as a Diabetes Mellitus SNOMED CT value set and a Diabetes Mellitus ICD-10-CM value set. They then define a Diabetes Mellitus Grouping value set that spools the two code system-specific value sets and then use the grouping value set for the measure data element. The VSAC allows only one level of grouping; a grouping value set cannot include another grouping value set. Therefore, the MAT measure logic must be used when the combination of two value set groupings is needed. As an example, the codes for all patients with hematologic malignancies may be linked into one value set grouping with SNOMED CT and ICD-10-CM values and patients with primary immunodeficiencies and those with human immunodeficiency virus (HIV) infection might be similarly grouped. MAT measure logic clauses would then be used to identify patients that fall into any of these groups.
17.2.4.1 Value Set Metadata

When creating a value set, the author must specify value set metadata that describe what the value set is meant to represent so that it can be used properly in measures and so that others can find it and reuse it in different measures, where appropriate. The value set metadata must include a clear and complete name (remember the value set will not exist solely in the context of the original measure), the identifier, and the value set purpose (a required element) that should describe the scope or breadth of concepts that should be included or excluded, in text for other humans to read.
The required metadata for each value set includes:

- **Name**—developed based on guidelines presented on the [VSAC Authoring Best Practices](#) website that includes naming conventions (i.e., recommendations about how to create a name for a value set).
- **Clinical focus**—a free text statement describing the general focus of the value set as it relates to the intended semantic space, which can be the information about clinical relevance or a statement about the general focus of the value set (e.g., description of types of messages, payment options, geographic locations).
- **Data element scope**—a free text statement describing how the data element in the intended information model (e.g., QDM) defines the concepts to be selected for inclusion in the value set.
- **Inclusion criteria**—a free text statement that defines the concepts or codes to be included and why.
- **Exclusion criteria**—a free text statement that defines the concepts or codes to be excluded and why.

### 17.2.4.2 Including “Historical” Codes

Some value sets will need to include concepts that are no longer active concepts in the code system of choice, which will often occur when the value set is included in a measure clause that requires a “look-back period” that extends back more than a year or the length of time between code system updates, due to the fact that newly retired codes were entered into patient records when they were still active codes. No measure developer should assume that old patient records will update content to use current codes. Therefore, value sets that are to be used to identify patients based on old record content will need to include inactive legacy codes in the value set expansions and document the need for including such content in the purpose statements of the value set metadata section. Measure developers need to notify NLM about any retired codes they need to use in a value set. NLM will then include the measure developers’ specified retired codes as “legacy codes” within the expansion profile calculation to be applied to the specified eCQM Program Release.

### 17.2.5 Value Set Mapping

Value sets use standard terminologies and use of standard terminologies in quality measures is essential to support interoperability and the ability to compare measure results between clinicians or healthcare organizations. Most standard terminologies have regularly scheduled updates in which new concepts may be added and other concepts deprecated. For deprecated concepts, measure developers must determine whether to keep legacy codes or identify new concepts to replace those codes. Value sets often have more than one standard terminology that can be used to express a data element, most commonly SNOMED CT and ICD-10-CM/PCS. There may be a need to map concepts in one terminology with the same concept in another terminology. Cross-terminology mapping can be complex and time-consuming. Matching concepts also occurs on the implementation side. Health IT vendors may use and/or end users may have different terminologies in use; drug databases are a good example.

Historically, cross-terminology mapping was a completely manual process. Automated mapping tools (e.g., Apelon Distributed Terminology System (DTS) and Usagi) have been developed and are constantly being refined. New techniques and tools are also being developed (e.g., Solor). However, no tool can connect 100% of the concepts, so some level of manual processing is necessary.
17.3 **VALUE SET REUSE**

To the extent possible, use existing value sets when developing eCQMs. The measure developer should examine the existing library of value sets to determine if any exist that define the clinical concepts described in the measure. If so, these value sets should be used, rather than creating a new value set, which promotes harmonization and decreases the time needed to research the various terminologies to build a new list.

A measure may reuse existing value sets or define new value sets. When selecting a value set for reuse, measure developers should be confident that the set of codes included is fit for their purpose. When considering reuse, a measure developer may contact the value set author of a reusable value set to determine planned maintenance and clarify common needs.

When multiple value sets appear to represent the same concept, harmonization should be attempted. Harmonization may require rethinking what was intended and an assessment of the expertise needed to define and maintain the value set content. Discussion regarding the content and work toward a harmonized single usable value set should occur within the VSAC Collaboration Tool.

DRCs may also be reused. VSAC provides access to DRCs in current use to enable such reuse and collaboration. Value set authors can obtain a Representational State Transfer (REST) application programming interface (API) Uniform Resource Locator (URL) for DRCs, for insertion into the MAT, in the Detail View of Browse Code Systems accessible from any VSAC web page.

17.4 **VALUE SET AUTHORITY CENTER**

The VSAC is provided by NLM, in collaboration with the ONC and CMS. Requiring a free UMLS license, the VSAC provides searchable and downloadable access to all official versions of value sets used by each of the eCQMs releases used in CMS, as well as other programs (e.g., HL7 C-CDA for clinical document interchange) and other non-eCQM programs. Each value set consists of the codes (i.e., concept identifiers from specified code systems) and human-readable names (i.e., descriptions or terms), drawn from standard codes systems such as SNOMED CT, RxNorm, LOINC, and ICD-10-CM, which are used to identify specific patient populations used in CQMs (e.g., patients with diabetes, clinical visit). The VSAC Support Center provides online information about VSAC access, value set lifecycles and work flow, author and steward roles, and best practices for value set development. In addition, the VSAC Support Center offers archived users forums and release notes, and provides links to VSAC publications.

VSAC is the only authoritative tool to author value sets for eCQMs. VSAC now includes intensional definition functionality for value set authors to intensionally define value sets using logical criteria, or statements, which will greatly increase value set accuracy and maintenance efficiency, and also greatly reduce burden on value set authors. For example, value set authors can specify all descendant codes of a hierarchical concept and expand them into an enumerated list of codes and terms. When defining value set code members using logical criteria like this, also known as an intensional definition, it is important to include this logic statement in the Purpose Statement section of the Value Set Metadata in VSAC. Tools other than VSAC exist, such as the CDC Public Health Information Network Vocabulary Access and Distribution System (PHIN VADS) and some proprietary offerings, to help build and maintain quality measure value sets.

The key benefits of using VSAC include:

- Serves as the authority and central repository for the official versions of value sets that support eCQMs adopted by CMS programs.
• Provides search, retrieval, and download capabilities through a web interface and APIs.
• Provides authoring and validation tools for creating new and revising published value sets.
• Hosts up-to-date versions of source vocabularies. The representative source vocabularies (not exhaustive) include SNOMED CT, RxNorm, LOINC, ICD-9-CM, and ICD-10-CM.
• Includes a required purpose statement for each value set, composed of clinical focus, data element scope, inclusion criteria, and exclusion criteria. These purpose statements ensure clear clinical intent of value set and building criteria, which should be used for evaluating the validity and accuracy of codes contained in the value set.
• Offers complete value set authoring guidance.

For more details, refer to the VSAC website.

17.5 VSAC Collaboration Tool
The VSAC Collaboration Tool is a companion tool to the VSAC value set authoring and maintenance environment. CMS expects that VSAC authors will use the VSAC Collaboration Tool for most value sets to obtain input into the value set content from participating users and clinical expert colleagues. This tool provides a central site where value set authors and stewards can post value sets either as single, multiple, or per CMS measure from the VSAC authoring environment into the VSAC Collaboration Tool for collaborative discussion and value set quality assurance analysis in addition to workflow and document management by stewards and their invited external collaborators. VSAC authors and stewards can use their posted value set spreadsheets in the VSAC Collaboration Tool to review the terminology changes in value sets as they are expanded using newer expansion profile calculations that contain newer code system versions. Although this functionality exists in VSAC for every value set, this feature in the VSAC Collaboration Tool enables quality review for all value sets within a measure. This feature supports the maintenance activity required of authors to keep value set content aligned with code system version changes and clinical knowledge enhancements. The VSAC Collaboration Tool requires a UMLS License Agreement for each user and NLM provides an extensive Help section.

It is important to note that all value sets can benefit from utilization of the VSAC Collaboration Tool, which is not focused solely on eCQM development.

17.6 Unique Situations for Value Set Use in eCQMs
17.6.1 Use of Value Sets to Address Negation Rationale in eCQMs
The concept of negation has several aspects: assurance that a certain condition does not exist (e.g., assertion that the patient has no known allergies, or that the patient takes no medications); lack of evidence that a condition does not exist (e.g., no information is available in the EHR about allergies or about medications taken); and that an action was intentionally not performed, with or without a reason (e.g., medication was not prescribed due to interaction with other medications a patient is taking). Generally, absence of information about a desired action in the medical record is assumed to indicate the action did not occur. Thus, intentional decisions to avoid actions must be asserted, sometimes called action negation. QDM calls this intentional assertion negation rationale; it addresses only the third definition.

QDM addresses negation rationale as an attribute of 73 QDM datatypes. By specifying negation rationale in a measure expression, the measure developer provides logic criteria requiring assertion that an action was not performed, most often requiring at least one of a set of specific reasons. Generally, measure developers use negation rationale as measure exceptions (i.e., if there was a documented
medical reason not to perform an action, the patient may be removed from the denominator. Negation rationale could potentially be used as denominator criteria as well (i.e., what numerator criteria should be expected if there is a medical reason to avoid the most common treatment).

The QDM documentation on the eCQI Resource Center describes how a measure developer can use negation rationale. Negation rationale is part of the current production version of QDM. A discussion about negation and methods for expressing negation with CQL can be found on the CQL Formatting and Usage Wiki.

Negation rationale is presented in measure logic as not done. Using negation requires two value sets or DRCs, one is the value set of concepts that would be expected to have been done (usually, this is literally the same value set or DRC used in the measure to identify patients with the applied therapy/action) and the other is a value set (or DRC) of the acceptable reason for avoiding the expected action.

Example

Medication, Administered Not Done: Angiotensin Converting Enzyme (ACE) Inhibitor for Medical Reason for Avoiding ACE Inhibitors

ACE Inhibitor is the value set of the expected medications for which documentation asserts was not administered.

Medical Reason for Avoiding ACE Inhibitors is the value set of acceptable reasons for avoiding administration of the expected medication. Only reasons included in the value set will meet the criteria for the measure.

17.6.2 Value Set Review

Value sets can be shared with reviewers in several ways. The most straightforward way is to email the list of value set codes in a Microsoft Excel spreadsheet. The VSAC Collaboration Tool can also be used to review and gather feedback. It should be noted that value set content often includes codes from code systems that are covered by intellectual property rights and reviewers should be notified of this fact.

There are rules for value set development in the VSAC. All value sets are required to have specific metadata that describe the purpose and content they contain (refer to Section 3, Chapter 17.2.4.1, Value Set Metadata). Assuring value sets have such descriptive metadata will significantly improve the likelihood that reviewers and later users of the value sets will understand the intent and provide valuable feedback about the content.

Reviewers evaluating value set content (i.e., the code descriptions) can best provide valuable feedback regarding the validity of the value set if the metadata provided are detailed. Evaluation based on the value set name alone is insufficient. The advantage of providing a Microsoft Excel spreadsheet for value set review is that it is simple, easy to browse for codes, and easy for the reviewer to document comments and feedback. Value sets exported from the VSAC include the metadata described above. The disadvantage of providing a Microsoft Excel spreadsheet to review value sets is that a flat list of codes does not reflect the hierarchical structure of codes residing in their code system. Reviewers need to look up codes in their original code systems in order to understand the parent-child relationship and determine semantic relevance to the intent. VSAC provides a powerful tool called Browse Code Systems, located at the top of every VSAC web page. The Browse Code Systems tool enables any VSAC user (i.e., author or non-author) with a UMLS License to look up codes for reviewing a code’s terms, synonyms, children, parents, and siblings, as well as view a graphical interface tree view for browsing hierarchical code systems. VSAC now has intensional definition functionality. When value set authors
define their value set intensionally, they may download a Microsoft Excel spreadsheet from the VSAC Authoring Tool that includes the logical definition (e.g., all descendants of a code) as well as the expanded code list of that value set, which will be extremely helpful for reviewers to understand the value set.

To overcome the disadvantages of a Microsoft Excel spreadsheet, links to the value sets directly within the NLM’s VSAC system are recommended not only for authoring value sets, but also for reviewing value sets among developers, reviewers, and stewards. The features of VSAC are described in the VSAC Support Center’s Help Section, which has no login requirements and is available to anyone, although access to the VSAC Repository, Authoring Tool, Collaboration Tool, and Browse Code Systems requires that all users obtain a UMLS Metathesaurus License to ensure that each user acknowledges and abides by code system licensing requirements. There is no charge for registration, and it is available for any user independent of nationality. With a UMLS license, value set access and creation are available to all VSAC users. Implementers must ensure that they comply with any specific code system implementation/use requirements. Some reviewers may find the registration process cumbersome for reviewing a small value set, but it is valuable.

17.6.3 Pre-MAT Value Set Review

The purpose of a value set review is to validate correct code selections meeting the clinical intent as well as the correct hierarchy in the code system. A value set review can be conducted by the quality assurance team, internal or external terminologist, and steward as the NLM is not involved in this review. The entity responsible for the final selection of codes based on the existing data and feedback from clinicians is considered the value set steward. The quality assurance team and terminologist will focus on the areas listed in Table 18 and take the appropriate remedial action. It is acceptable to use a Microsoft Excel spreadsheet to capture and distribute the value set and codes as long as measure developers understand the limitations previously noted. It is recommended, however, to use the NLM VSAC to author and review value sets.

Table 18. Value Set Review Areas and Remedial Actions

<table>
<thead>
<tr>
<th>Area</th>
<th>Remedial Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value set duplication</td>
<td>Duplicate value sets should be replaced by normalized value sets.</td>
</tr>
<tr>
<td>Clinical validity</td>
<td>Value sets must correspond to the intent and purpose of the clinical perspective.</td>
</tr>
<tr>
<td>Code list completeness</td>
<td>A value set should contain all the relevant codes for a specific data element.</td>
</tr>
<tr>
<td>Metadata completeness</td>
<td>Apply a common desirable pattern, implementing NLM VSAC guidance for value set</td>
</tr>
<tr>
<td></td>
<td>metadata. Always include the required Purpose Statements.</td>
</tr>
<tr>
<td>Alignment of code system to the standards</td>
<td>Value sets must use recommended terminology systems for an extensional value set.</td>
</tr>
<tr>
<td></td>
<td>Update code sets from transition vocabularies to those ideally desired.</td>
</tr>
<tr>
<td>Terminological correctness</td>
<td>Attention to selection of appropriate semantic type(s) is important based on the data element the value set is to be used by, for example, SNOMED CT disease versus morphologic abnormality. Code system combinations should use a grouping value set approach.</td>
</tr>
<tr>
<td>Single and multiple concepts</td>
<td>Value sets may be reused and combined with other value sets to create a grouping value set for representing a more general set of ideas. Nested grouping value sets are not allowed.</td>
</tr>
<tr>
<td>Impact to measure logic size</td>
<td>Repeating identical logic that is a consequence of multiple similar value sets may be replaced by using grouping value sets, which combine appropriate value sets.</td>
</tr>
</tbody>
</table>
17.6.4 Post-MAT Value Set Review – eCQM Annual Update Only\textsuperscript{41}

The VSAC Collaboration Tool provides a quality assurance report for value sets associated with each measure. This report, located in the Value Set Site Content dashlet of a measure’s collaboration site, is based on technical analysis of value sets provided before and after MAT development is completed. Value set review at this stage is not focused on the clinical intent representation. Rather, it is focused on verification that all codes have been successfully captured in VSAC for use by the MAT. New changes to value sets due to harmonization could be introduced during the post-MAT value set review; therefore, these value sets need ad hoc reviews to ensure the proper changes are in place.

17.6.5 NLM Value Set Review – eCQM Annual Update Only\textsuperscript{42}

During the pre-eCQM release quality assurance process and to improve value set authorship, curation, and delivery, NLM performs quality assurance checks to compare the validity of value set codes and terms with the latest source vocabularies. As value set authors and measure developers create their value sets within the VSAC Authoring Tool, the tool interactively assesses the code validity within a code system, as well as other quality assurance parameters. Measure developers should take proper actions as specified by NLM based on the analysis outcome. If the VSAC or NLM quality assurance teams identify value set deficiencies, measure developers should correct the value sets using the VSAC Authoring Tool.

\textsuperscript{41} This activity only occurs as part of the program release process for the eCQM Annual Update.

\textsuperscript{42} This activity only occurs as part of the program release process for the eCQM Annual Update.
**Measure Harmonization**

Differences in measure *specifications* limit comparability across settings. Multiple measures with essentially the same focus create burden and confusion in choosing measures to implement and when interpreting and comparing the measure results. This chapter addresses the concepts of harmonization and defines key terms related to the process of harmonizing measures. CMS measure developers are expected to consider harmonization as one of the core measure evaluation criteria that are applied throughout the Measure Lifecycle. NQF also requires consideration of measure harmonization as part of its endorsement processes.

Measure harmonization is defined as standardizing specifications for related measures when they:

- Have the same measure focus (i.e., *numerator* criteria)
- Have the same *target population* (i.e., *denominator* criteria)
- Apply to many measures (e.g., age designation for children).

Harmonized measure specifications are standardized so that they are uniform or compatible, unless differences are justified because the differences are dictated by the evidence.

The dimensions of harmonization can include numerator, denominator, exclusion, calculation, and data source and collection instructions. The extent of harmonization, per *Changes to NQF’s Harmonization and Competing Measures Process: Information for Measure Developers*, depends on the relationship of the measures, evidence for the specific measure focus, and differences in data sources.

Measure alignment is defined in *Changes to NQF’s Harmonization and Competing Measures Process: Information for Measure Developers* as “Encouraging the use of similar standardized performance measures across and within public and private sector efforts” (p. 6). Harmonization is related to measure alignment because measures of similar concept that are harmonized can then be used in multiple CMS programs and care settings. CMS seeks to align measures across programs, with other federal programs, and with private sector initiatives as much as is reasonable.

When quality initiatives are aligned across CMS programs and with other federal partners, information for patients and consumers is clarified. A core set of measures increases “signal” for public and private recognition and payment programs (*Conway et al., 2013*). When harmonized measures are selected by CMS across programs, it becomes possible to compare the care that is provided in different settings. For example, if the influenza immunization rate measure is calculated the same way in hospitals, nursing homes, and other settings, it is possible to compare the achievement for population health across the multiple settings. If functional status measurement is harmonized and the measure use aligned across CMS programs, it would be possible to compare gains across the continuum of care. Consumers and payers are enabled to choose based on measures calculated in similar ways. In these and other ways, harmonization promotes:

- Coordination across settings in the continuum of care
- Comparisons of population health outcomes
- Clearer choices for consumers and payers.

Measure developers should consider both harmonization and alignment throughout the Measure Lifecycle and whether to respecify an existing measure, adopt an existing measure, or develop a new measure.
QCDRs are also encouraged to share and/or harmonize similar measures unless there is a compelling reason not to do so. Harmonization between QCDRs provides clinicians with a larger cohort for comparison for performance scoring and benchmarking.

Harmonization should be considered when:

- Developing measure concepts by:
  - Conducting a thorough environmental scan to determine whether there are appropriate existing measures on the topic.
  - Consulting with a TEP and obtaining public input on the topic and the measures.
- Developing measure specifications by examining technical specifications for opportunities to harmonize.
- Conducting measure testing by assessing whether the harmonized specifications will work in the new setting or with the expanded population or data source.
- Implementing measures by proposing the harmonized measure for use in new programs.
- Conducting ongoing measure monitoring and evaluation by continuing environmental surveillance for other similar measures.

Table 19 summarizes ways to identify whether measures are related, competing, or new, and indicates the appropriate action based on the type of harmonization issue.

Table 19. Harmonization Decisions during Measure Development

<table>
<thead>
<tr>
<th>Measure</th>
<th>Harmonization Issue</th>
<th>Action</th>
</tr>
</thead>
</table>
| Numerator: Same measure focus Denominator: Same target population | Competing measures | • Use existing measure (adopted) or justify development of additional measure  
• A different data source will require new specifications that are harmonized (e.g., respecified) |
| Numerator: Same measure focus Denominator: Different target population | Related measures | • Harmonize on measure focus (i.e., respecified)  
• Justify differences  
• Respecify existing measure by expanding the target population |
| Numerator: Different measure focus Denominator: Same target population | Related measures | • Harmonize on target population  
• Justify differences |
| Numerator: Different measure focus Denominator: Different target population | New measures | • Develop measure |

18.1 **Respecified Measures**

If the measure developer changes an existing measure to fit the current purpose or use or changes the data source to an EHR, the measure is considered respecified. This process includes revising a measure to meet the needs of a different care setting, data source, or population. Alternatively, it may require changing the numerator or denominator, or adding new specifications to fit the new use. An example of this type of respecification would be altering the pressure ulcer quality measure used in nursing homes for use in other post-acute settings such as long-term acute care hospitals or IRFs.

In respecifying a measure to a different setting, the measure developer needs to consider accountability, attribution, and data source of the new setting. Measures that are being respecified for use in a different setting or a different unit of analysis may not need to undergo the same level of comprehensive testing or evaluation compared to a newly developed measure. However, when
respecifying a measure for use in a new setting, a new population, or with a new data source, the newly respecified measure must be evaluated and tested. Before deciding to respecify a measure already in existence, consider the questions:

- If the existing measure is NQF-endorsed, are the changes to the measure significant enough to require resubmission or an ad hoc review for continued NQF endorsement?
- Will the measure steward be agreeable to the changes in the measure specifications that will meet the needs of the current project?
- If a measure is copyright protected, are there issues relating to the measure’s copyright that need to be considered?

Discuss these considerations with the COR and the measure steward. NQF endorsement status may need to be discussed with NQF. After making any changes to the numerator and denominator statement to fit the specific use, new detailed specifications will be required.

The first step in evaluating whether to respecify a measure is to assess the applicability of the measure focus to the population or setting of interest or data source:

- Is the focus of the existing measure applicable to the quality goal of the new measure population, setting, or data source?
- Does it meet the importance criterion for the new population or setting?

For example, if the population changes or if the type of data is different, new measure specifications would have to be developed and properly evaluated for reliability, validity, and feasibility before a determination regarding use in a different setting can be made. Empirical analysis may be required to evaluate the appropriateness of the measure for a new purpose. The analysis may include, but is not limited to, evaluation of changes in the:

- Relative frequency of critical conditions used in the original measure specifications when applied to a new setting or population (e.g., when there is a dramatic increase in the occurrence of exclusionary conditions).
- Importance of the original measure in a new setting. An original measure addressing a highly prevalent condition may not show the same prevalence in a new setting; or, evidence that large disparities or suboptimal care found based on the original measure may not exist in the new setting or population.
- Applicability of the original measure (e.g., when the original measure composite contains preventive care components that are not appropriate in a new setting such as hospice care).
- Feasibility of the data when changing the data source to an EHR.
- Validity of the measure specifications (e.g., certain codes in the claims from commercial health plans may not be valid or payable under Medicare). Therefore, the measure as specified is no longer capturing the intended numerator or denominator when the measure is applied to a different setting.

18.2 ADOPTED MEASURES

Adopted measures must have the same numerator, denominator, and data source as the parent measure. In this case, the only information that would need to be provided is specific to the measure’s implementation use (e.g., data submission instructions, as they may be different from the original). In most cases, if the parent measure is NQF-endorsed and no changes are made to the specifications, the adopted measure is considered endorsed by NQF. An example of an adopted measure would be an
ambulatory program adopting the core hypertension measure, NQF 0018, Controlling High Blood Pressure.

When considering the adoption of an existing measure for use in a CMS program, investigate whether the measure is currently used in another CMS program.

### 18.3 New Measures

Decide whether to develop a new measure by first conducting an environmental scan for similar or related measures already in existence or in the CMS Measures Inventory Pipeline (in development or planned for development). If there are no existing or related measures that can be respecified or adopted, then it may be appropriate to develop a new measure. The material in Section 3, Chapter 9, Information Gathering, provides details on this process. Consult with the appropriate COR if the environmental scan reveals a similar measure is being developed by another measure developer. The Measures Manager can also help identify potential harmonization opportunities and help prevent duplication of measure development efforts.

If the information gathering process and input from the TEP determine that no existing or related measures apply to the contract objectives, then consider a new measure.

### 18.4 Harmonization during Measure Maintenance

CMS promotes the use of the same measure or harmonized measures across its programs as much as possible. Harmonization and alignment work are parts of both measure development and measure maintenance. This discussion is about procedures for harmonization and alignment after the measure is in use and is being maintained. These four steps (described in Sections 18.4.1 through 18.4.4) applied during measure maintenance will help ensure that measures continue to be harmonized after they are implemented.

#### 18.4.1 Decide Whether Harmonization is Indicated

Conduct an environmental scan for similar measures already in existence and measures in development that are similar or related. The COR and Measures Manager can help measure developers identify similar measures in development. Although this step may have been done during initial measure development, the related measures may no longer be harmonized because specifications were changed.

Table 20 describes harmonization issues and actions based on the numerator and denominator specifications.
### Table 20. Harmonization Decisions during Measure Maintenance

<table>
<thead>
<tr>
<th>Measure</th>
<th>Harmonization Issue</th>
<th>Action</th>
</tr>
</thead>
</table>
| Numerator: Same measure focus   | Competing measures   | • Use existing measure (i.e., adopted) or justify development of additional measure  
| Denominator: Same target population |                      | • A different data source will require new specifications that are harmonized (e.g., respecified) |
| Numerator: Same measure focus   | Related measures     | • Harmonize on measure focus (i.e., respecified)                       |
| Denominator: Different target population |                      | • Justify differences                                                  |
|                                  |                      | • Respecify existing measure by expanding the target population         |
| Numerator: Different measure focus | Related measures   | • Harmonize on target population                                       |
| Denominator: Same target population |                      | • Justify differences                                                  |
| Numerator: Different measure focus | No harmonization issue | • Develop measure – harmonization not appropriate                        |
| Denominator: Different target population |                      |                                                                         |

#### 18.4.2 Implement Harmonization Decisions

After evaluating for harmonization, the possible outcomes are:

- Retain the measure with minor updates and provide justification if there are related measures
- Revise the measure specifications to harmonize
- Retire the measure and replace it with a different measure.

#### 18.4.3 Test Scientific Acceptability of Measure Properties

If harmonization results in changes to the measure specifications, testing of the scientific acceptability, including re-analysis of reliability, validity, and exclusion appropriateness, is usually necessary.

#### 18.4.4 NQF Evaluates for Harmonization during Measure Maintenance

NQF will evaluate the measure for harmonization potential during the maintenance review of the measure. There may be instances when the measure developer may be unaware of newly developed similar or related measures until they have been submitted to NQF for review. If similar or related measures are identified by NQF and harmonization has not taken place, or reasons for not doing so are adequately justified, the NQF Steering Committee reviewing the measures can request that measure developers create a harmonization plan addressing the possibility and challenges of harmonizing certain aspects of their respective measures. NQF will consider the response and decide whether to recommend the measure for continued endorsement.
19  **RISK ADJUSTMENT**

Risk adjustment refers to the inclusion of risk factors associated with a measure score in a statistical model of provider performance. These risk factors may be captured at the person, facility, community, or other levels. Increasingly, risk adjustment is a consideration when developing measures. This section offers insight into currently used risk adjustment models.

Risk adjustment at the person-level, also referred to as case-mix adjustment, aims to answer the question: “How would the performance of various units compare if hypothetically they had the same mix of patients?” (NQF, 2014, p. 28) Thus, the purpose of risk adjustment is to increase the likelihood of fair comparison of provider performance, which is to compare apples with apples. It involves controlling for confounding factors, meaning systematic differences within the patient population of interest, be it clinical (e.g., types, number, or severity of conditions), demographic (e.g., age, gender), or socioeconomic (e.g., race, ethnicity, income), in the modeling of provider performance.

Taking confounding factors into account could prevent the model from being mis-specified or the estimates for performance scores from being biased. It would also increase the chance that healthcare providers and consumers can receive a more accurate picture of provider performance. Considering confounding factors becomes even more important as performance scores are often used as a basis for calculating the amount of incentives or penalties for value-based purchasing and many APMs.

Based on the input of the multi-stakeholder expert panel, NQF recommended in *Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors* that outcome or process of care measures be risk-adjusted for person-level socioeconomic status if there is a conceptual or methodological basis for doing so.

For example, there might be qualitative or empirical evidence in the literature documenting the impact of social risk factors on care delivery, health outcomes, or costs. This would provide a conceptual basis for social risk adjustment. Similarly, during the exploratory analysis for measure development, preliminary data analysis might reveal a statistically significant confounding association between a social risk factor and a performance measure of interest. Under this circumstance, it would be appropriate to include the social risk factor in the multi-variate statistical model.

Developers should consider whether their measure is appropriate to adjust for social risk factors because not all measures should be adjusted for social risk. Developers should examine each measure individually to determine the appropriateness for social risk adjustment, taking a measure’s empirical relationship with individual social risk factors into consideration. Also, it is crucial to conduct research to examine whether any clinical and social risk factors not yet included in the model should be considered for adjustment to improve the ability of a statistical model to tell the differences in performance between providers (Joynt et al., 2017).

Multiple studies have indicated that providers who serve a higher proportion of poorer and sicker population groups, like safety-net hospitals, may be more likely to perform poorly in terms of health outcomes, even after adjustment for person-level clinical risk factors, and have been disproportionally penalized by various Medicare value-based purchasing programs and APMs (Thirukumaran et al., 2019; Roberts et al., 2018; Johnston et al., 2019). It has been shown, in particular, that health outcomes like readmissions (Joynt Maddox et al., 2019a) and mortality (Charles et al., 2018) have been found to be closely associated with social risk factors such as household income, education, white-collar occupation (Eapen et al., 2015), dual-eligible status, housing instability, and neighborhood characteristics. Effects of these factors remained statistically significant—even after clinical risk factors were taken into account.
Risk-adjusting for social factors was seen to reduce the difference between safety-net providers and non-safety-net providers in the amount of penalties. While there were cases in which social risk adjustment did not significantly impact performance scores (Martsolf et al., 2016), it remains important to note that, in some cases, the adequacy of social risk adjustment could reduce the likelihood of unintended consequences, especially if high penalties leave some of these providers with fewer resources for quality improvement activities.

In response to the concern about social risk adjustment masking disparities, the NQF 2014 report stated that disparities between provider units in a performance score is a function of differences within and between units. Within-unit differences refer to disparities between patients of different socioeconomic status treated under the same unit, be it the same hospital, clinician practice, etc. Between-unit differences refer to disparities between patients treated by different hospitals or clinician practices, even though the patients at Hospital A have similar clinical and social characteristics as those at Hospital B. Risk adjusting for social factors at the person-level is appropriate because the impact of these person-level risk factors on person-level outcome or cost is well documented. However, social risk adjustment at the person-level is not linked to performance scores at the provider level. Because of that, risk adjusting for social factors at the patient-level would not mask health disparities in the comparison of provider performance.

To further examine whether the absence of social risk adjustment could aggravate health disparities, NQF implemented a 2-year Social Risk Trial between April 2015 and April 2017 (or the first NQF Social Risk Trial). New measures submitted for endorsement or previously endorsed measures submitted for re-endorsement (or maintenance) were included in the trial period (NQF, 2017a). Measures for which there was a conceptual or methodological basis for social risk adjustment had to go through the trial. Developers of these measures were required to submit testing results on how social risk factor were selected, data on measure performance with and without social risk adjustments, and specifications for stratification by social risk factors that showed evidence of disparities. The social risk factors examined during this trial included race, ethnicity, Medicaid status, the AHRQ Socioeconomic Status (SES) Index, insurance type, education level, percent of household under the Federal Poverty Level (FPL), etc. (NQF, 2017a). By the end of the trial, the NQF CSAC recommended:

- Availability of social risk factors should be considered as part of the annual update process.
- NQF should focus on adjustment based on social risk as well as unmeasured clinical complexity.
- Unresolved issues and concerns regarding risk adjustment approaches, including potential for adjustment at the hospital and community levels, should continue to be addressed.

Key recommendations from the first NQF Social Risk Trial are consistent with the guidance by the Risk Adjustment Expert Panel in the 2014 report in two important aspects. Firstly, socioeconomic risk factors should be considered for risk adjustment when there is a conceptual or empirical association between these factors and provider performance, be it health outcomes or processes of care. Secondly, the guidelines apply for selecting both clinical and social risk factors, including conceptual association with the outcome of interest, variation in prevalence of the factor across the measured entities, present at start of care, resistant to gaming and manipulation, potential improvement of the risk model, and face validity and acceptability. Moreover, each measure should be evaluated individually for the appropriateness of social risk adjustment.

The first NQF Social Risk Trial also presented several methodological challenges with social risk adjustment. For example, developers often found that data sources for potential social risk factors were limited and lacking in granularity. Also, while the 2014 NQF Risk Adjustment Expert Panel recommended the use of race as a social risk factor, the appropriateness of race/ethnicity as social risk factors has
remained inconclusive. Moreover, social risk factors that were included in risk adjustment models based on conceptual or empirical association with provider performance sometimes had minimal effect on the performance of the model. Developers also varied in how they determined whether a statistically significant social risk factor was included in the final model. Stakeholders called for guidance to enhance consistency in the conceptual and empirical analysis across developers and standardization of variables tested. To allow time for in-depth investigation into these issues and development of guidance, beginning in May 2018, NQF implemented the second phase of the Social Risk Trial with funding from CMS. A final report on the findings of the trial will be publicly available by the end of this 3-year trial (i.e., in May 2021).

To further examine the issue of social risk adjustment and its impact on Medicare value-based purchasing programs and health disparities, the Office of the Assistant Secretary for Program and Evaluation (ASPE) examined the issue. In ASPE’s 2016 Report to Congress, social risk factors examined included dual enrollment in Medicare and Medicaid as a marker for low income, residence in low-income area, race (i.e., African-American), ethnicity (i.e., Hispanic), and residence in a rural area. Consistent with existing literature, the report showed that Medicare beneficiaries with social risk factors had worse outcomes on many quality measures in general, and dual enrollment status was the most powerful predictor of poor outcomes. More importantly perhaps, for all five Medicare value-based purchasing programs, providers that disproportionally served beneficiaries with social risk factors tended to have worse performance on quality measures—even after accounting for case mix. The report provided several recommendations to address these issues. ASPE’s recommendations related to quality measurement and reporting for beneficiaries with social risk factors are:

- Enhance data collection and develop statistical techniques to allow measurement and reporting of performance for beneficiaries with social risk factors on key quality and resource use measures.
- Develop and introduce health equity measures or domains into existing payment programs to measure disparities and incentivize a focus on reducing them.
- Examine measures to determine whether adjustment for social risk factors is appropriate; this determination will depend on the measure and its empirical relationship to social risk factors.
- Continue studying program measures to determine whether differences in health status might underlie the observed relationships between social risk and performance, and whether better adjustment for health status might improve the ability to differentiate true differences in performance between providers.

Section 15002 of the Cures Act requires the Secretary of HHS to take into account the findings of the 2016 ASPE report and use dual-eligible status in the HRRP for risk adjustment. Specifically, the Cures Act requires HHS to implement a stratification scheme based on the proportion of full-benefit, dual-eligible beneficiaries served by each hospital. Performance scores shall be reported based on this stratification scheme. Risk adjustment with stratification is discussed further in the next section.

The risk adjustment requirement of the Cures Act has resulted in a reduction in penalties among safety-net hospitals with higher proportion of dual-eligible patients, and an increase in penalties among those with lower proportion of these patients (Joynt Maddox et al., 2019b). At the same time, there has been a call for exploring potential inclusion of covariates such as prior service utilization, functional status,

---

43 This included the HRRP, the Hospital Value-Based Purchasing Program, the HACRP, the Physician Value-Based Payment Modifier Program (which sunset in 2018), and the SNF VBP.
and frailty for improving risk adjustment models, especially for vulnerable population groups to minimize gaming (Wadhera et al., 2019).

At this time, however, there is not a single best practice for risk stratification for SES factors, and this should be evaluated on a case-by-case basis. CMS continues to examine SES risk adjustment in value-based incentives programs.

19.1 Risk Adjustment Strategies – When to Use Risk Adjustment with Stratification

Risk stratification refers to reporting outcomes separately for different groups, unadjusted by the risk factor associated with the grouping. For example, if we use a variable that represents the social risk factor of FPL, and this variable has four levels (e.g., under 100% of the FPL, 100% to under 200% of the FPL, 200% to under 300% of the FPL, and 300% of FPL or above), then risk stratification is conducted by running the statistical model of provider performance without the FPL variable on each of the four levels of FPL.

There are two circumstances when risk adjustment in combination with stratification might be the most appropriate risk adjustment strategy (i.e., where the goal of the strategy is fair comparisons). The first circumstance is when patient factors are not independent of the quality construct. The second circumstance is when there is treatment heterogeneity, which is another case where the patient factors and quality construct are not independent, but for legitimate clinical reasons. As part of a risk adjustment strategy, NQF recommends use of risk models in conjunction with risk stratification when use of a risk model alone would result in obscuring important healthcare disparities. Risk adjustment is appropriate when the patient factors are correlated with the outcome and not correlated with (i.e., independent of) the quality construct. Table 21 provides a high-level framework for risk adjustment strategies.

Table 21. Framework for Risk Adjustment Strategies

<table>
<thead>
<tr>
<th>Relationship of patient factors and quality construct</th>
<th>Measurement of Patient Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observable</td>
<td>Non-observable</td>
</tr>
<tr>
<td>Patient group stratification</td>
<td>Peer group stratification</td>
</tr>
<tr>
<td>Risk adjustment</td>
<td>Reliability adjustment</td>
</tr>
</tbody>
</table>

Developers should also consider what method would be most appropriate for accounting for social risk factors (e.g., risk adjustment, stratification by groups within a measure, stratification at the measure level). Stratification at the measure level may be similar to peer group stratification, in which patient factors are unobserved and correlated with observed provider characteristics. The exploration of a risk adjustment strategy (i.e., use of a statistical risk adjustment model, and, if necessary, risk stratification for selected populations) is required for measures developed using the Blueprint for outcome measures and other measures, when indicated in the contract. For an outcome measure to be accepted by CMS and endorsed by NQF, the measure developer must demonstrate appropriate use of a risk adjustment strategy and risk stratification. Rationale and strong evidence must be provided if a risk adjustment model or risk stratification is not used for an outcome measure.

Consequently, it is the measure developer’s responsibility to determine whether variation in factors intrinsic to the patient should be accounted for before outcomes can be compared and to determine how to best apply these factors in the measure specifications. Vogel and Chen (2018) noted “failure to address risk adjustment in an adequate manner can lead to biased conclusions that may adversely impact decision-making in both research and policy contexts” (p.1). The purpose of this chapter is to
provide guidance to CMS measure developers regarding the nature and use of a risk adjustment model in quality measurement.

Risk adjustment methodology currently cannot be modeled in the HQMF. Risk adjustment methodology may be described in the metadata. The measure data may be used post hoc to risk-adjust. Variables for risk adjustment should be represented as supplemental data elements because they are needed for every patient. The logic or algorithm for risk adjustment should be included in the risk adjustment section of the HQMF. In the future, risk adjustment methodology may be modeled using CQL.

19.2 FEATURES OF RISK ADJUSTMENT MODELS

The measure developer must evaluate the need for a risk adjustment strategy (i.e., risk adjustment, stratification, or both) for all potential outcome measures and statistically assess the adequacy of any strategies used. In general, a risk adjustment model possesses certain features. Some of these features are listed in Table 22, which was partially derived from a description of preferred features of models used for publicly reported outcomes (Krumholz et al., 2006). Each attribute listed in Table 22 is described in detail in Section 3, Chapter 19.2, Attributes of Risk Adjustment Models, subsections 19.2.1 through 19.2.6.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample definition</td>
<td>Sample(s) should be clearly defined, clinically appropriate for the measure’s risk adjustment, and large enough for sufficient statistical power and precision.</td>
</tr>
<tr>
<td>Appropriate time intervals</td>
<td>Time intervals for model variables should be clearly defined, sufficiently long to observe an outcome, and recent enough to retain clinical credibility.</td>
</tr>
<tr>
<td>High data quality</td>
<td>Data should be reliable, valid, complete, comprehensive, and rely on as few proxy measures as can be accomplished with due diligence.</td>
</tr>
<tr>
<td>Appropriate variable selection</td>
<td>Selected adjustment or stratification variables should be clinically meaningful.</td>
</tr>
<tr>
<td>Appropriate analytic approach</td>
<td>The analytic approach must be scientifically rigorous and defensible, and consider multilevel or clustered organization of data (if necessary).</td>
</tr>
<tr>
<td>Complete documentation</td>
<td>Risk adjustment and/or stratification details and the model’s performance must be fully documented, and all known issues disclosed.</td>
</tr>
</tbody>
</table>

19.2.1 Defining the Appropriate Measure Development

Measure development samples (i.e., the population used to develop the model) are intended to be microcosms such that the distributions of characteristics and their interactions should be representative of the overall population on which the risk model is to be applied. The sample(s) should be clearly and explicitly defined. All inclusion and exclusion criteria used to select the sample should be defined. Risk adjustment models generalize well (i.e., fit the parent population) to the extent that the samples used to develop, calibrate, and validate them appropriately represent the parent population. Measure developers need to explain their rationale for using selected samples and offer justification of the sample’s appropriateness.

19.2.2 Appropriate Time Intervals

The time interval is the time frame used to determine cases for inclusion in the population of interest and outcome of interest and includes an index event and a period of time. The criteria used to formulate decisions regarding the selection of the time interval should be clearly stated and explained in the
measure documentation. The time interval criteria used to identify risk factors for the stated outcomes should be clinically appropriate and clearly stated (e.g., the risk factor occurs within 24 hours of admission). Risk factors should be present at the start of care to avoid mistakenly adjusting for factors arising due to deficiencies in care being measured, unless person-time adjustments are used. Outcomes should occur soon enough after care to establish that they are the result of that care. For example, renal failure is one of the comorbidities that may be used for risk adjustment of a hospital mortality measure. If poor care received at the hospital caused the patient to develop renal failure after admission, it would be inappropriate to adjust for renal failure for that patient.

The evaluation of outcomes must also be based on a standardized period of assessment if person-time adjustments are not used. If the periods of the outcome assessments are not standardized, such as the assessment of events during hospitalization, the evaluation may be biased because healthcare providers have different practice patterns (e.g., varying lengths of stay).

19.2.3 High Data Quality

The measure developer must ensure that data used for risk adjustment are of high-quality. Considerations in determining the quality of data include:

- Data are collected in a reliable way. That is, the method of collection must be reproducible with minimal variation between one collection and another if the same population is the source.
- Data are sufficiently valid for their purpose. Validation ultimately rests on the strength of the logical connection between the construct of interest and the results of operationalizing their measurement, recording, storage, and retrieval.
- Data are sufficiently comprehensive to limit the number of proxy measures required for the model. Obtaining the actual information is sometimes impossible, so some proxy measures might be inevitable for certain projects.
- Data are as recent as possible. If the measure developer used 1990 data in a model designed to be used tomorrow, many people would argue that the healthcare system has changed so much since 1990 that the model may not be relevant.
- Data are as complete as possible. Data should contain as few missing values as possible. Missing values are difficult to interpret and lower the validity of the model.
- Data sources, including when the data were collected, if and how the data were cleaned and manipulated, and the data’s assumed quality, are documented and fully disclosed.

19.2.4 Appropriate Variable Selection

The risk adjustment model variables should be clinically meaningful or related to variables that are clinically meaningful. When developing a risk-adjusted model, the clinical relevance of included variables should be apparent to SMEs. When variables are clearly clinically relevant, two purposes are served: the clinical relevance contributes to the face validity of the model and the likelihood that the model will explain variation identified by healthcare professionals, and/or the literature as being important to the outcome is increased. Parsimonious models and their outcome are likely to have the highest face validity and be optimal for use in a model. The strengths of the associations required to retain adjustment factors ultimately depend on the conceptual model, but are rarely a factor included in a model that is not substantively associated with the outcome variable. The measure developer must determine which risk factors should be retained in the risk adjustment model (e.g., retain variables that are clinically relevant and statistically significantly associated with the outcomes).
Occasionally, proxy variables may be included in the risk adjustment model based on prior research. This situation may arise when direct assessment of a relevant variable is not possible, and the use of a substitute or proxy variable is required. However, the relevance of these substitute variables should be empirically appropriate for the clinical topic of interest. For example, medications taken might be useful as a proxy for illness severity or progression of a chronic illness, provided practice guidelines or prior studies clearly link the medication patterns to the illness severity or trajectory. Similarly, inclusion of variables previously shown to moderate the relationship between a risk adjustor and the measure may be included. Moderating variables are generally interaction terms that are sometimes included in a model to understand complex information structures among variables (e.g., a prior mental health diagnosis may be only weakly associated with a measured outcome, but it may interact with another variable to strongly predict the outcome). Moderating variables and interaction terms, when needed, require specialized data coding and interpretation.

19.2.5 Appropriate Analytic Approach

An appropriate statistical model is determined by many factors. Logistic regression or hierarchical logistic regression is often used when the outcome is dichotomous; however, in certain instances, the same data may be used to develop a linear regression model when key statistical assumptions are not violated. Selecting the correct statistical model is imperative because an incorrect model can lead to erroneous or misleading results. The analytic approach should also consider any multilevel and/or clustered organization of data, which is typically present when assessing institutions such as hospitals from widespread geographic areas.

Risk factors retained in the model should account for substantive and significant variation in the outcome. Overall differences between adjusted and unadjusted outcomes should also be pragmatically and clinically meaningful. Moreover, risk factors should not be related to the stratification factors, when stratifying. A statistician can guide the measure development team and recommend the most useful variable formats and appropriate models.

19.2.6 Complete Documentation

Transparency is one of the key design principles in the Blueprint. When researchers do not disclose the steps that were used to create a risk adjustment model, others cannot understand or fully evaluate the model. The Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement on the endorsement of proprietary measures promotes the full disclosure of all aspects of a risk adjustment model used in measure development.

HHS policies emphasize transparency; therefore, measure developers should fully describe the risk adjustment method used; performance of the risk adjustment model, its components, and its algorithms; sources of the data and methods used to clean or manipulate the data; and the code (e.g., SAS) and documentation for how to run the calculation code. Documentation should be sufficient to enable others to reproduce the findings. Measure documentation is expected to incorporate statistical and methodological recommendations from a knowledgeable statistician to explain the model that was chosen and why it was used.

44 There is no intention to suggest that logistic regression is appropriate to model continuous manifest variables (i.e., available data). Nonetheless, various forms of logistic regression are used to model latent traits (i.e., inferred variables modeled through related observations) that are assumed to be continuous, except where the available data are dichotomous, such as the probability of receiving a specified healthcare service.
19.3 **Risk Adjustment Procedure**

Seven (7) steps are recommended in the development of a risk adjustment model:

- Choose and define an outcome.
- Define the conceptual model.
- Identify risk factors and timing.
- Acquire data (sample, if necessary).
- Model the data.
- Assess the model.
- Document the model.

Some models may not lend themselves appropriately to these steps, and an experienced statistician and clinical expert can determine the need for each step. A list of requirements associated with risk adjustment is provided in Figure 37.

### 19.3.1 Choose and Define an Outcome

When selecting outcomes that are appropriate for risk adjustment, the time interval for the outcome must be meaningful, the definition of the outcome must clearly define what is counted and not counted, and one must be able to collect the outcome data reliably. An appropriate outcome has clinical or policy relevance. It should occur with sufficient frequency to enable statistical analysis, unless the outcome is a preventable and serious healthcare error that should never happen. Outcome measures should be evaluated for both validity and reliability, as described in Section 2, Chapter 4, Measure Testing. Whenever possible, clinical experts, such as those participating in the TEP, should also be consulted to help define appropriate and meaningful outcomes. Finally, as discussed in Section 3, Chapter 13, Person and Family Engagement, patients should be involved in choosing which outcomes are appropriate for quality measurement. They are the ultimate experts on what is meaningful to their experience and what they value.

 риск variables in risk-adjusted outcome eCQMs are currently represented as supplemental data elements and are represented as measure observation in the MAT.

### 19.3.2 Define the Conceptual Model

A clinical hypothesis or conceptual model about how potential risk factors relate to the outcome should be developed a priori. The conceptual model serves as a map for development of a risk adjustment model. It defines the understanding of the relationships behind the variables and, as such, helps to identify which risk factors, patients, and outcomes are important, and which can be excluded. Because the cost of developing a risk adjustment model may be prohibitive if every potential risk factor is included, the conceptual model also enables the measure developer to prioritize among risk factors and to evaluate the cost and benefit of data collection. An in-depth literature review can greatly enhance this process. Alternatively, the existence of large databases and computational approaches such as machine learning allow for statistical analyses to explore the data for relationships between outcomes and potential adjustment factors that might not yet be clinically identified, but empirically exist. Measure developers are reminded to be aware of the potential for spurious relationships.

---

**Figure 37. Risk Adjustment Requirements**

<table>
<thead>
<tr>
<th>Risk Adjustment Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Risk Adjustment Methodology Report that includes full documentation of the risk adjustment model or rationale and data to support why no risk adjustment or stratification is needed.</td>
</tr>
<tr>
<td>2. MIF (Deliverable 3-3) with completed risk adjustment sections for each measure (Deliverable 4-4).</td>
</tr>
<tr>
<td>3. For eCQMs, HQMF, and human-readable HTML files that include instructions where the complete risk adjustment methodology may be obtained.</td>
</tr>
</tbody>
</table>
The first step in developing or selecting the conceptual model is identifying relationship among variables. This process should include:

- Conducting a review of clinical literature and canvassing expert opinion to establish variable relationships and to identify patient factors that are measurable and related to the outcome and either independent of the quality construct (i.e., for risk adjustment) or not independent (i.e., for stratification).
- Obtaining expert opinion from healthcare providers with clinically relevant specialties, experienced statisticians and research methodologists, and relevant stakeholders such as patient advocates. A TEP may be used if diverse input is sought. Section 3, Chapter 12, Technical Expert Panel, covers the standardized process for convening a TEP.
- Conducting empirical analyses to further support variable selection (when appropriate data are available) to identify potential factors for consideration by SMEs.

19.3.3 Identify the Risk Factors and Timing

Use of a conceptual model and clinical expertise promotes selection of risk factors with the following criteria:

- Clinically relevant in the case of clinical risk factors
- Reliably collected
- Validly operationalized
- Sufficiently comprehensive
- Associated with the outcome
- Clearly defined
- Identified using appropriate time frames.

In addition to these attributes, risk factors should also align with NQF policies for endorsed measures.

19.3.4 Acquire Data (Sample, if Necessary)

Healthcare data can be acquired from many sources, but the three most frequently used are claims data, patient record data, and survey data. Of these, the most common source of data for developing risk adjustment models is claims data reported by the provider. Once data sources are acquired, relevant databases may need to be linked and various data preparation tasks performed, including an assessment of the data reliability and validity, if not previously confirmed. If samples are to be used, they should be drawn using predefined criteria and methodologically sound sampling techniques. Testing to determine the suitability of data sources and testing for differences across data sources may also be necessary. The alpha and beta testing discussion in Section 3, Chapter 22, Measure Testing provides details of the processes.

19.3.5 Analyze the Data

In addition to clinical judgment used to define the conceptual model and candidate variables, empirical analysis should also be conducted to help determine risk factors to include or exclude. Several concerns exist in data analysis and should be considered when developing an appropriate risk adjustment model.

19.3.5.1 Sufficient Data

When creating a risk adjustment model, there should be enough data available to ensure a stable model. Different statistical rules apply to different types of models. For example, a model with an outcome that is more common may require more than 30 cases per patient factor to consistently return
the same model statistics across samples. If the outcome is uncommon, then the number of cases required could be much larger (Iezzoni, 2013). Other factors may also affect the size needed for a sample, such as a lack of variability among risk factors for a small sample that results in partial collinearity among risk factors and a corresponding decrease in the stability of the parameter estimates. A statistician can provide guidance to determine the appropriate sample sizes based on the characteristics of the sample(s) and the requirements of the types of analyses being used.

19.3.5.2 Model Simplicity

Whenever possible, fitting a model with as few variables as possible to explain the most variance possible is preferred. This is often referred to as model simplicity or model parsimony, whereby a smaller number of variables accomplish approximately the same goal as a model with a larger number of variables. This principle of preferring parsimony captures the balance between errors of underfitting and overfitting inherent in risk adjustment model development. For example, developing a model with many predictors can result in model variables that primarily explain incremental variance unique to a data source or available samples (i.e., overfitting) and can also result in reduced stability of parameters due to increased multi-collinearity among a larger number of predictors. In contrast, a model with fewer predictors may reduce the amount of explained variance possible for the measure (i.e., underfitting).

When evaluating these models, determination of the preferred model may depend on the availability of other samples to validate findings and detect overfitting and the degree of multi-collinearity among predictors. However, in general, the simpler model may provide a more robust explanation, since it uses fewer variables to explain nearly the same observed variability. In addition, simpler models are likely to reduce the cost of model development by collecting fewer variables and may be less likely to show signs of model overfitting. Parsimonious models are often achieved by omitting statistically significant predictors that offer minimal improvement in predictive validity or overall model fit and by combining clinically similar conditions to improve performance of the model across time and populations. However, in situations with high visibility or potentially widespread fiscal repercussions, CMS has employed some of the most sophisticated models available, such as Hierarchical Generalized Linear Models as described in Statistical Issues in Assessing Hospital Performance.

19.3.5.3 Methods to Retain or Remove Risk Adjustors

When developing a risk adjustment model, the choice of variables to be included often depends on estimated parameters in the sample rather than the true value of the parameter in the population. Consequently, when selecting variables to retain or exclude from a model, the idiosyncrasies of the sample, as well as factors such as the number of candidate variables and correlations among the candidate variables, may determine the final risk adjustors retained in a model (Harrell, 2015). Improper model selection or not accounting for the number of, or correlation among the candidate variables may lead to risk adjustment models that include suboptimal parameters or overestimated parameters—making them too extreme or inappropriate for application to future datasets. This outcome is sometimes referred to as model overfitting, particularly when the model is more complicated than needed and describes random error instead of an underlying relationship.

Given these possibilities, it is advisable to consider steps to adjust for model overfitting, such as selection of model variables based on resampling methods and assessment of the model in multiple/diverse samples (refer also to Section 3, Chapter 19.3.5.4, Generalizability). Consultation of clinical expertise, ideally used during candidate variable selection, is also strongly recommended when examining the performance of candidate variables in risk adjustment models. This expertise may help inform relationships among model parameters and may help justify decisions to retain or remove variables.
### 19.3.5.4 Generalizability

Steps to ensure findings can be generalized to target populations should also be taken when developing the model. Researchers often use two datasets in building risk adjustment models: a development (i.e., calibration) dataset and a validation dataset. The development/calibration dataset is used to develop the model or calibrate coefficients, and the validation dataset is used to determine the extent to which the model can be appropriately applied to parent populations. When assessing generalizability to the population from which the development dataset was derived, the two datasets may be collected independently—which can be costly—or one dataset may be split using random selection.

Either of these methods enables evaluation of the model’s generalizability to the population and helps avoid any model features that arise from idiosyncrasies in the development sample. Additional validation using samples from different time periods may also be desirable to examine stability of the model over time.

### 19.3.5.5 Multilevel (Hierarchical) Data

The potential for observations to be “nested” within larger random groupings or levels frequently occurs in healthcare measurement (e.g., patients may be nested under physician groups, which may in turn be nested under hospitals). The risk adjustment model should account for these multi-level relationships, when present, and risk adjustment development should investigate theoretical and empirical evidence for potential patterns of correlation in this multi-level data. For example, patients in the same IRF may tend to have similar outcomes based on a variety of factors, and this should be addressed by the risk adjustment model.

Such multi-level relationships are often examined by building models designed to account for relationships between observations within larger groups. Terms for these types of models include multi-level model, hierarchical model, random-effects model, random coefficient model, and mixed model. These terms all refer to models that explicitly model the “random” and “fixed” variables at each level of the data. In this terminology, a “fixed” variable is one that is assumed to be measured without error, where the value or characteristic being measured is the same across samples (e.g., male vs. female, nonprofit vs. for-profit facility) and studies. In contrast, “random” variables are assumed to be values drawn from a larger population of values (e.g., a sample of IRFs), where the value of the random variable represents a random sample of all possible values of that variable.

Traditional statistical methods (i.e., linear regression and logistic regression) require observations (e.g., patients) in the same grouping to be independent. When observations co-vary based on the organization of larger groupings, these methods fail to account for the hierarchical structure, and assumptions of independence among the observations are violated. This situation may ultimately lead to underestimated standard errors and incorrect inferences. Attempts to compensate for this problem by treating the grouping units as fixed variables within a traditional regression framework are generally undesirable, as the grouping units must be treated as a fixed variable, which does not allow for generalization to any other groupings beyond those grouping units in the sample.

Multi-level models overcome these issues by explicitly modeling the grouping structure and by assuming that the groups reflect random variables (usually with a normal distribution) sampled from a larger population. They consider variation at different grouping levels and allow modeling of hypothesized factors at these different levels. For example, a multi-level model may allow modeling patient-level risk factors along with the facility-level factors. If the measure developer has reason to suspect hierarchical structure in the measurement data, these models should be examined. The models can be applied within common frameworks used for risk adjustment (e.g., ordinary least squares regression for...
continuous outcomes, logistic regression for binary outcomes), as well as less common longitudinal frameworks such as growth (i.e., change) modeling.

Developments in statistics are enabling researchers to improve both the accuracy and the precision of nested models using computer-intensive programs. These models include estimation of clustering effects independent of the main effects of the model to better evaluate the outcome of interest. For example, the use of precision-weighted empirical Bayesian estimation has been shown to produce more accurately generalizable coefficients across populations than methods that rely on the normal curve for estimation (e.g., linear regression). Hierarchical factor analysis and structural equation modeling have also been used. Recently, CMS has moved toward using the Hierarchical Generalized Linear Model for monitoring and reporting hospital readmissions.

19.3.6 Assess the Model

This step is required for a newly developed risk adjustment model. It is also required when using an “off-the-shelf” adjustment model because an existing risk adjustment model may perform differently in the new measure context. When multiple data sources are available (e.g., administrative and chart-based data), it is strongly recommended that model performance be assessed for each data source to allow judgment regarding the adequacy and comparability of the model across the data sources.

Assess any model developed to ensure that it does not violate underlying model assumptions (e.g., independence of observations or assumptions about underlying distributions) beyond the robustness established in the literature for those assumptions. Models must also be assessed to determine predictive ability, discriminant ability, and overall fit of the model. Justification of the types of models used must be provided to the COR and documented in the Risk Adjustment Methodology report. Some examples of common statistics used in assessing risk adjustment models include the $R^2$ statistic, receiver-operating characteristic (ROC) curve, and Hosmer-Lemeshow test (HL test). However, several other statistical techniques can enable measure developers to assess different aspects of model fit for different subpopulations as well as for the overall population. Use of an experienced statistician is critical to ensure the most appropriate methods are selected during model development and testing.

19.3.6.1 $R^2$ Statistic

A comparison of the $R^2$ statistic with and without selected risk adjustment is frequently used to assess the degree to which specific risk-adjusted models predict, explain, or reduce variation in outcomes unrelated to an outcome of interest. The statistic can also be used to assess the predictive power of risk-adjusted models overall. In that case, values for $R^2$ describe how well the model predicts the outcome based on values of the included risk factors.

The $R^2$ value for a model can vary, and no firm standard exists for the optimal expected value. Experience or previously developed models may inform which $R^2$ value is considered reasonable. In general, the larger the $R^2$ value, the better the model. However, clinical expertise may also be needed to help assess whether remaining variation is primarily related to differences in the quality being measured. Extremely high $R^2$ values can indicate that something is wrong with the model.

19.3.6.2 ROC Curve, Area Under the Curve (AUC), and C-statistic

The ROC curve is often used to assess models that predict a binary outcome (e.g., a logistic regression model), where responses are classified into two categories. The ROC curve can be plotted as the proportion of target outcomes correctly predicted (i.e., a true positive) against the proportion of outcomes incorrectly predicted (i.e., a false positive). The curve depicts the tradeoff between the model’s sensitivity and specificity.
An example of ROC curves is shown in Figure 38. Curves approaching the 45-degree diagonal of the graph represent less desirable models (Curve A) when compared to curves falling to the left of this diagonal that indicate higher overall accuracy of the model (Curves B and C). A test with nearly perfect discrimination will show a ROC curve that passes through the upper-left corner of the graph, where sensitivity equals 1, and 1 minus specificity equals zero (Curve D).

The power of a model to correctly classify outcomes into two categories (i.e., discriminate) is often quantified by the ROC AUC. The AUC, sometimes referred to as the c-statistic, is a value that varies from 0.5 (i.e., discriminating power not better than chance) to 1.0 (i.e., perfect discriminating power). The AUC can be interpreted as the percent of all possible pairs of observed outcomes in which the model assigns a higher probability to a correctly classified observation than to an incorrect observation. Most statistical software packages compute the probability of observing the model AUC found in the sample when the population AUC equals 0.5 (i.e., the null hypothesis). Both non-parametric and parametric methods exist for calculating the AUC, and this varies by statistical software.

**Figure 38. Example of ROC Curves**

19.3.6.3 **The Hosmer-Lemeshow Test**

Although the AUC/c-statistic values provide a method to assess a model’s discrimination, the quality of a model can also be assessed by how closely the predicted probabilities of the model agree with the actual outcome (i.e., whether predicted probabilities are too high or too low relative to true population values). This is sometimes referred to as calibration of a model. It is often assessed using the HL test of goodness-of-fit, which assesses the extent to which the observed values/occurrences match expected event rates in subgroups of the model population. The HL test identifies subgroups of ordered observations based on the predicted model values or other factors external to the model associated with the outcome risk. The subgroups can be formed for any reasonable grouping, but often, deciles or...
quintiles are used. Generally, a model is considered well calibrated when the expected and observed values agree for any reasonable grouping of the observations. Yet, high-risk and low-frequency situations pose special problems for these types of comparison methodologies; therefore, they should be addressed by an experienced statistician.

A statistician with experience in this methodology can determine the adequacy of any model. It is expected that the measure development team will employ the services of a statistician to accurately assess the appropriateness of a risk-adjusted model. Determining the best risk-adjusted model may involve multiple statistical tests that are more complex than those cited here. For example, a risk adjustment model may discriminate very well based on the c-statistic, yet still be calibrated poorly. Such a model may predict well at low ranges of outcome risk for patients with a certain set of characteristics (e.g., the model produces an outcome risk of 0.2 when roughly 20% of the patients with these characteristics exhibit the outcome in population), but predict poorly at higher ranges of risk (e.g., the model produces an outcome risk of 0.9 for patients with a different pattern of characteristics when only 55% of patients with these characteristics show the outcome in population). In this case, one or more goodness-of-fit indices may need to be consulted to identify a superior model and careful analysis of different subgroups in the sample may also be needed to further refine the model. Additional steps to correct for bias in estimators, improving confidence intervals, and assessing any violation of model assumptions may also be required. Moreover, differences across groups for measures that have not been risk-adjusted may be clinically inconsequential when compared to risk-adjusted outcomes. Clinical experts in the subject matter are also expected to be consulted to provide an assessment of both the risk adjustors and utility of the outcomes.

19.3.7 Document the Model

A Risk Adjustment Methodology report is a required deliverable and is expected at the conclusion of a measure development project. This report ensures relevant information about the development and limitations of the risk adjustment model are available for review by consumers, purchasers, and providers. The report also enables these parties to access information about the factors incorporated into the model, the method of model development, and the significance of the factors used in the model. Typically, the report will contain:

- Identification or review of the need for risk adjustment of the measure(s).
- A description of the sample(s) used to develop the model, including criteria used to select the sample and/or number of sites/groups, if applicable.
- A description of the methodologies and steps used in the development of the model or a description of the selection of an off-the-shelf model.
- A listing of all variables considered and retained for the model, the contribution of each retained variable to the model’s explanatory power, and a description of how each variable was collected (e.g., data source, time frames for collection).
- A description of the model’s performance, including any statistical techniques used to evaluate performance and a summary of model discrimination and calibration in one or more samples.
- Delineation of important limitations such as the probable frequency and influence from misclassification when the model is used (e.g., classifying a high-outcome provider as a low one or the reverse) (Austin, 2008).
- Enough summary information about the comparison between unadjusted and adjusted outcomes to evaluate whether the model’s influence is clinically significant.
• A section discussing a recalibration schedule for the model to accommodate changes in medicine and in populations; such schedules are normally first assigned based on the experience of clinicians and the literature’s results and then later updated as needed.

All measure specifications, including the risk adjustment methodology, must be fully disclosed. The risk adjustment method, data elements, and algorithm should be fully described in the Risk Adjustment portion of the MIF. Attachments or links to websites should be provided for coefficients, equations, codes with descriptors, and definitions and/or specific data collection items/responses used in the risk adjustment. Documentation should comply with the open-source requirements of NQF’s Conditions for Consideration, and all applicable programming code should be included. If calculation requires database-dependent coefficients that change frequently, the existence of such coefficients and the general frequency that they change should be disclosed, but the precise numerical values assigned need not be disclosed, as they vary over time.
20  **COST AND RESOURCE USE MEASURE SPECIFICATION**

The NQF notes in the National Voluntary Consensus Standards for Cost and Resource Use, it is important to submit instructions and analytic steps for aggregating data when designing cost and resource use measures. Specifications should include the types of data that are required, the time periods relevant to the measures, and who is included in the measurement. For example, if certain services are carved out from the claims for certain health plans and not for others, comparison of costs between the plans could be misleading. Most cost and resource use measures use administrative claims data. However, if coding practices vary, data reliability and validity can be compromised. These issues should be addressed during measure development and maintenance.

Resource use measures can be developed for different units of analysis:

- Per capita-population and per capita-patient
- Per encounter
- Per admission
- Per procedure
- Per visit.

20.1  **MEASURE CLINICAL LOGIC**

Measures are usually identified as resource use measures for acute conditions, chronic conditions, or preventive services, which often affects the clinical logic. Analytic steps are designed to create appropriately homogeneous units for measurement.

20.2  **MEASURE CONSTRUCTION LOGIC**

20.2.1  **Time Intervals**

Decisions about when to start or end a measurement period must be specified for each measure. These time intervals may be identified through clinical or evidence-based guidelines, expert opinion, or empirical data. Typically, the time interval for measure reporting is the calendar year.

20.2.2  **Assigning and Triaging Claims**

Some examples of decisions that need to be addressed in managing claims data include:

- How to use different claims that provide information for the same event—especially those that result in an inflation of resource use amounts
- When and how to map or feed claims from different sources into the same measure
- When and which services trump other services
- Identifying units of resource use.

The units of health services or resource use units must be identified and defined. Measure specifications must clearly define and provide detailed instructions on how to identify a single health service unit, including the relevant codes, modifiers, or approaches to identify the amount.
20.3 **ADJUSTING FOR COMPARABILITY**

20.3.1 Define **Risk Adjustment** Approach

Risk adjustment is designed to reduce any negative or positive consequences associated with caring for patients of higher or lower health risk or propensity to require health services. Resource use measures, including episode-based measures, generally risk-adjust as part of the steps to address differences in patient characteristics and disease severity or stage.

20.3.2 Define **Stratification** Approach

Another type of adjustment is stratification, which is important if known disparities exist or if there is a need to expose differences in results so that stakeholders can take appropriate action. In addition to exposing disparities, a measure may specify stratification of results within a major clinical category (e.g., diabetes) by severity or other clinical differences.

20.3.3 Define Costing Methodology

The following costing methods may be used, depending on the intended perspective:

- Count of services
- Actual amount paid
- Standardized prices.

20.4 **MEASURE REPORTING**

20.4.1 Attributing Resource Use Measures

Resource use measures are used to attribute the care provided as part of an episode of illness, care of a population, and event to a provider (e.g., physician, physician groups) or other entity (e.g., health plan), in combination with quality or health outcome performance. It is easier to identify the appropriate provider for attribution when the topic is narrowly defined, such as for a specific procedure. Measures for an episode of care or per capita measures are broader and often involve multiple providers, making valid attribution more difficult.

Care can be attributed to a single provider or multiple providers. Single attribution identifies the decision-maker, perhaps the primary care physician, and holds this individual responsible for all care rendered. Multiple attribution acknowledges that the decision-maker, if there is one, has incomplete control over treatment by other physicians or specialists, even if the decision-maker referred the patient to those other physicians.

20.4.2 Peer Group Identification and Assignment

Unlike quality measures, which normally compare performance to an agreed-upon standard (e.g., providing flu vaccinations to a percentage of eligible patients) and direction for improvement (e.g., higher or lower performance is better), preferred resource use amounts often are not standardized; and it is not always clear whether higher or lower resource use is preferable. Instead, resource use measures are used to compare a physician’s or entity’s performance to the average performance of their peers. For this reason, it is essential to identify an appropriate peer group for comparison.
20.4.3 Calculating Comparisons

The observed-to-expected (O/E) ratio compares the value for each resource use measure attributed to a physician or entity (i.e., observed amount) and divides it by the average resource use within the identified peer group (i.e., expected amount—the amount of resource use expected if the entity measured was performing at the mean).

More sophisticated statistical approaches (e.g., multilevel regression) also are used.

20.4.4 Setting Thresholds

After estimating the value of a resource use measure and to provide more context for the values, the measure developer should determine whether to apply thresholds or remove outliers. Outliers can be the result of inappropriate treatment, rare or extremely complicated cases, or coding error. Users often do not completely discard outliers, but rather examine them separately. These actions should be documented so users can understand the full context.

20.4.5 Providing Detailed Feedback

After the analytic steps are completed, users of resource use measures must decide which analytic results to publicly report or include in provider feedback.

20.4.6 Reporting with Descriptive Statistics

It is critical to choose the right statistics when reporting resource use measure results. Factors influencing this choice include whether the results will be used for public reporting or for feedback to providers. Well-crafted descriptive analytic results can provide the detailed information necessary to make feedback actionable for all stakeholders. However, it is important to balance detailed reporting with the possibility of information overload.
21 COMPOSITE MEASURE TECHNICAL SPECIFICATIONS

Although technical specifications of all components of the composite may already be documented, they should also be completed for the composite. The MIF and MJF are aligned with the requirements of NQF measure submission and guide the measure developer to ensure technical specifications are sufficient and complete. Composite measure technical specifications should be included with other measure documentation forms for submission to the COR for approval.

Even though the component measures may not meet all the evaluation criteria, the composite performance measure as a whole must meet evaluation criteria. The criteria for composite performance measures are described in Section 3, Chapter 23, Measure Evaluation.

The methodology and considerations for scoring include ensuring the weighting and scoring of components support the goal that is articulated for the measure. Then, using a specified method, the component scores are combined into one composite.

Descriptions of five common types of composite performance measure scoring are provided in Table 23. This list is not intended to be an exhaustive list of the only scoring methods allowed. Some advantages and disadvantages for each type with examples of measures in the category are included. The five types discussed are:

- All-or-none
- Any-or-none
- Linear combinations
- Regression-based composite performance measures
- Opportunity scoring.
### Table 23. Types of Composite Measure Scoring

<table>
<thead>
<tr>
<th>Type of Scoring</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Examples/Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All-or-One (Defect-free Scoring)</strong></td>
<td>Promotes a high standard of excellence. Patient-centric. Fosters a system perspective. Offers a more sensitive scale for assessing improvements. Especially useful for those conditions for which achieving a desired clinical outcome empirically requires reliable completion of a full set of tasks (i.e., when partial completion does not gain partial benefit).</td>
<td>May waste valuable information. May weight common but less important processes more heavily than infrequent but important processes. The provider who achieved four of five measures appears the same as the provider who achieved none of five measures. The all-or-none approach will amplify errors of measurement (e.g., one unreliable component measure will contaminate the whole score), so it is essential that each of the component measures be well designed.</td>
<td>Minnesota Community Measurement Optimal Diabetes Care measure. IHI Bundles: ventilator, central line. STS Perioperative Medical Care, a process bundle of four medications: preoperative beta blockade and discharge anti-platelet, beta blockade, and lipid-lowering agents. Study using Premier Surgical Care Improvement Project (SCIP) (Stulberg, Delaney, Neuhauser, Aron, Fu, &amp; Koroukian, 2010) data; adherence measured through a global all-or-none composite infection-prevention score was associated with a lower probability of developing a postoperative infection. However, adherence reported on individual SCIP measures was not associated with a significantly lower probability of infection.</td>
</tr>
<tr>
<td><strong>Any-or-One Process or Outcome Measures</strong></td>
<td>Promotes a high standard of excellence. Useful when component measures are rare events.</td>
<td>Particularly problematic when rare but important outcomes are mixed with common but relatively unimportant outcomes because the composite is likely to be dominated by the outcome that occurs most frequently.</td>
<td>STS Postoperative Risk-Adjusted Major Morbidity, which is any of the following: renal failure, deep sternal wound infection, re-exploration, stroke, and prolonged ventilation/ intubation. This is an “any-or-none” measure requiring the absence of all such complications.</td>
</tr>
<tr>
<td>Type of Scoring</td>
<td>Advantages</td>
<td>Disadvantages</td>
<td>Examples/Evidence</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Linear Combinations</td>
<td>Simplicity.</td>
<td>Does not account for potential differences in the validity, reliability, and importance of the different individual measures (Peterson et al., 2010).</td>
<td>Premier/CMS Hospital Quality Incentive (HQI) Demonstration uses a composite of process and outcome measures to measure quality for coronary artery bypass graft (CABG). The composite quality score (CQS) was based on an equally weighted combination of seven measures (i.e., four process measures and three outcome measures). The publicly reported data suggest that the CQS was more heavily influenced by process measures than would have been expected by the apparent 4:3 weighting.</td>
</tr>
<tr>
<td></td>
<td>Transparency.</td>
<td>Equal weighting may be undesirable if there is a considerable imbalance in the numbers of measures from different domains.</td>
<td>The U.S. News &amp; World Report Index of Hospital Quality for heart and heart surgery is a linear combination of three equally weighted components: reputation, risk-adjusted mortality, and structure. Although the three components are weighted equally, a hospital’s reputation score has the highest correlation with its overall score. In comparison, the Mortality Index appears to have much less influence.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Different stakeholders have different priorities; one weighting method may not meet the needs of all potential users (Peterson et al., 2010).</td>
<td>The AHRQ Patient Safety Indicators (PSI) composite performance measure (i.e., PSI 90) uses a weighted average of various individual component measures. The weighting was determined by an expert panel.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>When items with a small standard deviation are averaged with items with a large deviation, items with the large standard deviation tend to dominate the average.</td>
<td>Leapfrog developed surgical “survival predictor” composite measures to forecast hospital performance based on prior hospital volumes and prior mortality rates. An empirical Bayesian approach was used to combine mortality rates with information on hospital volume at each hospital. The observed mortality rate is weighted according to how reliably it is estimated, with the remaining weight placed on hospital volume.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>If items are combined that are not positively or negatively correlated with one another (i.e., co-vary), the resulting composite score may not possess reasonable properties to enable meaningful differentiation among patients and may not measure a single construct. This issue can be mitigated by pursuing latent factor analysis strategies to ensure that items cohere to form a reasonable single score for a construct.</td>
<td></td>
</tr>
</tbody>
</table>

| Regression-based Composite Performance Measures | The weight assigned to each item is directly related to its reliability and the strength of its association with the gold standard end point. Regression-based weighting may be appropriate for predicting specific end points of interest. | Weighting may not be optimal for objectives such as motivating healthcare professionals to adhere to specific treatment guidelines. | Leapfrog developed surgical “survival predictor” composite measures to forecast hospital performance based on prior hospital volumes and prior mortality rates. An empirical Bayesian approach was used to combine mortality rates with information on hospital volume at each hospital. The observed mortality rate is weighted according to how reliably it is estimated, with the remaining weight placed on hospital volume. |
### Opportunity Scoring

Opportunity scoring counts the number of times a given care process was performed (numerator) divided by the number of chances a provider had to give this care correctly (denominator). Unlike simple averaging, each item is implicitly weighted in proportion to the percentage of eligible patients, which may vary from provider to provider.

<table>
<thead>
<tr>
<th>Type of Scoring</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Examples/Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opportunity Scoring</td>
<td>Provides alternative to simple averaging often used for aggregating individual process measures. Increases the number of observations per unit of measurement, potentially increasing the stability of a composite estimate, particularly when the sample size for individual measures is not adequate.</td>
<td>Rate is influenced by the most common care processes, regardless of whether they are the most important methods.</td>
<td>The opportunity model was developed for the Hospital Core Performance Measurement Project for the Rhode Island Public Reporting Program for Health Care Services in 1998. CMS/Premier HQI Demonstration project used the opportunity scoring method for the process composite rate for each of five clinical areas. The sum of all numerators is divided by the sum of all denominators in each clinical area.</td>
</tr>
</tbody>
</table>

The MAT currently supports composite measures within the metadata section, but limits measure scoring to All or Nothing, Opportunity, or Patient-level Linear. Currently, users who are defining a composite measure can indicate the measure type as composite and can then identify measures in the metadata that are component measures.
22 **MEASURE TESTING**

Testing is performed iteratively and provides an opportunity to refine draft specifications before they are finalized; augment or reevaluate earlier judgments about the measure’s importance; and assess feasibility, usability, and scientific acceptability of the measure. For more information on measure testing as it relates to evaluation criteria, refer to Section 3, Chapter 23, Measure Evaluation.

This chapter provides an overview of the types of testing expected to be undertaken to assess measure criteria and outlines the process for development, implementation, and reporting of test plans, results, and associated artifacts. Information in this chapter is not meant to be prescriptive or exhaustive; other approaches to testing that employ appropriate methods and rationale may be used. Measure developers should always select testing that is appropriate for the measure being developed and always provide empirical evidence for importance to measure and report, feasibility, scientific acceptability, and usability and use.

### 22.1 ALPHA AND BETA TESTING

Initial testing during development (i.e., pilot testing) is generally conducted within the framework of alpha and beta tests. Although both alpha and beta testing are considered part of measure testing, alpha testing may occur as early as information gathering and is repeated iteratively during development of measure specifications. Developers should conduct testing early and often.

#### 22.1.1 Alpha Testing

Alpha testing (i.e., formative testing) is of limited scope since it usually occurs before detailed specifications are fully developed. Alpha testing, particularly regarding feasibility of the concept in the context of the data source, may be conducted as part of information gathering empirical analysis and may also be performed concurrently with development of the technical specifications as part of an iterative process. The alpha tests include methods to determine whether individual data elements are available and whether the form in which they exist is consistent with the intent of the measure. The types of testing done in an alpha test vary widely and often depend on the measure’s data source or uniqueness of the measure specifications. Measures that use data sources similar to existing measures may require minimal alpha testing. In contrast, measures that address areas for which specifications have never been developed may require multiple iterations of alpha testing. For example, an alpha test may include a query to a large integrated delivery system database to determine how specific data are captured, where they originate, and how they are currently expressed. Results can impact decisions about measure specifications.

#### 22.1.2 Beta Testing

Beta testing (i.e., field testing) generally occurs after initial technical specifications have been developed and is usually larger in scope than alpha testing. In addition to gathering further information about feasibility, beta tests serve as the primary means to assess scientific acceptability and usability of a measure. Beta tests can also be used to evaluate the measure’s suitability for risk adjustment or stratification and help expand previous importance and feasibility evaluations. When carefully planned and executed, beta testing helps document measure properties with respect to the evaluation criteria.

#### 22.1.3 Comparing Alpha and Beta Testing

Attributes of each phase of testing are shown in Table 24, and these may be used as considerations when developing a work plan for alpha or beta tests.
Table 24. Features of Alpha and Beta Testing

<table>
<thead>
<tr>
<th></th>
<th><strong>Alpha Testing</strong></th>
<th><strong>Beta Testing</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Timing</strong></td>
<td>• Usually conducted prior to completion of technical specifications</td>
<td>• Conducted after the measure developer’s detailed and precise technical specifications are developed</td>
</tr>
<tr>
<td></td>
<td>• May be conducted multiple times in quick succession</td>
<td></td>
</tr>
<tr>
<td><strong>Scale</strong></td>
<td>• Typically, smaller scale</td>
<td>• Strives to achieve representative sample sizes</td>
</tr>
<tr>
<td></td>
<td>• Only enough records to ensure data set contains all elements needed for the measure</td>
<td>• Requires appropriate sample selection protocols</td>
</tr>
<tr>
<td></td>
<td>• Only enough records to identify common occurrences or variation in the data</td>
<td>• May require evaluation of multiple sites in a variety of settings depending on the data source (e.g., administrative, medical record)</td>
</tr>
<tr>
<td><strong>Sampling</strong></td>
<td>• Convenience sampling</td>
<td>• Sufficient to allow adequate testing of the measure’s scientific acceptability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Representative of the target population</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Representative of the people, places, times, events, and conditions important to the measure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• If based on administrative data, uses the entire eligible population</td>
</tr>
<tr>
<td><strong>Specification Refinement</strong></td>
<td>• Permits early detection of problems in technical specifications (e.g., identification of additional inclusion and exclusion criteria)</td>
<td>• Used to assess or revise the complexity of computations required to calculate the measure</td>
</tr>
<tr>
<td><strong>Importance</strong></td>
<td>• Designed to look at volume, frequency, or costs related to a measure topic (e.g., cost of treating the condition, costs related to procedures measured)</td>
<td>• Allows for enhanced evaluation of a measure’s importance, including evaluation of performance thresholds, disparities analysis, and outcome variation</td>
</tr>
<tr>
<td></td>
<td>• Establishes, on a preliminary basis, that the measure can identify gaps in care</td>
<td>• Evaluates opportunities for improvement in the population, which aids in evaluation of the measure’s importance (e.g., obtaining evidence of substantial variability among comparison groups, obtaining evidence that the measure is not topped-out where most groups achieve similarly high performance levels approaching the measure’s maximum possible value)</td>
</tr>
<tr>
<td></td>
<td>• Provides support for further development of the measure</td>
<td></td>
</tr>
<tr>
<td><strong>Scientific Acceptability</strong></td>
<td>• Limited in scope if conducted during the formative stage</td>
<td>• Assesses measure reliability and validity</td>
</tr>
<tr>
<td></td>
<td>• Usually occurs later in development</td>
<td>• Reports results of analysis of exclusion (if any used)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Tests results of the risk adjustment model, quantifying relationships between and among factors</td>
</tr>
<tr>
<td><strong>Feasibility</strong></td>
<td>• Provides initial information about the feasibility of collecting required data and calculating measures using technical specifications</td>
<td>• Provides enhanced information regarding feasibility, including greater determination of barriers and provider burden to implementation and costs associated with measurement</td>
</tr>
<tr>
<td></td>
<td>• Identifies barriers to implementation</td>
<td>• Evaluates feasibility of stratification factors based on occurrences of target events in the sample</td>
</tr>
<tr>
<td></td>
<td>• Offers an initial estimate of costs or burden of data collection and analysis</td>
<td></td>
</tr>
<tr>
<td><strong>Usability</strong></td>
<td>• No formal analytic testing at this stage; may use qualitative testing with patients and providers during alpha testing</td>
<td>• Identifies unintended consequences, including susceptibility to inaccuracies and errors</td>
</tr>
<tr>
<td></td>
<td>• TEP may be used to assess potential usability of the measure</td>
<td>• Reports strategies to ameliorate unintended consequences</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• May consist of focus groups or similar means of assessing usefulness of the measure by consumers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• May not be in the scope of measure development contract</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• TEP may also be used to assess potential usability</td>
</tr>
</tbody>
</table>
22.2 **THE MEASURE TESTING PROCESS**

*Section 2, Chapter 4, Measure Testing*, provides additional details on interrelationships and chronology of the measure testing steps:

- Develop Measure Testing Plan
- Submit plan and obtain COR approval
- Implement plan
- Analyze test results
- Refine measure
- Retest refined measure
- Compile deliverables and submit them to the COR
- Support the COR during NQF endorsement process.

> For more information about special considerations pertaining to eCQM testing, refer to *Section 3, Chapter 24.2, eCQM Testing*.

### 22.2.1 Develop the Testing Work Plan

Measure testing can be conducted for a single measure or a set of measures. If the testing targets a set of measures, construct a work plan that describes the full measure set. The work plan for alpha testing is usually prepared early in the measure development process; therefore, the exact number of measures to be tested may not be known, and many of the work plan areas listed below may not be appropriate. In contrast, the work plan for a beta test should be prepared after measure specifications have been developed and should include sufficient information to help the COR understand how the sampling and planned analyses aim to meet scientific acceptability, usability, and feasibility criteria required for approval by CMS and endorsement by NQF.

The testing plan should contain:

- Name(s) of measure(s)
- Type of testing
- Study objective(s)
- Timeline for testing and report completion
- Data collection methodology
- Description of test population, including number and distribution of test sites/data sets, when available
- Description of data elements to be collected
- Sampling methods to be used, if applicable
- Description of strategy to recruit providers/obtain test data sets, if multiple sites or data sets are used
- Analysis methods planned and description of test statistics to be used to support assessment. This will be less extensive for an alpha test. For a beta test, methods and analysis should address these evaluation criteria:
  - **Importance**—including analysis of opportunities for improvement such as reducing variability in comparison groups or disparities in healthcare related to race, ethnicity, age, or other classifications
  - Scientific acceptability—including analysis of reliability, validity, and exclusion appropriateness
Feasibility—including evaluation of reported costs or perceived burden, frequency of missing data, and description of data availability

Usability—including planned analyses to demonstrate that the measure is meaningful and useful to the target audience. This may be accomplished by the TEP reviewing the measure results (e.g., means and detectable differences, dispersion of comparison groups). More formal testing, if requested by CMS, may require assessment via structured surveys or focus groups to evaluate usability of the measure (e.g., clinical impact of detectable differences, evaluation of the variability among groups).

- Description and forms documenting patient confidentiality and description of Institutional Review Board (IRB) compliance approval or steps to obtain data use agreements (if necessary).
- Methods to comply with the PRA, if relevant.
- Training and qualification of staff who:
  - Manage the project (and their qualifications)
  - Conduct testing (and their qualifications)
  - Conduct or oversee data abstraction
  - Conduct or oversee data processing
  - Conduct or oversee data analysis.

22.2.2 Perform Sampling

The need for sampling often varies depending on the type of test (i.e., alpha or beta) and the type of measure. For example, measures that rely on administrative data sources (e.g., claims) can sometimes be tested by examining data from the entire eligible population, with limited drain on external resources, depending on the nature of the analysis. However, to test some measures, it is necessary to collect information from service providers or beneficiaries directly, which can become burdensome to measure developers, service providers, and beneficiaries. As noted above, alpha testing frequently uses a sample of convenience; however, beta testing may involve measurement of a target population, which requires careful construction of samples to support adequate testing of the measure’s scientific acceptability. The analytic unit of the specific measure (e.g., physician, hospital, home health agency) determines the sampling strategy. In general, samples used for reliability and validity testing should have the following characteristics:

- Represent the full variety of entities whose performance will be measured (e.g., large and small hospitals). This is especially critical if the measured entities volunteer to participate, which limits generalizability to the full population.
- Include adequate numbers of observations to support reliability and validity analyses using the planned statistical methods. When possible, observations should be randomly selected.
- Be high-quality. Measure developers must ensure data used for risk adjustment are of high-quality. Refer to Section 3, Chapter 19.2.3, High Data Quality for specific considerations.

When determining the appropriate sample size during testing, it is necessary to evaluate the burden placed on providers and/or beneficiaries to collect the information. The PRA mandates that all federal government agencies obtain approval from the OMB before collection of information that will impose a burden on the public. However, with the passage of the MACRA, data collection for quality measure development is now exempt from PRA requirements. Measure developers should consult with the COR about ramifications of the PRA and MACRA exemption before requesting information from the public.
22.2.3 Submit the Plan and Obtain COR Approval
Submit the work plan to the COR with any necessary supporting documents. Revise as necessary to obtain the COR’s approval.

22.2.4 Implement the Plan
Following COR review and approval, execute the approved work plan.

22.2.5 Analyze the Test Results
Once all data are gathered from the test sites, the measure developer conducts a series of analyses to characterize the evaluation criteria of the measures. The findings of all testing analyses will be presented in a final summary report and discussed with the COR.

22.2.6 Refine the Measure
The measure developer may need to modify the measure specifications, data collection instructions, and calculation of measure results based on analysis of testing results.

Examples:
- Following alpha testing, measure respecification or efforts to overcome implementation barriers are often undertaken.
- Following beta testing, changes in definition of the population or adjustments to the comparison group definition may occur.
- If changes are made to the measure, consultation with the TEP is recommended prior to retesting the measure.

22.2.7 Retest the Refined Measure
Measure testing is an iterative process. Continue to refine and retest measures as deemed necessary by the measure developer and the COR.

22.3 Compile and Submit Deliverables to COR
Communicate findings of the measure testing with revised measure specifications to the COR for review. Update the MIF with revised specifications, and update the MJF with new information obtained during testing, including additional information about importance, such as variability in comparison groups and opportunities for improvement; reliability, validity, and exclusion results; risk adjustment or stratification decisions; usability findings; and feasibility findings.\(^{45}\)

Based on beta testing results, prepare a Measure Evaluation Report to summarize how well each measure meets each of the evaluation criteria and subcriteria. The updated Measure Evaluation Report can be included as part of the Measure Testing Summary Report.

22.3.1 Measure Testing Summary Report
For each measure or set of measures, complete the required summary reports and submit them to the COR. Following the analysis of information acquired during testing, the measure developer must

\(^{45}\) The NQF submission may be acceptable for this deliverable.
summarize measure testing findings. The goal of these summaries is to document sufficient evidence to support approval by COR and possible endorsement by NQF.

When reporting measure testing results, assessment of each of the four measurement criteria is a matter of degree. For example, not all revisions will require extensive reassessment for all testing criteria, and not all previously endorsed measures will be strong—or equally strong—among each set of criteria. Assessment is often a matter of judgment and expertise. In addition to clinical experts, given the difficulty of assessment, measure developers are expected to contract or employ experienced statisticians and methodologists to provide expert judgment when reporting measure reliability and validity, and summarize expert findings/consensus with respect to measure, including importance, acceptability, usability, and feasibility.

Recommendations for the content of the measure Testing Summary Report are provided. However, these recommendations are not intended to be exhaustive, and not all recommendations will apply to each measure, depending on the type of testing and characteristics of the measure.

The summary of testing may include:

- Name of measure or measure set
- Executive summary of tests and resulting recommendations
- Type of testing conducted (i.e., alpha or beta), and overview of testing scope
- Description of any deviation from the work plan along with rationale for deviation
- Data collection and management method(s):
  - Description of test population(s) and description of test sites, if applicable
  - Description of test data elements, including type and source
  - Data source description (and export/translation processes, if applicable)
  - Sampling methodology, if applicable
  - Description of exclusion, if applicable
  - Medical record review process, if applicable, including abstractor/reviewer qualifications and training, and process for adjudication of discrepancies between abstractors/reviewer
- Detailed description of measure specifications and measure score calculations
- Description of the analysis conducted, including:
  - Qualifications of analysts performing tests
  - Summary statistics (e.g., means, medians, denominators, numerators, descriptive statistics for exclusion)
  - Importance—specific analyses demonstrating importance, such as suboptimal performance for a large proportion of comparison groups and analysis of differences between comparison groups
  - Scientific acceptability
    - Reliability—description of reliability statistics and assessment of adequacy in terms of norms for the tests, and rationale for analysis approach
    - Validity—specific analyses and findings related to any changes observed relative to analyses reported during the prior assessment/endorsement process, or changes observed based on revisions to the measure; these may include assessment of adequacy in terms of norms for the tests conducted, panel consensus findings, and rationale for analysis approach
Exclusion/Exception—discussion of the rationale, which may include listing citations justifying exclusion; documentation of TEP qualitative or quantitative data review; changes from prior assessment findings such as summary statistics and analyses, which may include changes in frequency and variability statistics; and sensitivity analyses

- Analysis of the need for risk adjustment and stratification as described in Section 3, Chapter 19, Risk Adjustment

- Usability—if the measure has been materially changed, a summary of findings related to measure interpretability and methods used to provide a qualitative and quantitative usability assessment is recommended (e.g., TEP review of measure results; or, in rare situations, use of a CMS-requested focus group or survey)

- Feasibility—discussion of feasibility challenges and adjustments that were made to facilitate obtaining measure results, and description of estimated costs or burden of data collection

- Any recommended changes to the measure specifications and an assessment as to whether further testing is needed

- Detailed discussion of testing results compared to NQF requirements, including whether NQF requirements are sufficiently met or whether additional testing is required

- Examples of limitations of the alpha or beta testing:
  - Sample limited to two sites or three EHR applications
  - Sample used registry data from one state, and registry data are known to vary across states
  - Testing was formative alpha test only and was not intended to address validity and reliability.

- Recommending approval of a candidate measure for further development

- Recommending approval of a fully tested and refined measure for implementation

- Conducting comprehensive reevaluation.
23 Measure Evaluation

23.1 Measure Evaluation Criteria and Subcriteria

Measure developers should apply standardized evaluation criteria to their measure throughout the development process. The more effectively the measure properties meet evaluation criteria, the more likely the measure will be approved for use by CMS and endorsed by NQF. Measure developers should strive to identify weaknesses in the justification for their measure—through applying the evaluation criteria—and revise and strengthen the measure during development. The MJF is intended to provide information demonstrating that the evaluation criteria have been met. The form should be updated continuously with any information demonstrating strength of the measure. CMS and NQF use these criteria when evaluating a measure, although CMS may have additional criteria based on program statutes and priorities:

- **Importance** to Measure and Report, including evidence and performance gap, and priority (i.e., impact)
- **Scientific acceptability** of measure properties, including reliability and validity
- **Feasibility**
- **Usability** and Use
- Comparison to related or competing measures—harmonization.

Measure evaluation does not end when the measure is fully developed. The measure must also be continuously reevaluated during maintenance, with reports submitted at specified periods. Although there may be differing evaluation details for specific reevaluations, the general principles are the same.

The Measure Evaluation criteria descriptions in the *Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement*, the guidance from NQF on applying the criteria, and the Measure Evaluation Report form facilitate a systematic approach for applying the measure evaluation criteria, rating the strength of the measure, and tracking results. Results help the measure developer identify how to refine and strengthen the measure as it moves through the development and evaluation process. These documents function as a grading rubric, enabling measure developers to anticipate the evaluation the measure may receive when submitted to NQF and/or CMS. Although measure evaluation occurs throughout measure development, formal reports of the measure developer’s self-evaluation of the measure must be submitted to CMS as specified in the contract deliverables. The reports inform CMS of what it would take (e.g., pros/cons, costs/benefits) to increase the measure’s evaluation rating versus the risks if it is left unchanged.

23.2 Applying Measure Evaluation Criteria

Throughout measure development, the measure is evaluated to determine the degree to which the measure is consistent with the standardized evaluation criteria. The resulting evaluation information is used to determine how the measure can be modified to increase the importance, scientific acceptability, usability and use, and feasibility of the measure. Figure 39 is a diagram of the process of applying the measure evaluation criteria.
Measure evaluation criteria are applied:

- During information gathering to guide the search for appropriate measures and measure concepts
- During TEP meetings to inform TEP members and contribute to meaningful deliberation
- During testing and refinement of specifications to strengthen the measure
- During development of a testing plan and to help evaluate results derived from execution of the testing plan
- When preparing the deliverables: Measure Evaluation Report, MIF, and MJF

**Figure 39. Applying Measure Evaluation Criteria**

### 23.3 Timing of Measure Evaluation

While a formal Measure Evaluation Report is required in only three of the Measure Lifecycle phases, evaluating the measure and completing a Measure Evaluation Report may be useful during all phases of the Measure Lifecycle. If a new full report is not needed each time, an updated report will be useful so that corrections can be made or weaknesses strengthened at each point rather than waiting for the formal reporting time.

---

46 The NQF submission may be acceptable for these deliverables.
Measure Conceptualization (Section 2, Chapter 2):

- Provide the TEP with an analysis of how the measure(s) might perform by applying the measure evaluation criteria to candidate measure(s)
- Use the criteria when refining the candidate measure list (formal report required)

Measure Specification (Section 2, Chapter 3):

- Report how the measure’s proposed technical specifications function
- Evaluate how the risk model works for outcome measures

Measure Testing (Section 2, Chapter 4):

- Apply the evaluation criteria when analyzing test results
- Review updated measure specifications and justification according to the evaluation criteria (formal report required)

Measure Implementation (Section 2, Chapter 5):

- Respond (during endorsement consideration) to questions or suggestions made by the NQF Steering Committee by updating the report
- Support CMS by providing requested information on the business case during the MAP deliberations

Measure Use, Continuing Evaluation, and Maintenance (Section 2, Chapter 6):

- Apply the evaluation criteria during comprehensive reevaluation to review performance (formal report required)
- Update measure specifications and justification based on the evaluation.

It is important to evaluate the measure as objectively as possible, to anticipate any issues that may arise when the measure is submitted to NQF for endorsement. The measure developer communicates any anticipated risks associated with endorsement and presents plans to strengthen any weaknesses identified to CMS using the Measure Evaluation Report. For example, if the measure’s feasibility is difficult to test broadly with actual patient or facility data, then it should be evaluated through pilot testing. Testing results are reported in the Measure Evaluation Report. It is important for CMS to be fully informed of the pros/cons and costs/benefits for improving the rating, and the risks if the weaknesses cannot be corrected.

The COR will work with the measure developer to identify points that are appropriate to the specific measure to conduct a formal measure evaluation. The Measure Evaluation Report can be modified as appropriate for specific types of measures such as eCQMs, composite measures, and cost and resource use measures.

23.4 Testing and Measure Evaluation Criteria

Results of measure testing are used to demonstrate a measure’s alignment with most of the measure evaluation criteria. Because testing is often an iterative process, both alpha and beta testing findings may provide information that address measure evaluation criteria:

- Alpha testing often supplies information that demonstrates the feasibility of the measure’s implementation.
• The findings from one or more beta tests are often used to demonstrate scientific acceptability and usability, as well as augment previously obtained information on the importance and feasibility of the measure.

Refer to Section 3, Chapter 22, Measure Testing, and Section 3, Chapter 24, Testing for Special Types of Measures for more information on measure testing.

Application of the testing results to each of the four measurement areas—importance, scientific acceptability, usability, and feasibility—is discussed.

23.4.1 Importance

Information from testing often provides additional empirical evidence to support prior judgments of a measure’s importance generated earlier during the measure development process. In particular, beta testing results may reveal that a measure assesses an area with substantial opportunities for improvement. Testing can also uncover that the measure addresses a high-impact or meaningful aspect of healthcare. Examples of empirical evidence for importance or improvement opportunities derived from testing data include:

• Quantifying the frequency or cost of measured events to demonstrate that rare or low-cost events are not being measured
• Identifying substantial variation among comparison groups or suboptimal performance for a large proportion of the groups
• Demonstrating that methods for scoring and analysis of the measure allow for identification of statistically significant and practically/clinically meaningful differences in performance
• Showing disparities in care related to race, ethnicity, gender, income, or other classifiers
• Identifying evidence that a measure is associated with consistent delivery of effective processes or access that lead to improved outcomes.

Reported data to support the importance of a measure may include:

• Descriptive statistics such as means, medians, standard deviations, confidence intervals for proportions, and percentiles to demonstrate the existence of gaps or disparities
• Analyses to quantify the amount of variation due to comparison groups such as rural versus urban through $R^2$ or intraclass correlation.

23.4.2 Scientific Acceptability

With respect to CMS and NQF review for endorsement, scientific acceptability of a measure refers to the extent to which the measure produces reliable and valid results about the intended area of measurement. These qualities determine whether the measure can be used to draw reasonable conclusions about care in a given domain. Because many measure scores are composed of patient-level data elements (e.g., blood pressure, laboratory values, medication, surgical procedures) that are aggregated at the comparison group level (e.g., hospital, nursing home, physician), evidence of reliability and validity is often needed for both the measure score and measure elements, and the measure developer should ensure both are addressed. Following are some examples of common measure testing and reporting errors:

• Reporting limited to descriptive statistics. Descriptive statistics demonstrate that data are available and can be analyzed but do not provide evidence of reliability or validity.
• **Lack of testing of respecified measures.** When respecifying a measure (e.g., using similar process criteria for a different population or denominator), the newly respecified measures still require testing to obtain empirical evidence of reliability and validity.

• **Inadequate evidence of scientific acceptability for commonly used measure elements.** Measure elements (e.g., diagnosis codes, EHR fields) that are in common use still require testing or evidence of reliability and validity within the context of the new measure specifications (e.g., new population, new setting).

• **Inadequate analysis or use of clinical guidelines for justifying an exclusion.** Analyses and/or clinical guidelines justifying an exclusion or demonstrating reliability should be reported for different methods of data collection.

Since reliability and validity are not all-or-none properties, many issues may need to be addressed to supply adequate evidence of scientific acceptability. *However, the complexity of different healthcare environments, data sources, and sampling constraints often preclude ideal testing conditions. As such, judgments about a measure’s acceptability are often a matter of degree.* Therefore, determination of adequate measure reliability and validity is always based on review of the testing data by qualified experts. It is assumed that a measure developer will contract or employ experienced methodologists, statisticians, and SMEs to select testing that is appropriate and feasible for the measure(s) under consideration and ensure demonstration of measure reliability and validity.

Although not replacing the expert judgment of the measure development team, the next subsections describe the general considerations for evaluating reliability and validity of both a measure score and its component elements.

### 23.4.2.1 Reliability

**Reliability testing** demonstrates that measure results are repeatable and the measurement error is acceptable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

#### 23.4.2.1.1 Types of Reliability

Depending on the complexity of the measure specifications, one or more types of reliability may need to be assessed. Several general classes of reliability testing are described in the following paragraphs.

**Inter-rater (i.e., inter-abstractor) reliability.** This test assesses the extent to which ratings from two or more observers are congruent with each other when rating the same information, often using the same methods or instruments. It is often employed to assess reliability of data elements used in exclusion specifications, as well as the calculation of measure scores when review or abstraction is required by the measure. The extent of inter-rater/abstractor reliability can be quantitatively summarized, and concordance rates and Cohen’s *Kappa* with confidence intervals are acceptable statistics to describe inter-rater/abstractor reliability. However, additional reliability testing is needed if using inter-rater/abstractor reliability. More recent analytic approaches are also available that involve calculation of intraclass correlations for ratings on a scale, where variation between raters is quantified for raters randomly selected to rate each occurrence.

**eCQMs** that are implemented as direct queries to EHR databases may not use abstraction. Therefore, inter-rater reliability may not be needed for eCQMs.

**Form equivalence reliability.** This test is sometimes called parallel-forms reliability and assesses the extent to which multiple formats or versions of a test yield the same results. It is often used when testing comparability of results across more than one method of data collection, across automated data
extraction from different data sources, or agreement between the known values from a simulated data set and the elements obtained when the specifications are applied to the data set. It may be quantified using a coefficient of equivalence, where a correlation between the forms is calculated. As part of the analysis, reasons for discrepancies between methods (i.e., mode effects) should also be investigated and documented (e.g., when the results from a telephone survey are different from the results when the same survey is mailed).

**Test-retest reliability.** This test is sometimes called temporal reliability and assesses the extent to which a measurement instrument elicits the same response from the same respondent across two measurement time periods. The coefficient of stability may be used to quantify the association for the two measurement occasions. It is generally used when assessing information that is not expected to change over a short or medium interval of time. Test-retest reliability is not appropriate for repeated measurement of disease symptoms and is not appropriate for measuring intermediate outcomes that follow an expected trajectory for improvement or deterioration. Test-retest reliability should be assessed when there is a rationale for expecting stability—rather than change—over the time period.

**Internal consistency reliability.** Testing of a multiple item test or survey assesses the extent that the items designed to measure a given construct are inter-correlated. Cronbach’s alpha has been used to evaluate internal consistency reliability for several decades (Cronbach, 1951). It is often used when developing multiple survey items that assess a single construct. Other internal consistency analysis approaches may involve the use of exploratory or confirmatory factor analysis.

**Other approaches to reliability.** Across each type of reliability estimation described above, the shared objective is to ensure replication of measurements or decisions. In terms of comparisons of groups, reliability can be extended to assess stability of the relative positions of different groups or determination of significant differences between groups. These types of assessments address the proportion of variation in the measure attributable to the group. This proportion can also be described as true differences (or “signal”) relative to variation in the measure due to other factors, including chance variation (or “noise”). Measures with a relatively high proportion of signal variance are considered reliable because of their power for discriminating among providers and the repeatability of group-level differences across samples. Provided that the number of observations within groups is sufficiently large, these questions can be partially addressed using methods such as analysis of variance (ANOVA), calculation of intraclass correlation coefficients (ICC), estimation of variance components within a hierarchical mixed (i.e., random-effects) model, or bootstrapping simulations. Changes in group ranking across multiple measurements may also add to an understanding of the stability of group-level measurement.

23.4.2.1.2 Measure Data Elements versus Measure Score

Because many measures are composed of multiple data elements, reliability testing ideally applies to both the data elements comprising the measure and the computed measure score. However, for measures that rely on many data elements, testing of the individual data elements is sometimes only conducted for critical elements that contribute most to the computed measure score, rather than all the data elements. Similarly, commonly used data elements for which reliability can be assumed (e.g., gender, age, date of admission) are also occasionally excluded from reliability testing, although some mistakes can happen there as well. NQF does not require data element reliability testing if data element validity is demonstrated.
Flexibility in the reliability testing of data elements contrasts with assessment of the measure score. The measure score under development should always be assessed for reliability using data derived from testing.

23.4.2.2 Validity

In measure development, the term validity has a specific application known as test validity, which refers to the degree to which evidence, clinical judgment, and theory support interpretations of a measure score. Stated more simply, test validity is empirically demonstrated and indicates the ability of a measure to record or quantify what it purports to measure; it represents the intersection of intent (i.e., what is being assessed) and process (i.e., how it is assessed).

23.4.2.2.1 Types of Validity

Validity testing of a measure score can be assessed in many ways. Although some view all types of validity as a special case of construct validity, researchers commonly reference the types of validity separately: construct validity, discriminant validity, predictive validity, convergent validity, criterion validity, and face validity (Messick, 1994). A description of each type is shown in the following paragraphs.

Construct validity. This type refers to the extent to which the measure quantifies what the theory says it should. Construct validity evidence often involves empirical and theoretical support for the interpretation of the construct. Evidence may include statistical analyses such as confirmatory factor analysis of measure elements to ensure they cohere and represent a single construct.

Discriminant validity/contrasted groups. This type examines the degree to which a test of a concept is not highly correlated with other tests designed to measure theoretically different concepts. It may also be demonstrated by assessing variation across multiple comparison groups (e.g., healthcare providers) to show that the measure can differentiate between disparate groups that it should theoretically be able to distinguish.

Predictive validity. This type refers to the ability of measure scores to predict scores of other related measures at some point in the future, particularly if these scores predict a subsequent patient-level outcome of undisputed importance, such as death or permanent disability. Predictive validity also refers to scores on the same measure for other groups at the same point in time.

Convergent validity. This type refers to the degree to which multiple measures/indicators of a single underlying concept are interrelated. Examples include measurement of the correlations between a measure score and other indicators of processes related to the target outcome.

Reference strategy/Criterion validity. This type refers to verification of data elements against some reference criterion determined to be valid (i.e., the gold standard). Examples include verification of data elements obtained through automated search strategies of EHRs compared against manual review of the same medical records (i.e., the gold standard).

Face validity. This type is the extent to which a measure appears to reflect that which it is supposed to measure “at face value.” It is a subjective assessment by experts about whether the measure reflects what it is intended to assess. Face validity for a CMS quality measure may be adequate if accomplished through a systematic and transparent process, by a panel of identified experts, where formal rating of the validity is recorded and appropriately aggregated. The expert panel should explicitly address whether measure scores provide an accurate reflection of quality, and whether they can be used to distinguish between good and poor quality. Because of the subjective nature of evaluating the face
validity of a measure, special care should be taken to standardize and document the process used. NQF has recommended in Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties that a formal consensus process be used for the review of face validity such as a modified Delphi approach in which participants systematically rate their agreement, and formal aggregating and consensus failure processes are followed. NQF in Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement allows the use of face validity in lieu of empirical testing for new measures if a systematic assessment is performed and targeted to reflect the accuracy of the targeted care measured. Since this is the weakest form of validity testing, the recommendation is that the experts involved in the measure development should be different from the ones who perform face validity. This type of formal process can also be used when addressing whether specifications of the measure are consistent with medical evidence. Empirical validity testing is required at maintenance review. Justification is necessary if empirical validity testing is not possible.

23.4.2.2 Measure Data Elements versus Performance Measure Score

Patient-level data elements are the building blocks for a performance measure and should be assessed for reliability and validity. Although patient-level data elements are important, it is the computed measure scores that are used to draw conclusions about the targeted aspect of care. NQF in Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement will accept data element and/or measure score validity testing. Performance score validity testing is needed for instrument-based and composite measures.

Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Some examples of validity testing using comparative analysis measure data elements include:

- Claims data that have codes used to represent primary clinical data (e.g., ICD, CPT) can be compared to manual abstraction from a sample of patient medical records.
- Standardized patient assessment instrument information (e.g., minimum data set [MDS], OASIS, registry data) that is not abstracted, coded, or transcribed can be compared with “expert” assessor evaluation (conducted at approximately the same time) for a sample of patients.
- EHR information extracted using automated processes based on measure technical specifications can be compared to manual abstraction of the entire EHR.

23.4.2.3 Prior Evidence of Reliability and Validity for Measure Elements

According to NQF’s Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties, when prior evidence of reliability or validity of the data elements comprising the measure exists, it can sometimes be used in place of updated or additional testing of the measure’s data elements. In contrast to a measure’s data elements, although prior evidence can augment findings for a calculated measure score under development, prior evidence cannot be used to demonstrate score-level reliability or validity. Commonly used measure elements should always be assessed for reliability and validity within the context of the new measure specifications using data derived from the beta test. Data from prior validity or reliability testing of data elements from the same data source may be used to calculate the measure score or computed measure score since the two concepts are both mathematically and conceptually related. Prior evidence of reliability or validity testing may include published or unpublished testing results of same data elements, same data type, and/or a representative sample of sufficient size.
NQF states in their Guidance:

- Prior evidence of validity of data elements can be used if the measure under development uses the same data elements and data type and obtains a representative sample of sufficient size.
- Separate reliability testing of the data elements is not required if validity testing was conducted on the data elements. If patient scores from an instrument/scale are used in the measure under development, usually the reliability of the scale has been tested and documented and can be used as evidence of data element reliability. If validity testing of the data elements was not conducted, prior evidence of reliability of data elements can be used.

Refer to the NQF Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement for more information and guidance on validity testing.

23.4.2.4 Testing of Exclusions/Exceptions

The review of measure exclusions and exceptions should be based on the testing data and should include, at a minimum:

- Evidence of sufficient frequency of occurrence of the exclusions/exceptions
- Evidence that measure elements (e.g., codes) used to identify exclusions/exceptions are valid.

Review may also include evidence that measure results are distorted without the exclusions/exceptions. For example, evidence that an exclusion distorts a measure may include variability of an exclusion across comparison groups and sensitivity analyses of the measure score with and without the exclusion.

Additional review is required when patient preference or other individual clinical judgment based on unique patient conditions is allowed as an exception category. Analyze whether the exception will make a major change to the measure results. Consider whether patient preference represents a clinical exception to eligibility or if it can be influenced by provider intervention. These measures should always be evaluated with and without the exception, and the proportion of exception should be included for any group-level tabulations.

23.4.2.5 Risk Adjustment and Stratification

Beta testing should be used to evaluate an evidence-based risk adjustment strategy when the measure being developed is an outcome measure. Risk adjustment typically is not needed for process measures.

Empirical evidence must be provided for the adequacy of risk adjustment or rationale that risk adjustment is not necessary to ensure fair comparisons.

Information should include the analytic methods used and evidence of meaningful differences. If stratification is used, the stratification results should be included. More information about stratification is provided in Section 3, Chapter 19, Risk Adjustment.

23.4.3 Usability

Formal usability testing may not be required, and a review of measure characteristics (e.g., descriptive statistics, dispersion of comparison groups) may be conducted by the TEP to determine usability of the measure for performance improvement and decision-making. When more formal testing is required by CMS to assess the understandability and decision-making utility of the measure with respect to intended audiences (e.g., consumers, purchasers, providers, policy makers), a variety of methods are available:
• Focus groups
• Structured interviews
• Surveys of potential users.

These different methods often focus on the discriminatory ability of the measure and the meaning of the score as applied to evaluation of comparison groups or decision-making. For example, a survey of potential users may be used to rate the clinical meaningfulness of the performance differences detectable by the measure or to assess the congruence of decisions based on measure summary data from a sample.

23.4.4 Feasibility

Testing can be used to assess measure feasibility and determine the extent to which the required data are available and retrievable without undue burden, and the extent to which they can be implemented for performance measurement. Some feasibility information may be obtained when assessing the validity of the measure score or measure elements (e.g., quantifying the frequency of absent diagnosis codes when a target condition is present). Other feasibility information can be obtained using systematic surveys (e.g., survey of physician practices tasked with extracting the information). More in-depth information may be gathered by conducting focus groups composed of professionals who may be responsible for a measure’s implementation.

Feasibility assessments should address:

• Availability of data (e.g., evidence that required data, including any exclusion criteria, are routinely generated and used in care delivery)
• Extent of missing data, measure susceptibility to inaccuracies, and the ability to audit data to detect problems
• Estimate of the costs or burden of data collection and analysis
• Barriers encountered in implementing performance measure specifications, data abstraction, measure calculation, or performance reporting
• Ability to collect information without violation of patient confidentiality, including circumstances where measures based on patient surveys or the small number of patients may compromise confidentiality
• Identification of unintended consequences.

 العاصفة eCQMs have a separate feasibility scorecard assessing data availability, data accuracy, data standards, and workflow.

23.5 Evaluation during Measure Maintenance

As they did during measure development, the measure developers, TEP members, and other stakeholders involved in measure maintenance work toward ensuring sound measures that can be used to drive healthcare quality improvement and inform consumer choice. During measure maintenance, the measure developer must continue to evaluate the measures and provide strong evidence that the measures are constructed in a sound manner, and continue to add value to quality reporting programs. These two steps help CMS ensure that its measures retain NQF endorsement.

23.5.1 Apply Measure Evaluation Criteria

Each measure undergoes an update at least annually and a rigorous, comprehensive reevaluation every 3 years to assess its continued value, based on the most current set of measure evaluation criteria.
Evaluation during maintenance should also document how the measure is performing compared to the trajectory that was projected in the business case during measure development. Throughout the measure evaluation process, developers update justification for the measure and any changes to the technical specifications to demonstrate:

- Aspects of care included in the specifications continue to be highly important to measure and report, supply meaningful information to consumers and healthcare providers, and drive significant improvements in healthcare quality and health outcomes.
- Data elements, codes, and parameters included in the specifications are the best ones to use to quantify the specific measure, and data collection still does not cause undue burden on resources.
- Calculations included in the specifications represent a clear and accurate reflection of the variation in the health outcome of interest, or the quality or efficiency of the care delivered.

### 23.5.2 Report Results of Evaluation

Measure Evaluation Criteria and subcriteria are detailed in the Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement. A blank Measure Evaluation Report is found with the Blueprint forms. It is important during maintenance to document how the measure is performing. Submit a separate Measure Evaluation Report for each measure to CMS during maintenance in the following situations:

- When recommending disposition of the measure after a comprehensive reevaluation
- When recommending disposition of the measure after an ad hoc review.

When completing the Measure Evaluation Report during maintenance, the current rating of each subcriterion should be compared to the prior measure evaluation.

The prior Measure Evaluation Report may have been prepared during measure development or during the last maintenance review.
24  TESTING FOR SPECIAL TYPES OF MEASURES

24.1  RESPECIFIED MEASURES
When respecifying a measure for use in a new domain (e.g., new setting or population) or using a different data source (e.g., EHR data), construct the measure testing to detect important changes in the functionality or properties of the measure. As applicable, review changes in:

- Relative frequency of critical conditions used in the original measure specifications when applied to a new setting/population (e.g., dramatic increase in the occurrence of exclusionary conditions).
- Importance of the original measure in a new setting (e.g., an original measure addressing a highly prevalent condition may not show the same prevalence in a new setting, or evidence that large disparities or suboptimal care found using the original measure may not exist in the new setting/population).
- Location of data or the likelihood that data are missing (e.g., an original measure that uses an administrative data source for medications in the criteria specification, when applied to Medicare patients in an inpatient setting, may need to be modified to use medical record abstraction because Medicare Part A claims do not contain medication information due to bundling).
- Frequency of codes observed in stratified groups when the measure is applied to a new setting or subpopulation.
- Risk adjustment model or changes that make the previous risk adjustment model inappropriate in the new setting/population.

24.2  eCQM TESTING
When evaluating an eCQM’s readiness for implementation and adoption, eCQM testing assesses the extent to which an eCQM meets the measure properties of feasibility, validity, and reliability. Testing measure properties is an iterative process, with the purpose of refining and revising the eCQM until all quality issues are resolved. The goal is to produce a reliable and valid eCQM ready for implementation. eCQM testing is possible once the eCQM specification is completed in the MAT, and the eCQM package has been exported and provided to the testing team.

Early feasibility testing is recommended prior to electronic specification in the MAT to test the reasonableness of collecting expected data elements during common workflow practice, and determining whether data elements are captured within an EHR system. Post-MAT, validity and reliability are tested to confirm that the electronically specified measure has achieved its intended purpose; the measure produces consistent, repeatable results; and the logic is not ambiguous.

24.2.1  Types of eCQM Testing
As EHR systems become more generally available and more integrated, additional documented clinical information may also become widely available for measure use. However, a multitude of EHR systems are in use today—particularly in the ambulatory care setting—and this diversity must be managed when measure specifications are developed for use across EHR systems. To address this issue, CMS requires new measures (or measures respecified for EHRs) to be specified using HQMF, which is a standard for representing a CQM for use with an electronic data source.
In alignment with this format, measure developers are expected to author eCQMs in the MAT and specify measures using the QDM. The use of the MAT, CQL, and QDM promote measures that are standard based, consistent, reliable, and valid when extracted across diverse, certified EHR systems. However, standards also raise new considerations when testing measures that include EHR specification accuracy, EHR validity testing, measure score and data element testing, testing of respecified measures, and feasibility testing. The different types of testing uncover different information about the extent of feasibility, reliability, and validity of the measure properties. Testing identifies ambiguities in the measure logic, potential barriers to implementation, and reasonableness of the data elements specified in the measure.

24.2.2 Feasibility

The feasibility assessment may include discussions with SMEs such as vendors and implementers of EHR systems and evaluation of how data are captured in an active clinical setting. The measure developer must assess feasibility of the measure concept at the time the measure is conceived and definitely prior to drafting initial eCQM specifications to ensure that the data elements are available in a usable structured format and can be coded using standard terminologies within the EHR. This process is critical to ensure that a developed measure passes feasibility assessments during beta (i.e., field) testing and to avoid re-expressing measure concepts or replacing the measure after a considerable amount of work has been completed.

In addition to information obtained from SMEs, empirical analysis can also be used to test the feasibility of data elements required for a measure. Feasibility considerations include:

- Data availability (including standardization)
- Accuracy of the information in the data
- Maturity of standards
- Standard terminologies
- Extent to which data are collected and encoded where necessary as part of the normal workflow and the measure specifications and calculation logic.

When testing feasibility, it is important to understand the intent of the measure, because the intent can influence which data must be collected. General information on feasibility assessment is provided in the feasibility subsection of Section 3, Chapter 22, Measure Testing.

Feasibility is more than a demonstration by an EHR vendor of the system’s ability to capture a data element. Feasibility testing evaluates the reasonableness of collecting the expected data elements during a typical clinical workflow in an EHR system, evaluates the burden on clinicians, and determines whether the data elements are captured by the system. When developing the feasibility testing plan, careful consideration should be made in determining the threshold for feasibility. Refer to the Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement and the NQF eCQM Feasibility Scorecard for more information and guidance. NQF requires the Feasibility Scorecard for endorsement of eCQMs.

24.2.3 Validity

Validity testing for the eCQM confirms the intent of the measure, ensures eCQM logic is not ambiguous and expected test patients fall in the correct populations, determines whether data elements are aligned with national standards, and checks calculated scores from automated extraction for accuracy. These endpoints are in addition to the validity testing criteria for other types of measures, which evaluate whether the measure assesses what it purports to measure.
Ideally, CEHRT will use clinical information recorded in discrete computer-readable fields, which potentially reduces errors in measure elements arising from manual abstraction or coding errors. However, even under these circumstances, measures need to be evaluated during measure testing. Some examples that can affect validity include:

- Complex specifications may make a measure more susceptible to varying data field interpretation by different users.
- Users may enter information into EHR fields other than those from which the vendor extracts data for measure reporting.
- Even small errors in the measure specifications, such as omission of codes for commonly documented concepts in value sets, can reduce the capture of appropriate patients in the measure’s denominator.

Measures originally specified using data sources other than an EHR (i.e., chart abstraction or claims data) can be respecified for use with EHRs. However, even if these measures were previously approved by CMS and show adequate reliability and validity in the original measure, the eCQM should be assessed for reliability and validity.

A subjective evaluation of the human-readable document of the eCQM should be conducted to confirm that the intent of the measure is unchanged. An example of a subjective evaluation includes confirmation by the steward for a respecified measure that the eCQM preserves the intent of the original paper or claims-based measure equivalent “at face value.” A subjective evaluation for a de novo measure includes confirmation by a clinical working group or TEP that eCQM concepts reflect the intent. Measure level (i.e., face) validity testing may involve iterative discussions with the measure steward or clinical working group/TEP to ensure the original intent of the measure concept is maintained in the eCQM.

Refer to the NQF Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement for more information and guidance on validity testing.

**24.2.3.1 Measure Logic Validity**

Perform an objective evaluation of measure logic to confirm whether the measure can correctly identify patients intended to be included in or excluded from the numerator, denominator, and other relevant populations of the eCQM. The test aims to ensure that the logic of the eCQM is expressed without ambiguity, so the same patients are categorized by the relevant patient populations. Testing may identify potential differences in the interpretation of measure logic encoded in the eCQM. Bonnie (refer to Section 3, Chapter 7, eCQM Standards-Based Guidance and Tools) is a measure testing tool that provides feedback on the behavior of eCQM logic.

**24.2.3.2 Data Element Validity**

Conduct an objective evaluation of whether data elements electronically extracted from an EHR are comparable to similar data elements visually abstracted by the reviewers. The vocabulary file containing the relevant value sets is the baseline for the automatic extraction. This testing method applies to respecified and de novo measures.

Data elements from test site EHRs will be collected through electronic extraction and compared to a gold standard EHR abstract to assess validity of the electronic extraction. This comparison will be performed to determine whether the eCQM provides the same results for numerator inclusion/exclusion and denominator inclusions as the reviewers. Where discrepancies are identified, the visual
review of the manually abstracted data elements will be presumed correct, serving as the “gold standard.”

This design is guided by the rationale that electronic extraction of EHR data cannot detect values entered as free text as opposed to structured data, while visual review will usually capture both free text and structured data and, therefore, would be more complete and accurate. Data elements demonstrating a pattern of disagreement between the results from visual abstraction and electronic extraction may arise either because some data required for the measure are documented in the EHR in a format that the electronic extraction did not capture, or there are problems with the way the eCQM query was written.

For measure data elements, demonstration of validity is considered adequate if either:

- Adequate agreement is observed between data elements electronically extracted and data elements manually abstracted from the EHR
- Complete agreement is observed between the known values from a simulated QDM-compliant data set and the elements obtained when the eCQM specifications are applied to the data set.

NQF guidance further clarifies that reliance on data from structured data fields is expected. Otherwise, unstructured data must be shown to be both reliable and valid.

24.2.3.3 Standards Conformance Validation

To help ensure accuracy of data elements, measure developers are expected to validate the content of the XML, which is often achieved using three types of validation:

- Syntactic validation—This method of accuracy validation ensures that the XML content follows (i.e., conforms to) specific constraints required by the HL7 HQMF Standard and the XML patterns based on the QDM. These quality-checking processes are built into the MAT application. The MAT uses the CQL-to-ELM Translator for validation of syntactically correct CQL content. The Translator provides validation of CQL expressions based on the CQL grammar files, which are part of the HL7 CQL standard. If no syntax errors exist in a CQL file, the Translator converts the CQL file into the respective ELM XML and JSON content based on the ELM XML Schema.

To help a measure developer avoid certain pitfalls that would violate conformance requirements in CQL-based HQMF, the MAT provides various levels of validation within the tool to help guide users in creating syntactically correct CQL-based HQMF before they package their eCQMs. These various preventions that are built into MAT include, but are not limited to:

- Provision of correct model (i.e., QDM) and version within the CQL workspace CQL Syntax Error and Warning checking with highlighting. Note: MAT does not allow a user to package a measure if CQL Errors are present.
- Provision of default CQL expressions: “Measurement Period” parameter and the four CMS Supplemental Data Element definitions (i.e., Ethnicity, Race, ONC Administrative Sex, and Payer). Based on the CQL-based HQMF IG, these expressions must meet certain requirements.
- Duplicate identifier checking. Note: No two library-level identifiers (e.g., . definition names, function names, local identifiers for codes and value sets) may have the same name within a library.
Filtering of definitions for population workspace based on the user-provided patient-based indicator for the measure.  
Character checks for library-level identifiers, function arguments, etc., to ensure that users are providing the correct form of identifiers based on the CQL-based HQMF IG.  
Population grouping help/validation to ensure that users may only use correct type and number of populations within a measure group based on the user-provided measure score (i.e., Cohort, CV, Proportion, or Ratio).  
Expression Return Type validation.  
- HL7 ISO-based Schematron—This method is a possible mechanism for validating XML that is written outside the MAT; however, it may not include all components that are now built into the MAT. Additional resources for information on the ISO Schematron, including technical specifications, may be found at the ISO website.  
- Narrative validation—The MAT output includes a human-readable document that can be viewed in a standard web browser in HTML. When viewed in a web browser, the measure author can assess the extent to which the machine-generated criteria correctly reflect the original measure criteria under development. When the measure author validates correctness of the human-readable format, this is considered narrative validation.

### 24.2.4 Reliability

Testing for reliability involves experts assessing the human-readable format (i.e., HTML) of the eCQM for clarity and alignment to standard specifications. A reliable measure is reproducible and can be implemented consistently within and across organizations. Reliability allows for comparability of results. Three ways of testing reliability of an eCQM are to evaluate the measure for clarity, logic ambiguity, and data element alignment with standard specifications that support consistent implementations. This testing is in addition to and does not replace statistical reliability testing.

### 24.2.5 Testing Multiple Sites and Multiple EHRs

Testing multiple sites for feasibility, validity, and reliability is important to address potential variability in reporting based on differences in local workflow process. Even multiple sites using the same EHR vendor product may show different results since the local workflow may vary, and data may not be consistently entered into the fields expected by the vendor. Variances in results from such testing at multiple sites should be evaluated to determine whether changes are needed in the measure logic or definition. Testing must encompass at least two EHR products.

### 24.3 Composite Measures

A composite measure is a combination of two or more individual measures, each of which reflects quality of care, into a single measure that results in a single score. The use of composite measures creates unique issues associated with measure testing. The NQF criteria for endorsement of composite measures are outlined in the Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement. Chapters 24.3.1–24.3.4 provide recommendations for testing a composite measure in support of submission to CMS for approval and NQF for endorsement.

#### 24.3.1 Component and Composite Reliability and Validity Testing

Demonstration of reliability and validity is recommended for the composite and the components of the composite. However, demonstration of the reliability of the individual components is insufficient. It is possible for individual components to contribute to the reliability of the composite without being
Independently reliable. Validity must be empirically demonstrated for the composite score. Much like validity testing for single measures, validity testing for the composite should also include reporting overall frequency of missing data and distribution across providers. It is ideal to report the effect of alternative rules for handling missing data and the rationale for the approach that was selected. Discuss the pros and cons of the approaches and the rationale for the rules that were selected.

### 24.3.2 Component Coherence

Per NQF’s Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement, testing is recommended to determine whether components of a composite measure adequately support the goals articulated in the constructs for the measure. In addition, reliability of the components can be tested using correlation analyses or confirmatory factor analysis methods. If components are coherent, the component items meet the intent of the measure construct.

### 24.3.3 Composite-Specific Testing

Components of a composite measure should support the overall goal of the measure. If components are correlated, testing analysis should be based on shared variance such as factor analysis, Cronbach’s alpha, item-total correlation, and mean inter-item correlation. If components are not correlated, testing should demonstrate the contribution of each component to the composite score.

Examples:

- Change in a reliability statistic such as ICC, with and without the component measure
- Change in validity analyses, with and without the component measure
- Magnitude of regression coefficient in multiple regression with composite score as a dependent variable
- Clinical justification demonstrating correlation of the individual component measures to a common outcome measure.

### 24.3.4 Appropriateness of Aggregation Methods

When aggregating components for a composite measure to explain an outcome, measure developers should identify the method used to estimate the composite score and test the validity of the score. Once a score is obtained, present the results with justification of the methods used to estimate the composite score because the method selected for combining components may influence interpretation of a composite measure result.

#### 24.3.4.1 Selecting Appropriate Method to Test for Composite Validity

Testing should include an examination of the appropriateness of the method(s) used to combine the components into an aggregate composite score. For example, the testing (i.e., assessment) of a weighting methodology for process measures may include examining the adequacy of all-or-none, any-or-none, if/then, or opportunity scoring approaches used to create the composite. For a composite outcome that uses differential weighting of the components, the documented support for the weighting methodology might include a regression of a “gold standard” outcome upon the components. When a linear combination is used to create a composite, the components of the composite should be assessed for their contribution to the validity of the overall composite score. Linear combination alone does not imply equal or differential weighting or the appropriateness of retained components within a composite score.
### 24.3.4.2 Justification of Methodology Selected

Regardless of whether the components are combined with equal or unequal weighting, the composite development methodology needs to include a justification for why each contributing component is included, or “retained,” in the composite. Developers should provide specific explanations for the decisions surrounding both weighting and component retention. In addition, assessment methods should include a description of how the composite’s components relate to one another regarding the decisions on component retention and weighting.

If most of the composite’s variation is the result of only a subset of the components used for the composite, also provide information (e.g., a table) on the contribution of each of the components to the composite (e.g., regression coefficients or factor loadings) to address which subset of components is contributing to the majority of the aggregate’s variation. The variation (i.e., information content) of a composite might be conveyed in a variety of ways, such as through reporting of regression results, factor loadings, and percentages of shared variation explained from a principal components analysis.

The results of the composite evaluation process might not be well aligned with the separate results for each of the components in the composite measure, as the composite may primarily reflect a minority of the components of the composite. For example, group differences on an emergency room composite measure may be largely determined by ED wait times because variability for this component may be large relative to the variability of all remaining composite components. This issue can be resolved by providing tables showing the weights or loading for each composite such that a reader can determine the impact of differential weighting on the meaning of the overall composite measure.

Information should also be provided for variable or component-within-composite retention decisions. For example, when using a stepwise regression model, one often selects the default values for entering and removing variables (i.e., for entry, \( p < 0.05 \); for removal, \( p < 0.10 \)). When using composites created through principal component analysis or other factor analytic models, a table should show the item loadings (i.e., a type of weighting) and contain a note if other inclusion or exclusion criteria were used.

The appropriateness of methods to address component missing data when creating the composite score should also be assessed. This analysis of missing component scores should support the specifications for scoring and handling missing component scores.

Examples of resources for methodology:


---

47 This is not an exhaustive list, nor is it making a judgment that these sources are the best sources.
24.3.5 **Feasibility** and **Usability** of Composite Components

Measure testing may also demonstrate that the measure can be consistently implemented across organizations by quantifying comparable variation for individual components, that the measure can be deconstructed into its components at the group/organization level to facilitate transparency, and that the measure can be understood by the intended measure audience.
25 EVALUATION FOR SPECIAL TYPES OF MEASURES

Certain types of measures require additional considerations when applying the Measure Evaluation criteria. The criteria for these special types are included in Measure Evaluation Criteria and Instructions with the other criteria descriptions and guidance, when applicable.

25.1 EVALUATING COMPOSITE MEASURES

NQF’s Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement defines a composite performance measure as a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score. There are two primary types of composite performance measures:

- Measures of two or more individual performance areas scored using an algorithm that produces as single score as its only output. With this type of composite, the individual components cannot produce individual scores.
- Measures with two or more individual component measures assessed separately and then aggregated into one score. Component elements of this type of composite stand alone, but their combination produces a richer representation of the target construct (e.g., pain management).

Single performance measures—even if the data are patient scores from a scale or tool with more than one item—are not composites. NQF endorses measures, not the tools from which a score is derived. Measures with multiple linked steps in a care process are also not considered composites. Measures that combine information from other factors for risk adjustment are not composites.

There are unique issues associated with composite approach that require additional evaluation. The validity of the component measures, the appropriateness of the methods for scoring/aggregating and weighting the components, and interpretation of the composite score all require evaluation. Both the composite and its component measures need to be evaluated to determine the suitability of the composite measure. The measure evaluation criteria and subcriteria include special considerations to be used when evaluating composite measures.

25.1.1 Considerations for Evaluating Composite Measures

Information from the Composite Performance Measure Evaluation Guidance (NQF, 2013b) describes NQF’s approach to evaluation.

A coherent quality construct and rationale for the composite performance measure are essential for determining:

- What components are included in a composite performance measure
- How components are aggregated and weighted
- What analyses should be used to support components and demonstrate reliability and validity
- Added value over that of individual measures alone.

Reliability and validity of the individual components do not guarantee reliability and validity of the constructed composite performance measure. Reliability and validity of the constructed composite performance measure should be demonstrated. Consider:
• When evaluating composite performance measures, both the quality construct itself and the empirical evidence for the composite (i.e., supporting the method of construction and methods of analysis) should be considered.

• Each component of a composite performance measure should provide added value to the composite as a whole—either empirically (because it contributes to the validity or reliability of the overall score) or conceptually (for evidence-based theoretical reasons). Choose the smallest set of component measures possible. However, including measures from all necessary performance domains may be conceptually preferable to eliminating measures because they do not contribute as much statistically.

• Individual components in a composite performance measure may or may not be correlated, depending on the quality construct.

• Aggregation and weighting rules for constructing composite performance measures should be consistent with the quality construct and rationale for the composite. A related objective is methodological simplicity. However, complex aggregation and weighting rules may improve the reliability and validity of a composite performance measure, relative to simpler aggregation and weighting rules.

• Standard NQF measure evaluation criteria apply to composite performance measures.

• NQF only endorses performance measures that are intended for use in both performance improvement and accountability applications.

25.2 Evaluating Cost and Resource Use Measures

The resource use measure evaluation criteria are grounded in the standard NQF Evaluation Criteria and Guidance, keeping the major evaluation criteria in place but modifying the subcriteria as appropriate to reflect the specific needs of resource use measure evaluation.

Resource use measures are broadly applicable and comparable measures of input counts (i.e., in terms of units or dollars) applied to a population or population sample. Resource use measures count the frequency of specific resources. These resource units may be monetized as appropriate. The approach to monetizing resources varies and often depends on the perspective of the measurer and those being measured. Monetizing resource use permits aggregation across resources.

25.2.1 Considerations for Evaluating Resource Use Measures

• Well defined, complete, and precise specifications for resource use measures include measure clinical logic and method, measure construction logic, and adjustments for comparability as relevant to the measure.

• Data protocol steps are critical to the reliability and validity of the measure.

• Examples of evidence that an exclusion distorts measure results include, but are not limited to, frequency or cost of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusion across providers.

• Some measures may specify the exclusion of some patients, events, or episodes that are known or determined to be high-cost. For example, a patient with active cancer may be excluded from a COPD resource use measure because cancer is considered the dominant medical condition with known high costs.

• Testing for resource use measure exclusions should address the appropriate specification steps (i.e., clinical logic, thresholds, and outliers).

• For those exclusions not addressed, justification for and implications of not addressing them is required.
25.3 **Evaluating eCQMS**

The EHR holds significant promise for improving the measurement of healthcare quality. It can make available a broad range of reliable and valid data elements for quality measurement with a lower burden of data collection. Because clinical data are extracted directly from standardized computer-readable fields, the EHR will be considered the authoritative source of clinical information and legal record of care. Measures developed based on data extracted from the EHR must still meet the evaluation criteria, just like any other measure.

25.4 **Evaluating Patient-Reported, Outcome-Based Performance Measures**

Evaluation for PRO-based performance measures is a special case of overall outcome measure evaluation. In its January 2013 report, *Patient-Reported Outcomes in Performance Management*, NQF outlined criteria specific to PRO-based performance measures. Their overarching principle was that these measures should put the patient foremost. Measures designed to capture performance on PROs should be:

- Psychometrically sound—In addition to the usual validity and reliability criteria, cultural and language considerations, and patients’ burden of responding should be considered.
- Person-centered—Measures should reflect collaboration and shared decision-making with patients. Patients become more engaged when they can give feedback on outcomes important to them.
- Meaningful—Measures should capture impact on health-related quality of life, symptom burden, experience with care, and achievement of personal goals.
- Amenable to change—Outcomes of interest must be responsive to specific healthcare services or intervention.
- Implementable—Data collection directly from patients involves challenges of burden to patients, health literacy of patients, cultural competence of providers, and adaptation to computer-based platforms. Evaluation should address how these challenges are managed.
26 NQF ENDORSEMENT AND MAINTENANCE

The NQF currently serves as the CBE regarding performance measurement for HHS. To the extent feasible, CMS uses measures that have been endorsed by NQF in CMS public reporting and value-based purchasing programs. This section explains actions and responsibilities of the measure developer in the measure submission process to NQF and the measure developer’s role during the NQF measure endorsement process. NQF endorses measures only if they pass five measure evaluation criteria—importance to measure and report, scientific acceptability of measure properties, feasibility, usability and use, and related and competing measures.

26.1 MEASURE SUBMISSION TO NQF

26.1.1 When to Submit Measures to NQF

The NQF requires measure stewards or developers to submit an Intent to Submit form at least 3 months prior to the designated cycle’s measure submission deadline. Deadlines for the Intent to Submit form are generally August 1 (i.e., fall) and January 2 (i.e., spring) every year. This form will notify the NQF of the measure steward’s or measure developer’s readiness to submit measures for endorsement consideration and will allow adequate opportunity for technical assistance prior to submitting measures for evaluation. Up-to-date testing information and full measure specifications must also be submitted to NQF along with the Intent to Submit form. Measure developers should obtain approval from their COR before initiating online submission of a measure.

26.1.2 The COR Decides to Submit a Measure to NQF

The measure developer should confirm the list of measures with the COR and begin preparing the measures for submission. The measure developer and COR must inform the Measures Manager of an upcoming measure submission. The measure developer should review the NQF website for updated forms and resources, including directions on completing the online submission. Contact appropriate NQF project staff for technical assistance.

With the introduction of the Promoting Interoperability Programs, formerly the EHR Incentive programs, there is a movement toward the development and/or respecifying of measures specified for use with EHRs (i.e., eCQMs). The COR will provide guidance about which eCQMs are candidates for NQF submission. These eCQMs, which are encoded in the HQMF, must meet specific NQF submission criteria. Measure developers are responsible to monitor NQF’s eCQM requirement policies prior to submission and to contact NQF staff for technical assistance. Note: While many measures can be calculated from EHR data, only some of them are considered eCQMs.

26.1.3 The Measure Developer Completes the NQF Measure Submission

Measure developers must submit their measures via an online measure submission, which is available on the NQF website and enables users to:

- Gain secure access to the submission form from any location with an Internet connection
- Save a draft version of the form and return to complete it at their convenience
- Print a copy of the submission form for reference or other uses, if desired.

When initiating an online measure submission, the measure developer may contact NQF and request access for additional users to enter data in the online form, which enables the measure developer to assign sections of the form to appropriate staff and facilitates internal review. The COR may also be
listed as a user to facilitate ongoing and final review of the form. The measure developer should inform
the COR of this option. However, users must coordinate the timing at which they save their respective
edits, or their edits could be overwritten.

The MIF and MJF are CMS forms designed to present measures in a standardized way. The CMS MIF and
MJF have been aligned, as appropriate, with the most recent NQF measure submission requirements
available at the time of the Blueprint publication. Both forms were designed to guide the measure
developer throughout the measure development process in gathering the information needed for a
successful NQF submission and organizing it to minimize rework.

The measure developer is responsible for completing the NQF measure submission and ensuring that
the information is sufficient to meet NQF’s requirements. The measure submission is the developer’s
presentation of the measure to the Standing Committee and others to demonstrate that the measure
meets the criteria for endorsement. A measure submission form is required for each measure submitted
for endorsement consideration. Tips for successful submissions are:

- Contact NQF project staff for technical assistance with the submission and evaluation process.
- Answer every part of the NQF measure submission clearly and concisely.
- Provide substantive, practical responses to each item.
- Ensure that the form is complete, with enough information that it can be understood as a
  standalone document.
- Attachments, references, and URLs are considered only supplementary and should include
  specific page numbers, table numbers, specific links, etc.
- Submit attachments or URLs, as needed, for long lists of codes or other data elements used in
  the measure, details of a risk adjustment model, and the calculation algorithm.
- Provide any pilot test data available, even if it does not satisfy NQF’s entire testing
  requirements.
- Identify all possible endorsement roadblocks in advance and address them in the measure
  submission.
- Document the rationale for all decisions made in the specifications.
- Document the rationale for all measure exclusions.
- Discuss any controversies about the science behind the measure and why the measure was built
  as it was.
- Double-check the document to ensure no questions are left unanswered (i.e., no fields should
  be left blank and all questions should have a response).

Measure developers can contact the Measures Manager for content questions while completing the
online submission. For technical questions about the online submission, contact the appropriate NQF
project manager/director or measuremaintenance@qualityforum.org. Questions about the content or
information required by the online submission form should be directed to the NQF project manager/director
whose name and contact information appear on the project’s Information page on the NQF
website. Measure developers are expected to have worked closely with the Measures Manager
throughout development—and provided all measure development deliverables—to ensure that no
duplication of measure development occurs and to identify potential harmonization opportunities prior
to NQF submission.

The search for related and competing measures should be conducted early during the information
gathering phase of development and again just prior to submission to NQF. Before NQF will even
consider a measure that is submitted, the measure developer must attest that harmonization with
related measures and issues with competing measures have been considered and addressed, as appropriate. Measure Evaluation Criterion 5, Comparison to Related or Competing Measures, found in NQF’s Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement, is the standard by which NQF evaluates harmonization. The measure developer should work closely with the Measures Manager to identify potential related and/or competing measures that may be in development.

A measure developer may not discover that measures exist for the same condition, process of care, outcome, or care setting until after the measures are submitted to NQF. If that happens, the NQF Standing Committee reviewing the measures could select a superior measure or request that both responsible measure developers create a harmonization plan that addresses the possibility and challenges of harmonizing their respective measures. The Standing Committee would consider the harmonization plan and decide whether to recommend the measure for endorsement.

26.1.4 The COR Approves the NQF Measure Submission

The measure developer will refer the completed measure submission to the COR for approval before submitting it to NQF.

The measure developer should be aware that the COR may seek additional reviews of the completed measure submission before approving it. These reviews may come from the Measures Manager and other experts within CMS. Therefore, measure developers should account for that review period in their submission timeline.

The measure developer will then make any necessary changes to obtain COR approval.

26.1.5 The Measure Developer or the COR Submits the Measure to NQF According to NQF Processes

Once the COR has approved the measure submission, the measure developer or the COR submits it to the NQF using the online process.

The NQF follows a standardized measure review process, delineated in Section 3, Chapter 26.2, NQF Endorsement Process, to consider granting endorsement to the measure.

26.2 NQF ENDORSEMENT PROCESS

The NQF follows a Consensus Development Process. Measure developers are encouraged to monitor the NQF Submitting Standards website for new information, as the NQF has recently changed their processes and timelines. Measure developers may now begin their measure submission at any time. There are two evaluation cycles per year. Measure developers should refer to the NQF web page, Submitting Standards, for more information on submitting new measures, and refer to the Maintenance of NQF-Endorsed Performance Measures web page for information about measure maintenance reviews.

26.3 MEASURE DEVELOPER’S ROLE DURING NQF EVALUATION

During its evaluation, NQF may have questions about the submitted measure. These questions may come from NQF staff or from the project Steering Committee. To facilitate answering any questions, the measure developer is encouraged to be actively involved in the NQF process while the measure is being considered. A member of the measure developer’s team who is prepared to explain and defend the measure should attend the Standing Committee and CSAC meetings while the measure is being discussed. By attending the meetings, the measure developer will gain better understanding of NQF’s
approach to the overall project as well as the specific measures being considered. This level of active involvement better positions the measure developer to answer NQF’s questions. During discussions, the measure developer should be prepared to defend the importance of the clinical topic, the scientific basis for each measure, the construction of the measure, and measure testing results.

For eCQMs, the measure developer, with support from a HQMF or eCQM standards SME, will communicate and collaborate with NQF during the evaluation.

Questions may also arise during the NQF public comment period and may also need to be reviewed by the TEP used by the measure developer to develop or reevaluate the measure. The measure developer proposes responses to the questions, which the COR reviews and approves, before the measure is submitted to NQF. During its evaluation, NQF may suggest changes to the measure to make it more acceptable, to harmonize with other measures, or both. If this occurs, the measure developer may then consult with the TEP used to develop or reevaluate the measures. With the COR’s approval, the measure developer will make the changes and submit the revised measure to NQF.

26.4 **Trial Use Approved Measures**

In some cases, the NQF may grant an eCQM Approval for Trial Use status. Measures with trial use approval will lose that status after 3 years. The measure developer can submit the measure for endorsement prior to or after the end of the 3-year Trial Use period. The NQF Trial Use Approval Policy can be found on the NQF website along with the Measure Testing Form for Trial Use Approval.

26.5 **Measure Maintenance for NQF**

Once NQF has endorsed a measure, the measure developer supports ongoing maintenance of the endorsement of the measure if it is part of the scope of work for that measure developer. The measure developer is responsible for being familiar with NQF’s current measure endorsement maintenance processes as described on NQF’s website. NQF’s endorsement maintenance processes are designed to ensure that NQF continues to endorse only measures that meet the current NQF evaluation criteria. NQF endorsement maintenance reviews are separate from CMS MMS maintenance reviews, but Figure 40 depicts the way the two processes parallel. The CMS scheduled maintenance reviews are in the top row with the parallel NQF maintenance submissions listed as sub-processes below. Figure 41, from the 2018 Measure Developer Guidebook report, depicts the timeline. More information on the CMS scheduled maintenance reviews can be found in Section 3, Chapter 30, Measure Maintenance Reviews.
Figure 40. Measure Review Cycle Timeline

Figure 41. NQF Consensus Development Process
27 MEASURE SELECTION

Depending on the CMS program, there are different paths that a measure can take for selection and implementation. Not all measures go through the pre-rulemaking and rulemaking processes. Contractors should discuss the need to participate in these activities with their CORs.

27.1 PRE-RULEMAKING PROCESS

Section 3014 of the ACA mandated the establishment of a federal pre-rulemaking process for selecting quality and efficiency measures for specific programs within HHS. The pre-rulemaking process requires HHS to consider multi-stakeholder input on quality and efficiency measure selection. To meet these requirements, CMS develops a MUC list. The NQF-convened MAP is the multi-stakeholder group described in Section 3014, and it provides input to HHS on the list of measures for use in a specified program. By statute, HHS and CMS must consider MAP input and publish the rationale for selecting any measure (i.e., in proposed or final rules) that is not NQF-endorsed.

27.1.1 Measures Under Consideration

Over the past few years, CMS has articulated a number of measure selection criteria in its federal rules for various programs. The term “measure selection” typically applies to determining whether a measure should be included in a measure set for a specific program, while “measure evaluation” applies to assessing the merits of an individual measure, not in the context of a specific program. CMS has established a set of measure selection criteria so that HHS can develop the MUC list for qualifying programs and make it available publicly by December 1 each year. These selection criteria are operationalized by CMS program staff and leadership to decide which measures to place on the MUC list to be reviewed by the MAP.

Figure 42 contains a sample timeline for the pre-rulemaking process.

![Figure 42. Sample Pre-rulemaking Process Timeline](image)

After opening an ONC Project Tracking System (Jira) project for new candidate measures, CMS publishes an MUC User Guide found on the CMS Pre-Rulemaking website. Refer to Appendix B for more information on Jira. CMS hosts a series of educational webinars to kick off the official MUC season. More information on the CMS Pre-Rulemaking process is available on the CMS Pre-Rulemaking website.
27.1.2 CMS Measure Selection Criteria

CMS measure selection criteria include:

- Measure is responsive to specific program goals and statutory requirements.
- Measure addresses an important condition or topic with a performance gap and has a strong scientific evidence base to demonstrate that the measure when implemented can lead to the desired outcomes and more affordable care. This requirement corresponds to NQF’s importance criterion.
- Measure addresses one of the Meaningful Measurement areas.
- Measure selection promotes alignment with CMS program attributes and across HHS programs.
- Measure reporting is feasible, and measures have been fully developed and tested. In essence, measures must be tested for reliability and validity.
- Measure results and performance should identify opportunities for improvement. CMS will not select measures when evidence already identifies high levels of performance with minimal opportunity for improvement (i.e., measures that are topped out).
- Potential use of the measure in a program does not result in negative unintended consequences (e.g., reduced length of stay, overuse or inappropriate use of treatment, limiting access to care).
- Measure should not duplicate another measure currently implemented in programs.
- eCQMs must be fully developed and tested.
- eCQMs must be created using the MAT and expressed in HQMF.

Applying the measure selection criteria listed above, CMS develops the MUC list. Measure developers may be asked to provide details on the measures to help CMS develop the MUC list. CMS then provides this list to the MAP.

27.1.3 MAP Recommendations

The MAP input to HHS on the list of quality and efficiency MUC by the Medicare program is due by February 1 of each year as a recommendation report. Each annual report can be found on the MAP pages on the NQF website and on the CMS Pre-Rulemaking website. To be fully enfranchised in this process, measure developers are strongly encouraged to attend the MAP.

27.1.4 CMS Considers MAP Input for Final Selection

After CMS receives the MAP input, a deliberation process begins to determine which measures will be included in the federal rulemaking processes. The measure selection criteria used during development of the MUC list, and identified above, are the same criteria used for federal rulemaking. HHS must consider MAP input and publish the rationale for selecting any measure for use in a CMS program—in proposed or final rules—that is not endorsed by NQF.

27.2 CMS RULEMAKING PROCESSES

After CMS completes the pre-rulemaking process and selects measures for potential inclusion in rulemaking, the next steps in the cycle are:

1. Proposed rules—CMS writes the proposed rules and publishes them in the Federal Register. A proposed rule is generally available for public comment for 60 days.
2. Final rules—CMS considers the comments that were received and publishes the final rules in the Federal Register.
27.3 **Rollout, Production, and Monitoring of Measures**

When measures are finalized in the rule, CMS prepares plans for implementation, including the initial rollout, data management and production, audit and validation, provider education, dry runs, and appeals processes. Lessons learned and other important information gathered from these processes should be conveyed to the CMS staff leading the measure priorities planning task.

27.4 **Measure Maintenance**

After measures are implemented, the measure developer monitors performance of the measures, responds to ongoing feedback, and continuously scans the environment regarding the measures. For example, for eCQMs, the ONC Project Tracking System (Jira) is one method for collecting and monitoring feedback on measure implementation.

In addition, there are two measure maintenance activities that apply to every measure: annual update and a triennial comprehensive reevaluation. A third activity, the ad hoc review, occurs only if there are significant unforeseen problems with the measure, such as a major change in the measure’s scientific evidence base. A full description of these reviews is found in Section 2, Chapter 6, Measure Use, Continuing Evaluation, and Maintenance. The following five outcomes are possible following maintenance review of CMS measures:

1. Retain—Keep the measure active with its current specifications and minor changes (refer to Figure 43).
2. Revise—Update the measure’s current specifications to reflect new information (refer to Figure 43).
3. Retire—Cease to collect or report the measure indefinitely. This applies to measures that are not owned or maintained by any measure steward. If it is necessary to retire a measure from a set, consider that there may be other replacement measures to complement the remaining measures in the set (refer to Figure 44).
4. Suspend—Temporarily cease to report a measure. Data collection and submission may continue, as directed by CMS. This option may be used by CMS for topped-out measures where there is concern that rates may decline after data collection or reporting ceases.
5. Remove—A measure is no longer included in a specific CMS program set for one or more reasons. This does not imply that other payers/purchasers/programs should cease using the measure. If CMS is the measure steward and another CMS program continues to use the measure, CMS will continue maintaining the specific measure. If another entity is the steward, the other payers/purchasers/programs that may be using the measure are responsible for determining if the steward is continuing to maintain the measure (refer to Figure 45).

Figure 43, Figure 44, and Figure 45 are adapted from the “Standard CMS Measure Implementation Determination Criteria” and lists the criteria CMS uses to make decisions regarding the various dispositions.48

---

### Revise or Retain

<table>
<thead>
<tr>
<th>Core Criteria</th>
<th>Optional Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure is responsive to specific program goals and statutory requirements.</td>
<td>eCQMs must be fully developed and tested, entered in the MAT and created in the HQMF format. eCQMs must pass feasibility testing. eCQMs must undergo reliability and validity testing, including review of the logic and value sets by the CMS partners.</td>
</tr>
<tr>
<td>Measure addresses an important condition/topic with a performance gap and has a strong scientific evidence base to demonstrate that the measure when implemented can lead to the desired outcomes and/or more affordable care.</td>
<td>Should be electronically specified whenever possible. Should be aligned with the EHR Incentive Programs where applicable.</td>
</tr>
<tr>
<td>Measure addresses one or more of the six NQS priorities and the CMS Quality Strategy.</td>
<td></td>
</tr>
<tr>
<td>Measure selection promotes alignment with CMS program attributes and across HHS programs.</td>
<td></td>
</tr>
<tr>
<td>Measure reporting is feasible and measures have been fully developed and tested.</td>
<td></td>
</tr>
<tr>
<td>Measure results and performance should identify opportunities for improvement. CMS will not select measures that are topped out.</td>
<td></td>
</tr>
<tr>
<td>Use of the measure in a program does not result in negative unintended consequences (e.g., reduced lengths of stay, overuse or inappropriate use of care or treatment, limiting access to care).</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 43. CMS Criteria for Measure Disposition: Revise or Retain**

### Retire

<table>
<thead>
<tr>
<th>Core Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure is owned by CMS, and CMS will no longer maintain the measure.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Optional Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>No longer adds value commensurate with the cost of data collection and reporting. Performance or improvement on a measure does not result in better outcomes. Collection or public reporting of a measure leads to negative unintended consequences other than patient harm. Does not align with current clinical guidelines or practice. Measure performance is so high and unvarying that meaningful distinctions and improvements in performance can no longer be made. The availability of a better measure that is more (1) broadly applicable (across settings, populations, or conditions); (2) proximal in time to desired outcomes for the particular topic; (3) strongly associated with desired outcomes for the particular topic; or (4) aligned with other CMS/HHS programs.</td>
</tr>
</tbody>
</table>

**Figure 44. CMS Criteria for Measure Disposition: Retire**
27.5 **Impact Assessment of Medicare Quality Measures**

Also mandated by Section 3014 of the ACA, the Secretary of HHS must provide a publicly available assessment of the impact of all Medicare quality measures (i.e., measures that are implemented, measures that are planned for implementation, and measures that are included in the [MUC list](#)) once every 3 years. This triennial report assesses how well CMS, using quality measures, has achieved their quality priorities. The report also evaluates the impact of CMS quality measures by assessing the measures’ reach, their effectiveness, and issues associated with their adoption, implementation, and maintenance. The first CMS Measure Impact Assessment report was published in March 2012 and included findings for the measures implemented in CMS programs. The report examined trends over time, including how much the measure results declined, remained unchanged, or increased. The [March 2015 report](#) greatly expanded from the trend data reported in 2012. The [most recent report](#) was published in February 2018. Several positive patient and cost impacts were noted in the 2018 Impact Report, such as:

- 670,000 additional patients with controlled blood pressure (2006–2015)
- $6.5 billion–$10.4 billion estimated in costs avoided for fewer patients with poor diabetes control (2006–2015)
- Performance on most measures (60%, including 55% of [outcome measures](#)) improved.

While the 2018 Impact Report shows the likely contribution of CMS quality measures to improving quality and reducing expenditures within the healthcare system, there are challenges and barriers that suggest the need for additional progress. As an example, [healthcare disparities](#) continue to persist among select [populations](#). CMS commits to taking steps to address these challenges, such as minimizing the burden of reporting quality measures through the evaluation of CMS measures by targeting only the highest impact areas through CMS’s work on the Meaningful Measures Initiative.
28 MEASURE ROLLOUT

When CMS decides to start data collection at a national level, the measure is considered rolled out. Measure developers should note that it is possible (in certain circumstances) that a measure could be implemented prior to full nationwide rollout. For example, a measure might be used for facility-level quality improvement before it is rolled out for national use as a publicly reported accountability measure.

The work described in this chapter, as with all parts of the Blueprint, will comply with requirements of the Data Quality Act, incorporated into Public Law 106-554, which provides “policy and procedural guidance to federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by federal agencies” (section 515). The measure development and maintenance procedures detailed in this Blueprint also comply with HHS’s Guidelines for Ensuring the Quality of Information Disseminated to the Public.

Figure 46 depicts the process of measure implementation as well as associated responsibilities for related tasks when rolling out measures approved by the COR.

---

**Figure 46. Overview of the Measure Rollout Process**

Once the COR has approved a measure for use in a specific program, several tasks must be completed for rollout. Perform these eight steps outlined in Sections 28.1 through 28.8, simultaneously whenever possible to achieve an efficient timeline.

### 28.1 MEASURES ARE SELECTED BY CMS

Once the measure(s) are developed, CMS selects a measure for use in one or more of its programs utilizing the process described in Section 3, Chapter 27, Measure Selection. The measure developer develops the coordination and rollout plan, which includes:

---

49 CMS may not require all eight steps. Check with your COR and contract requirements.
• Timeline for quality measure implementation
• Plan for stakeholder meetings and communication
• Anticipated business processes model\(^50\)
• Anticipated data management processes
• Audit and validation plan
• Plans for any necessary education.

The coordination and rollout plan is referenced in the MIDS USOW. The anticipated business processes model and anticipated data management processes together represent the implementation referenced in the MIDS USOW.

The COR is responsible for overseeing the plan to inform stakeholders during rollout. The measure developer is responsible for coordinating and actively participating in stakeholder meetings, open door forums, or other means by which the public is informed of upcoming measure revisions. Stakeholders may include, but are not limited to:

• State agencies
• Other CMS divisions
• Office of Information Services
• Software vendors
• Providers and provider organizations.

In addition to coordination with groups and individuals, the measure developer coordinates the implementation with other timelines, including the federal rulemaking process and the NQF measure review cycle.

Communication about the rollout may vary by program and measure. Some of the factors influencing the types of communication include the number of providers affected, the impact of the measures on the providers, and the newness of the measure or program. Examples of communication strategies include:

• Announcement to the QIN-QIO community by a Healthcare Quality Information System (HCQIS) memo or to stakeholders by email
• Presentations at conferences or scientific society meetings
• Publication of articles in peer-reviewed journals
• Publication in the Federal Register through the full rulemaking process
• National provider calls
• Press releases from CMS or CMS partners
• Notices in major media outlets
• Town hall meetings with prominent CMS officials in various major cities
• Open door forums
• Other strategies as determined by the COR.

A measure developer must consider these communication activities when developing the initial timeline for quality measure implementation. The timeline is then reviewed for approval by the COR.

---

\(^{50}\) According to Williams (1967), business process models are graphical and textual representations are used to compare current and future business processes, which is similar in many ways to the Lean Kaizen approach of current, future, and ideal states. Using this methodology, measure developers can compare and contrast the proposed influence of a measure through quantifiable means, including process improvements and burden, as well as in terms of economics, morbidity, and mortality.
For some measures, measure developers should develop an implementation algorithm (i.e., calculation algorithm), which is an ordered sequence of data element retrieval and aggregation through which numerator and denominator events or CV values are identified in the measure specifications. The algorithm is documented in the MIF under Measure Specifications as the Calculation Algorithm/Measure Logic. This documented process is expected to begin with the submission of data by the providers (i.e., for measures based on data abstraction) or the initiation of data collection (i.e., for measures based on administrative data), and end with the posting of the measures for public reporting. The measure developer should consult with the COR if uncertain about the need for an implementation algorithm.

Before implementing any process that involves the collection of new data, the measure developer should consult with the COR regarding the PRA requirements. OMB approval is required before requesting most types of information from the public. Section 2, Chapter 2.3, Stakeholder Engagement, includes a discussion of these requirements.

### 28.2 Implement the Rollout Plan

This step applies primarily to a new measure or a new use for an existing measure. The ultimate intended use of the measure will be a major factor in determining what is required for the rollout plan. Examples of activities that may be conducted during this step include:

- Develop work processes and tools for data collection, rate calculation, and reporting.
- Develop process for responding to questions about the measure
- Identify CMS divisions that need to be involved to ensure adequate resources are available when the measure is fully implemented
- Determine relevant program rules, such as how eligibility for payment will be evaluated in a value-based purchasing or pay-for-reporting program
- Develop process for documenting questions and answers so they can be monitored for trends and used to inform measure maintenance activities.

### 28.3 Implement the Data Management Process

The data management processes that were created and tested during measure development must now be adapted for measures that are in use.

Major tasks in this step include:

- Translating the algorithm used with hypothetical or test data into one that can be used with actual data
- Developing protocols and tools to receive data
- Performing parallel processing of data through the analysis program to ensure accuracy of the interpretation of the algorithm
- Developing measure data collection quality control processes.

### 28.4 Develop the Auditing and Validation Plan

The measure developer must provide an audit and validation plan to the COR for approval before the measure is put into production. The primary consideration when conducting audit and validation is determining exactly what is being audited and validated (i.e., the full measure or individual data elements).
When auditing and validating data element results, consider the following questions:

- Have the data been collected correctly?
- Were the algorithm and all auxiliary instructions followed correctly? This is a concern for data that are abstracted from hard copy patient medical records where sampling methodologies and data hierarchies may be involved.
- Have the data been transmitted correctly?
- Are the standards for each data field maintained throughout the data transmission process? For example, abstraction instructions may require that dates be consistently expressed in mm/dd/yyyy format, but one or more mediating computer programs may employ yy/mm/dd formatting. If the calculation program relies on the first format, it may misread the second and adversely affect the provider’s rate.
- Do the incoming data make sense? For example, a record might be suspect if it indicates a male received a hysterectomy or a female was diagnosed with prostate cancer.

When auditing and validating measure function, consider the following questions:

- If there are multiple databases used to calculate the rates, were they correctly linked?
- Was the sampling methodology correct?
- Were data elements linked appropriately according to the measure specifications?
- Was the calculation algorithm programmed correctly?
- Do the measure results make sense? For example, rates greater than 100% may indicate an error in the calculation algorithm or in the calculation programming. Similarly, unexpectedly low rates may indicate a problem as well.

28.5 **DEVELOP AN APPEALS PROCESS**

Before implementing a measure, CMS will determine whether providers can appeal either the audit results or measure rates. The measure developer may be required to help develop and design these processes.

28.6 **IMPLEMENT EDUCATION PROCESSES**

Providers will likely need to be educated on exactly what is being measured and how to interpret results. For example, QIN-QIOs may need to be informed about the measure and its meaning. For measures relying on abstracted data, abstractors must be trained to consistently identify correct data and qualifying cases. Methods for education include, but are not limited to:

- User guides and training manuals
- Conference calls and recordings of the calls
- Web-based presentations and recordings of the presentations, including answering follow-up questions from the viewers
- Workshops at conferences or scientific society meetings
- Train-the-trainer events
- Other venues as determined by the COR.
28.7 **CONDUCT THE DRY RUN**

The dry run is the final stage of measure testing and the second-to-last stage in measure rollout. The final stage in measure rollout is the first use of a measure in a CMS program or first results reporting. In the dry run, data are collected from all relevant providers across the country.

The purpose of the dry run is to finalize all methodologies related to case identification/selection, data collection (for measures using patient medical records), and measurement calculation. The dry run will verify that the measure design works as intended and will begin to identify unintended consequences such as gaming or misrepresentation. The dry run also familiarizes relevant entities such as CMS, the QIN-QIOs, and the providers with the reports of the measure results and provides the COR the opportunity to communicate and collaborate with these entities to improve the usability of the reports before actual implementation and to identify and respond to questions and concerns. The dry run also identifies any issues with the report production processes so they can be improved to avoid problems when the measure is implemented.

Rates from a dry run are not publicly reported or used for payment or other reward systems, although CMS may decide to use them as the baseline measurement.

The dry run may not be a discrete step in the implementation of the measure. At the COR’s direction, this step may be skipped. Skipping this step means that the first round of data collection and results reporting may serve as the de facto dry run. If problems arise during the dry run, those problems must be addressed and resolved before the measure is fully implemented.

28.8 **SUBMIT REPORTS**

CMS may request reports summarizing the rollout processes. These reports may include:

- Reports describing the business processes
- Results of any education that was conducted
- Results of the dry run, including, but not limited to:
  - Analysis of the measure’s success in meeting CMS’s intentions for it
  - Recommendations regarding:
    - Measure specifications\(^\text{51}\)
    - Business processes model
    - Data management processes
    - Audit and validation processes
    - Educational processes for either data collectors or users of the measure results.

---

\(^{51}\) Recommendation for changes to the measure specifications should clearly document the proposed changes and also address (at a minimum) whether the changes are material, whether the changes requires public comment or publication in the Federal Register, and whether the changes affect other harmonized measures.
29 Measure Production and Monitoring

Measure production and monitoring includes the ongoing tasks necessary to use the measure over time. These tasks are described in this overview, but refer to other chapters for more detailed instructions, where applicable. The process of measure production and monitoring varies significantly from one measure set to another depending on a number of factors, which may include, but are not limited to:

- Scope of measure implementation
- Healthcare provider being measured
- Data collection processes
- Ultimate use of the measure (e.g., quality improvement, public reporting, pay-for-reporting, value-based purchasing)
- Program in which the measure is used.

The intensity or amount of effort involved in each of these tasks may vary and be affected by the factors listed above. Work conducted as part of measure production and monitoring should comply with the requirements of the Data Quality Act as well as with HHS’s Guidelines for Ensuring the Quality of Information Disseminated to the Public available online with instructions. Figure 47 lists the requirements associated with Measure Production and Monitoring.

Figure 48 shows a diagram of the overall production and monitoring components of a measure that have been implemented in a CMS program. Depending on the scope of the contract and program requirements, measure developers may be required to perform various tasks associated with ongoing implementation and production. Some examples of these steps include, but are not limited to, the seven steps discussed in the next subsections.
### Conduct Data Collection and Ongoing Surveillance

Once measure development is complete and any problems that surfaced during the dry runs are resolved, the measure will be fully implemented (i.e., data are being collected, calculated, and publicly reported). As the measure is being used, the measure maintenance contractor, which may or may not be the same as the measure developer, should continue environmental scans of the literature about the measure. In addition to publications in medical and scientific publications, the general media should be scanned for articles and commentaries about the measure. This process should be continuous, with periodic reports to CMS. The information collected during the past 3 years will be summarized and included in the comprehensive reevaluation. Information obtained may also trigger an ad hoc review if the concern needs immediate action. Ongoing information surveillance is similar to the information gathering stage of measure development as covered in Section 2, Chapter 1.2, Information Gathering. Similar analyses should be conducted of the literature, with reports submitted as required by the contract.

As the measure is being used, new studies may be published that address the soundness of the measure. The measure developer needs to pay attention to any organizations that issue relevant clinical practice guidelines, especially for process measures. If the measure is based on a specific set of guidelines, monitor the guideline writers closely for any indication that they are planning to make changes to their guidelines. If the measure is not based on guidelines, monitor the scientific and clinical literature for reports that would impact the scientific basis of the measure. These guideline changes or other statements may cause an ad hoc review.

After data collection begins, the measure developer monitors for unintended consequences the measure might have on clinical practice or outcomes. Look for articles or studies in the literature that identify unusual trends in data suggesting unintended consequences. If significant unintended consequences are identified, especially if patient safety is the concern, do not wait for a scheduled annual or comprehensive review. An ad hoc review may be necessary and requested.
29.2 Respond to Questions about the Measure

The measure maintenance contractor may also be responsible for reviewing any stakeholder feedback and responding to it in a timely manner. This stakeholder feedback may include questions or comments about the measure or the program in which the measure is being used. This feedback may be submitted electronically or by other means. Assuming the submitter has provided contact information, the measure developer receiving the feedback should reply immediately, alerting the submitter that feedback has been received and is being reviewed. Within 2 weeks of the submission date, the measure developer should provide either a final response to the submitter or a status update to let the submitter know what is happening regarding the feedback. All responses will be reviewed by the COR unless the COR makes other arrangements.

The ONC Project Tracing System (Jira) is used to collect and respond stakeholder feedback about eCQMs.

Comments and questions may also have been submitted as part of the federal rulemaking process as measures were selected for implementation. Those comments and questions should be reviewed by the maintenance contractor for indications that the measure may need to be refined. These comments may identify areas that need clarification. They may also identify feasibility issues and possible unintended consequences.

If the measure developer is not responsible for responding to questions, the measure maintenance contractor should obtain reports and review them on a regular basis. As with the other components of the environmental scan, stakeholder feedback may identify the need for an ad hoc review.

29.3 Produce Preliminary Reports

For public reporting programs, results will be released to providers before they are released to the public. Providers will be allowed time (i.e., usually 30 days) to review and respond to the measure results.

Preliminary reports should be monitored for unusual trends both by CMS and by its measure developers. Investigate any trends that are discovered, rerunning reports to check for errors in calculation. If unexpected results are not due to errors in calculation, the cause should be investigated and reported to CMS. If necessary, CMS has the option of suppressing some or all the data from appearing publicly for a given reporting period (e.g., quarter, year). Data suppression might be necessary due to known problems with a given measure or measure set or data collection issues with a specific provider or group of providers. The decision to suppress data may apply to:

- All measures in a given measure set(s)
- A specific measure(s)
- A group of providers (e.g., state, region)
- A specific provider(s).

29.4 Report Measure Results

Once measure results are calculated and providers have reviewed them (i.e., for public reporting or value-based purchasing programs), the results are released. Depending on the program, the reporting process will vary. For quality improvement programs, individual results will be released to the providers, often with other provider results included for comparison. The COR will determine whether the other provider results are to be reported anonymously. The process by which the information is to be shared

---

Blueprint 15.0 September 2019  Page 253
with providers and others, the format of the reports, and support for questions from providers should be established in the rollout plan prior to implementation.

If measure results are to be posted on an appropriate website, an announcement of the updated site may be made. The display of the measure results on the website may require collaboration with quality alliances and will require consumer testing. Other considerations include compliance with Section 508 of the Rehabilitation Act and the HHS Section 508 website, which require federal agencies’ electronic information to be accessible to people with disabilities.

For value-based purchasing programs, results will be shared with appropriate areas within CMS responsible for calculating provider payments in addition to any requirements for public reporting of the data.

29.5 **Monitor and Analyze the Measure Rates and Audit Findings**

Measure performance rates and audit findings will be monitored and analyzed periodically and at least once a year for:

- Overall performance trends
- Variations in performance, gaps in care, and extent of improvement
- Disparities in resulting rates by race, ethnicity, age, social risk factors, income, region, gender, primary language, disability, or other classifications
- Frequency of use of exclusion or exception and how they influence rates
- Discretionary exclusion, which should be evaluated carefully for gaming, unintended consequences, and uneven application that could influence comparability
- Patterns of errors in data collection or rate calculation
- Changes in practice that may adversely affect rates
- Impact of measurement activities on providers
- Correlation of the data to either confirm the measure’s efficacy or identify weaknesses in the measure.

Ongoing monitoring should continually assess a measure’s linearity; any marked departures may be cause for concern. If performance targets were predicted as recommended, the measure developer should investigate any measure whose performance over time falls short of its target. This information is reported during reevaluation as described in Section 3, Chapter 30.2, Comprehensive Reevaluation.

29.6 **Perform Measure Maintenance or Ad Hoc Review, When Appropriate**

As measures are in production and their performance is monitored, they need to be maintained on a schedule. Each measure is reviewed at least annually to ensure that the codes used to identify the populations (i.e., denominator, numerator, and exclusions) are current, and to address other minor changes that may be needed. The standardized process for annual update is described in Section 3, Chapter 30, Measure Maintenance Reviews. Each measure is also fully reevaluated every 3 years to ensure that it still meets the measure evaluation criteria. The standardized process is described in Section 3, Chapter 30.2, Comprehensive Reevaluation.

Situations may also arise in which a measure must be reviewed before the scheduled annual update or comprehensive reevaluation. In this case, an ad hoc review is conducted. The standardized process is described in Section 3, Chapter 30.3, CMS Ad Hoc Review, including the process for determining when an ad hoc review is necessary. For endorsed measures, a request for ad hoc review may also come from
the NQF if there is evidence to justify such review. The outcome of the ad hoc review will be incorporated into the monitoring cycle at the appropriate place, based on the decision approved by the COR.

The outcome of the reevaluation will determine CMS’s decision about continued use of a specific measure. Those decisions are described in Section 3, Chapter 30.5, Possible Outcomes of Maintenance Reviews, and include whether to retain, revise, retire, remove, or suspend the measure in a program.

If the NQF has endorsed the measure, results of the maintenance review will be reported to the NQF to reevaluate its endorsement at the time of NQF maintenance review. The outcome of the NQF review may influence whether CMS continues using a specific measure in a program.

### 29.7 Provide Information that CMS Can Use in Measure Priorities Planning

Lessons learned from the measure rollout, the environmental scan, and ongoing monitoring of the measure should be conveyed to CMS. Section 3, Chapter 2, Measure Prioritization and Planning, describes how CMS uses this input. CMS leadership may find information from measures monitoring valuable for setting priorities and planning future measurement projects. CMS may request an evaluation of current measures and sets used in the programs or initiatives and recommendations for ways to accommodate cross-setting use of the measures. The evaluation may also include options for alternative ways to interpret measures and measure sets through the continuum of care. Performance trends of the measure can be used by the MAP to evaluate use of the same or similar measures in other settings or programs. This evaluation may be performed as part of the pre-rulemaking process for the MUC list.
30 **MEASURE MAINTENANCE REVIEWS**

This chapter describes three types of maintenance reviews, including deliverables and the steps required for each:

- Annual update
- Comprehensive reevaluation
- Ad hoc reviews.

### 30.1 ANNUAL UPDATE

The first type of measure reevaluation is the annual update, which is usually a limited review of the precision of the measure’s specifications—completed annually (or semiannually, in some cases). Annual updates ensure the procedure, diagnostic, and other codes (e.g., CPT, ICD-10-CM, LOINC) used within the measure are updated when code systems change. However, this is also the time to review and address feedback received about the measure’s specifications, reliability, and validity.

Consider feedback from the field to address feasibility concerns (for eCQMs) and implement code changes suggested from the field to address validity.

The measure developer reviews the measure for opportunities for harmonization as part of the annual update.

During the 2 years when an endorsed measure is not being reevaluated for continued NQF endorsement, measure stewards will submit the online, annual update form as required by NQF for continued endorsement. This submission will either reaffirm that the measure specifications remain the same as those at the time of endorsement or last update, or outline any changes or updates made to the endorsed measure.

If changes occur to a measure at any time during the 3-year endorsement period, the measure steward is responsible for informing the NQF immediately of the timing and purpose of the changes. An NQF ad hoc review will be conducted if the changes materially affect the measure’s result (e.g., changes to the population being measured, changes in what is being measured, inclusion of new data sources, expansion of the level of analysis or care settings).

The annual update process ensures that measures are updated as the code sets on which the measures rely are updated. Any comments and suggestions that were collected after implementation are also considered during annual updates to determine whether revision is needed beyond updating the codes. Figure 49 contains a list of Annual Update Requirements:

1. An updated MIF (Deliverable 3-3) showing all recommended changes to the measure. If there are changes relevant to the MJF (Deliverable 3-4), that form should be updated as well.
2. For measure developers maintaining eCQMs, the revised specifications (i.e., HQMF file, human-readable HTML file, and value sets) must be submitted detailing the changes to the measure.
3. A document summarizing changes made such as Release Notes, if not included in the updated MIF.
4. NQF Annual Update online submission, regardless of whether any change was made to the measure.
5. NQF submission documentation for any material changes to the measure.

**Figure 49. Annual Update Requirements**
requirements (Deliverable 4-9) associated with Annual Updates or Maintenance Reevaluation.

The annual update process involves three parts, divided into eight steps outlined in Chapter 30.1.2, Procedure:

- Gathering information that has been generated since the last review (i.e., comprehensive reevaluation, annual update, or measure development—whichever occurred most recently)
- Recommending action
- Approving and implementing the action(s).

30.1.1 Potential for Harmonization

Whenever a measure is evaluated or reevaluated, it must be compared to related or competing measures, assessing for the possibility of harmonization. Measures need to be aligned as much as possible for many reasons. For example, harmonized measures can reduce burden on providers, focus on priority topics with the most potential to improve healthcare, and bring other benefits of parsimony. An annual update is a good time to consider harmonization opportunities.

If related measures are found, consider ways the measure being updated could be aligned with the related measures. If there are competing measures, either justify why the measure being updated is best in class or give a rationale for continuing with possibly duplicative measures.

30.1.2 Procedure

Follow the steps detailed in the subsections below to conduct measure maintenance.

30.1.2.1 Review the Measure’s Code Systems

Review the code systems used by the measure to determine whether:

- New codes have been added to or deleted from the code systems that may affect the measure
- Codes have been changed so that their new meaning affects their usefulness within the measure.

If the measure has not been specified with ICD-10 codes, convert any ICD-9 codes to ICD-10 unless needed for a look-back period. On January 16, 2009, HHS released the final rule mandating that everyone covered by the Health Insurance Portability and Accountability Act (HIPAA) must implement ICD-10 for medical coding by October 1, 2013. However, on April 1, 2014, the “Protecting Access to Medicare Act of 2014” H.R. 4302 bill was signed, which delayed the compliance date for ICD-10 from October 1, 2014, to October 1, 2015.

When maintaining eCQM value sets, it is important to align with the vocabulary recommendations made by Health IT Standards Committee Clinical Quality Working Group and Vocabulary Task Force. Section 3, Chapter 17, Codes, Code Systems, and Value Sets, provides more information on the procedure.

30.1.2.2 Gather Information

The measure developer/measure maintenance contractor is expected to continually conduct environmental scans. This includes reviewing and managing comments on the measure and reviewing literature pertinent to the measure. All new information should be considered during the annual update; however, the most important information is any evidence of unforeseen adverse consequences.
or any controversies that have arisen surrounding the measure. This surveillance may result in an ad hoc review by the NQF or CMS.

If stakeholder feedback can be resolved with minimal change to the measure, consider doing so. If the feedback indicates a serious scientific concern with the clinical practice underlying the measure, incorporate an ad hoc review into the annual update. Details of the procedure are discussed in Section 3, Chapter 30.3, CMS Ad Hoc Review. Evaluate the feasibility and impact of changing measure specifications if feedback during the review recommends modifications.

Conduct a limited review of measure performance, including:

- National performance rates
- State and regional performance rates
- Variations in performance rates
- Validity of the measure and its constituent data elements, as directed by the COR
- Reliability of the measure and its constituent data elements, as directed by the COR.

### 30.1.2.3 Determine the Recommended Disposition of the Measure

Criteria that form the basis for the disposition decision for each measure and description of the possible outcomes are discussed at the end of this chapter under Section 3, Chapter 30.5, Possible Outcomes of Maintenance Reviews.

The possible dispositions are:

- Retain
- Revise
- Remove
- Retire
- Suspend.

### 30.1.2.4 The COR Reviews the Recommendation for Approval

Forward the recommendations to the COR, along with an updated MIF (and eCQM files, if appropriate), and value sets, and Summary of Changes/Release Notes. If significant changes were made, an updated MJF and Measure Evaluation Report may be necessary.52

The COR reviews the annual update documentation. If the recommendation is not approved, the COR documents the approved course of action and instructs the measure developer/measure maintenance contractor as necessary. If the recommendation is approved, the COR notifies the measure developer of the approved course of action.

### 30.1.2.5 Implement the Approved Action

For measures that are proposed to be revised, suspended, removed, or retired, evaluate the impact of the decision on the program using the measure when developing the implementation plan. If there are relevant regulatory or rulemaking schedules, include them in the implementation plan.

After the review, the measure developer/measure maintenance contractor may be responsible to help CMS implement the chosen measure disposition. Communicate and collaborate with the COR to determine what deliverables and actions that are necessary (e.g., making announcements through usual communication modes for the project, arranging for reprogramming, notifying other CMS measure

---

52 The NQF submission may be acceptable for this deliverable.
developers, notifying the NQF if a measure is currently endorsed, re-educating providers). Notify the Measures Manager to ensure that the CMS Measures Inventory is updated appropriately regardless of the disposition decision.

30.1.2.6 Assist the COR in Notifying NQF of the Updated Measure

After a measure is endorsed by the NQF, CMS (as the measure steward) is required to submit a status report of the measure specifications to the NQF annually. This report either affirms that the detailed measure specifications of the endorsed measure have not changed or, if changes have been made, it provides details and underlying reason(s) for the change(s). If changes occur to a measure at any time in the 3-year endorsement period, the measure steward must inform NQF immediately of the timing and purpose of the changes. Some measure maintenance contracts may require updates to the measure more than once a year. In this case, the measure developer/measure maintenance contractor may need to notify NQF of the changes each time they occur. If no changes are made, only the annual update is required.

The NQF provides a standardized template for submission of an annual measure maintenance update that is prepopulated with measure information. CMS will direct the NQF regarding the appropriate measure developer to contact for the annual update. The measure developer/measure maintenance contractor is responsible to prepare this report for the NQF. The measure developer/measure maintenance contractor must also obtain COR review and approval before submitting the report to the NQF. If the changes materially affect the measure’s original result, the NQF may conduct its own ad hoc review. The measure developer/measure maintenance contractor responsible for measure maintenance should be aware of the NQF’s measure maintenance schedule and when the annual update is due to the NQF. The due date for the measure developer/measure maintenance contractor’s measures updates should be confirmed annually with NQF because schedules may change. The measure developer should also inform the NQF of any contact information changes so that notifications can be sent to the correct recipients.

30.1.2.7 Consider Measures Not Stewarded by CMS

Regarding measures for which CMS is not the measure steward (i.e., not ultimately responsible for maintaining the measure), the measure developer/measure maintenance contractor will be responsible for monitoring the maintenance of the measure. This includes ensuring that the measure is revised periodically in response to updates in the underlying code systems (e.g., CPT, ICD-10, LOINC) and that the measure is reevaluated in a manner consistent with (though not necessarily identical to) the reevaluation requirements discussed in Section 3, Chapter 30.2, Comprehensive Reevaluation. The measure developer/measure maintenance contractor is also responsible for updating any CMS documentation of the measure to reflect changes made by the measure steward and discussing those changes with CMS to ensure CMS wants to continue using the measure. Changes cannot be made to a measure that is copyright protected without the steward’s consent. The measure developer/measure maintenance contractor will also be responsible for ongoing surveillance of the literature addressing the measure and alerting the COR to possible issues.

30.1.2.8 Submit the NQF Annual Status Update Report

The measure developer/measure maintenance contractor prepares the annual update report of the measure specifications, submits it to the COR for approval, and then submits it online to NQF. Some measure maintenance contracts may require updates more than once a year. In those cases, the measure developer/measure maintenance contractor should notify the NQF of the changes as often as appropriate.
The NQF stagger deadlines for annual maintenance submissions throughout the year. The NQF assigns each newly endorsed measure to a quarter (i.e., Q1, Q2, Q3, Q4) for annual maintenance submission, and that schedule remains the same through subsequent years. However, measure developers/measure maintenance contractors may request a different quarter for their annual updates.

Confirm the deadline for each annual update with the NQF. These update requirements also appear on measure developers’ NQF dashboards. It is the responsibility of the measure developers/measure maintenance contractor to visit their NQF dashboards periodically and track when updates are due. It is the responsibility of the maintenance contractor to ensure the updates are submitted in a timely manner. The measure developer/measure maintenance contractor and COR may seek guidance from the Measures Manager during any stage of this process.

### 30.2 Comprehensive Reevaluation

Measure developers/measure maintenance contractors are required to conduct a thorough review of the measure every 3 years. In many ways, the comprehensive reevaluation process parallels the measure development process. Details of the comprehensive reevaluation process are described in the following paragraphs. These processes are updated periodically to stay aligned with NQF requirements.

A comprehensive reevaluation consists of information gathering (including a literature review of recent studies and guidelines), analysis of measure performance rates, and synthesis of all feedback received. A TEP is also usually convened and consulted for the comprehensive review.

Generally, CMS and the measure developer/measure maintenance contractor can align the schedules of the reevaluations, so that the comprehensive reevaluation immediately precedes the NQF 3-year maintenance review. This allows CMS time to review the findings and recommendations prior to submission to the NQF.

Figure 50 lists requirements associated with comprehensive reevaluation (Deliverable 4-10).

The comprehensive reevaluation process ensures that the CMS measures continue to be of the highest caliber possible. By periodically reviewing the measures against standard measure evaluation criteria, the measure developer/measure maintenance contractor helps CMS maintain the best measures over time.

The comprehensive reevaluation process includes nine steps, outlined in Section 30.2.3, Procedure, that fall into three phases:

- Gathering information generated since the measure’s development or since the last comprehensive reevaluation, whichever occurred most recently
- Evaluating the measure and recommending action based on the evaluation
- Approving and implementing the action.

The comprehensive reevaluation process assumes that the measure developer/measure maintenance contractor has been monitoring the scientific literature and clinical environment related to the measure, including relevant clinical guidelines.
30.2.1 Measures in CMS Programs for Which CMS is Not the Steward or Providing Maintenance Through a CMS Measure Developer/Measure Maintenance Contractor

CMS measure developers and contractors working with CMS on their programs that use measures not developed or maintained by CMS should monitor information from the measure steward for updates. If there is no steward for a measure (i.e., contracted or non-contracted), CMS will decide whether resources can be allocated to conduct the measure maintenance. If NQF has identified that an endorsed measure is no longer being maintained by its steward, and CMS determines the measure is needed for a program, CMS may take over stewardship and assign the work to a measure developer.

30.2.2 Harmonization

During comprehensive reevaluation, the measure developer makes full consideration to determine whether there are related and/or competing measures available on the same topic. If measure specifications need to be altered so they can harmonize with other measures, the changes could be substantive. The comprehensive reevaluation period may be the best time to make these changes. The process for deciding whether similar existing measures are related or competing is described in Section 3, Chapter 30.1.1, Potential for Harmonization. It is part of Meaningful Measures to foster alignment of performance measures as much as possible, so these considerations are particularly important during comprehensive review.

30.2.3 Procedure

30.2.3.1 Develop a Work Plan

The work plan for measures under comprehensive review should reflect the Blueprint processes as directed by the measure developer’s scope of work. This work plan gives the COR evidence that the
measure developer has a strategy for executing the measurement system processes. Refer to the contract scope of work for the work plan due date.

When developing the work plan, two other schedules should be considered:

- Rulemaking cycle for any regulatory process governing the measure set in question
- NQF’s measure maintenance schedule.

30.2.3.2 Gather Information

During measure monitoring, ongoing surveillance is conducted. Summarize the findings of the environmental scan and update the MJF. The ongoing environmental scan should focus on information published or otherwise available since the last time the measure was evaluated.

At a minimum, this synthesis should include:

- Changes to clinical guidelines on which the measure is based
- Relevant studies that might change clinical practice, which in turn, might affect the underlying assumptions of the measure
- Relevant studies that document unintended consequences of the measure
- Relevant studies that document continued variation or gaps in the care being measured
- Technological changes that might affect how data are collected, calculated, or disseminated
- Similar measures based on their structure, clinical practices, or conditions that could offer an opportunity for harmonization or might serve as replacement measures
- Relevant information gathered from the TEP or interviews with SMEs or measurement experts
- Patients’ perspective on the measure
- Reevaluation of the business case supporting the measure
- Feedback that has been received since the measure was last evaluated (i.e., the initial evaluation or the last comprehensive reevaluation, whichever is most recent).

Obtain measure performance information including, but not limited to:

- Current aggregate national and regional measurement results
- Measurement results trended across the years since the measure’s initial implementation
- Comparison to the trajectory predicted in the business case
- Current distribution of measurement results by provider types (e.g., rural vs. urban, for-profit vs. nonprofit, facility bed size)
- Analysis of the measure’s reliability, stability, and validity since implementation
- Results of audit and data validation activities
- Analysis of any disparities in quality of care based on race, ethnicity, age, social risk factors, income, region, gender, primary language, disability, or other classifications, including a determination of any disparities identified earlier that are being reduced or eliminated
- Analysis of unintended consequences that have arisen from the use of the measure
- Validation and analysis of the exclusion, including, but not limited to:
  - Analysis of variability of use
  - Implications of rates
- Other performance information that CMS has collected or calculated, as available.

The measure developer compares the information gathered with projections made in the original business case and report the measure performance and the impact of the measure. Update the business case as appropriate and make projections for the next evaluation period.
30.2.3.3 **Convene a TEP**

During comprehensive reevaluation, a TEP is usually convened to assess the measure. It is best to continue with the TEP that worked on measure development. However, review the membership to ensure an appropriate breadth of expertise and diversity is still represented in the membership. [Section 3, Chapter 12, Technical Expert Panel](#), provides details of the standardized process of issuing a call for nominations and convening a TEP.

The measure developer presents the results of the environmental scan, literature review, and empirical data analysis of the measure performance data, patients’ perspective, and analysis of ongoing feedback received. If the patient perspective was not obtained by other means, patient representation should be recruited for this TEP. Develop recommendations on the disposition of the measure using the measure evaluation and selection criteria. [Section 3, Chapter 23, Measure Evaluation](#), describes the measure evaluation criteria. Measure selection criteria are discussed in [Section 2, Chapter 5, Measure Implementation](#) and [Section 3, Chapter 27, Measure Selection](#).

Summarize the TEP’s recommendations in the TEP report. Consider the TEP’s input to update the Measure Evaluation Report and make recommendations to CMS on the disposition of the measure.

30.2.3.4 **Identify and Document Changes that will Be Recommended**

For each measure, the measure developer compiles the information gathered in these steps using the measure evaluation criteria to update the [MIF](#). Complete the [Measure Evaluation Report](#) and compare the strengths and weaknesses of each measure to the previous evaluation.

If the measure has not been specified with ICD-10 codes, convert all ICD-9 codes to ICD-10 unless a look-back period is necessary. On January 16, 2009, HHS released the final rule mandating that everyone covered by the HIPAA must implement ICD-10 for medical coding by October 1, 2013. However, on April 1, 2014, the “Protecting Access to Medicare Act of 2014” H.R. 4302 bill was signed, which delayed the compliance date for ICD-10 from October 1, 2014, to October 1, 2015.

When maintaining [eCQM](#) value sets, it is important to align with the vocabulary recommendations made by Health IT Standards Committee Clinical Quality Working Group and Vocabulary Task Force. More information on these requirements is provided in [Section 3, Chapter 17, Codes, Code Systems, and Datasets](#).

The measure developer updates the [MIF](#) (i.e., HQMF, CQL, ELM, JSON, HTML, and [value sets](#) for eCQMs) with any new measure specifications and coding. All changes to measure specifications should be described in the MIF or in a separate summary of changes and release notes document. Any material or substantive changes should be identified and the purpose of the changes explained. A material or substantive change is one that changes the specifications of an endorsed measure to affect the original measure’s concept or logic, the intended meaning of the measure, or the strength of the measure relative to the measure evaluation criteria.

30.2.3.5 **Determine the Preliminary Recommended Disposition of the Measure**

Criteria that form the basis for the disposition decision for each measure and description of the possible outcomes are discussed in [Section 3, Chapter 30.5, Possible Outcomes of Maintenance Reviews](#).
The possible dispositions include:

- Retain
- Revise
- Remove
- Retire
- Suspend.

30.2.3.6 Test Measures as Necessary

For the first comprehensive reevaluation, the measure will require evaluation of reliability and validity beyond what occurred during measure testing at the time of development. If the measure is not in use, it will require expanded testing. The extent of measure testing or reevaluation of validity and reliability for measures in use and not in use are outlined Table 25.

Table 25. Extent of Measure Evaluation as a Function of Prior Comprehensive Evaluation and Measure Use

<table>
<thead>
<tr>
<th>Measure in Use</th>
<th>Measure Not in Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>First comprehensive reevaluation</td>
<td>Measure developer/measure maintenance contractor should conduct expanded testing</td>
</tr>
<tr>
<td></td>
<td>relative to the initial testing conducted during development (e.g., expand number</td>
</tr>
<tr>
<td></td>
<td>of groups/patients included in testing compared to prior testing used to support</td>
</tr>
<tr>
<td></td>
<td>the measure’s initial development and submission for endorsement).</td>
</tr>
<tr>
<td>Subsequent comprehensive reevaluations</td>
<td>If measure has not materially changed, measure developer/measure maintenance</td>
</tr>
<tr>
<td></td>
<td>contractor may submit prior testing data when past results demonstrated adequate</td>
</tr>
<tr>
<td></td>
<td>reliability and validity of the measure.</td>
</tr>
</tbody>
</table>

If the measure requires testing, develop a plan. The components of a testing plan are described in Section 2, Chapter 4, Measure Testing.

30.2.3.7 Obtain Public Comment on the Measure

If there have been substantive changes to a measure as the result of comprehensive reevaluation, public comment should be sought on those changes. The measure developer consults with the COR for approval to release the measure for public comment. If the comprehensive reevaluation results in a recommendation to retain the measure with only minor changes, it likely is not necessary to seek public comment. The process for obtaining public comment is found in Section 3, Chapter 14, Public Comment.

The measure developer analyzes the comments received and refines the measure as indicated. Document any changes in the MIF (i.e., HQMF, CQL, ELM, JSON, HTML, and value sets for eCQMs). If necessary, update the MJF and Measure Evaluation Report, as appropriate. Depending on the extent of measure revisions, it may be necessary to retest the measure iteratively as deemed necessary by the measure developer and the COR. Submit the revised measure and related documentation to the COR for approval.

30.2.3.8 Implement the Approved Action

After review, the measure developer/measure maintenance contractor may be responsible to help CMS implement the chosen measure disposition. For measures that are proposed to be revised, suspended, removed, or retired, evaluate the impact of the decision on the program using the measure when developing the implementation plan. If there are relevant regulatory or rulemaking schedules, include
them in the implementation plan. Communicate and collaborate with the COR to determine any deliverables and actions that are necessary (e.g., making announcements through usual communication modes for the project, arranging for reprogramming, notifying other CMS contractors, re-educating providers). Notify the Measures Manager to ensure that the CMS Measures Inventory is updated appropriately regardless of the disposition decision.

30.2.3.9 Maintain NQF Endorsement

NQF requires comprehensive review every 3 years to maintain continued endorsement. Endorsed measures are reevaluated against the NQF’s Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement and are reviewed alongside newly submitted (but not yet endorsed) measures. This head-to-head comparison of new and previously endorsed measures fosters harmonization and helps ensure NQF is endorsing the best available measures. The deliverables used for comprehensive reevaluation should be used to complete NQF maintenance submissions. NQF describes its maintenance requirements, including the schedule, on the NQF website.

Ideally, the comprehensive reevaluation should precede NQF scheduled review, so that CMS, along with the measure developers, can determine the outcome of the reevaluation and address any harmonization issues identified. Measure developers will need to factor the time required for testing significant changes into the timing of the comprehensive reevaluation.

NQF will notify CMS before a measure’s endorsement is due to expire. The notification will also appear on the measure developer’s NQF dashboard. The Measures Manager or the COR will confirm with the appropriate measure developer that the measure developer received NQF notice. NQF usually sends reminders and email notifications about the maintenance review due date; however, measure developers must be aware of NQF endorsement expiration dates and seek advice from their COR or NQF if they have not received notification of an endorsement maintenance review.

NQF will send a standardized online submission template for the 3-year endorsement maintenance review to the measure steward of record. The form will be prepopulated with information from the original or most recent annual update submission. CMS notifies NQF regarding the appropriate measure developer contact for the 3-year endorsement maintenance review.

The 3-year maintenance review report documents the review of the current evidence and guidelines and provides information about how the measure still meets the criteria for NQF endorsement. The measure developer will use information from the most recent comprehensive reevaluation, subsequent annual updates, and ongoing surveillance to complete the NQF submission form. Following COR approval, the measure developer submits the report to NQF.

30.3 CMS Ad Hoc Review

A CMS ad hoc review is a limited examination of the measure based on new information. If evidence comes to light that may have a significant, adverse effect on the measure or its implementation, an ad hoc review must be conducted. Ad hoc reviews must be completed as quickly as possible regardless of annual or 3-year scheduled comprehensive reviews because of the nature of the triggering information. The ad hoc review process ensures that the CMS measures remain balanced between the need for measure stability and the reality that the measure environment is constantly shifting. The urgency of an ad hoc review reflects those shifts; to preserve measure stability, it should be reserved for only those instances when new evidence indicates that a very significant revision may be required.
Ad hoc review specifically does not include the process of adapting or harmonizing a measure for use with a broader or otherwise different population. Requirements associated with an ad hoc review (Deliverable 4-11) are listed in Figure 51.

30.3.1 Trigger for an Ad Hoc Review

The potential ad hoc review begins when the measure developer becomes aware of evidence that may have a significant, adverse effect on the measure or its implementation. The evidence may be obtained through the measure developer’s ongoing surveillance of the scientific literature, or from the Measures Manager, CMS, and other stakeholders.

If the measure is NQF-endorsed, the NQF may have received a request for an ad hoc review and may have contacted CMS because it is the steward. CMS may then ask the measure developer to investigate the situation and conduct its own ad hoc review—even if NQF has declined to conduct an ad hoc endorsement review. If NQF has decided to conduct an ad hoc review, the measure developer will be asked to help CMS assess the situation and provide information for NQF review. NQF ad hoc reviews may also be initiated at the request of CMS for specific situations, such as the need to significantly change measure specifications outside of the usual maintenance cycle. For example, NQF has required CMS to harmonize a measure before the next maintenance review; however, CMS needs the revised measure prior to that time due to program or legislative requirements and must use an endorsed measure.

CMS reserves the right to conduct an ad hoc review for any reason, at any time, on any measure. Nothing in this Blueprint is intended to limit the options CMS may exercise.

30.3.2 Deferring an Ad Hoc Review

The measure developer should postpone an ad hoc review to the next scheduled review if that is reasonable. The timing of the ad hoc review will be influenced by the presence of any accompanying patient safety concerns associated with the changes to the endorsed measure. If the measure will be updated or reevaluated in the near future, the information received should be incorporated into that update or reevaluation. For example, if the measure is due for a comprehensive reevaluation or an annual update within the next 120 days, the information should be referred to the team conducting the review, and that team should incorporate the ad hoc review process into its work.

Because measures are used in specific programs that may have their own schedules (e.g., hospital measures that are governed by specific rulemaking schedule requirements), a decision may take some time to be implemented in all of the programs using a given measure.

---

**Ad Hoc Review Requirements**

**Ad Hoc Review Requirements (Deliverable 4-11)**

1. Updated MIF (Deliverable 3-3), if the ad hoc review results in changes to the measure specifications.
2. For measure developers maintaining eCQMs, the updated HQMF, CQL, ELM, JSON, HTML, and value sets, if the ad hoc review results in changes to the measure specifications.
3. Updated MJF (Deliverable 3-4), reflecting the new information that triggered the review, any additional information used in the decision-making process, and the rationale for the outcome of the review.
4. Updated Measure Evaluation Report (Deliverable 4-2), if the review results in a change to the measure’s strengths and/or weaknesses.
30.3.3 Procedure

The CMS measure developer continues to be responsible for monitoring the maintenance performed by the steward—even if the measure developer is not the measure steward and, therefore, is not ultimately responsible for maintaining the measure. This includes ensuring that the measure is updated periodically in response to changes in the underlying code systems (e.g., CPT, ICD-10-CM, LOINC) and is reevaluated in a manner consistent with the Blueprint. The CMS measure developer will also be responsible for ongoing surveillance of the literature addressing the measure and alerting the COR to possible issues.

If a significant concern is identified about a measure for which CMS is not the steward, the measure developer responsible for monitoring the measure should alert the COR to determine what action, if any, is necessary. CMS may contact the steward to determine whether the steward is aware of the concern and what action is being taken. If the measure is NQF-endorsed, CMS may consider requesting that NQF conduct an ad hoc maintenance review. CMS has the option of suspending data collection pending the outcome of any action by the steward and NQF, or CMS may choose to remove the measure from the program.

The ad hoc review process includes seven steps comprising three primary subparts:

- Determining whether an ad hoc review should be conducted
- Conducting the review and recommending an outcome
- Approving and implementing the approved outcome.

30.3.3.1 Determine Whether the Concern Is Significant

If the clinical practice underlying the measure is causing harm to the patients, the measure should be revised, suspended, removed, or retired. This includes harm caused by unintended consequences of the measure. Although there is no defined schedule for this process, CMS or NQF may require the measure developer/measure maintenance contractor to give the concern urgent attention. If measure revision is not feasible in the time frame necessary, the measure should be suspended or retired.

If no such harms are projected, only the strongest concerns will result in an ad hoc review. The measure developer monitoring the measure should consider first whether the issue is significant and then may engage the TEP most recently involved with the measure. If the measure developer does not have access to the TEP, then the measure developer may contact a professional association closely associated with the measure for input regarding the significance of the issue raised. The NQF may also be the source of the request for urgent ad hoc review depending on the nature and source of the concerns.

If experts determine that the issue is significant or if they cannot agree on its significance, the measure developer should notify CMS of the situation and propose conducting a full ad hoc review (i.e., the remaining steps). If the measure maintenance contractor is not the measure developer monitoring the measure, the measure maintenance contractor is responsible for the review.

If experts determine that the issue is not significant, the issue should be documented for consideration at the next scheduled review.

30.3.3.2 Conduct Focused Information Gathering

Unlike environmental scans conducted during measure development, ongoing surveillance, or comprehensive reevaluation, the scan performed for an ad hoc review is limited to new information directly related to the issue that triggered the review. Not all aspects of the measure must be investigated—only the aspect that generated concern.
The measure developer conducts a literature review to determine the extent of the issues involved and to identify significant areas of controversy if they exist. Guidance for conducting and documenting the environmental scan (including literature review) is detailed in Section 3, Chapter 9, Information Gathering.

### 30.3.3.3 Consult with the Experts, Especially the TEP

The measure developer consults with the TEP that contributed to the most recent comprehensive reevaluation or measure development, if that is feasible.

If the issue generating the concern relates to clinical guidelines, ask the organization responsible for the guidelines about its plans for updating the guidelines or issuing interim guidelines. The professional organization most closely related to the measure may also be consulted.

Ask the experts (e.g., TEP, guideline writers, professional organizations) about the:

- Significance of the issue, to confirm that they consider it important as well
- Risk of possible patient harm if the measure remains in use, including harm from unintended consequences
- Feasibility of implementing measure revisions, including cost and time.

### 30.3.3.4 Determine Whether It Is Feasible to Change the Measure

The feasibility of changing a measure should include consideration of the cost of resources associated with data collection, measure calculation, and reporting systems, including those requiring updates to vendor systems. Depending on the resources available and the time involved in making the changes necessary, the measure may be either revised immediately or suspended until the systems can be updated with the measure’s updated specifications.

### 30.3.3.5 Recommend a Course of Action to the COR

Criteria that form the basis for the disposition decision for each measure and description of the possible outcomes are discussed in Section 3, Chapter 30.5, Possible Outcomes of Maintenance Reviews.

Depending on the findings from the previous steps, the recommendation may be:

- Retain
- Revise
- Remove
- Retire
- Suspend.

The measure developer submits the recommendation along with supporting documentation and the updated MIF and Measure Evaluation Report (if recommending immediate revision or suspension until revision is possible) to the COR.

### 30.3.3.6 The COR Reviews the Recommendation for Approval

The measure developer forwards the recommendations to the COR, with the updated MIF and summary of changes or Release Notes as indicated. If significant changes were made, an updated MJF and Measure Evaluation Report may be necessary.\(^{53}\)

---

\(^{53}\) The NQF submission may be acceptable for these deliverables.
The COR will review the submitted documentation. If the recommendation is approved, the COR notifies the measure developer of the approved course of action. If the measure developer’s recommendation is not approved, the COR documents an approved course of action and instructs the measure developer as necessary.

30.3.3.7 Implement the Approved Action

For measures that are proposed to be revised, suspended, removed, or retired, evaluate the impact of the decision on the program using the measure when developing the implementation plan. If there are relevant regulatory or rulemaking schedules, include them in the implementation plan.

After review, the measure maintenance contractor may be responsible to help CMS implement the chosen measure disposition. Communicate and collaborate with the COR to determine any deliverables and actions that are necessary (e.g., making announcements through usual communication modes for the project, arranging for reprogramming, notifying other CMS measure developers, re-educating providers). Notify the Measures Manager to ensure that the CMS Measures Inventory is updated appropriately regardless of the disposition decision.

30.4 NQF Ad Hoc Reviews

The COR will notify NQF of all relevant activities and changes to the measure. If the CMS ad hoc review process results in retirement, removal, or measure revision, the COR will or direct the measure developer to notify the NQF. Any significant changes made to a measure could also trigger an NQF ad hoc endorsement review. The measure developer should be available to answer NQF questions about the ad hoc review process and results. Refer to the NQF website for current NQF measures maintenance policies that apply to updated measures.

NQF also has a process for initiating and conducting an ad hoc review of its own. These can come from requests received by NQF and must meet one or more criteria:

- The evidence supporting the measure, practice, or event has changed, and it no longer reflects updated evidence.
- There is evidence that implementation of the measure or practice may result in unintended consequences.
- Use of the measure or practice may result in inappropriate or harmful care.
- Measure performance scores may yield invalid conclusions about quality of care (e.g., misclassification or incorrect representation of quality).
- Material changes have been made to a currently endorsed measure.

NQF will notify the measure steward of the request and evidence presented by the requestor and will indicate the response and format required.

NQF ad hoc reviews may be requested at any time by any party. Adequate evidence to justify the review and under which criterion the review is requested must be submitted when seeking an ad hoc maintenance review. NQF reviews the request and initiates an ad hoc review if there is adequate evidence justifying a review. The timing of the ad hoc review will be determined by the presence of any accompanying safety concerns associated with the changes to the endorsed measure.

If NQF has received a request for an ad hoc review, they will notify the steward whether they have determined that there is sufficient evidence to conduct the ad hoc review. If an NQF ad hoc review is requested for a measure supported by the measure developer, the measure developer is expected to
help CMS respond to the request from NQF. NQF currently does not use a standardized form for the ad hoc review. The measure developer and CMS should meet with NQF to discuss the request and clarify the types of information that should be submitted and the timeline for the ad hoc maintenance review.

30.5 Possible Outcomes of Maintenance Reviews

Based on recommendations made as a result of any of the three maintenance review types discussed in this chapter, CMS has five measure dispositions from which to choose, as outlined in Section 3, Chapter 27.4, Measure Maintenance.

If the measure continues to meet the measure evaluation criteria (Section 3, Chapter 23, Measure Evaluation) and the measure selection criteria (Section 3, Chapter 27, Measure Selection) used by CMS to place it in a program, it will be retained or revised with minor changes and updates. If a measure is going to be retired or removed, the measure developer should consider recommending other available qualifying measures as replacements.
Section 4. Forms and Templates
1 **ENVIRONMENTAL SCAN OUTLINE**

An environmental scan outline includes:

1. **Cover Page**, including the task order title and contract number, measure developer contact information, and COR’s name and contact information

2. Table of Contents

3. Executive Summary

4. **Background and Significance**, including a description of the problem addressed, purpose of measurement, and anticipated outcome of measurement

5. Literature Review, including:
   - Search methods, including a complete explanation of all research tools used:
     - All online publication directories
     - Sources selected from traditional journals and grey literature (e.g., website, conference proceedings)
     - Keyword combinations
     - Boolean logic used to find studies and clinical practice guidelines
   - Complete literature citations
   - Level of evidence and rating scheme used
   - Characteristics of reviewed studies:
     - Population
     - Study size
     - Data sources
     - Study type
     - Methods
     - Identification of measure evaluation criteria the study supports (i.e., importance, scientific acceptability, usability, and feasibility)
     NOTE: Sorting the literature review by these criteria will facilitate development of the MJF in the later phases of measure development or reevaluation.
   - Information gathered to build the business case for the measure:
     - Incidence/prevalence of a condition in Medicare population
     - Major benefits of the process or intermediate outcome under consideration for the measure
     - Untoward effects of process or intermediate outcome and likelihood of their occurrence
     - Cost statistics relating to cost of implementing the measured process, as well as savings that result from implementing the process, and costs of treating any complications that may arise
     - Current performance of process or intermediate outcome and identification of gaps in performance
     - Size of improvement that is reasonable to anticipate
   - Summary of findings
   - Other pertinent information, if applicable
6. **Summary of Clinical Practice Guidelines Review**, including the following information (by measure set; or, if needed, provide for individual measures in the set):

- Guideline name
- Developer
- Year published
- Summary of major recommendations
- Level of evidence
- If multiple guidelines exist, note inconsistencies and rationale for using one guideline over another

7. **Review of Regulations and their Implications on Measurement**, limited to new regulations affecting measurement (e.g., MACRA):

- Regulation or rule name
- Agency responsible
- Law to which it responds
- Year published
- Summary of major implications
- If multiple regulations exist, enumerate them by Act, agency, and year

8. **Review of Existing Measures, Related Measures, and Gap Analysis Summary**, including a summary of findings and measurement gaps:

- Existing related measures, including stewards
- Gap analysis
- Opportunities for harmonization

9. **Empirical Data Analysis Summary**

- New measures:
  - Data source(s) used, if available
  - Time period
  - Methodology
  - Findings
- Measure reevaluation; use the Measure Evaluation Form

10. **Expert Input**

- TEP
  - List of members and attendees of all meetings
  - Meeting summaries, any individual discussions, and additional pertinent information (e.g., Delphi results)
  - Recommendations
- Other experts
  - List of additional experts and purposes for their input
  - Manner of interaction (e.g., telephone call, face-to-face meeting, survey)
  - Summary of findings with recommendations
• Stakeholders, highlighting persons and family
  o List of stakeholders and their relevance to the project
  o Manner of interaction (e.g., telephone call, face-to-face meeting, survey, focus groups, online panels)
    (Note: Names of individuals solicited for TEPs, interviews, and other forms of expert input are not required for patients, family members, and caregivers if revealing names would inhibit participation.)
  o Summary of findings with recommendations
  o Summary of solicited and structured interviews, if applicable (might refer to any of the expert types)
    ▪ Summarize overall findings from the input received
    ▪ Name of the person(s) interviewed, type of organization(s) represented, date(s) of interview, the area of quality measurement expertise if the input was from patients or other consumers, etc.
    ▪ List of interview questions
    ▪ Qualitative evaluation of findings with implications for measurement and overall recommendations

11. Conclusion, with overall discussion of measurement implications, perhaps including future and ideal states
2 BUSINESS CASE FORM INSTRUCTIONS

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, CMS’s Section 508 Guide for Microsoft Word 2013 may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

<PLEASE DELETE THIS SECTION BEFORE SUBMISSION.>

This form is a guide for measure developers to use when documenting the business case. The form is not required, but it is provided to help measure developers fulfill the deliverable requirement of submitting an adequate business case for the measure under development or being reevaluated during maintenance.

The form includes instructions for making a business case that the measure:

- Contributes to better health
- Promotes better care
- Leads to more affordable care.

Project Title:
<List the project title as it should appear.>

Project Overview:
The Centers for Medicare & Medicaid Services (CMS) <has contracted with / provided funding to> <measure developer name> to develop <measure (set) name or description>. The <contract / cooperative agreement / other (please specify)> name is <insert contract name>. The <contract / cooperative agreement / other (please specify)> number is <project number.>

Date:
Information included is current on <insert date>.

Measure Description:
Use the Measure Title, as it is listed in the MIF. It should be brief and include the measure focus and the target population.

Numerator Statement:
All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets should be presented here.
Denominator Statement:

Provide a brief, narrative description of the target population being measured.

Business Case Report Executive Summary:

Summarize the case—what is being measured and how the measure will contribute to better health, promote better care, and lead to more affordable care. The executive summary of business case conclusions should be presented here.

Incidence and prevalence data should be presented, highlighting any disparities that may exist.

The purpose of providing these data is to determine the size of the population to be included in the denominator of the proposed measure. These data can be found from the literature and from empirical analysis of available data sources. Particular attention should be given to disparities. If the incidence and prevalence vary by sociodemographic factors, include those statistics as well. The purpose of providing information on disparities is to determine the current baseline of the measure and demonstrate that there are gaps in performance. Mortality and morbidity statistics relating to the process or outcome under consideration should be reported. If disparities are found, describe the current performance by subpopulations. Use the references obtained through information gathering.54

Measure Uses (select all that apply):

Check all the current and planned uses for the measure.

- Public Reporting
- Public Health/Disease Surveillance
- Payment Program
- Regulatory and Accreditation Programs Payment and Network Selection
- Professional Certification or Recognition
- Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
- Quality Improvement (internal to the specific organization)
- Not in use.

Current Performance (including any disparities):

The purpose of this item is to determine the current baseline of the measure and demonstrate that there are gaps in performance. Mortality and morbidity statistics relating to the process or outcome under consideration should be reported. If disparities are found, describe the current performance by subpopulations. Use the references obtained through information gathering.55

Measure Impact on Care and Health Outcomes:

Estimate the expected performance of the measure on the quality of care and on health outcomes. If improvement is expected in certain subpopulations, use stratified estimates. Quantify the size of improvement that is reasonable to expect based on literature, performance of similar measures, and construction of the measure. Provide a time frame and trajectory for the anticipated improvements.

54 Concepts incorporated from Business Case sample prepared by Yale for THA/TKA resource use measure in 2014.
55 Concepts incorporated from Business Case sample prepared by Yale for THA/TKA resource use measure in 2014.
During measure maintenance, compare the actual performance to the estimates and report the
differences with analysis and recommendations.\textsuperscript{56}

\textbf{Measure Impact on Healthcare Costs (if any):}

Estimate the expected performance of the measure on healthcare costs. Follow the approach detailed
for Measure Impact on Care.

\textbf{Influencing Factors:}

There may be factors that influence adoption, implementation, and endorsement of a measure; quality
of care; and outcomes resulting from the measure. This may include legislation and regulation,
endorsements, competitive market pressures, data infrastructure, stakeholder inputs, and technical
assistance. Anticipated influencing factors should be discussed, and data should be provided to
document any observed influencing factors affecting measure implementation and/or performance.

\textbf{Resources Required for Measure Implementation:}

There may be costs to capture and report measure data, including the use of staff time, software, etc.
These costs should be estimated, calculated, and reported in the business case.

\textbf{Costs of Clinical Care:}

There may be a cost of clinical care required to improve performance. For process measures
of underuse, the additional cost of receiving the recommended care should be included in the discussion.
This may also apply to outcome measures if additional care is needed to improve outcomes. These and
other related costs should be estimated, calculated, and reported in the business case.\textsuperscript{57}

\textbf{Potential Unintended Consequences of the Measure (if any):}

Document the incidence of untoward effects of the process being measured as reported in the literature
initially and during maintenance. Report the costs of treating potential unintended complications.\textsuperscript{58}

\textbf{Description of Model(s) and Formulas Used:}

Describe the assumptions, variables, and formulas used to construct the business case.\textsuperscript{59}

\textbf{Limitations of Analysis:}

Describe any limitations in the data or the assumptions used in the business case.\textsuperscript{60}

\textsuperscript{56} Concepts incorporated from Business Case sample prepared by QP Rhode Island for Antipsychotic Polypharmacy and Warfarin Monitoring in 2009.
\textsuperscript{57} Concepts incorporated from Business Case sample prepared by FMQAI for Glycemic Control measures in 2014.
\textsuperscript{58} Concepts incorporated from Business Case sample prepared by FMQAI for Glycemic Control measures in 2014.
\textsuperscript{59} Concepts incorporated from Business Case sample prepared by FMQAI for Glycemic Control measures in 2014.
\textsuperscript{60} Concepts incorporated from Business Case sample prepared by FMQAI for Glycemic Control measures in 2014.
**Net Benefit:**

Describe the anticipated (or for maintenance, realized) benefits associated with the measure. Net benefits include, but are not limited to:

- Lives saved
- Functional status improvements
- Patient experience and perception improvements
- Reduced complications, readmissions, etc.
- Cost savings to Medicare, patients, providers, or other stakeholders\(^{61}\)

---

\(^{61}\) Concepts incorporated from Business Case samples prepared by FMQAI for Glycemic Control measures in 2014 and by QP Rhode Island for Antipsychotic Polypharmacy and Warfarin Monitoring in 2009.
3 BUSINESS CASE TEMPLATE

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, CMS’s Section 508 Guide for Microsoft Word 2013 may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

<PLEASE DELETE THIS SECTION BEFORE SUBMISSION.>

<Instructions for completing this form: Measure developers should submit an INITIAL Business Case during the Measure Conceptualization process and present a FINAL business case before Measure Implementation begins. Although some of the data and details may be limited at first, it is expected that the measure developer will provide detailed information in the final submission.>

<CMS has intentionally aligned this form with NQF submission forms when appropriate. In some cases, a measure developer may be able to use text from their NQF submission to complete this form and vice versa. This practice is accepted and encouraged by CMS. To facilitate this practice, this template indicates when a field is also required for an NQF submission.>

<The business case is about more than the possible financial benefits of a measure. Include all benefits expected to result from the measure.>

Project Title:

<List the project title as it should appear.>

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) has <contracted with / provided funding to> <measure developer name> to develop <measure (set) name or description>. The <contract / cooperative agreement / other (please specify)> name is <insert contract name>. The <contract / cooperative agreement / other (please specify)> number is <project number>.

Date:

Information included is current on <insert date>.

Measure Description:

Use the Measure Title, as it is listed in the MIF. It should be brief and include the measure focus and the target population.

<Field part of NQF Measure Submission Form, Field De.3. Text from NQF submission may be inserted here if available.>
**Numerator Statement:**

All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets.

<Field part of NQF Measure Submission Form, Field S.4. Text from NQF submission form may be inserted here if available.>

**Denominator Statement:**

Provide a brief narrative description of the target population being measured.

<Field part of NQF Measure Submission Form, Field S.6. Text from NQF submission may be inserted here if available.>

**Business Case Report Executive Summary:**

Summarize the business case, including what is being measured and how the measure will contribute to better health, promote better care, and lead to more affordable care.

Incidence and prevalence data should be presented, highlighting any disparities that may exist.

<Information may be detailed in NQF Measure Submission Form, fields De.3 and 1b.1. Text from NQF submission may be inserted here if available.>

**Measure Uses (select all that apply):**

- Public Reporting
- Public Health/Disease Surveillance
- Payment Program
- Regulatory and Accreditation Programs Payment and Network Selection
- Professional Certification or Recognition
- Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
- Quality Improvement (internal to the specific organization)
- Not in use

<Information may be detailed in NQF Measure Submission Form, Usability and Use, field 4.1. Text from NQF submission may be inserted here if available.>

**Current Performance (including any disparities):**

The purpose of this item is to determine the current baseline of the measure and demonstrate that there are gaps in performance. Mortality and morbidity statistics relating to the process or outcome under consideration should be reported. If disparities are found, describe the current performance by subpopulations. Use the references obtained through information gathering.\(^\text{62}\)

<Information can be compiled from fields in NQF Measure Submission Form, including fields 1b.2-1b.5. Text from NQF submission may be inserted here if available.>

---

\(^{62}\) Concepts incorporated from Business Case sample prepared by Yale for THA/TKA resource use measure in 2014.
**Measure Impact on Care and Health Outcomes:**

Describe the linkages and steps between the measure focus and anticipated improvements to care and health outcomes provided and received. Include details on structure, process, intermediate outcome(s), health outcome(s), and any evidence to support this. Provide supporting evidence as appropriate.

*Information can be compiled from fields in NQF Measure Evidence Form. Text from NQF submission may be inserted here if available.*

**Measure Impact on Healthcare Costs (if any):**

If no anticipated impact, state “none.” Include a brief explanation of why there is still a strong business case for measure (e.g., lives saved, improved health outcomes).

**Influencing Factors:**

Describe factors that may influence adoption, implementation, and endorsement of a measure; quality of care; and outcomes resulting from the measure. This may include legislation and regulation, endorsements, stakeholder feedback, competitive market pressures, data infrastructure, and technical assistance. If there are any concerns about the feasibility of implementing a measure, or for the healthcare organizations to respond to the measure, those should be explicitly stated in this section.

**Resources Required for Measure Implementation:**

Describe and quantify the resources necessary to implement the measure, including staff time and other direct and indirect costs.

*Information may be detailed in NQF Measure Submission Form, Data Collection Strategy, fields 3c.1 and 3c.2. Text from NQF submission may be inserted here if available.*

**Costs of Clinical Care:**

Describe and quantify the resources necessary to implement the changes in care the measure seeks to change, including staff time and other direct and indirect costs.

**Potential Unintended Consequences of the Measure (if any):**

*Field part of NQF Measure Submission Form, field 4b2.1. Text from NQF submission may be inserted here if available.*

**Description of Model(s) and Formula(s) Used:**

*Field part of NQF Measure Submission Form, Measure Specification, including fields S.12-22. Text from NQF submission may be inserted here if available.*

**Limitations of Analysis:**

*Information provided as part of NQF Measure Submission Form. Text from NQF submission may be inserted here if available.*
**Net Benefit:**

Report benefits (i.e., anticipated benefits if measure not yet implemented) and progress toward improvement in health outcomes, quality of care, and/or lower costs of care.

*Field part of NQF Measure Submission Form, including fields 4b2.1 and 1b2.2 and 1b.1.*
4 CALL FOR MEASURES WEB POSTING

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, CMS’s Section 508 Guide for Microsoft Word 2013 may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

<PLEASE DELETE THIS SECTION BEFORE SUBMISSION.>

Project Title:

<List the project title as it should appear on the Web posting.>

Dates:

The Call for Measures period opens on <list the date> and closes on <list the date>.

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) has <contracted with / provided funding to> <measure developer name> to develop <measure (set) name or description>. The <contract / cooperative agreement / other (please specify)> name is <insert contract name>. The <contract / cooperative agreement / other (please specify)> number is <project number>. As part of its measure development process, <measure developer name>requests that interested parties submit candidate or concept measures that may be suitable for this project.

Project Objectives:

<List the contract objectives.>

Instructions:

When submitting measures and/or concepts for consideration, include all required documentation using the instructions below:

- All submissions must be received by the end of the Call for Measures period.
- Email the completed form and any attachments to: <insert email address>.
- Submit the candidate measures and/or concepts with relevant information from the MIF for each.
- If you are submitting fully developed or endorsed measures, attach any additional measure information to the email.
- Provide appropriate contact information with the submission.
5 **MEASURE INFORMATION FORM INSTRUCTIONS**

This form is a guide for your use when submitting measures. You may use it to draft your responses for each section in preparation for entering them in the measure submission. The MIF tracks very closely to the NQF online measure submission, Version 7.1, and references corresponding fields from that submission form in the parentheses. Developers may submit the NQF submission form in lieu of the MIF.

Information from the MIF and the MJF may be used for other purposes, so developers may be asked to complete the MIF for measures that are not submitted to NQF.

**1. Measure Name/Title (NQF Submission Form De.2.)**

Briefly convey as much information as possible about the measure focus and target population.

**2. Descriptive Information**

2.1 Measure Type (NQF Submission Form De.1.)

Identify a measure type from among the list. PROs include health-related quality of life, functional status, symptom burden, experience with care, and health-related behavior:

- **Process**
- Process: Appropriate Use
- **Outcome**
- Outcome: PRO
- Cost / **Resource Use**
- **Efficiency**
- **Structure**
- Intermediate Outcome
- **Composite**.

2.2 Brief Description of Measure (NQF Submission Form De.3.)

This description should also be concise, but include type of score, measure focus, target population, and time frame.

2.3 If Paired or Grouped (NQF Submission Form De.4.)

Provide the reason why the measure must be reported with other measures to appropriately interpret results.

**3. Measure Specifications**

These items follow the NQF requirements for measure submission and provide information required for measure evaluation.

3.1 Measure-specific Web Page (NQF Submission Form S.1.)

Provide a URL link to a web page where current, detailed specifications can be obtained, including code lists, **risk adjustment** model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.
3.2 If this is an eCQM (NQF Submission Form S.2a.):

HQMF specifications and Bonnie testing must be attached. Attach the zipped output from the MAT and Bonnie tool. If the MAT or Bonnie was not used, contact NQF staff. Use the specification fields from the online form for the plain language description of the specifications.

3.3 Data Dictionary, Code Table, or Value Sets (NQF Submission Form S.2b.)

Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). The preferred file format is either .xls or .csv. If those are not used, contact the NQF for further directions.

3.4 For an instrument-based measure (NQF Submission Form S.2c and S.2d):

Attach copy of instrument if available.

Indicate the responder (i.e., patient, family or other caregiver, clinician, or indicate if not an instrument-based measure).

3.5 Updates since last submission (NQF Submission Form S.3.1 and S.3.2)

Are there changes to the specifications since the last updates/submission? If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

Briefly describe any changes to the measure specifications since the last endorsement date and explain the reasons for the changes.

3.6 Numerator Statement (NQF Submission Form S.4.)

Briefly describe the measure focus or what is being measured about the target population—cases from the target population with the target process, condition, or event based on the evidence.

For example:

Patients in the target population who received/had [measure focus] [during [time frame] if different than for target population

Do not include the rationale for the measure.

For outcome measures, state the outcome being measured. Describe calculation of the risk-adjusted outcome later in the calculation algorithm (NQF Submission Form S.14).

3.7 Numerator Details (NQF Submission Form S.5.)

Include all information necessary to identify and calculate the cases from the target population with the target process, condition, event, or outcome. Provide definitions and specific data collection items and responses. For measures based on a coded data set, identify the code set, the specific codes, and the code descriptors. (If using this to format submissions to the NQF and the list of codes and descriptors exceeds one page, provide the list in a Microsoft Excel or .csv file.)

For outcome measures, describe how the observed outcome is identified and counted. The calculation algorithm should also describe how to calculate the risk adjustment (NQF Submission Form S.14).

Provide the time period in which data will be aggregated for the measure (e.g., 12 months, 3 years, another specified look-back period).
3.8 **Denominator Statement** (NQF Submission Form S.6.)

Proved a narrative description of the broadest population (based on the evidence) for which the target process, condition, event, or outcome is applicable. Include the time period in which data will be aggregated for the measure, if different than the numerator.

Example:

Patient’s [age] with [condition] in [setting] during [time frame]

For outcome measures, state the target population for the outcome. The **calculation algorithm** should also describe how to calculate the risk adjustment (NQF Submission Form S.14).

3.9 **Denominator Details** (NQF Submission Form S.7.)

Provide all definitions and instructions needed to identify and calculate the target population/denominator (e.g., definitions, time period for data collection, specific data collection items/responses, codes/value sets). For measures based on a coded data set, identify the code set, the specific codes, descriptors, definitions, and specific data collection items as appropriate. (If using this to format submissions to NQF and the list of codes and descriptors exceeds one page, provide the list in an .xls or .csv file in the required format listed in NQF Submission Form S.2b.).

For outcome measures, describe how the target population is identified. The **calculation algorithm** should also describe how to calculate the risk adjustment (NQF Submission Form S.14).

3.10 **Denominator Exclusions** (NQF Includes “Exception” in the “Exclusion” Field) (NQF Submission Form S.8.)

Identify patients in the target population who should not receive the process (i.e., medical treatment), or are not eligible for the outcome for some other reason, particularly if their inclusion may bias results. Exclusion should be evidence-based.

Example:

Patients in the [target population] who [have some additional characteristic, condition, procedure]

3.11 **Denominator Exclusion Details** (NQF Includes “Exception” in the “Exclusion” Field) (NQF Submission Form S.9.)

All information needed to identify and calculate exclusions from the denominator (e.g., definitions and/or specific data collection items and responses). For measures based on a coded data set, identify the code set, specific codes, descriptors, definitions, and specific data collection items for the codes as appropriate. Lists of individual codes with descriptors that exceed one page should be provided in an .xls or .csv file in the required format listed in S.2b.

3.12 **Stratification Details/Variables** (NQF Submission Form S.10.)

Provide instructions for calculating the measure by category (e.g., age), including the stratification variables, all codes, logic, definitions, specific data collection items/responses, and the risk model covariate and coefficients for the clinically adjusted version of the measure, when appropriate. Lists of individual codes with descriptors that exceed one page should be provided in an .xls or .csv file in the required format listed in S.2b.
3.13 Risk Adjustment Type (NQF Submission Form S.11.)

Select risk adjustment type. Provide specifications for risk stratification in 3.14 (NQF Submission Form S.12.) and for statistical model in 3.16-3.17 (NQF Submission Form S.14.–15.):

- No risk adjustment or risk stratification
- Stratification by risk category/subgroup
- Statistical risk model
- Other (S.13.a.).

3.14 Type of Score (NQF Submission Form S.12.):

- Count
- Rate/proportion
- Ratio
- Categorical (e.g., yes or no)
- CV (e.g., an average)
- Other (specify).

3.15 Interpretation of Score (NQF Submission Form S.13.)

Provide interpretation that classifies whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score.

3.16 Calculation Algorithm/Measure Logic (NQF Submission Form S.14.)

Describe the sequence of steps necessary to calculate the measure score, including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; time period of data; and any other calculations.

You may provide a diagram of the Calculation Algorithm/Measure Logic described in S.14. at a measure-specific web page URL identified in S.1. or in an attached appendix.

3.17 Sampling (NQF Submission Form S.15.)

If the measure is based on a sample or survey, provide instructions for obtaining the sample and conducting the survey, along with minimum response rate required. If the measure is an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

3.18 Survey/Patient-Reported Data (NQF Submission Form S.16.)

If the measure is based on a survey or instrument, provide instructions for data collection and guidance on the minimum response rate. If the measure is a PRO-PM, specify calculation of response rates to be reported with performance measure results.

Specify how missing data are handled (e.g., imputation, delete case). This item is required for composite measures and PRO-PMs.

3.19 Data Source (NQF Submission Form S.17.)

Indicate all sources for which the measure is specified and tested:

- Administrative data
- Claims data
• Patient medical records (i.e., paper-based or electronic)
• Electronic clinical data
• Registries
• Standardized patient assessments
• Patient-reported data and surveys
• Non-medical data
• Other—describe in 3.20 (NQF Submission Form S.18.).

3.20 Data Source or Collection Instrument (NQF Submission Form S.18.)
Identify the specific data source/data collection instrument (e.g., name of database, clinical registry, collection instrument). If the measure is instrument-based (e.g., PRO-PM), identify the specific tools/instruments being used to collect the measure information and standard methods, modes, and languages of administration.

3.21 Data Source or Collection Instrument (Reference) (NQF Submission Form S.19.)
Provide the reference for the data source or collection instrument. Either attach a copy or specify the URL where it can be found.

3.22 Level of Analysis (NQF Submission Form S.20.)
Indicate only the levels for which the measure is specified and tested:
• Clinician: Individual
• Clinician: Group/Practice
• Facility
• Health Plan
• Integrated Delivery System
• Population: Community, County, or City
• Population: Regional and State
• Other.

3.23 Care Setting (NQF Submission Form S.21.)
Indicate only the settings for which the measure is specified and tested:
• Ambulatory Surgery Center
• Clinician Office/Clinic
• Outpatient Rehabilitation
• Urgent Care – Ambulatory
• Behavioral Health: Inpatient
• Behavioral Health: Outpatient
• Dialysis Facility
• Emergency Medical Services/Ambulance
• Emergency Department
• Home Health
• Hospice
• Hospital
• Hospital: Critical Care
• Hospital: Acute Care Facility
- Imaging Facility
- Laboratory
- Pharmacy
- Nursing Home / SNF
- IRF
- Long-Term Acute Care
- Birthing Center
- No Applicable Care Setting
- Other.

3.24 Composite Performance Measure (NQF Submission Form S.22.)

This section is for additional specifications as needed. Use it for aggregation and weighting rules or calculation of individual performance measures if they were not individually endorsed.
6 BLANK MEASURE INFORMATION FORM TEMPLATE

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, CMS’s Section 508 Guide for Microsoft Word 2013 may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

<PLEASE DELETE THIS SECTION BEFORE SUBMISSION.>

Project Title:

<List the project title as it should appear.>

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) has <contracted with / provided funding to> <measure developer name> to develop <measure (set) name or description>. The <contract / cooperative agreement / other (please specify)> name is <insert contract name>. The <contract / cooperative agreement / other (please specify)> number is <project number.>

Date:

Information included is current on <insert date>.

1. Measure Name (Measure Title De.2.)
2. Descriptive Information
   2.1 Measure Type (NQF Submission Form De.1.)
   2.2 Brief Description of Measure (NQF Submission Form De.3.)
   2.3 If Paired or Grouped (NQF Submission Form De.4.)
3. Measure Specifications
   3.1 Measure-specific Web Page (NQF Submission Form S.1.)
   3.2 If this is an eCQM (NQF Submission Form S.2a.)
   3.3 Data Dictionary, Code Table, or Value Sets (NQF Submission Form S.2b.)
   3.4 For Instrument-Based Measure (NQF Submission Form S.2c)
   3.5 For Endorsement Maintenance (NQF Submission Form S.3.1. and S.3.2.)
   3.6 Numerator Statement (NQF Submission Form S.4.)
   3.7 Numerator Details (NQF Submission Form S.5.)
   3.8 Denominator Statement (NQF Submission Form S.6.)
3.9  Denominator Details (NQF Submission Form S.7.)
3.10 Denominator Exclusions (NQF Includes “Exception” in the “Exclusion” Field) (NQF Submission Form S.8.)
3.11 Denominator Exclusion Details (NQF Includes “Exception” in the “Exclusion” Field) (NQF Submission Form S.9.)
3.12 Stratification Details/Variables (NQF Submission Form S.10.)
3.13 Risk Adjustment Type (NQF Submission Form S.11.)
3.14 Type of Score (NQF Submission Form S.12.)
3.15 Interpretation of Score (NQF Submission Form S.13.)
3.16 Calculation Algorithm/Measure Logic (NQF Submission Form S.14.)
3.17 Sampling (NQF Submission Form S.15.)
3.18 Survey/Patient-Reported Data (NQF Submission Form S.16.)
3.19 Data Source (NQF Submission Form S.17.)
3.20 Data Source or Collection Instrument (NQF Submission Form S.18.)
3.21 Data Source or Collection Instrument (Reference) (NQF Submission Form S.19.)
3.22 Level of Analysis (NQF Submission Form S.20.)
3.23 Care Setting (NQF Submission Form S.21.)
3.24 Composite Performance Measure (NQF Submission Form S.22.)
7 **MEASURE JUSTIFICATION FORM INSTRUCTIONS**

This form is a guide for use when submitting measures. Use it to draft responses for each section in preparation for entering them in the measure submission. The MJF tracks very closely to the NQF online measure submission. Version 7.1, and references corresponding fields from that submission in the parentheses. The numbers used throughout this form correspond to the same numbered items on the NQF submission. Developers may submit the NQF submission form in lieu of the MJF.

Information from the MIF and the MJF may be used for other purposes, so developers may be asked to complete the MJF for measures that are not submitted to NQF.

1. **Measure Name/ Title (NQF Submission Form De.2.)**

Provide the measure name that was used for the MIF. It should be brief and include the measure focus and the **target population**.

2. **Type of Measure (NQF Submission Form De.1., NQF Evidence Attachment 1a.1.)**

Identify a measure type from the listed items. PROs include health-related quality of life, functional status, symptom burden, experience with care, and health-related behaviors. Use the same type that was identified on the MIF:

- Process
- Process: Appropriate Use
- **Outcome**
- Cost/Resource Use
- **Efficiency**
- Outcome: PRO-PM
- Structure
- Outcome: Intermediate Outcome
- Composite.

3. **Importance (NQF Importance Tab)**

3.1 Evidence to Support the Measure Focus (for reference only) (NQF Evidence Attachment Subcriterion 1a).

The measure focus is evidence-based, demonstrate as follows:

- Health outcome—a rationale supports the relationship of the health outcome to processes or structures of care.
- Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
- Intermediate outcome—a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence that the measured intermediate outcome leads to a desired health outcome.
- The preferred systems for grading the evidence are the USPSTF grading definitions and methods, or GRADE guidelines.
• Measures derived from patient reports—evidence that the measured aspects of care are those valued by patients and for which the patient is the best and/or only source of information, or that patient experience with care is correlated with desired outcomes.
• Efficiency—evidence for the quality component implied in experience with care. Measures of efficiency combine the concepts of resource use and quality (i.e., NQF’s Measurement Framework: Evaluating Efficiency Across Episodes of Care).

For NQF submission of the subcriteria information on importance, a template is available on the NQF Submitting Standards website. Items from that document are reproduced here for reference and for other submission purposes.

3.1.1 This is a Measure of: (should be consistent with type of measure entered in NQF Measure Submission Form De.1) (NQF Evidence Attachment 1a.1)
- Process: Name the process.
- Process: Appropriate Use: Name the appropriate use being measured.
- Outcome: Name the outcome.
- PRO. Outcome: PRO: PROs such as health-related quality of life, functional status, symptom or burden, experience with care, and health-related behaviors.
- Cost /Resource Use: Name the cost/resource.
- Efficiency: Name the efficiency.
- Structure: Name the structure.
- Intermediate outcome: Name the intermediate outcome.
- Composite: Name what is being measured.

3.1.2 Logic Model (NQF Evidence Attachment 1a.2)
Briefly state or diagram the steps between the healthcare structures and processes (e.g., interventions, services) and the patient’s health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process, or outcome being measured.

3.1.3 Value and Meaningfulness (NQF Evidence Attachment 1a.3)
If this measure is derived from patient report, provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful. (Describe how and from whom their input was obtained.)

3.1.4 Empirical Data (for outcome measures) – as applicable (NQF Evidence Attachment 1a.2)
Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

3.1.5 Systematic Review of the Evidence (for intermediate outcome, process, or structure performance measures, include those that are instrument-based) – as applicable (NQF Evidence Attachment 1a.3)
What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar, but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (Institute of Medicine, 2011b)
• Clinical Practice Guideline recommendation (with evidence review)
• USPSTF recommendation
• Other systematic review and grading of the body of evidence (e.g., Cochrane Collaboration, AHRQ Evidence Practice Center)
• Other.

To include more than one systematic review, add additional tables.

<table>
<thead>
<tr>
<th>Source of Systematic Review (SR):</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Title</td>
</tr>
<tr>
<td>• Author</td>
</tr>
<tr>
<td>• Date</td>
</tr>
<tr>
<td>• Citation, including page number</td>
</tr>
<tr>
<td>• URL</td>
</tr>
</tbody>
</table>

Quote the guideline or recommendation verbatim about the process, structure, or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.

Grade assigned to the evidence associated with the recommendation with the definition of the grade.

Provide all other grades and definitions from the evidence grading system.

Grade assigned to the recommendation with definition of the grade.

Provide all other grades and definitions from the recommendation grading system.

Body of evidence:
• Quantity – how many studies?
• Quality – what type of studies?

Estimates of benefit and consistency across studies.

What harms were identified?

Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?

3.1.6 Other Source of Evidence – as applicable (NQF Evidence Attachment 1a.4)

If source of evidence is not from a clinical practice guideline, USPSTF, or SR, describe the evidence on which performance measure is based.

3.1.6.1 Briefly Synthesize the Evidence (NQF Evidence Attachment 1a.4.1)

A list of references without a summary is not acceptable.

3.1.6.2 Process Used to Identify the Evidence? (NQF Evidence Attachment 1a.4.2)

Identify guideline recommendation number and/or page number and quote verbatim the specific guideline recommendation.
3.1.6.3 Citation(s) for the Evidence (NQF Evidence Attachment 1a.4.3)

Grade assigned to the quoted recommendation with definition of the grade.

For NQF Maintenance of Endorsement (NQF Evidence Attachment 1a.1)

Update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.

3.2 Performance Gap – Opportunity for Improvement (NQF Measure evaluation criterion 1b)

3.2.1 Rationale (NQF Submission Form 1b.1.)

Briefly explain the rationale for this measure (i.e., the benefits or improvements in quality envisioned by use of this measure).

If the measure is a composite, a combination of component measure scores, all-or-none, or any-or-none, skip this question and provide the rationale for the composite in question 3.2.3 (NQF Submission Form 1c.3.)

3.2.2 Performance Scores (NQF Submission Form 1b.2.)

Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. (This is required for NQF endorsement maintenance of endorsement. Include the mean, standard deviation, minimum, maximum, interquartile range, and scores by decile. Describe the data source, including number of measured entities, number of patients, dates of data, and, if a sample, characteristics that the entities include.) This information also will be used to address the subcriterion on improvement (NQF Submission Form 4b.1. under Usability and Use.)

3.2.3 Summary of Data Indicating Opportunity (NQF Submission Form 1b.3.)

If no or limited performance data on the measure as specified is reported in 3.2.2 (NQF Submission Form 1b.2.), then provide a summary of data from the literature that indicates opportunity for improvement or overall less-than-optimal performance on the specific focus of measurement. Include citations.

3.2.4 Disparities (NQF Submission Form 1b.4.)

Provide data on how the measure, as specified, addresses disparities—current and over time—by population group (i.e., race or ethnicity, gender, age, insurance status, socioeconomic factors, and disability). (This is also required for NQF maintenance of endorsement. Describe the data source, including number of measured entities, number of patients, and dates of the data. If the data are from a sample, include characteristics of the entities.) For measures that show high levels of performance (i.e., “topped out”), disparities data may demonstrate an opportunity for improvement/gap in care for certain subpopulations. This information also will be used to address the subcriterion on improvement 6.2.1 (NQF Submission form 4b) under Usability and Use.

3.2.5 Provide summary of data if no or limited data (NQF Submission Form 1b.5.)

If no or limited data on disparities from the measure as specified is reported in 3.2.4 (NQF Submission Form 1b.4.), then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement and include citations. This is not necessary if performance data is provided in 3.2.4 (NQF Submission Form 1b.4.).
4. **Scientific Acceptability** *(NQF Scientific Acceptability Tab)*

4.1 Data Sample Description *(NQF Testing Attachment 1.)*

This description should be the same as what is submitted with the MIF.

4.1.1 What Types of Data Were Used for Testing? *(NQF Testing Attachment 1.1.)*

Note all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation. If different data sources are used for the numerator and denominator, indicate “numerator” or “denominator” with each source.

Measure specified to use data from must be consistent with data sources entered in 3.19 *(NQF Submission Form S.17.):*

- Abstracted from paper record
- Administrative claims
- Clinical database/registry
- Abstracted from EHR
- eCQM *(HQMF)* implemented in EHRs
- Other (please describe).

Measure tested with data from:

- Abstracted from paper record
- Administrative claims
- Clinical database/registry
- Abstracted from EHRs
- eCQM *(HQMF)* implemented in EHRs
- Other (please describe).

4.1.2 Identify the Specific Dataset *(NQF Testing Attachment 1.2.)*

If an existing dataset was used, identify the specific dataset; the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured (e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

4.1.3 What Are the Dates of the Data Used in Testing? *(NQF Testing Attachment 1.3.)*

Enter the date range.
4.1.4 What Levels of Analysis Were Tested? (NQF Testing Attachment 1.4.)

Testing must be provided for all the levels specified and intended for measure implementation (e.g., individual clinician, hospital, health plan).

Measure specified to measure performance of (must be consistent with data sources entered in 3.22) (NQF Submission Form S.20):

- Individual clinician
- Group/practice
- Hospital/facility/agency
- Health plan
- Other (please describe).

Measure tested at level of:

- Individual clinician
- Group/practice
- Hospital/facility/agency
- Health plan
- Other (please describe).

4.1.5 How Many and Which Measured Entities Were Included in the Testing and Analysis? (NQF Testing Attachment 1.5.)

Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.

4.1.6 How Many and Which Patients Were Included in the Testing and Analysis? (NQF Testing Attachment 1.6.)

Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample.

4.1.7 Sample Differences, if Applicable (NQF Testing Attachment 1.7.)

If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusion, risk adjustment), identify how the data or sample are different for each aspect of testing reported.

4.1.8 What were the social risk factors that were available and analyzed? (NQF Testing Attachment 1.8.)

Describe social risk factors; for example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g., census tract), or patient community characteristics (e.g., percent vacant housing, crime rate), which do not have to be a proxy for patient-level data.

Measures must be tested for all the data sources and levels of analyses that are specified.

Information on scientific acceptability should be sufficient for CMS and external stakeholders to understand to what degree the testing results for the measure meet evaluation criteria for testing.
4.2** Reliability Testing** (for reference only) (NQF Testing Attachment 2a.2.)

Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high percentage of the time when assessed in the same population in the same time period, and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and **composite performance measures**, reliability should be demonstrated for the computed performance score.

Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to, inter-rater/abstractor or intra-rater/abstractor studies, **internal consistency** for multi-item scales, and **test-retest** for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

If accuracy/correctness (i.e., validity) of data elements was empirically tested, separate reliability testing of data elements is not required—in 4.2.1 (NQF Testing Attachment 2a2.1.), check critical data elements; in 4.2.2 (NQF Testing Attachment 2a2.2.), enter “refer to section 4.4 (NQF Testing Attachment 2b2) for validity testing of data elements”; and skip 4.2.3 and 4.2.4 (NQF Testing Attachment 2a2.3. and 2a2.4.)

4.2.1 Level of Reliability Testing (NQF Testing Attachment 2a2.1.)

What level of reliability testing was conducted? (May be one or both levels):

- Critical data elements used in the measure (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)
- Performance measure score (e.g., signal-to-noise analysis).

4.2.2 Method of Reliability Testing (NQF Testing Attachment 2a2.2.)

Describe the method of reliability testing for each level used—from 4.2.1 (NQF Testing Attachment 2a2.1.). Describe the steps—do not just name a method. What type of error is it testing? What statistical analysis was used?

4.2.3 Statistical Results from Reliability Testing (NQF Testing Attachment 2a2.3.)

What were the statistical results from reliability testing for each level—from 4.2.1 (NQF Testing Attachment 2a2.1.)? Examples include percent agreement and kappa for the critical data elements, and distribution of reliability statistics from a signal-to-noise analysis. Provide reliability statistics and assessment of adequacy in the context of norms for the test conducted.

4.2.4 Interpretation (NQF Testing Attachment 2a2.4.)

What is your interpretation of the results in terms of demonstrating reliability? What do the results mean and what are the norms for the test conducted?

4.3** Validity Testing** (for reference only) (NQF Testing Attachment 2b1.)

Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument-based measures, including PRO-PMs and **composite performance measures**, validity should be demonstrated for the computed performance score.
Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to, testing hypotheses that the measures scores indicate quality of care (e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method); correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

4.3.1 Level of Validity Testing (NQF Testing Attachment 2b1.1.)

What level of validity testing was conducted? (Note: It may be more than one level.):

- Critical data elements (Note: Data element validity must address all critical data elements.)
- Performance measure score:
  - Empirical validity testing
  - Systematic assessment of face validity of performance measure score as an indicator of quality or resource use (i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance).

Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

4.3.2 Method of Validity Testing (NQF Testing Attachment 2b1.2.)

For each level tested, describe the method of validity testing and what it tests. Describe the steps—do not just name a method; and what was tested (e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected, statistical analysis used).

4.3.3 Statistical Results from Validity Testing (NQF Testing Attachment 2b1.3.)

Provide statistical results and assessment of adequate validity (e.g., correlation, t test).

4.3.4 Interpretation (NQF Testing Attachment 2b1.4.)

What is your interpretation of the results in terms of demonstrating validity? What do the results mean and what are the norms for the test conducted?

4.4 Exclusions Analysis (for reference only) (NQF Testing Attachment 2b2.)

Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure. Examples of evidence that an exclusion distorts measure results include, but are not limited to, frequency of occurrence, variability of exclusion across providers, and sensitivity analyses (with and without the exclusion). If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
If there are no exclusions, indicate that this section is not applicable and skip the next section.

4.4.1 Method of Testing Exclusions (NQF Testing Attachment 2b2.1.)

Describe the method of testing the exclusions and what it tests. Describe the steps—do not just name a method; what was tested (e.g., whether the exclusions affect overall performance scores); and what statistical analysis was used.

4.4.2 Statistical Results From Testing Exclusions (NQF Testing Attachment 2b2.2.)

What were the statistical results from testing the exclusions? Include overall number and percentage of individuals excluded, frequency distribution of the exclusions across measured entities, and impact on performance measure scores.

4.4.3 Interpretation (NQF Testing Attachment 2b2.3.)

What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results (i.e., the value outweighs the burden of increased data collection and analysis)? If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent (e.g., scores with and without the exclusion).

4.5 Risk Adjustment or Stratification for Outcome or Resource Use Measures (for reference only) (NQF Testing Attachment 2b3.)

For outcome measures and other measures when indicated (e.g., resource use): an evidence-based risk adjustment strategy (e.g., risk model, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; and has demonstrated adequate discrimination and calibration.

Risk factors that influence outcomes should not be specified as exclusions. Developers should consider both stratification and risk adjustment of measures by social risk factors, which include but are not limited to, income, education, race and ethnicity, employment, disability, community resources, and social support (certain factors of which are also sometimes referred to as SES factors or sociodemographic status [SDS] factors).

If this is not applicable, describe the rationale/data support no risk adjustment/stratification.

If the measure is not an intermediate, or health outcome, or PRO performance measure, or resource use measure, skip to the next section.

4.5.1 Method of Controlling for Differences (NQF Testing Attachment 2b3.1.)

What method of controlling for differences in case mix is used?

- No risk adjustment or stratification
- Statistical risk model with (specify number) risk factors
- Stratification by (specify number) risk categories
- Other (please describe).

If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions (NQF Testing Attachment 2b3.1.1).
4.5.2 Rationale Why Risk Adjustment Is Not Needed (NQF Testing Attachment 2b3.2.)

If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

4.5.3 Conceptual, Clinical, and Statistical Methods (NQF Testing Attachment 2b3.3.a.)

Describe the conceptual, clinical, and statistical methods and criteria used to select patient factors (i.e., clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p < 0.10; correlation of r or higher; patient factors should be present at the start of care and not related to disparities). Also, discuss any “ordering” of risk factor inclusion; for example, are social risk factors added after all clinical factors?

4.5.4 Conceptual Model of Impact of Social Risks (NQF Testing Attachment 2b3.3b.)

How was the conceptual model of how social risk impacts this outcome developed? Select all that apply:

- Published literature
- Internal data analysis
- Other (please describe).

4.5.5 Statistical Results (NQF Testing Attachment 2b3.4a.)

Describe the statistical results of the analyses used to select risk factors.

4.5.6 Analyses and Interpretation in Selection of Social Risk Factors (NQF Testing Attachment 2b3.4b.)

Describe the analyses and interpretation resulting in the decision to select social risk factors (e.g., prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects). Also, describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.

4.5.7 Method Used to Develop the Statistical Model or Stratification Approach (NQF Testing Attachment 2b3.5.)

Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach; describe the steps—do not just name a method—and what statistical analysis was used.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix). If stratified, skip to 4.5.11 (NQF Testing Attachment 2b3.9).

4.5.8 Statistical Risk Model Discrimination Statistics (e.g., c-statistic, R²) (NQF Testing Attachment 2b3.6.)

4.5.9 Statistical Risk Model Calibration Statistics (e.g., Hosmer-Lemeshow statistic) (NQF Testing Attachment 2b3.7.)

4.5.10 Statistical Risk Model Calibration—Risk decile plots or calibration curves (NQF Testing Attachment 2b3.8.)

4.5.11 Results of Risk Stratification Analysis (NQF Testing Attachment 2b3.9.)
4.5.12 Interpretation (NQF Testing Attachment 2b3.10.)

What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix) (i.e., what do the results mean and what are the norms for the test conducted?)?

4.5.13 Optional Additional Testing for Risk Adjustment (NQF Testing Attachment 2b3.11.)

This testing is not required but would provide additional support of adequacy of the risk model, (e.g., testing of risk model in another data set, sensitivity analysis for missing data, other methods that were assessed).

4.6 Identification of Meaningful Differences in Performance (for reference only) (NQF Testing Attachment 2b.54.)

Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% vs. 75%) is clinically meaningful, or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 vs. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

You may also describe the evidence of overall less-than-optimal performance. The intent of this section is to go beyond demonstrating a performance gap and address statistical significance, if possible.

4.6.1 Method (NQF Testing Attachment 2b4.1.)

Describe the method for determining whether statistically significant and clinically or practically meaningful differences in performance measure scores among the measured entities can be identified. Describe the steps—do not just name a method. What statistical analysis was used? Do not just repeat the information provided related to performance gap in the section on importance 3.2 (NQF Testing Attachment 1b.) Performance Gap.

4.6.2 Statistical Results (NQF Testing Attachment 2b4.2.)

What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? For example, was there a different than expected number and percentage of entities with scores significantly varying from the mean or some benchmark? How was meaningful difference defined?

4.6.3 Interpretation (NQF Testing Attachment 2b4.3.)

What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? What do the results mean in terms of statistical and meaningful differences?

4.7 Comparability of Multiple Data Sources/Methods (for reference only) (NQF Testing Attachment 2b5.)

If there is only one set of specifications, skip this section.
If multiple data sources/methods are specified, there is demonstration that they produce comparable results.

This item is directed to measures that are risk-adjusted—with or without social risk factors—or to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eCQMs). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

4.7.1 Method (NQF Testing Attachment 2b5.1.)
Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources or specifications. Describe the steps—do not just name a method. What statistical analysis was used?

4.7.2 Statistical Results (NQF Testing Attachment 2b5.2.)
What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications (e.g., correlation, rank order)?

4.7.3 Interpretation (NQF Testing Attachment 2b5.3.)
What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources or specifications? What do the results mean and what are the norms for the test conducted?

4.8 Missing Data Analysis and Minimizing Bias (for reference only) (NQF Testing Attachment 2b6.)
Analyze and identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

4.8.1 Method (NQF Testing Attachment 2b6.1)
Describe the testing method conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias. Describe the steps—do not just name a method. What statistical analysis was used?

4.8.2 Missing Data Analysis (NQF Testing Attachment 2b6.2)
What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data (e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse)? If no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each.

4.8.3 Interpretation (NQF Testing Attachment 2b6.3)
What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and
how the specified handling of missing data minimizes bias? What do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted? If no empirical analysis was conducted, provide rationale for the selected approach for missing data.

5. **Feasibility** (NQF Feasibility Tab)

This criterion assesses the extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement.

5.1 Data Elements Generated as Byproduct of Care Processes (NQF Measure evaluation criterion 3a./3a.1)

How are the data elements generated that are needed to compute measure scores?

Data used in the measure are: (list all that apply)

- Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, laboratory value, diagnosis, depression score).
- Coded by someone other than the person obtaining original information (e.g., Diagnosis-Related Group [DRG], ICD-10 codes on claims).
- Abstracted from a record by someone other than the person obtaining original information (e.g., chart abstraction for quality measure or registry).
- Other (please describe).

5.2 Electronic Sources (NQF Measure evaluation criterion 3b.)

5.2.1 Data Elements Electronic Availability (NQF Submission Form 3b.1.)

To what extent are the data elements needed for the measure available electronically (i.e., elements that are needed to compute performance measure scores are in defined, computer-readable fields)?

- All data elements are in defined fields in EHRs.
- All data elements are in defined fields in electronic claims.
- All data elements are in defined fields in electronic clinical data such as clinical registry, nursing home MDS, and home health OASIS.
- All data elements are in defined fields in a combination of electronic sources.
- Some data elements are in defined fields in electronic sources.
- No data elements are in defined fields in electronic sources.
- Data is patient/family reported information; may be electronic or paper.

5.2.2 Path to Electronic Capture (NQF Submission Form 3b.2.)

If all the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, or provide a rationale for using other than electronic sources.

5.2.3 **eCQM** Feasibility (NQF Submission Form 3b.3.)

If this is an eCQM, provide a summary of the feasibility assessment in an attached file or make it available at a measure-specific URL.
5.3 Data Collection Strategy (NQF Measure evaluation criterion 3c.)

5.3.1 Data Collection Strategy Difficulties (optional) (NQF Submission Form 3c.1.)
Describe difficulties as a result of testing or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, and other feasibility or implementation issues.

If the measure is instrument-based, consider the implications of burden for both individuals providing the data (e.g., patients, service recipients, respondents) and those whose performance is being measured.

5.3.2 Fees, Licensing, Other Requirements (NQF Submission Form 3c.2.)
Describe any fees, licensing, or other requirements to use any aspect of the measure as specified, such as the value or code set, the risk model, programming code, or algorithm.

6. Usability and Use (NQF Usability and Use Tab)
This criterion evaluates the extent to which intended audiences such as consumers, purchasers, providers, and policy makers can understand the results of the measure and are likely to find them useful for decision-making. NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to being used for performance improvement.

6.1 Use (NQF Measure evaluation criterion 4a.)

6.1.1 Current and Planned Use (NQF Submission Form 4.1.)
Select all the uses that apply. Identify whether the use is current or planned. If the measure is in current use, provide the specific program name and URL for that specific program.

- Public reporting
- Public health or disease surveillance
- Payment program
- Regulatory and accreditation programs
- Professional certification or recognition program
- Quality improvement with external benchmarking to multiple organizations
- Quality improvement internal to a specific organization
- Not in use
- Use unknown.

For each current use listed, provide (NQF Submission Form 4a.1.):

- Name of the program and sponsor
- Purpose
- Geographic area
- Number and percentage of accountable entities and patients included
- Level of measurement
- Setting.
6.1.1.1 Reasons for Not Publicly Reporting or Use in Other Accountability Application (NQF Submission Form 4a.1.2.)

If the measure is not currently publicly reported or used in at least one other accountability application such as payment program, certification, or licensing, what are the reasons? Are there policies or actions of the developer and steward or accountable entities that restrict access to performance results or impede implementation?

6.1.1.2 Plan for Implementation (NQF Submission Form 4a.1.3.)

If not currently publicly reported or used in at least one other accountability application, provide a credible plan for implementation within the expected time frames (i.e., any accountability application within 3 years and publicly reported within 6 years of initial endorsement). (Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified time frames. A plan for accountability applications addresses mechanisms for data aggregation and reporting.)

6.1.2 Feedback on the measure by those being measured or others (NQF Measure evaluation criterion 4a2)

6.1.2.1 Technical Assistance Provided During Development or Implementation (NQF Submission Form 4a2.1.1.)

Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

6.1.2.2 Technical Assistance with Results (NQF Submission Form 4a2.1.2.)

Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

6.1.2.3 Feedback on Measure Performance and Implementation (NQF Submission Form 4a2.2.1.)

Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1. Describe how feedback was obtained.

6.1.2.4 Feedback from Providers being Measured (NQF Submission Form 4a2.2.2.)

Summarize the feedback obtained from those being measured.

6.1.2.5 Feedback from Other Users (NQF Submission Form 4a2.2.3.)

Summarize the feedback obtained from other users.

6.1.2.6 Consideration of Feedback (NQF Submission Form 4a2.3.)

Describe how the feedback described in 6.1.2.3 (NQF Submission Form 4a2.2.1) has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.
6.2 **Usability** (NQF Measure evaluation criterion 4b)

6.2.1 Improvement (NQF Measure evaluation criterion 4b.1.)

Refer to data provided in 3.2 Performance Gap (NQF Submission Form 1b.), but do not repeat here. Discuss or document progress on improvement, such as trends in performance results; number and percentage of people receiving high-quality healthcare; and geographic area and number and percentage of accountable entities and patients included.

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

6.2.2 Unexpected Findings (NQF Measure evaluation criterion 4b2., NQF Submission Form 4b2.1.)

Explain any unexpected findings—positive or negative—during implementation of this measure including unintended impacts on patients.

6.2.3 Unexpected Benefits (NQF Submission Form 4b2.2.)

Explain any unexpected benefits from implementation of this measure.

7. **Related and Competing Measures** (NQF Related and Competing Measures Tab)

If a measure meets other criteria and there are endorsed or new related measures (either same measure focus or target population) or competing measures (both the same measure focus and same target population), the measures are compared to address harmonization and/or selection of the best measure.

7.1 Relation to Other NQF-Endorsed Measures (NQF Measure evaluation criterion 5, NQF Submission Form 5)

Are there related measures or competing measure?

- Yes
- No

If there are related measures (i.e., conceptually related by same measure focus or same target population) or competing measures (i.e., same measure focus and same target population), list the NQF number and title of all related and/or competing measures.

7.2 Harmonization (NQF Submission Form 5a., 5a.1., 5a.2.)

If this measure conceptually addresses either the same measure focus or the same target population as NQF-endorsed measure(s): Are the measure specifications harmonized to the extent possible?

If the measure specifications are not completely harmonized: identify the differences, rationale, and impact on interpretability and data collection burden.
7.3 Competing Measures (NQF Submission Form 5b., 5b.1.)

If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality), or provide a rationale for the additive value of endorsing an additional measure. Provide analyses when possible.

**Additional Information (NQF Additional Information Tab)**

Appendix

Supplemental materials may be provided in an appendix.

All supplemental materials, such as data collection instrument or methodology reports, should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and measure testing attachment. There is no guarantee that supplemental materials will be reviewed. Indicate whether supplemental materials are available at a measure-specific web page (URL identified in 3.1 [NQF Submission Form S.1]), available in attached file, or no supplemental materials.

**Contact Information**

Co.1. Measure Steward Point of Contact

Co.1.1. Organization

Co.1.2. First Name

Co.1.3. Last Name

Co.1.4. Email Address

Co.1.5. Phone Number ( ) ext.

Co.2. Developer Point of Contact (indicate if same as Measure Steward Point of Contact)

Co.2.1. Organization

Co.2.2. First Name

Co.2.3. Last Name

Co.2.4. Email Address

Co.2.5. Phone Number ( ) ext.

Other Additional Information

Ad.1. Working Group/Expert Panel Involved in Measure Development

List the working group/panel members’ names and organizations

Describe the members’ role in measure development
Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2. Year the Measure Was First Released
Ad.3. Month and Year of Most Recent Revision
Ad.4. What is your frequency for review/update of this measure?
Ad.5. When is your next scheduled review/update for this measure?
Ad.6. Copyright Statement
Ad.7. Disclaimers
Ad.8. Additional Information/Comments
8 **BLANK MEASURE JUSTIFICATION FORM TEMPLATE**

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, CMS’s Section 508 Guide for Microsoft Word 2013 may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

<PLEASE DELETE THIS SECTION BEFORE SUBMISSION.>

**Project Title:**

<List the project title as it should appear.>

**Project Overview:**

The Centers for Medicare & Medicaid Services (CMS) has <contracted with / provided funding to> <measure developer name> to develop <measure (set) name or description>. The <contract / cooperative agreement / other (please specify)> name is <insert contract name>. The <contract / cooperative agreement / other (please specify)> number is <project number.>

**Date:**

Information included is current on <insert Date>.

1. **Measure Name/Title (NQF Submission Form De.2)**

2. **Type of Measure (NQF Submission Form De.1., NQF Evidence Attachment 1a.1.)**

3. **Importance (NQF Importance Tab)**

   3.1 Evidence to Support the Measure Focus *(for reference only)* (NQF Evidence Attachment Subcriterion 1a.)

   3.1.1 This is a Measure of: (should be consistent with type of measure entered in NQF Measure Submission Form De.1) (NQF Evidence Attachment 1a.1)

   3.1.2 Logic Model (NQF Evidence Attachment 1a.2)

   3.1.3 Value and Meaningfulness (NQF Evidence Attachment 1a.3)

   3.1.4 Empirical Data (for outcome measures) – as applicable (NQF Evidence Attachment 1a.2)

   3.1.5 Systematic Review of the Evidence (for intermediate outcome, process, or structure performance measures, include those that are instrument-based) – as applicable (NQF Evidence Attachment 1a.3)
3.1.6 Other Source of Evidence – as applicable (NQF Evidence Attachment 1a.4)

3.1.6.1 Briefly Synthesize the Evidence (NQF Evidence Attachment 1a.4.1)

3.1.6.2 Process Used to Identify the Evidence? (NQF Evidence Attachment 1a.4.2)

3.1.6.3 Citation(s) for the Evidence (NQF Evidence Attachment 1a.4.3)

3.2 Performance Gap – Opportunity for Improvement (NQF Measure evaluation criterion 1b)

3.2.1 Rationale (NQF Submission Form 1b.1.)

3.2.2 Performance Scores (NQF Submission Form 1b.2.)

3.2.3 Summary of Data Indicating Opportunity (NQF Submission Form 1b.3.)

3.2.4 Disparities (NQF Submission Form 1b.4.)

3.2.5 Provide summary of data if no or limited data (NQF Submission Form 1b.5.)

4. Scientific Acceptability (NQF Scientific Acceptability Tab)

4.1 Data Sample Description (NQF Testing Attachment 1.)

4.1.1 What Type of Data Were Used for Testing? (NQF Testing Attachment 1.1.)

4.1.2 Identify the Specific Dataset (NQF Testing Attachment 1.2.)

4.1.3 What Are the Dates of the Data Used in Testing? (NQF Testing Attachment 1.3.)

4.1.4 What Levels of Analysis Were Tested? (NQF Testing Attachment 1.4.)

4.1.5 How Many and Which Measured Entities Were Included in the Testing and Analysis? (NQF Testing Attachment 1.5.)

4.1.6 How Many and Which Patients Were Included in the Testing and Analysis? (NQF Testing Attachment 1.6.)

4.1.7 Sample Differences, if Applicable (NQF Testing Attachment 1.7.)

4.1.8 What Social Risk Factors Were Available and Analyzed? (NQF Testing Attachment 1.8.)

4.2 Reliability Testing (for reference only) (NQF Testing Attachment 2a.2.)

4.2.1 Level of Reliability Testing (NQF Testing Attachment 2a.2.1.)

4.2.2 Method of Reliability Testing (NQF Testing Attachment 2a.2.2.)

4.2.3 Statistical Results from Reliability Testing (NQF Testing Attachment 2a.2.3.)

4.2.4 Interpretation (NQF Testing Attachment 2a.2.4.)

4.3 Validity Testing (for reference only) (NQF Testing Attachment 2b1.)

4.3.1 Level of Validity Testing (NQF Testing Attachment 2b1.1.)

4.3.2 Method of Validity Testing (NQF Testing Attachment 2b1.2.)

4.3.3 Statistical Results from Validity Testing (NQF Testing Attachment 2b1.3.)
4.3.4 Interpretation (NQF Testing Attachment 2b1.4.)

4.4 Exclusions Analysis *(for reference only)* (NQF Testing Attachment 2b2.)

4.4.1 Method of Testing Exclusions (NQF Testing Attachment 2b2.1.)

4.4.2 Statistical Results From Testing Exclusions (NQF Testing Attachment 2b2.2.)

4.4.3 Interpretation (NQF Testing Attachment 2b2.3.)

4.5 Risk Adjustment or Stratification for Outcome or Resource Use Measures *(for reference only)* (NQF Testing Attachment 2b3)

4.5.1 Method of Controlling for Differences (NQF Testing Attachment 2b3.1.)

4.5.2 Rationale Why Risk Adjustment Is Not Needed (NQF Testing Attachment 2b3.2.)

4.5.3 Conceptual, Clinical, and Statistical Methods (NQF Testing Attachment 2b3.3.a.)

4.5.4 Conceptual Model of Impact of Social Risks (NQF Testing Attachment 2b3.3b.)

4.5.5 Statistical Results (NQF Testing Attachment 2b3.4a.)

4.5.6 Analyses and Interpretation in Selection of Social Risk Factors (NQF Testing Attachment 2b3.4b.)

4.5.7 Method Used to Develop the Statistical Model or Stratification Approach (NQF Testing Attachment 2b3.5.)

4.5.8 Statistical Risk Model Discrimination Statistics (e.g., c-statistic, $R^2$) (NQF Testing Attachment 2b3.6.)

4.5.9 Statistical Risk Model Calibration Statistics (e.g., Hosmer-Lemeshow statistic) (NQF Testing Attachment 2b3.7.)

4.5.10 Statistical Risk Model Calibration—Risk decile plots or calibration curves (NQF Testing Attachment 2b3.8.)

4.5.11 Results of Risk Stratification Analysis (NQF Testing Attachment 2b3.9.)

4.5.12 Interpretation (NQF Testing Attachment 2b3.10.)

4.5.13 Optional Additional Testing for Risk Adjustment (NQF Testing Attachment 2b3.11.)

4.6 Identification of Meaningful Differences in Performance *(for reference only)* (NQF Testing Attachment 2b.54.)

4.6.1 Method (NQF Testing Attachment 2b4.1.)

4.6.2 Statistical Results (NQF Testing Attachment 2b4.2.)

4.6.3 Interpretation (NQF Testing Attachment 2b4.3.)

4.7 Comparability of Multiple Data Sources/Methods *(for reference only)* (NQF Testing Attachment 2b5.)

4.7.1 Method (NQF Testing Attachment 2b5.1.)

4.7.2 Statistical Results (NQF Testing Attachment 2b5.2.)
4.7.3 Interpretation (NQF Testing Attachment 2b5.3.)

4.8 Missing Data Analysis and Minimizing Bias (for reference only) (NQF Testing Attachment 2b6.)

4.8.1 Method (NQF Testing Attachment 2b6.1)

4.8.2 Missing Data Analysis (NQF Testing Attachment 2b6.2)

4.8.3 Interpretation (NQF Testing Attachment 2b6.3)

5. **Feasibility (NQF Feasibility Tab)**

5.1 Data Elements Generated as Byproduct of Care Processes (NQF Measure evaluation criterion 3a./3a.1)

5.2 Electronic Sources (NQF Measure evaluation criterion 3b.)

5.2.1 Data Elements Electronic Availability (NQF Submission Form 3b.1.)

5.2.2 Path to Electronic Capture (NQF Submission Form 3b.2.)

5.2.3 eCQM Feasibility (NQF Submission Form 3b.3.)

5.3 Data Collection Strategy (NQF Measure evaluation criterion 3c.)

5.3.1 Data Collection Strategy Difficulties (optional) (NQF Submission Form 3c.1.)

5.3.2 Fees, Licensing, Other Requirements (NQF Submission Form 3c.2.)

6. **Usability and Use (NQF Usability and Use Tab)**

6.1 Use (NQF Measure evaluation criterion 4a.)

6.1.1 Current and Planned Use (NQF Submission Form 4.1.)

6.1.1.1 Reasons for Not Publicly Reporting or Use in Other Accountability Application (NQF Submission Form 4a.1.2.)

6.1.2 Feedback on the measure by those being measured or others (NQF Measure evaluation criterion 4a2)

6.1.2.1 Technical Assistance Provided During Development or Implementation (NQF Submission Form 4a2.1.1.)

6.1.2.2 Technical Assistance with Results (NQF Submission Form 4a2.1.2.)

6.1.2.3 Feedback on Measure Performance and Implementation (NQF Submission Form 4a2.2.1.)

6.1.2.4 Feedback from Providers being Measured (NQF Submission Form 4a2.2.2.)

6.1.2.5 Feedback from Other Users (NQF Submission Form 4a2.2.3.)

6.1.2.6 Consideration of Feedback (NQF Submission Form 4a2.3.)
6.2 Usability (NQF Measure evaluation criterion 4b)
6.2.1 Improvement. (NQF Measure evaluation criterion 4b1.)
6.2.2 Unexpected Findings (NQF Measure evaluation criterion 4b2., NQF Submission Form 4b2.1.)
6.2.2.2 Unexpected Benefits (NQF Submission Form 4b2.2.)

7. Related and Competing Measures (NQF Related and Competing Measures Tab)
7.1 Relation to Other NQF-Endorsed Measures (NQF Measure evaluation criterion 5, NQF Submission Form 5)
7.2 Harmonization (NQF Submission Form 5a., 5a.1., 5a.2.)
7.3 Competing Measures (NQF Submission Form 5b., 5b.1.)

Additional Information (NQF Additional Information Tab)

Appendix
Supplemental materials

Contact Information
Co.1.—Measure Steward Point of Contact
Co.1.1. Organization
Co.1.2. First Name
Co.1.3. Last Name
Co.1.4. Email Address
Co.1.5. Phone Number ( ) ext.
Co.2.—Developer Point of Contact (indicate if same as Measure Steward Point of Contact)
Co.2.1. Organization
Co.2.2. First Name
Co.2.3. Last Name
Co.2.4. Email Address
Co.2.5. Phone Number ( ) ext.
Other Additional Information
Ad.1. Working Group/Expert Panel Involved in Measure Development
List the working group/panel members’ names and organizations
Describe the members’ role in measure development
Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2. Year the Measure Was First Released
Ad.3. Month and Year of Most Recent Revision
Ad.4. What is your frequency for review/update of this measure?
Ad.5. When is your next scheduled review/update for this measure?
Ad.6. Copyright Statement
Ad.7. Disclaimers
Ad.8. Additional Information/Comments
9 **Measure Evaluation Criteria and Instructions**

Many of these elements are found in the MIF and MJF and are expanded in the Measure Evaluation Report.

It is important for measure developers to self-evaluate their measures iteratively throughout the Measure Lifecycle. This form is designed to help measure developers compare the specifics of their measure against the criteria by which they will be evaluated. Section 3, Chapter 23, Measure Evaluation describes the process in detail.

Many measure developers plan from the beginning to submit their measures to NQF for endorsement. Therefore, this material is adapted from and follows the NQF Measure Evaluation Criteria very closely.

Additional instruction for evaluating the specific measure types—Composite Performance Measure Evaluation Guidance, Cost-Resource Use Measures, eCQMs, and Patient-Reported Outcomes in Performance Measurement—are also included and noted where applicable. If there are any questions regarding a measure type, consult NQF or Measures Management staff.

Specific criteria for evaluating different types of measures are included where they vary from or are added to the general criteria.

1. **Evidence and Performance Gap—Importance to Measure and Report (NQF Measure evaluation criterion 1)**

Extent to which the specific measure focus is evidence-based and important to making significant gains in healthcare quality where there is variation in or overall less-than-optimal performance. Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.

Cost and Resource Use Measures

For cost and resource use measures, the candidate consensus standards must be judged to be important to measure and report in order to be evaluated against the remaining criteria.

*Patient-reported Outcome-based Performance Measures*

For PRO measures, patients must be involved in identifying PROs for performance measurement. The measures must be person-centered and meaningful.

1.1 Evidence to Support the Measure Focus (NQF Submission Form 1a.)

Health outcomes are often the preferred focus of a measure because they integrate the influence of multiple care processes and disciplines involved in the care. The measure focus is evidence-based, demonstrated as:

- **Outcome**—a rationale supports the relationship of the health outcome to processes or structures of care.
- **Intermediate outcome**—a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence that the measured intermediate outcome leads to a desired health outcome.
- **Process**—a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence that the measured process leads to a desired health outcome.
- **Structure**—a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence that the measured structure leads to a desired health outcome.
- **Patient-reported evidence** should demonstrate that the **target population** values the measured outcome, process, or structure and finds it meaningful.
- **Efficiency**—evidence is required for the quality component, but not required for the resource use component. Note: Measures of efficiency combine the concepts of resource use and quality ([NQF’s Measurement Framework](#): Evaluating Efficiency Across Patient-Focused Episodes of Care; American Quality Alliance’s Principles of Efficiency Measures).
- **Process Measures** incorporating Appropriate Use Criteria—refer to NQF’s guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well (refer to Guidance on Evaluating Evidence for Appropriate Use Measures).

Clinical care processes typically include multiple steps: assess —> identify problem or potential problem —> choose, plan intervention (with patient input) —> provide intervention —> evaluate its impact on health status. If the measure focus is one step in such a multi-step process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. A measure focused only on collecting **PROM** data is not a **PRO-PM**.

The preferred systems for grading the evidence are the USPSTF grade definitions and methods or GRADE guidelines.

Evidence for specific time frames or thresholds included in a measure should be presented. If evidence is limited, then literature regarding standard norms would be considered.

Examples of evidence to demonstrate that the target population for patient-reported measures values the measured outcome, process, or structure and finds it meaningful includes, but is not limited to, patient input in the development of the instrument, survey, or tool; and focus group input regarding the value of the performance measure derived from the instrument/survey/tool.

Current requirements for **structure** and **process measures** (i.e., a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence that the measured structure/process leads to a desired health outcome) also apply to patient-reported structure/ process measures.

Domains of PROs include health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.

**Instrument-based Measures, Including PRO-PMs**

Patients/persons must be involved in identifying structures, processes, or outcomes for performance measurement (i.e., person-centered; meaningful).

PRO-PMs should have the same evidence requirement as health outcomes (i.e., empirical data demonstrates the relationship of the health outcome to processes or structures of care).

Process or structure measures derived from data collected via instrument have the same evidence requirements as other structure or process measures (i.e., systematic assessment and grading of the quantity, quality, and consistency of the body of evidence linking the measured structure or process to a desired outcome).

Exceptions to the evidence requirement for performance measures focused solely on administering a particular instrument should be addressed the same way as for other measures based solely on conducting an assessment (e.g., order laboratory test, check blood pressure).
Cost and Resource Use Measures

For cost and resource use measures, the intent of the resource use and the measure construct should be clearly described. In addition, the service categories for resource uses (i.e., types of resources or costs) that are included in the resource use measure are consistent with and representative of the intent of the measure.

1.2 Performance Gap (NQF Submission Form 1b.)

It is not enough that the measure is merely related to an important broad topic area. Evaluate whether the measure focus is a quality problem, an opportunity for improvement with data showing considerable variation, overall less-than-optimal performance in the quality of care across providers, or disparities in care across population groups.

When assessing measure performance data for Performance Gap (1b), the following factors should be considered:

- Distribution of performance scores
- Number and representativeness of the entities included in the measure performance data
- Data on disparities
- Size of the population at risk, effectiveness of an intervention, likely occurrence of an outcome, and consequences of the quality problem.

Examples of data on opportunity for improvement include, but are not limited to, prior studies, epidemiologic data, or data from pilot testing or implementation of the proposed measure. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

Performance Gap (i.e., opportunity for improvement) should be considered differently for outcome measures such as mortality and patient safety events, where it may be appropriate to continue measurement even with low event rates. Process measures can reasonably reach near 100% performance with minimal opportunity for additional meaningful gains. For mortality and adverse events measures, however, it is less clear how low is attainable.

For measures that use ICD-10 coding:

- For CY2018 submissions, performance gap can be based on literature and/or data based on ICD-9 or ICD-10 coding.
- For CY2019 and beyond, gap information should be based on ICD-10 coded data.

Composite Performance Measures

The performance gap criterion, 3.2 (NQF Submission Form1b), must be met for the composite performance measure as a whole. The performance gap for each component also should be demonstrated. However, if a component measure has minimal opportunity for improvement, justification for why it should be included in the composite is required (e.g., increase reliability of the composite, clinical evidence).

Cost and Resource Use Measures

Cost and resource use measures must demonstrate that the information presented in this measurement area has a cost problem or that there is variation in resources across entities.
1.3 Explicit Logic for Composite Measures (NQF Composite Measure Submission Form 1c.)

Subcriterion 1.3 (NQF Composite Measure Submission Form 1c) must also be met for a composite performance measure to meet the must-pass criterion of importance to Measure and Report.

For composite performance measures, these items must be explicitly articulated and logical.

1.3.1 The quality construct, including the overall area of quality, included component measures, and the relationship of the component measures to the overall composite and to each other (NQF Composite Measure Submission Form 1c.2).

1.3.2 The rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually (NQF Composite Measure Submission Form 1c3).

1.3.3 How the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale (NQF Composite Measure Submission Form 1c4).

2. Reliability and Validity—Scientific Acceptability of Measure Properties (NQF Measure evaluation criterion 2)

Scientific acceptability is the extent to which the measure, as specified, produces consistent (i.e., reliable) and credible (i.e., valid) results about the quality of care when implemented. Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.

- eCQMs should be specified in the HQMF and must use the QDM, CQL, and value sets/DRCs posted on the NLM’s VSAC. eCQM specifications include data type from the QDM, value sets/DRCs and attributes, and measure logic from CQL. To be considered for NQF endorsement, all eCQMs must be tested for reliability and validity using the HQMF specifications.

- For eCQMs, beginning September 30, 2017, all respecified measure submissions for use in federal programs (previously known as “legacy” eMeasures) were required to the same evaluation criteria as respecified measures—the “Bonnie testing only” option no longer meets endorsement criteria.

- The minimum requirement is testing in EHR systems from more than one EHR vendor. Developers should test on the number of EHRs they feel appropriate. It is highly desirable that measures are tested in systems from multiple vendors.

- In the description of the sample used for testing, indicate how the eCQM specifications were used to obtain the data.

- eCQMs specified in older HQMF releases that have previously been endorsed do not need to be retested for maintenance. They may, however, need to be respecified to accommodate variations in the most current HQMF release. All newly developed measures should be tested in the most current HQMF release format.

- Reliance on data from structured data fields is expected; otherwise, unstructured data must be shown to be both reliable and valid.

- If testing of eCQMs occurs in a small number of sites, it may be best accomplished by focusing on patient-level data element validity (i.e., comparing data used in the measure to the authoritative source). However, as with other measures, testing at the level of the performance measure score is encouraged if data can be obtained from enough measured entities. The use of
EHRs and the potential access to robust clinical data provides opportunities for other approaches to testing.

- If the testing is focused on validating the accuracy of the electronic data, analyze the agreement between electronic data obtained using the eCQM specifications and those obtained through abstraction of the entire electronic record—not just the fields used to obtain the electronic data—using statistical analyses such as sensitivity and specificity, positive predictive value, and negative predictive value. The guidance on measure testing allows this type of validity testing to also satisfy the requirement for reliability testing.

- Note that testing at the level of data elements requires all critical data elements be tested—not just agreement of one final overall computation for all patients. At a minimum, the numerator, denominator, and exclusions (or exceptions) must be assessed and reported separately.

- Use of a simulated data set (e.g., Bonnie) is no longer accepted for testing validity of data elements, but is required for checking that the measure specifications and logic are working as intended.

2.1 **Reliability** (NQF Testing Attachment 2a2.)

The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allow for comparability.

Measure specifications include the target population (i.e., denominator) to whom the measure applies, identification of those from the target population who achieved the specific measure focus (e.g., numerator, target condition, event, outcome), measurement time window, exclusion, risk adjustment/stratification, definitions, data source, code lists with descriptors, sampling, and scoring/computation.

All measures that use the ICD classification system must use ICD-10-CM.

**Instrument-based Measures**

Specifications for instrument-based measures also include the specific instrument (e.g., PROM); standard methods, modes, and languages of administration; whether (and how) proxy responses are allowed; standard sampling procedures; handling of missing data; and calculation of response rates to be reported with the performance measure results.

**Composite Performance Measures**

Composite measure specifications include component measure specifications (unless individually endorsed); scoring rules (i.e., how the component scores are combined or aggregated); how missing data are handled (if applicable); required sample sizes (if applicable); and when appropriate, methods for standardizing scales across component scores and weighting rules (i.e., whether all component scores are given equal or differential weighting when combined into the composite).

**Cost and Resource Use Measures**

Cost and resource use measures are assessed on these items when evaluating the measure’s reliability:

- Construction Logic (i.e., detail the logic steps used to cluster, group, or assign claims beyond those associated with the measure’s clinical logic).

- Clinical Logic (i.e., detail any clustering and the assignment of codes, including the grouping methodology, assignment algorithm, and relevant codes for these methodologies).
• Adjustments for Comparability—Inclusion/Exclusion Criteria (related to clinical exclusion, claim-line or other data quality, data validation [e.g., truncation or removal of low- or high-dollar claim, exclusion of ESRD patients]).
• Adjustments for Comparability—Risk Adjustment (name the statistical method [e.g., logistic regression] and list all the risk factor variables).
• Adjustments for Comparability—Costing Method (detail the costing method, including the source of cost information; steps to capture, apply, or estimate cost information; and provide rationale for this methodology).
• Adjustments for Comparability—Scoring (i.e., classifies interpretation of a ratio score(s) according to whether higher or lower resource use amounts are associated with a higher score, a lower score, a score falling within a defined interval, or a passing score, etc.).

2.1.1 Reliability Testing (NQF Testing Attachment 2a2.)

Reliability testing demonstrates that the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period, and/or that the measure score is precise.

Reliability testing applies to the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to, inter-rater/abstractor or intra-rater/abstractor studies, internal consistency for multi-item scales, and test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

Samples used for testing:

• Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.
• The sample should represent the variety of entities whose performance will be measured. The 2010 Measure Testing Task Force recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.
• The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.
• When possible, units of measurement and patients within units should be randomly selected.

For measures that use ICD-10 coding: For CY2018 submissions, submit updated ICD-10 reliability testing if available; if not, testing based on ICD-9 coding will suffice. For CY2019 and beyond, reliability testing should be based on ICD-10 coded data.

Instrument-based Measures

Data collection instruments (i.e., tools) should be identified (e.g., specific instrument, scale, single-item). Reliability should be demonstrated for the computed performance score. If multiple data sources (e.g., instruments, methods, modes, languages) are used, then comparability or equivalency of performance scores should be demonstrated. Specifications should include standard methods, modes, languages of administration; whether (and how) proxy responses are allowed; standard sampling procedures; how missing data are handled; and calculation of response rates to be reported with the performance measure results.
Composite Performance Measures

For composite performance measures, reliability must be demonstrated for the composite measure score. Testing should demonstrate that measurement error is acceptable relative to the quality signal. Examples of testing include signal-to-noise analysis, inter-unit reliability, and ICC.

Demonstration of the reliability of the individual component measures is not sufficient. In some cases, component measures that are not independently reliable can contribute to reliability of the composite measure.

2.2 Validity (NQF Testing Attachment 2b.)

Evaluation of a measure’s validity involves an assessment of the consistency between measure specifications and a correct, credible reflection of the quality of care provided that adequately identifies differences in quality. Therefore, evaluation of a measure’s validity requires reviewing the measure specifications (e.g., numerator, denominator, exclusion, risk factors) and the evidence that supports them.

The measure specifications are consistent with the evidence presented to support the focus of measurement under criterion 1c. The measure is specified to capture the most inclusive target population indicated by the evidence, and exclusions are supported by the evidence.

Measure specifications include the target population (i.e., denominator) to whom the measure applies, identification of those from the target population who achieved the specific measure focus (e.g., numerator, target condition, event, outcome), measurement time window, exclusion, risk adjustment/stratification, definitions, data sources, code lists with descriptors, sampling, and scoring/computation.

2.2.1 Data Elements Correct (NQF Testing Attachment 2b1.)

Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

Validity testing applies to the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to:

- Testing hypotheses that the measure’s scores indicate quality of care (e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method).
- Correlation of measure scores with another valid indicator of quality for the specific topic.
- Relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures).

Face validity of the measure score as a quality indicator may be adequate if it is accomplished through a systematic and transparent process, by identified experts, and if the specifications explicitly address whether performance scores can be used to distinguish levels of quality. The degree of consensus and any areas of disagreement must be provided/discussed.

Composite Performance Measures

For composite performance measures, validity should be empirically demonstrated for the composite measure score. If empirical testing is not feasible at the time of initial endorsement, acceptable
alternatives include systematic assessment of content or face validity of the composite performance measure or demonstration that each of the component measures meet NQF subcriteria for validity. By the time of endorsement maintenance, validity of the composite performance measure must be empirically demonstrated. It is unlikely that a “gold standard” criterion exists, so validity testing generally will focus on construct validation—testing hypotheses based on the theory of the construct. Examples include testing the correlation with measures hypothesized to be related or not related, and testing the difference in scores between groups known to differ on quality assessed by some other measure.

eCQMs

For eCQMs, validity must be demonstrated at the data element level. If this is not possible, justification is required. Face validity is not sufficient for eCQMs.

Instrument-based Measures

Reliability and validity should be demonstrated for the data (i.e., instrument) and the performance measure score.

**Patient-reported Outcome Measures**

For PROs, response rates can affect validity and should be addressed in testing. Differences in individuals’ responses related to instruments or methods, modes, and languages of administration need to be analyzed and potentially included in risk adjustment.

2.2.2 Exclusions are Supported by Clinical Evidence (NQF Testing Attachment 2b2.)

Exclusions are supported by clinical evidence; otherwise, are of sufficient frequency to warrant inclusion in the specification so that results are distorted without the exclusion.

and

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately; denominator exclusion category computed separately).

Examples of evidence that exclusion distorts measure results include, but are not limited to:

- Frequency of occurrence
- Variability of exclusion across providers
- Sensitivity analyses with and without the exclusion.

Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

**Composite Performance Measures**

This criterion applies to the component measures and to the composite performance measures.
2.2.3 **Risk Adjustment** Strategy (NQF Testing Attachment 2b3.)

For outcome measures and other measures when indicated (e.g., resource use):

- An evidence-based, risk adjustment strategy is specified (e.g., risk models, or risk stratification)
- Is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care
- Has demonstrated adequate discrimination and calibration.

or

- Rationale/data support no risk adjustment/stratification.

Note: Risk factors that influence outcomes should not be specified as an exclusion.

In late 2014, the NQF Board of Directors approved, for a trial period, a change in the policy that prohibited the use of social factors in statistical risk models. During the trial period, the NQF policy that restricted use of social risk factors in risk adjustment approaches was suspended, and NQF implemented several of the Risk Adjustment Expert Panel’s recommendations. The SDS Trial concluded in Spring 2017. After review of the findings of the trial, NQF’s Board of Directors agreed to allow, for the present, use of social risk factors in risk adjustment approaches.

**Composite Performance Measures**

Risk adjustment strategy applies to outcome component measures (unless NQF-endorsed).

2.2.4 **Meaningful Differences** (NQF Testing Attachment 2b4.)

Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

or

There is evidence of overall less-than-optimal performance.

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of 1% point in the percentage of patients who received smoking cessation counseling (e.g., 74% vs. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 vs. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

**Composite Performance Measures**

Meaningful differences applies to composite performance measures.

2.2.5 **Comparable Results for Multiple Data Sources** (NQF Testing Attachment 2b5.)

If multiple data sources or methods are specified, there is demonstration that they produce comparable results.

**Composite Performance Measures**

Comparable results for multiple data sources applies to component measures of the composite.
Cost and Resource Use

Cost and resource use measures are assessed on these items when evaluating the measure’s validity:

- Adjustments for Comparability—Inclusion/Exclusion Criteria (related to clinical exclusion, claim-line or other data quality, data validation [e.g., truncation or removal of low- or high-dollar claim, exclusion of ESRD patients])
- Adjustments for Comparability—Risk Adjustment (name the statistical method—e.g., logistic regression and list all the risk factor variables)
- Significant Differences in Performance
- Comparability of Multiple Data Sources
- Validity Testing.

2.2.6 Frequency of Missing Data and Distribution (NQF Testing Attachment 2b6.)

Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

Composite Performance Measures

This section requires analyses of overall frequency of missing data and distribution across providers.

Ideally, include sensitivity analysis of the effect of various rules for handling missing data and the rationale for the selected rules; at a minimum, include a discussion of the pros and cons of the considered approaches and rationale for the selected rules.

2.3 Empirical Support for Composite Measures (NQF Composite Testing Attachment 2c.)

Composite Performance Measures

For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

- Component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible. (NQF Composite Testing Attachment 2d1.)
- Aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible. (NQF Composite Testing Attachment 2d2.)
- Estimate the extent of missing data and how the specified handling of missing data minimizes bias (i.e., achieves scores that are an accurate reflection of quality). (NQF Composite Testing Attachment 2b6.)

Subcriterion 2.3 (NQF Composite Testing Attachment 2c) must also be met for a composite performance measure to meet the must-pass criterion of Scientific Acceptability of Measure Properties.

If empirical analyses do not provide adequate results (or are not conducted), other justification must be provided and accepted for the measure to potentially meet the must-pass criterion of Scientific Acceptability of Measure Properties.
Examples of analyses:

- If components are correlated—analyses based on shared variance (e.g., factor analysis, Cronbach’s alpha, item-total correlation, mean inter-item correlation).
- If components are not correlated—analyses demonstrating the contribution of each component to the composite score (e.g., change in a reliability statistic), with and without the component measure; change in validity analyses with and without the component measure; magnitude of regression coefficient in multiple regression with composite score as dependent variable (Diamantopoulos & Winklhofer, 2001), or clinical justification (e.g., correlation of the individual component measures to a common outcome measure).
- Ideally, sensitivity analyses of the effect of various considered aggregation and weighting rules and the rationale for the selected rules; at a minimum, a discussion of the pros and cons of the considered approaches and rationale for the selected rules.
- Overall frequency of missing data and distribution across providers.

Composite measures need to be assessed as a whole in addition to the components; therefore, the specifications need to include scoring, aggregation, and weighting rules. Also, reliability and validity must be assessed for the composite rate. In some cases, components that might not be independently reliable may contribute to the overall reliability of the composite measure.

cesso eCQM-Specific Additional Subcriteria

In addition to the standard five measure evaluation criteria and subcriteria, there are additional or adapted subcriteria that are used to evaluate eCQMs:

- The measure is well defined and precisely specified so it can be implemented consistently within and across organizations, permits comparability, and has EHR measure specifications created using the HQMF specifications.
- eCQM specifications include data type from the QDM (i.e., value sets and attributes), value sets and/or DRCs, and measure logic from CQL.
- Validity demonstrated by analysis of agreement between data elements exported electronically and data elements abstracted from the entire EHR with statistical results within acceptable norms; or, complete agreement between data elements and computed measure scores obtained by applying the EHR measure specifications to a simulated test EHR data set with known values for the critical data elements.
- Analysis of comparability of scores produced by the respecified EHR measure specifications with scores produced by the original measure specifications demonstrated similarity within tolerable error limits.
- Crosswalk of the EHR measure specifications (i.e., QDM data elements, code lists, and measure logic) is needed if there is a case where a measure needs to be respecified. Note: Comparability is only an issue if maintaining two sets of specifications.

Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria. If a measure does not meet all subcriteria for reliability and validity, stop; the evaluation does not proceed.
3. **Feasibility** (NQF Measure evaluation criterion 3)

This criterion evaluates the extent to which required data are readily available, captured without undue burden, and can be implemented for performance measurement. Feasibility is important to the adoption and ultimate impact of the measure and needs to be assessed through testing or actual operational use of the measures.

For eCQMs, the definition is expanded to: “extent to which specifications and logic require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.”

**Instrument-based Measures**

The burden to respondents (i.e., people providing the data) should be minimized (e.g., availability and accessibility enhanced by multiple languages, methods, modes). There should be infrastructure to collect instrument-level data and integrate that collection into workflow and EHRs, as appropriate.

The burdens of data collection, including those related to use of proprietary instruments, are minimized and do not outweigh the benefit of performance improvement.

3.1 **Byproduct of care (clinical measures only)** (NQF Submission Form 3a.)

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3.2 **Data elements are available in EHRs or other electronic sources** (NQF Submission Form 3b.)

The required data elements are available in EHRs or other electronic sources. If the required data are not in EHRs or existing electronic sources, a credible, near-term path to electronic collection is specified.

3.3 **Data collection strategy can be implemented** (NQF Submission Form 3c.)

Demonstration how the data collection strategy (e.g., data source/availability, timing, frequency, sampling, patient-reported data, patient confidentiality) can be implemented (i.e., already in operational use or testing demonstrates that the strategy is ready to put into operational use).

All data collection must conform to laws regarding protected health information. Patient confidentiality is of particular concern with measures based on patient surveys and when there are small numbers of patients.

For eCQMs, a feasibility assessment is required; this feasibility assessment must address the data elements and measure logic and demonstrate that the eCQM can be implemented or that feasibility concerns can be adequately addressed. The feasibility assessment uses a standard score card. Bonnie testing should be used to demonstrate that the measure logic will work.

**Composite Performance Measures**

Criteria 3.1, 3.2, and 3.3 (NQF Submission Form 3a., 3b., and 3c.) apply to composite performance measures as a whole, considering all component measures.
4. **Usability and Use (NQF Measure evaluation criterion 4)**

Evaluation of a measure’s usability and use involves an assessment of the extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) could use or are using performance results for both accountability and performance improvement to achieve the goal of high-quality and efficient healthcare for individuals or populations.

Important **outcome measures** without an identified improvement may still be considered because they are expected to be useful by informing quality improvement. They inform quality improvement by identifying the need for stimulating new approaches to improvement.

*Composite Performance Measures*

NQF endorsement applies only to the composite performance measure as a whole, not to the individual component measures—unless they are submitted and evaluated for individual endorsement.

*Instrument-based Measures*

Provide adequate demonstration of the criteria that support usability, and ultimately, the use of an instrument-based measure for accountability and performance improvement.

An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

4.1 **Use (NQF Submission Form 4a.)**

4.1.1 **Accountability and Transparency (NQF Submission Form 4a1.)**

Performance results are used in at least one accountability application within 3 years after initial endorsement and are publicly reported within 6 years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified time frames is provided.

Transparency is the extent to which performance results about identifiable, accountable entities are disclosed and available outside of the organizations or practices whose performance is measured. Maximal transparency is achieved with public reporting defined as making comparative performance results about identifiable, accountable entities freely available (or at nominal cost) to the public (i.e., generally on a public website). At a minimum, the data on performance results about identifiable, accountable entities are available to the public (e.g., in an unformatted database). The capability to verify the performance results adds substantially to transparency.

Accountability applications are uses of performance results about identifiable, accountable entities to make judgments and decisions as a consequence of performance such as reward, recognition, punishment, payment, or selection (e.g., public reporting, accreditation, licensure, professional certification, health IT incentives, performance-based payment, network inclusion/exclusion). Selection is the use of performance results to make or affirm choices regarding providers of healthcare or health plans. Note: A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified time frames. A plan for accountability applications addresses mechanisms for data aggregation and reporting.

Measures that are included in finalized rule for federal public reporting programs will be considered publicly reported, even if not yet implemented.
4.1.2 Measure Feedback (NQF Submission Form 4a2.)

Feedback on the measure by those being measured or by others is demonstrated when:

- Those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data.
- Those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation.
- This feedback has been considered when changes are incorporated into the measure.

This guidance is not intended to be construed as favoring measures developed by organizations that are able to implement their own measures (e.g., government agencies, accrediting organizations) over equally strong measures developed by organizations that may not be able to do so (e.g., researchers, consultants, academics). Measure developers may request a longer time frame with appropriate explanation and justification.

Composite Performance Measures

Measure feedback applies to composite performance measures. To facilitate transparency, at a minimum, the individual component measures of the composite must be listed with use of the composite measure.

4.2 Usability (NQF Submission Form 4b.)

4.2.1 Improvement (NQF Submission Form 4b1.)

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement. Demonstrated progress toward achieving the goal of high-quality, efficient healthcare includes evidence of improved performance and/or increased numbers of individuals receiving high-quality healthcare. Exceptions may be considered with appropriate explanation and justification.

Demonstrated progress of the performance measure in facilitating progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Composite Performance Measures

Improvement applies to composite performance measures.

4.2.2 Benefits (NQF Submission Form 4b2.)

The benefits of the performance measure in facilitating progress toward achieving high-quality efficient healthcare outweigh the evidence of negative unintended consequences to individuals or populations (if such evidence exists).
Composite Performance Measures

Benefits applies to composite performance measures and component measures. If there is evidence of unintended negative consequences for any of the components, the measure developer should explain how that is handled or justify why that component should remain in the composite.

5. **Comparison to Related or Competing Measures (NQF Measure evaluation criterion 5)**

**Harmonization** should be considered from the beginning of the development of the measure, and CMS measure developers are expected to consider harmonization as one of the core measure evaluation criteria. Either the measure specifications must be harmonized with related measures so that they are compatible, or the differences must be justified.

Instrument-based Measures

Performance measures specified to use different instruments will be considered competing measures.

5.1 Related Measures (NQF Submission Form 5a.)

The specifications for this measure are harmonized with related measures.

or

The differences in specifications are justified.

Measure harmonization refers to the standardization of specifications for:

- Related measures with the same measure focus (e.g., influenza immunization of patients in hospitals or nursing homes)
- Related measures with the same target population (e.g., eye exam and HbA1c for patients with diabetes)
- Definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are justified (i.e., dictated by the evidence).

The dimensions of harmonization can include numerator, denominator, exclusion, calculation, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

5.2 Competing Measure (NQF Submission Form 5b.)

The measure is superior to competing measures (e.g., a more valid or efficient way to measure quality).

or

Multiple measures are justified.

Composite Performance Measures

Criteria 5.1 and 5.2 (NQF Submission Form 5a and 5b) apply to composite performance measures as a whole as well as the component measures.
10 **BLANK MEASURE EVALUATION REPORT TEMPLATE**

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, [CMS’s Section 508 Guide for Microsoft Word 2013](https://www.cms.gov/Regulations-and-Guidance/Guidance/Other-Information/Downloads/Section508WordGuide.pdf) may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

**<PLEASE DELETE THIS SECTION BEFORE SUBMISSION.>**

**Project Title:**

<List the project title as it should appear.>

**Project Overview:**

The Centers for Medicare & Medicaid Services (CMS) has <contracted with / provided funding to> <measure developer name> to develop <measure (set) name or description>. The <contract / cooperative agreement / other (please specify)> name is <insert contract name>. The <contract / cooperative agreement / other (please specify) number is <project number>.>

**Date:**

Information included is current on <insert Date>.

**Measure Name:**

**Measure Set (or Setting):**

**Measure Developer:**

**Instructions:** For each subcriterion, enter the rating assigned using the criteria from the chapter on Measure Evaluation, and the NQF guidance document. Use the supporting information provided in the MIF and MJF, as well as any additional relevant studies or data. For any less-than-satisfactory ratings, enter an improvement plan in the appropriate spaces. Make a summary determination for each criterion using the subcriteria ratings with statements to support the conclusions. Because many measures are planned when developed, to be submitted for endorsement, this material is adapted from and follows very closely the NQF [Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement](https://www.nqf.org/Content/CM/Documents/Measure-Evaluation-Criteria-and-Guidance-for-Evaluating-Measures-for-Endorsement.pdf).

### 1. Evidence, Performance Gap, and Priority (Impact)—Importance to Measure and Report

<table>
<thead>
<tr>
<th>Subcriteria</th>
<th>Anticipated NQF Rating</th>
<th>Rating Improvement Plan (if Low/Moderate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Evidence to Support the Measure Focus/Measure Intent (NQF Submission Form 1a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2 Performance Gap (NQF Submission Form 1b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3 Explicit Logic (Composite Measures only) (NQF Composite Measure Submission Form 1c)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary Rating for Importance:

Fail: At least one of the Impact subcriteria is not rated as high.

Pass: Measure is important; all the subcriteria are rated high.

(If the measure is to be submitted to NQF for endorsement, the measure must be judged to pass all subcriteria to pass this criterion, or NQF will not evaluate it against the remaining criteria.)

Brief statement of conclusions that support the Summary Rating:

2. Reliability and Validity—Scientific Acceptability of Measure Properties (NQF Measure Evaluation Criterion 2)

<table>
<thead>
<tr>
<th>Subcriteria</th>
<th>Anticipated NQF Rating</th>
<th>Rating Improvement Plan (if Low/Moderate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Reliability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1.1 Reliability Testing (NQF Testing Attachment 2a2.)</td>
<td>[H/M/L]</td>
<td></td>
</tr>
<tr>
<td>2.2 Validity (NQF Testing Attachment 2b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2.1 Data Elements Correct (NQF Testing Attachment 2b1.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2.2 Exclusions (NQF Testing Attachment 2b2.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2.3 Risk Adjustment (NQF Testing Attachment 2b3.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2.4 Meaningful Differences (NQF Testing Attachment 2b4.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2.5 Comparable Results (NQF Testing Attachment 2b5.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2.6 Missing Data (NQF Testing Attachment 2b6.)</td>
<td>[H/M/L]</td>
<td></td>
</tr>
<tr>
<td>2.3 Empirical Analysis (Composite Measures Only) (NQF Composite Testing Attachment 2c.)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Summary Rating for Scientific Acceptability of Measure Properties:

Pass: The measure rates moderate to high on all aspects of reliability and validity.

Fail: The measure rates low for one or more aspects of reliability or validity.

(If the measure is to be submitted to NQF for endorsement, the measure must be judged to pass all subcriteria for both reliability and validity to pass this criterion or NQF will not evaluate it against the remaining criteria.)
Brief statement of conclusions that support the Summary Rating:

3. Feasibility

<table>
<thead>
<tr>
<th>Subcriteria</th>
<th>Anticipated NQF Rating</th>
<th>Rating Improvement Plan (if Low/Moderate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Data are a Byproduct of Care (NQF Submission Form 3a.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2 Electronic Sources (NQF Submission Form 3b.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3 Data Collection Strategy (NQF Submission Form 3c.)</td>
<td>[H/M/L]</td>
<td></td>
</tr>
</tbody>
</table>

Summary Rating for Feasibility:

High/3 rating indicates that the predominant rating for most of the subcriteria is high.

Moderate/2 rating indicates that the predominant rating for most of the subcriteria is moderate.

Low/1 rating indicates that the predominant rating for most of the subcriteria is low.

Brief statement of conclusions that support the Summary Rating:

4. Usability and Use

<table>
<thead>
<tr>
<th>Subcriteria</th>
<th>Anticipated NQF Rating</th>
<th>Rating Improvement Plan (if Low/Moderate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Use (NQF Submission Form 4a.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1.1 Accountability and Transparency (NQF Submission Form 4a1.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1.2 Measure Feedback (NQF Submission Form 4a2.)</td>
<td>[Pass/No Pass]</td>
<td></td>
</tr>
<tr>
<td>4.2 Usability (NQF Submission Form 4b.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.2.1 Improvement (NQF Submission Form 4b1.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.2.2 Benefits (NQF Submission Form 4b2.)</td>
<td>[H/M/L]</td>
<td></td>
</tr>
</tbody>
</table>

Summary Rating for Usability:

High rating indicates that the predominant rating for most of the subcriteria is high.

Moderate rating indicates that the predominant rating for most of the subcriteria is moderate.

Low rating indicates that the predominant rating for most of the subcriteria is low.

Brief statement of conclusions that support the Summary Rating:

5. Comparison to Related or Competing Measures

<table>
<thead>
<tr>
<th>Subcriteria</th>
<th>Anticipated NQF Rating</th>
<th>Rating Improvement Plan (if Low/Moderate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Related Measure (NQF Submission Form 5a.)</td>
<td>[H/M/L]</td>
<td></td>
</tr>
<tr>
<td>5.2 Competing Measure (NQF Submission Form 5b.)</td>
<td>[H/M/L]</td>
<td></td>
</tr>
</tbody>
</table>
Summary Rating for Harmonization:

High rating indicates that the measure is completely harmonized with any related measures, and there are no competing measures.

Moderate rating indicates that there may be related measures, but there are justifications for differences. However, there is some risk that the measure may require further harmonization.

Low rating indicates that there may be other measures that are competing or not harmonized with this measure.

Rationale for Rating/Comments:

Preliminary Recommendation for Endorsement

Based on the individual rating of each of the five major criteria, provide an initial recommendation for endorsement based on the overall suitability of this measure by marking an X in the appropriate boxes.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>High</th>
<th>Medium</th>
<th>Low</th>
<th>Insufficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Importance to Measure and Report</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1 Overall Reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2 Overall Validity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Feasibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Usability and Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Comparison to Related or Competing Measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Recommendation:

Explanation:
11 CMS MEASURES MANAGEMENT SYSTEM WEB PAGE POSTING INSTRUCTIONS

Posting Submissions:

- Send submissions to the MMS Support inbox (MMSSupport@battelle.org) for posts on:
  - Public Comment web page
  - TEP web page.

Posting Time Frame:

- Allow at least 5 business days for processing your post. Posts may be posted prior to this time frame; if your post needs to be published on a specific date, note this in your email and we (the Measures Manager) will work with CMS to accomplish by this date/time.

Posting Format:

- Web posting document should be submitted in Microsoft Word format (every post must include a web posting document).
- All other documents/attachments to the post should be 508 compliant and submitted in PDF format. Note: Tables must have repeated headers on every page.

Posting Templates:

- All posts must follow the latest Blueprint templates to be compliant. If they do not, we (the Measures Manager) may ask you to revise them before submitting it as a final post. All templates are found under Section 4: Tools, Appendices, and Forms.

Public Comment Documents to Include with Each Post:

- Call for Public Comments:
  - Public Comment Call Web Posting document (Word format)
  - Other files, if any, to be included with the call for public comment (PDF format)
- Public Comment Summary report:
  - Public Comment Summary Web Posting document (Word format)
  - Public Comment Summary Report (PDF format)

TEP Documents to Include with Each Post:

- Call for TEP:
  - Technical Expert Panel (Call for TEP) Web Page Posting document (Word format)
  - TEP Nomination Form (PDF format)
  - TEP Charter (PDF format)
- TEP Composition (Membership List):
  - TEP Composition (Membership List) Web Page Posting document (Word format)
  - TEP Composition (Membership) List (PDF format)
- TEP Summary Report:
  - TEP Summary Web Page Posting document (Word format)
  - TEP Composition (Membership) List (PDF format)
  - TEP Summary Report (PDF format)
12 Public Comment Call Web Posting

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, CMS’s Section 508 Guide for Microsoft Word 2013 may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

<PLEASE DELETE THIS SECTION BEFORE SUBMISSION.>

Project Title: <List the project title as it should appear on the Web posting.>

Dates:
The Call for Public Comment period opens on <list the date> and closes on <list the date>.

Project Overview:
The Centers for Medicare & Medicaid Services (CMS) has <contracted with / provided funding to> <measure developer name> to develop <measure (set) name or description>. The <contract / cooperative agreement / other (please specify) name is <insert contract name>. The <contract / cooperative agreement / other (please specify)> number is <project number>. As part of its measure development process, <measure developer name> requests interested parties to submit comments on the candidate or concept measures that may be suitable for this project.

Project Objectives:
<List contract objectives>

Documents and Measures for Comment:
These documents are provided for your review and comment. The files are found in the Download section.

<Name the measures and documents for comment.>

Project Specific Instructions:
<List project specific instructions>

Send your comments to <insert email address>.

<Or, if a web-based tool such as Survey Monkey or Jira, often used for eCQMs, is being used to receive comments provide the link here.>
13 Public Comment Summary Web Posting

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, CMS's Section 508 Guide for Microsoft Word 2013 may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

<PLEASE DELETE THIS SECTION BEFORE SUBMISSION.>

Project Title: <List the project title as it should appear on the Web posting.>

Dates:
The Call for Public Comment period closed on <list the date>.

Project Overview:
The Centers for Medicare & Medicaid Services (CMS) has <contracted with / provided funding to> <measure developer name> to develop <measure (set) name or description>. The <contract / cooperative agreement / other (please specify)> name is <insert contract name>. The <contract / cooperative agreement / other (please specify)> number is <project number>. As part of its measure development process, <measure developer name> requested interested parties to submit comments on the candidate or concept measures that may be suitable for this project.

Project Objectives:
<List contract objectives>

Comment Summary:
These documents, including a summary of public comments and the original measures, are found in the Download section.
<List the document names.>

Expiration Notice:
This notice expires on <list the date>. 
14 Public Comment Summary Report Template

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, CMS’s Section 508 Guide for Microsoft Word 2013 may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

<PLEASE DELETE THIS SECTION BEFORE SUBMISSION.>

Project Title: <List the project title as it should appear.>

Dates:

The Call for Public Comment ran from <list the date> to <list the date>.

The Public Comment Summary was made on <list the date>.

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) has <contracted with / provided funding to> <measure developer name> to develop <measure (set) name or description>. The <contract / cooperative agreement / other (please specify)> name is <insert contract name>. The <contract / cooperative agreement / other (please specify)> number is <project number>. As part of its measure development process, <measure developer name> has requested interested parties to submit comments on the candidate or concept measures that may be suitable for this project.

Project Objectives:

<List contract objectives>

Information About the Comments Received:

Public comments were solicited by <methods used to notify stakeholders and the public of comment period>.

<Volume> responses were received on this topic.

Stakeholder Comments—General and Measure-Specific

General Stakeholder Comments:

<Include summary of general comments>

Measure-Specific Stakeholder Comments:

<Include summary of comments on specific measures>

Preliminary Recommendations

<Summarize recommendations>
Overall Analysis of the Comments and Recommendations

<Include a summary of the TEP discussion and changes to the list of candidate measures.>
Public Comment Verbatim Report

*This table may be attached as a separate file if necessary. The table is a template and contains optional fields to be deleted at the request of the COR.*

<table>
<thead>
<tr>
<th>Comment Number*</th>
<th>Date Posted/Received</th>
<th>Name, Credentials, and Organization of Commenter</th>
<th>Type of Organization*</th>
<th>Email Address*</th>
<th>Measure Set or Measure</th>
<th>Text of Comments</th>
<th>Response*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Optional

*Note: Developers may enter the text of comments verbatim without edits for spelling, punctuation, grammar, or any other reason and should ask their COR for specific guidance.*
15 **TECHNICAL EXPERT PANEL CALL FOR TEP WEB POSTING**

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in « > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, [CMS’s Section 508 Guide for Microsoft Word 2013](https://www.cms.gov/Regulations-and-Guidance/Guidance/Section508/Downloads/Section508GuideForMicrosoftWord2013.pdf) may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

**PLEASE DELETE THIS SECTION BEFORE SUBMISSION.**

**Project Title:** «List the project title as it should appear on the Web posting.»

**Dates:**

The TEP nomination period opens on «list the date» and closes on «list the date». Submit all nomination materials by the closing date.

**Project Overview:**

The Centers for Medicare & Medicaid Services (CMS) has «contracted with / provided funding to» «measure developer name» to develop «measure (set) name or description». The «contract / cooperative agreement / other (please specify)» name is «insert contract name». The «contract / cooperative agreement / other (please specify)» number is «project number». As part of its measure development process, «measure developer name» convenes groups of stakeholders and experts who contribute direction and thoughtful input to the measure developer during measure development and maintenance.

**Project Objectives:**

«List contract objectives»

**TEP Requirements:**

We are seeking a TEP of approximately «insert desired TEP size» individuals with differing perspectives and areas of expertise, such as:

**Subject matter expertise:**

Consumer/patient/family (caregiver) perspective (required unless approved by COR)

«Insert specific topic»

«Example topics include: Healthcare disparities, Performance measurement, Quality improvement, and Purchaser perspective»

Potential TEP members should understand that participation is voluntary and that their input will be recorded in the meeting minutes. Proceedings of the TEP will be summarized in a report that may be disclosed to the public. If a participant has chosen to disclose private, personal data, then related material and communications are not deemed to be covered by patient-provider confidentiality. Patient/caregiver participants may elect to keep their names confidential in public documents. «TEP organizers» will answer any questions about confidentiality.
All potential TEP members must disclose any current and past activities that may pose a potential conflict of interest for performing the tasks required of the TEP. All potential TEP members should be able to commit to the anticipated time frame needed to perform the functions of the TEP.

**Patient Nominees:**

*<measure developer name>* is seeking patients to participate on a TEP. We are seeking patients who *<description of type of patients needed>* to join the TEP. Patients who have *<of the type listed>* can provide unique and essential input on quality measures based on their own experience and perspective. Patient nominees should submit a completed and signed TEP Nomination Form and letter of interest as described below but are not required to submit a curriculum vitae.

**TEP Expected Time Commitment:**

*<anticipated meeting dates and time frame for measure development activities>*

*<Types of meetings (e.g., conference call, webinar, in person)>*

**Required Information:**

A completed and signed TEP Nomination form located in the download section below.

The nomination form includes a consent and confidentiality statement.

A letter of interest (not to exceed 2 pages) highlighting experience/knowledge relevant to the expertise described above and involvement in measure development. Consumer/patient/family (caregiver) applicants/nominees are not expected to have experience in measure development.

Curriculum vitae or a summary of relevant experience for a maximum of 10 pages. Consumer/patient/family (caregiver) applicants/nominees are not required to submit a curriculum vitae.

The Nomination forms and proposed TEP Charter are found in the download section below.

If you wish to nominate yourself or other individuals for consideration, complete the form and email it to: *<measure developer contact information>*
16  TECHNICAL EXPERT PANEL NOMINATION FORM TEMPLATE

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, CMS’s Section 508 Guide for Microsoft Word 2013 may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

<PLEASE DELETE THIS SECTION BEFORE SUBMISSION.>

Project Title: <List the project title as it should appear on the Web posting.>

Project Overview:
The Centers for Medicare & Medicaid Services (CMS) has <contracted with / provided funding to> <measure developer name> to develop <measure (set) name or description>. The <contract / cooperative agreement / other (please specify)> name is <insert contract name>. The <contract / cooperative agreement / other (please specify)> number is <project number>. As part of its measure development process, <measure developer name> convenes groups of stakeholders and experts who contribute direction and thoughtful input to the measure developer during measure development and maintenance.

Project Objectives:
<List contract objectives>

TEP Expected Time Commitment:
<Anticipated meeting dates and time frame for measure development activities >

<Types of meetings (conference call, webinar, in person, etc.)>

TEP Requirements:
A TEP of approximately <insert desired TEP size> individuals will recommend <insert objective>. The TEP will be composed of individuals with differing areas of expertise and perspectives, such as:

Subject Matter Expertise:
Consumer/patient/family (caregiver) perspective (required unless approved by COR)
<Insert specific topic>

<Example topics include: Healthcare disparities, Performance measurement, Quality improvement, and Purchaser perspective>

Instructions:
Applicants/nominees must submit these documents with this completed and signed form:
A letter of interest (not to exceed 2 pages) highlighting experience/knowledge relevant to the expertise described above and involvement in measure development. Consumer/patient/family (caregiver) applicants/nominees are not expected to have experience in measure development.

Curriculum vitae or a summary of relevant experience (including publications) for a maximum of 10 pages. Consumer/patient/family (caregiver) applicants/nominees are not required to submit a curriculum vitae.

Send this completed and signed TEP Nomination form, statement of interest, CV to <insert measure developer name> with “Nomination” in the subject line at <insert email address>. Due by close of business <insert date> Eastern Time.

Potential TEP members should understand that participation is voluntary and that their input will be recorded in the meeting minutes. Proceedings of the TEP will be summarized in a report that may be disclosed to the public. If a participant has chosen to disclose private, personal data, then related material and communications are not deemed to be covered by patient-provider confidentiality.

Patient/caregiver participants may elect to keep their names confidential in public documents. <TEP organizers> will answer any questions about confidentiality.

All potential TEP members must disclose any significant financial interest or other relationships that may influence their perceptions or judgment. It is unethical to conceal (or fail to disclose) conflicts of interest. However, the disclosure requirement is not intended to prevent individuals with particular perspectives or strong points of view from serving on the TEP. The intent of full disclosure is to inform the measure developer, other TEP members, and CMS about the source of TEP members’ perspectives and how that might affect discussions or recommendations.

**Applicant/Nominee Information (Self-Nominations Are Acceptable):**

Name:

Credentials:

Role:

Organizational Affiliation, if any:

City:

State:

Mailing address:

Telephone:

Email:

**Person Recommending the Nominee:**

Complete this section only if you are nominating a third party for the TEP. You must sign this form and attest that you have notified the nominee of this action and that they are agreeable to serving on the TEP. The measure developer may request the required information from the nominee.
Name:

Credentials:

Role:

Organizational Affiliation, if any:

City:

State:

Mailing address:

Telephone:

Email:

I attest that I have notified the nominee of this action and that the nominee is agreeable to serve on the TEP.

Signature: ______________________________________________ Date: _________________

The nominee must submit the remainder of the nomination package within the specified period for consideration.

Applicant/Nominee’s Disclosure:

This section addresses disclosure of any current and past activities that may indicate a conflict of interest. As a measure developer for the Centers for Medicare & Medicaid Services (CMS), <measure developer name> must ensure independence, objectivity, scientific rigor, and balance in its measure development activities.

Do you or any family members have a financial interest, arrangement, or affiliation with any corporate organizations that may create a potential conflict of interest? ☐Yes ☐No

If yes, describe (grant/research support, consultant, speaker’s bureau, and major stock shareholder, other financial or material support). Include the name of the corporation/organization.

Do you or any family members have intellectual interest in a study or other research related to the quality measures under consideration? ☐Yes ☐No

If yes, describe the type of intellectual interest and the name of the organization/group.

Applicant/Nominee’s Agreement:

If at any time during my service as a member of this TEP my conflict of interest status changes, I will notify the measure developer and the TEP chair.

It is anticipated that there will be <approximate time commitment that is required>. I am able to commit to attending the TEP meetings in person, by teleconference, or by mutually agreed-upon alternative means.

<Omit this paragraph for TEPs not focused on measure development.> If selected to participate in the TEP and the measures are submitted to a measure endorsement organization (such as the NQF), I will be
available to discuss the measures with the organization or its representatives and work with the measure developer to make revisions to the measures, if necessary.

I understand my participation is voluntary and that my input will be recorded in the meeting minutes. Proceedings of the TEP will be summarized in a report that may be disclosed to the public. If I chose to disclose private, personal data, then related material and communications are not deemed to be covered by patient-provider confidentiality. If I am a patient/caregiver participant, I may elect to keep my name confidential in public documents.

If selected to participate in the TEP, I will keep all materials and discussions confidential until such time that CMS authorizes their release.

I have read the above and agree to abide by it.

Signature: ________________________________ Date: ____________________

For patient participants only: I wish to keep my name confidential. ☐ Yes ☐ No
17  **TECHNICAL EXPERT PANEL COMPOSITION (MEMBERSHIP) LIST TEMPLATE**

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, [CMS’s Section 508 Guide for Microsoft Word 2013](#) may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

**PLEASE DELETE THIS SECTION BEFORE SUBMISSION.**

Project Title: <Insert Project Title>

Dates:

<List projected dates (or time period) of TEP meetings>

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) has <contracted with / provided funding to> <measure developer name> to develop <measure (set) name or description>. The <contract / cooperative agreement / other (please specify)> name is <insert contract name>. The <contract / cooperative agreement / other (please specify)> number is <project number>. As part of its measure development process, <measure developer name> convenes groups of stakeholders and experts who contribute direction and thoughtful input to the measure developer during measure development and maintenance.

These individuals were selected and have agreed to serve as the TEP for this project.

<table>
<thead>
<tr>
<th>Name, Credentials, Professional Role*</th>
<th>Organizational Affiliation, City, State*</th>
<th>Consumer/ Patient/ Family/ Caregiver Perspective*</th>
<th>Clinical Content</th>
<th>Performance Measurement</th>
<th>Coding and Informatics</th>
<th>Conflict of Interest Disclosure*</th>
</tr>
</thead>
</table>

<Note: The first three columns and the last column are required and cannot be changed. The other columns (italicized) should align with the Subject Matter Expertise listed in other TEP templates.>
18  **TECHNICAL EXPERT PANEL COMPOSITION (MEMBERSHIP)**

**LIST WEB POSTING**

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, [CMS’s Section 508 Guide for Microsoft Word 2013](#) may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

**PLEASE DELETE THIS SECTION BEFORE SUBMISSION.**

**Project Title:** *List the project title as it should appear on the Web posting.*

**Dates:**

The Call for TEP nomination period closed on *the closing date*.

**Documents:**

The TEP Membership List is posted below in the download section.

**Project Overview:**

The Centers for Medicare & Medicaid Services (CMS) has *contracted with / provided funding to* [measure developer name] to develop *measure (set) name or description*. The *contract / cooperative agreement / other (please specify)* name is *insert contract name*. The *contract / cooperative agreement / other (please specify)* number is *project number*. As part of its measure development process, [measure developer name] convenes groups of stakeholders and experts who contribute direction and thoughtful input to the measure developer during measure development and maintenance.

**Project Objectives:**

*List contract objectives*

**TEP Requirements:**

We sought a TEP of approximately *insert desired TEP size* individuals with differing perspectives and areas of expertise, such as:

**Subject matter expertise:**

Consumer/patient/family (caregiver) perspective (required unless approved by COR)

*Insert specific topic*

*Example topics include: Healthcare disparities, Performance measurement, Quality improvement, and Purchaser perspective*
**TEP Expected Time Commitment:**

<Anticipated meeting dates and time frame for measure development activities>

<Types of meetings (e.g., conference call, webinar, in person)>

**Expiration Notice:**

This notice expires on <list the date>.
19  TECHNICAL EXPERT PANEL CHARTER TEMPLATE

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, CMS’s Section 508 Guide for Microsoft Word 2013 may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

<PLEASE DELETE THIS SECTION BEFORE SUBMISSION.>

Project Title: <List the project title>

Dates:

<List projected dates (or time period) of TEP meetings>

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) has <contracted with / provided funding to> <measure developer name> to develop <measure (set) name or description>. The <contract / cooperative agreement / other (please specify)> name is <insert contract name>. The <contract / cooperative agreement / other (please specify)> number is <project number>. As part of its measure development process, <measure developer name> convenes groups of stakeholders and experts who contribute direction and thoughtful input to the measure developer during measure development and maintenance.

Project Objectives:

<List the contract objectives.>

TEP Objectives:

<Clearly state the goals for the TEP, such as whether the TEP is refining concepts, evaluating specifications, or re-evaluating during maintenance, etc.>

Scope of Responsibilities:

<Describe the TEP members’ roles, duties, and degree of authority. For example, the TEP’s role is to provide input and advice to the measure developer on the list of measures under development.>

Guiding Principles:

<Describe how the measure evaluation criteria will be applied and how decisions will be made by the group (e.g., by voting, by consensus). Describe how confidentiality will be handled in the TEP reports, especially for patient participants.>

Estimated Number and Frequency of Meetings:

Date Approved by TEP:

<List the date the charter was approved.>
TEP Membership:

<Attach the TEP Membership List.>
20  **TECHNICAL EXPERT PANEL SUMMARY WEB POSTING**

<INSTRUCTIONS: Instructions and language that may be changed are enclosed in < > and in italics. Headers cannot be changed. The templates, as published, are Section 508 compliant. Any alterations can affect the compliance. For guidance about 508 compliance, [CMS’s Section 508 Guide for Microsoft Word 2013](#) may be a helpful resource.>

<Reminder: Please use the CMS-approved template language and format provided to ensure all necessary information is included in the web posting and to bring consistency across posts. Any changes to the language could result in delays as it will need to be reviewed by CMS.>

<PLEASE DELETE THIS SECTION BEFORE SUBMISSION.>

**Project Title:**  *List the project title as it should appear on the Web posting.*

**Dates:**

The Call for TEP nomination period closed on *<the closing date>.*

The TEP met on *<list date(s) of the meetings>.*

**Documents:**

The TEP Membership List and TEP Summary are posted below in the download section.

**Project Overview:**

The Centers for Medicare & Medicaid Services (CMS) has *<contracted with / provided funding to> measure developer name*> to develop *<measure (set) name or description>*. The *<contract / cooperative agreement / other (please specify)> name* is *<insert contract name>*. The *<contract / cooperative agreement / other (please specify)> number* is *<project number>*. As part of its measure development process, *<measure developer name>* convenes groups of stakeholders and experts who contribute direction and thoughtful input to the measure developer during measure development and maintenance.

**Project Objectives:**

*<List contract objectives>*
Section 5. Glossary and Acronyms
GLOSSARY

Access Measure
A measure that focuses on a patient’s or enrollee’s attainment of timely and appropriate healthcare.

Adopted Measures
If a measure has the same numerator, denominator, data source, and care setting as its parent measure, and the only additional information that needs to be provided is particular to the measure’s implementation use (such as data submission instructions), the measure is considered adopted.

Alignment
Measure alignment is defined by NQF in its Changes to NQF’s Harmonization and Competing Measures Process as “encouraging the use of similar, standardized performance measures across and within public and private sector efforts.” (p. 6) Alignment is achieved when a set of measures works well across settings or programs to produce meaningful information without creating extra work for those responsible for the measurement. Alignment includes using the same quality measures in multiple programs when possible. It can also come from consistently measuring important topics across settings.

Appropriate Use Criteria
The appropriate use criteria are standards that are evidence-based (to the extent feasible) and assist professionals who order and furnish applicable services to make the most appropriate treatment decisions for a specific clinical condition (modified from CMS Appropriate Use Criteria Program).

Attribution
Assignment of the results of a measure to an individual, group, or organization responsible for the decisions, costs, and outcomes (Krumholz et al., 2008).

Audit
A systematic inspection of records or accounts to verify their accuracy.

Bootstrap Analysis
In risk adjustment models, bootstrapping generally refers to estimating properties of a model estimate or the stability of an estimate by sampling from an approximating distribution. This is often accomplished by constructing many resamples of equal size from the observed dataset (e.g., the development sample), where the resamples are smaller than the observed dataset. This technique allows estimation of the sample distribution of a statistic. It can also be used to construct hypothesis tests. In the case of a regression or logistic regression risk adjustment model, it can be used to provide additional guidance regarding the inclusion of risk factors in the model.

Business Case
A business case exists if the entity realizes a financial return on its investment in a reasonable time frame. This may be realized as profit, reduction in losses, or avoided costs. A business case may also exist if the investor believes that a positive indirect effect on organizational function and sustainability will accrue within a reasonable time frame (Leatherman et al., 2003). The business case for a process measure relies on the financial return on the investment necessary to implement the intervention.
advocated by the measure. The business case for other types of measures relies on the financial return resulting from improving the quality of care indicated by the measure.

**Calculation Algorithm**

An ordered sequence of data element retrieval and aggregation through which numerator and denominator events or CV values are identified by a measure. Also referred to as the performance calculation.

**Clinical Practice Guidelines**

Clinical practice guidelines are systematically developed statements to support practitioner and patient decisions about appropriate healthcare for specific clinical circumstances.

**Clinical Quality Language (CQL)**

CQL is an HL7 STU. It is part of the effort to harmonize standards between eCQMs and clinical decision support. CQL provides the ability to express logic that is human-readable yet structured enough for processing a query electronically. It replaces the logic expressions previously defined in the QDM and QDM for use with CQL includes only the method for defining the data elements (i.e., data model). For the most current information, refer to the CQL page on the eCQI Resource Center.

**Clinical Quality Language (CQL) Style Guide**

The CQL Style Guide provides standardize expressions of measure concepts across eCQMs and defines a uniform “look and feel” to eCQM logic using CQL. The guide focuses on a set of common best practices that have been implemented across CQL-based eCQMs in CMS quality reporting programs. The style guide also promotes the use of consistent language within the framework of CQL, including libraries, aliases, definitions, and functions, as well as guidance on other conventions, such as operator precedence. Measure stewards or developers who are developing or specifying eCQMs for future inclusion in CMS programs should align with these best practices.

**Clinical Quality Measure (CQM)**

A mechanism used for assessing the degree to which a provider competently and safely delivers clinical services that are appropriate for the patient in an optimal time frame. CQMs are a subset of the broader category of performance measures.

**CQL-to-ELM Translator Reference Implementation**

The CQL-to-ELM Translator is a specification that describes a formal mechanism for translating the high-level CQL syntax into the canonical called ELM representation. The reference implementation is intended to be used in support of clinical quality framework implementations as a tool to enable CQL output to be uniformly and automatically translated into ELM XML or JSON documents for sharing and distribution to support implementation, integration, translation, and execution of CQL-based artifacts.

**Code System**

A code system is a managed collection of concepts with each concept represented by at least one internally unique code and a human-readable description (e.g., SNOMED CT).
Comparable Data

The accuracy, reproducibility, risk-adjustability, and validity of the measure should not be affected by differences in IT infrastructure, architecture, vendor, or presentation format. Data collection must meet the minimum necessary criteria (i.e., measure definition) stated or intended by the developer, regardless of data source, to populate the fields for measure calculation.

Competing Measures

Competing measures address the same topic and the same population. This term is used when considering harmonization. Refer to Related Measures.

Composite Performance Measure

NQF defines composite performance measure in Composite Performance Measure Evaluation Guidance as “a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure, also called composite measures, with a single score.” (p. 4)

Conflict of Interest

A conflict of interest exists when an individual (or entity) has more than one motivation for trying to achieve an objective. In measure development, this situation arises when an individual has opportunities to affect specifications for quality measures that impact an interest with which the individual has a relationship.

Construct Validity

The extent to which the measure actually quantifies what the theory says it should. Construct validity evidence often involves empirical and theoretical support for the interpretation of the construct.

Continuous Variable (CV)

A measure score in which each individual value for the measure can fall anywhere along a continuous scale and can be aggregated using a variety of methods such as the calculation of a mean or median (e.g., mean number of minutes between presentation of chest pain to the time of administration of thrombolytics).

Convergent Validity (concurrent validity)

Refers to the degree to which multiple indicators of a single underlying concept are correlated.

Cost of Care

The total healthcare spending, including total resource use and unit price, by payer or consumer, for a healthcare service or group of healthcare services associated with a specified patient population, time period, and unit of clinical accountability.

Criterion

An accepted standard, principle, or rule used to make a decision or to inform an evaluator’s judgment.

C-statistic

Used to assess risk-adjusted models, it indicates the ability of the model to discriminate between one event and the other. If a model discriminates randomly, c = 0.5. If the risk factor modeling predicts the
outcome well, then discrimination increases. The higher the c-statistic, the better the predictive power of the model.

Data Aggregation
Combining data from multiple sources to generate performance information.

Data Element, Critical
Those elements that contribute most to the computed measure score, that is, account for identifying the greatest proportion of the target condition, event, or outcome being measured (numerator); the target population (denominator); population excluded (exclusion); and when applicable, risk factors with largest contribution to variability in outcome.

Data Element, Quality
A single piece of information that is used in quality measures to describe part of the clinical care process, including both a clinical entity and its context of use (e.g., diagnosis, active).

Data Sources
The primary source document(s) used for data collection (e.g., billing or administrative data, encounter form, enrollment forms, patient medical record).

Denominator
The denominator is a statement that describes the population evaluated by the performance measure and is the lower part of a fraction used to calculate a rate, proportion, or ratio. It can be the same as the initial population or a subset of the initial population to further constrain the population for the purpose of the measure. CV measures do not have a denominator, but instead define a measure population.

Denominator Exception
Those conditions that should remove a patient, procedure, or unit of measurement from the denominator of the performance rate only if the numerator criteria are not met. A denominator exception allows for adjustment of the calculated score for those providers with higher risk populations. A denominator exception also provides for the exercise of clinical judgment and should be specifically defined where capturing the information in a structured manner fits the clinical workflow. A denominator exception is used only in proportion measures.

These cases are removed from the denominator. However, the number of patients with valid exceptions may still be reported. Allowable reasons fall into three general categories:

- Medical reasons
- Patient reasons
- System reasons.

Denominator Exclusion
Patients who should be removed from the measure population and denominator before determining whether numerator criteria are met. Denominator exclusions are used in proportion and ratio measures to help narrow the denominator. For example, patients with bilateral lower extremity amputations would be listed as a denominator exclusion for a measure requiring foot exams.
De novo eCQM

New eCQM that is not based on an existing measure. De novo eCQMs must adhere to the NQF measure submission process and requirements for eCQM submissions outlined in the Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement – Requirements for Endorsing eMeasures.

Direct Reference Code (DRC)

A specific code that is referenced directly in the eCQM logic to describe a data element or one of its attributes. DRC metadata include the description of the code, the code system from which the code is derived, and the version of that code system.

Discriminant Validity

The degree to which a test of a concept (a quality measure) is not highly correlated with other tests designed to measure theoretically different concepts. It may be demonstrated by assessing variation across multiple comparison groups (such as healthcare providers) to show that a performance measure can differentiate between disparate groups that it should theoretically be able to distinguish.

Disparities in Healthcare

Differences in health outcomes and their determinants between segments of the population, as defined by social, demographic, environmental, and geographic attributes (Truman et al., 2011, p. 3).

Dry Run

Full-scale measure testing involving all providers/practitioners representing the full spectrum of the population being measured. The purpose is to finalize all methodologies related to case identification/selection, data collection, and measurement calculation; and to quantify unintended consequences.

Efficiency Measure

A measure that evaluates the resource use (or cost) associated with a specific level of performance with respect to the aims of quality. For example, a provider in the healthcare system would be efficient if it was able to maximize output for a given set of inputs or to minimize inputs used to produce a given output.

Electronic Clinical Quality Measure (eCQM)

CQM that is expressed and formatted to use data from EHRs and/or health IT systems to measure healthcare quality, ideally data captured in structured form during the process of patient care. For the eCQM to be reported from an EHR, the HQMF is used to format the eCQM content using CQL to express the logic and QDM to express the data elements needed to evaluate a provider or organization’s performance. These electronic specifications include:

- **HTML** – A web-facing, human-readable document of some of the XML file content so that the user can understand how the elements are defined and the logic used to calculate the measure. HTML is divided into two parts: the header, which provides narrative details on the measure itself; and the body, which contains the data criteria and logic for how the measure is calculated.
- **XML** – A computer-readable format that describes the logic content and allows for the creation of queries against an EHR (or other data store) for quality reporting. The XML conforms to HQMF standards.
• **Value Sets** and/or **DRCs** – Specific codes and code systems to capture clinical concepts and patient data in the EHR system. Value sets and DRCs provide definitions of the codes necessary to calculate the eCQM. The value sets and DRCs for each measure are stored by the NLM VSAC. Through the VSAC, providers, implementers, and developers can access the value sets and DRCs for each eCQM.

**Electronic Clinical Quality Measure (eCQM) Logic and Implementation Guidance**

The eCQM Logic and Implementation Guidance is a resource document which provides guidance for understanding, using, and/or implementing eCQMs.

**eCQM Union**

The eCQM union operator within the CQL is used to combine two or more data elements, arguments or expressions onto a list, such that any item contained in the combined list fulfills the criteria of the union.

**Electronic Health Record (EHR)**

Also known as the electronic patient record, electronic medical record, or computerized patient record. As defined by Healthcare Information Management and Systems Society, “the EHR is a longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting. Included in this information are patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, and imaging reports.”

**Empirical Evidence**

Data or information resulting from studies and analyses of the data elements and/or scores for a measure as specified, whether unpublished or published.

**Encounter**

Defined by the ASTM International as: “(1) an instance of direct provider/practitioner to patient interaction, regardless of the setting, between a patient and a practitioner vested with primary responsibility for diagnosing, evaluating or treating the patient’s condition, or both, or providing social worker services; and (2) a contact between a patient and a practitioner who has primary responsibility for assessing and treating the patient at a given contact, exercising independent judgment.” Encounter serves as a focal point linking clinical, administrative and financial information. Encounters occur in many settings—ambulatory care, inpatient care, emergency care, home healthcare, field and virtual (telemedicine).

**Environmental Scan**

The process of systematically reviewing and interpreting data to identify issues and opportunities that will influence prioritization of current or future plans.

**Exception**

Refer to [denominator exception](#).

**Exclusion**

Refer to [denominator exclusion](#) and [numerator exclusion](#).
Expert Consensus

Recommendations formulated by one of several formal consensus development methods such as consensus development conference, Delphi method, and nominal group technique.

Face Validity

The extent to which a test appears to cover the concept it purports to measure “at face value.” It is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with blood pressure < 140/90 is a marker of quality.)

Feasibility Criteria

Extent to which the specifications, including measure logic, require data that are readily available or that could be captured without undue burden and can be implemented for performance measurement.

Gaming

When providers exploit weaknesses in the measurement system to tweak the data to make their outcomes look better than they actually are. Includes limiting access to certain populations, neglecting care, or overuse of medications or services to ensure that the measure results are favorable.

Grey Literature

Unpublished or not commercially indexed material that can include any documentary materials issued by government, academia, business, and industry such as technical reports, working papers, and conference proceedings. For example, contributors to the New York Academy of Medicine Grey Literature website include the AHRQ, NQF, CDC, HHS, The Joint Commission, National Academy of Sciences, RAND, and RTI International.

Guide for Reading eCQMs

The Guide for Reading eCQMs is a resource to assist stakeholders in interpreting and understanding eCQMs. The guide provides information on eCQMs such as file naming conventions, understanding an eCQM human-readable rendition, value sets, QDM data criteria, and more.

Harmonization

The standardization of specifications for related measures with the same measure focus (e.g., influenza immunization of patients in hospitals or nursing homes); related measures for the same target population (e.g., eye exam and HbA1c for patients with diabetes); or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are justified (i.e., dictated by the evidence). The dimensions of harmonization can include numerator, denominator, exclusion, calculation, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources. Value sets used in measures (especially eCQMs) should be harmonized when the intended meaning is the same. Harmonization of logic in eCQMs is beneficial when the data source in the EHR is the same.

Health Information Technology (Health IT)

Per Section 3000 of the HITECH Act, the term ‘health information technology’ means “hardware, software, integrated technologies or related licenses, intellectual property, upgrades, or packaged
solutions sold as services that are designed for or support the use by healthcare entities or patients for the electronic creation, maintenance, access, or exchange of health information.”

**Health Information Technology for Economic and Clinical Health (HITECH) Act**

A provision within [ARRA](#) that authorizes incentive payments through Medicare and Medicaid to hospitals and clinicians toward meaningful use of EHRs.

**Health Level Seven International (HL7)**

A standards-developing organization that provides framework and standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery, and evaluation of health services.

**Health Quality Measure Format (HQMF)**

A standards-based representation of quality measures as electronic documents. A quality measure expressed in this way is referred to as an eCQM.

**Hosmer-Lemeshow test (HL test)**

The [HL test](#) is a goodness of fit test for logistic regression, especially for risk prediction models. A goodness-of-fit test tells you how well your data fits the model. Specifically, the HL test calculates if the observed event rates match the expected event rates in population subgroups.

The test is only used for [binary](#) response variables (i.e., a variable with two outcomes such as alive or dead, yes or no).

**Impact of a Measure (importance subcriterion)**

Now called High Priority by NQF. The measure topic addresses a specific national health goal or priority; affects large numbers of patients; is a leading cause of morbidity/mortality; high resource use and severity of patient/societal consequences of poor quality. For PROs, there is evidence that the target population values the PRO and finds it meaningful.

**Importance Criterion**

Extent to which the specific measure focus is important to making significant gains in healthcare quality (e.g., safety, timeliness, effectiveness, efficiency, equity, patient centeredness) and improving health outcomes for a specific high-impact aspect of healthcare where there is variation in or overall poor performance.

**Initial Population**

Refers to all events to be evaluated by a specific performance measure involving patients who share a common set of specified characteristics within a specific measurement set to which a given measure belongs. All patients counted (e.g., as numerator, as denominator) are drawn from the initial population.

**Intermediate Outcome**

An intermediate outcome is a (measured) change in physiologic state that leads to a long-term health outcome. There should be a body of evidence that the measured intermediate outcome leads to a desired health outcome.
**Internal Consistency Reliability Testing**

Testing a multiple item test or survey to assess the extent that the items designed to measure a given construct are inter-correlated. Pertains to survey type measures and to the data elements used in measures constructed from patient assessment instruments.

**Inter-Rater (inter-abstractor) Reliability Testing**

Assesses the extent to which observations from two or more human observers are congruent with each other.

**Inverse Measures**

Inverse measures are measures where a lower performance rate is better. For example, the National Healthcare Safety Network calculates most HAIs as a standardized infection ratio (SIR). The SIR compares the actual number of HAIs (i.e., the numerator) with the predicted number based on the baseline U.S. experience (e.g., standard population), adjusting for several risk factors that have been found to be most associated with differences in infection rates. The goal is to have the numerator equal to or very close to zero, thereby, having an SIR equal to or very close to zero.

**Jira**

A software application that tracks issues and bugs. It also allows users to quickly search issues that have or are currently being resolved. HHS Groups are using the ONC Project Tracking System (Jira) to track issues with quality measures.

**Kappa Coefficient**

A statistical measure of inter-rater agreement for qualitative (categorical) items. Cohen’s kappa can be thought of as a chance-corrected proportional agreement. Possible values range from +1 (perfect agreement), 0 (no agreement above that expected by chance) to -1 (complete disagreement).

**Lean Kaizen**

A Japanese phrase meaning continuous quality improvement by eliminating waste. The principles were implemented after World War II, influenced by American quality management teachers who visited Japan.

**Level of Analysis**

The performance measurement level (e.g. clinician, health plan, county populations).

**Logic**

The criteria used to define a quality measure and its key components.

**Material Change**

A material change is one that changes the specifications of an endorsed measure to affect the original measure’s concept or logic, the intended meaning of the measure, or the strength of the measure relative to the measure evaluation criteria.
**Measure (performance measure or quality measure)**

AHRQ defines a measure as a mechanism to assign a quantity to an attribute to enable comparisons among entities over time. A measure may stand alone or belong to a composite, subset, set, and/or collection of measures. NQF states a healthcare performance measure is a way to calculate whether and how often the healthcare system does what it should. Measures are based on scientific evidence about processes, outcomes, perceptions, or systems that relate to high-quality care.

**Measure Applications Partnership (MAP)**

An NQF-convened, multi-stakeholder group that provides input to HHS on the list of measures for use in a specified program. The MAP consists of multiple working groups, including Clinicians, Post-Acute Care/LTC, Hospitals, and Dual-Eligible Beneficiaries.

**Measure Authoring Tool (MAT)**

A publicly available, web-based tool for measure developers to create eCQMs; it should also reduce the time required to create new quality measures, and to convert existing paper-based measures into EHR-readable format.

**Measure Maintenance**

Periodic and consistent reviewing, evaluating, and updating of performance measures to ensure continued reliability, validity, feasibility, importance, usability, and currency with science. Also involves comparison to similar measures for potential harmonization.

**Measure Score**

The numeric result that is computed by applying the measure specifications and scoring algorithm. The computed measure score represents an aggregation of all appropriate patient-level data (e.g., proportion of patients who died, average lab value attained) for the entity being measured (e.g., hospital, health plan, home health agency, clinician). The measure specifications designate the entity that is being measured and to whom the measure applies.

**Measure Set**

A measure set is a group of measures related in some way such as measures addressing a specific condition, procedure, or specialty.

**Measure Steward**

An individual or organization that owns a measure and is responsible for maintaining the measure. Measure stewards are often the same as measure developers, but not always. Measure stewards are also the ongoing point of contact for people interested in a given measure.

**Measure Testing**

Empirical analysis to demonstrate the reliability and validity of the measure as specified, including analysis of issues that pose threats to the validity of conclusions about quality of care such as exclusions, risk adjustment/stratification for outcome and resource use measures, methods to identify differences in performance, and comparability of data sources/methods.
Measures Under Consideration (MUC)

A list of quality and efficiency measures HHS is considering adopting, through the federal rulemaking process, for use in the Medicare program. Made publicly available by December 1 each year for categories of measures that are described in section 1890(b)(7)(B)(i)(I) of the SSA as amended by Section 3014 of the ACA.

Medical Record (data source)

Data obtained from the records or documentation maintained on a patient in any healthcare setting (e.g., hospital, home care, LTC, practitioner office). Includes automated and paper medical record systems.

Metadata

Metadata is data that describes data.

Minor Change

A minor change does not change the process of data collection, aggregation, or calculation, nor does it change the intended meaning of the measure or the strength of the measure in terms of the measure evaluation criteria.

Morbidity

The rate of incidence of disease. For example, lumbar puncture, if improperly performed, may be followed by a significant morbidity. It also can refer to the relative incidence of a particular disease state or symptom.

Mortality

The number of deaths in a given time or place. The proportion of deaths to population. “Death rate” is also called “mortality rate.”

Multiple Chronic Conditions (MCC)

The NQF Multiple Chronic Conditions Measurement Framework defines MCC as “having two or more concurrent chronic conditions that collectively have an adverse effect on health status, function, or quality of life and that require complex healthcare management, decision-making, or coordination.” (pp. 7-8)

Non-parametric Methods

Non-parametric methods are a type of statistical test not involving the estimation of parameters of a statistical function. (Non-parametric, n.d.)

Null Performance Rate

If the measure is not applicable for all patients within the sample, the performance rate would be 0/0 (null) and would be considered satisfactorily reporting.

Numerator

The upper portion of a fraction used to calculate a rate, proportion, or ratio. Also called the measure focus, it is the target process, condition, event, or outcome. Numerator criteria are the processes or
outcomes expected for each patient, procedure, or other unit of measurement defined in the denominator. A numerator statement describes the clinical action that satisfies the conditions of the performance measure.

**Numerator Exclusion**

Defines instances that should not be included in the numerator data. Numerator exclusions are used only in ratio and proportion measures.

**Opportunity for Improvement**

Data demonstrates considerable variation or overall less-than-optimal performance, in the quality of care across providers, and/or there are disparities in care across population groups.

**Outcome Measure**

A measure that assesses the results of healthcare that are experienced by patients: clinical events, recovery and health status, experiences in the health system, and efficiency/cost.

**Paperwork Reduction Act (PRA)**

The PRA mandates that all federal government agencies must obtain approval from the OMB before collection of information that will impose a burden on the public. Measure developers should be familiar with the PRA before implementing any process that involves the collection of new data.

**Parametric Methods**

Parametric methods make certain assumptions about a data set; namely, that the data are drawn from a population with a normal distribution. Parametric methods generally have high statistical power. (Tyler, 2017)

**Patient-Reported Outcome (PRO)**

The FDA Guidance for Industry PRO Measures: Use in Medical Product Development to Support Labeling Claims defines PRO as “any report of the status of a patient’s health condition, health behavior, or experience with healthcare that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else.” (p. 2) This definition reflects the key domains of:

- Health-related quality of life (including functional status)
- Symptoms and symptom burden (e.g., pain, fatigue)
- Experience with care
- Health behaviors (e.g., smoking, diet, exercise).

**Patient-Reported Outcome Measure (PROM)**

The NQF defines PROM in PROs in Performance Measurement as an “instrument, scale, or single-item measure used to assess the PRO concept as perceived by the patient, obtained by directly asking the patient to self-report.” (p. 27)

---

Note that CMS and other HHS agencies define and use the term “domain” differently from one another. Therefore, within the Blueprint the term “domain” is defined differently in different contexts, depending on the relevant agency within the discussion.
Patient-Reported Outcome-based Performance Measure (PRO-PM)

The NQF defines PRO-PM in PROs in Performance Measurement as “a performance measure that is based on PROM data aggregated for an accountable healthcare entity.” (p. 27)

Pilot Testing

Measure testing (sometimes referred to as pilot testing) is divided into two main types:

- Alpha testing (also called formative testing)
- Beta testing (also called field testing).

Population

The total group of people of interest for a quality measure, sometimes called the initial population. The measure population is a defined subset appropriate to the measure set who are not excluded from the individual measure.

Predictive Validity

Ability of measure scores to predict scores on some other related valid measure. The degree to which the operationalization can predict (or correlate) with other measures of the same construct that are measured at some time in the future.

Process Measure

A measure that focuses on steps that should be followed to provide good care. There should be a scientific basis for believing that the process, when executed well, will increase the probability of achieving a desired outcome.

Proportion

A score derived by dividing the number of cases that meet a criterion for quality (i.e., the numerator) by the number of eligible cases within a given time frame (i.e., the denominator) where the numerator cases are a subset of the denominator cases (e.g., percentage of eligible women with a mammogram performed in the last year).

Public Domain

“The realm embracing property rights that belong to the community at large, are unprotected by copyright or patent, and are subject to appropriation by anyone”.

Qualified Clinical Data Registry (QCDR)

A QCDR is a CMS-approved vendor that is in the business of improving healthcare quality. These organizations may include specialty societies, regional health collaboratives, large health systems or software vendors working in collaboration with one of these medical entities.64

---

64 Beginning with the MIPS 2020 performance period a QCDR will be defined as an entity with clinical expertise in medicine and in quality measurement development that collects medical or clinical data on behalf of a MIPS eligible clinician for the purpose of patient and disease tracking to foster improvement in the quality of care provided to patients.
**Qualified Registry**

A Qualified Registry is a vendor that collects clinical data from an individual MIPS eligible clinician, group, or virtual group and submits it to CMS on their behalf.

**Quality Data Model (QDM)**

A QDM is an information model that defines relationships between patients and clinical concepts in a standardized format to enable electronic quality performance measurement. The model is the current structure for electronically representing quality measure concepts for stakeholders involved in electronic quality measurement development and reporting. The QDM provides the language that defines the criteria for clinical quality measurement. It allows the electronic definition of a clinical concept via its data elements and provides the vocabulary to relate them to each other. By relating attributes between data elements and using filtering functions, the QDM provides a method to construct complex clinical representations for eCQMs.

**Quality Measure (or performance measure)**

Numeric quantification of healthcare quality for a designated accountable healthcare entity, such as hospital, health plan, nursing home, or clinician. A healthcare performance measure is a way to calculate whether and how often the healthcare system does what it should. Measures are based on scientific evidence about processes, outcomes, perceptions, or systems that relate to high-quality care.

**Quality Reporting Document Architecture (QRDA)**

QRDA is a standard document format for the exchange of eCQM data. QRDA documents:

- Contain data extracted from EHRs and other Health IT systems
- Can be used to exchange eCQM data between systems
- Serve as the data submission standards for a variety of quality measurement and reporting initiatives
- Were adopted by the ONC as the standard to support both QRDA Category I (individual patient) and QRDA Category III (provider’s aggregate) data submission.

**R² Statistic**

Values for R² describe how well the outcome can be predicted based on the values of the risk factors or predictors. It is frequently used to assess the predictive power of specific types of risk-adjusted models.

**Ratio**

A score that is derived by dividing a count of one type of data by a count of another type of data (e.g., number of patients with central lines who develop infection divided by the number of central line days). The key to the definition of a ratio is that the numerator is not in the denominator.

**Receiver-Operating Characteristic (ROC) Curve**

The graph that provides the c-statistic value. The ROC curve graphs the predictive accuracy of a logistic regression model.

**Related Measures**

Measures that address either the same topic or the same population. This term is used when considering harmonization. Refer to Competing Measures.
Reliability (part of scientific acceptability)

Reflects that the measure is well defined and precisely specified so it can be implemented consistently within and across organizations and that it distinguishes differences in performance.

Reliability Testing

Evaluates whether the measure data elements are extracted over time, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. Reliability is often referred to as inter-rater or inter-observer reliability, which also applies to abstractors and coders. It can also refer to the amount of error associated with the computed measure scores (e.g., signal vs. noise).

Respecified Measure

An existing measure is changed to fit the current purpose or use. This may mean changing a measure to meet the needs of a different care setting, data source, or population; or, it may mean changes to the numerator, denominator, or adding specifications to fit the current use.

Resource Unit

Refers to the resources used to provide care to a patient or population. Resource units are generally identified through claims data and measured in terms of dollars, but they can also include resource not captured on a claim (e.g., nursing hours).

Resource Use Measures

Also called cost and resource use measures. Refers to broadly applicable and comparable measures of health services counts (in terms of units or dollars) applied to a population or event (broadly defined to include diagnoses, procedures, or encounters). A resource use measure counts the frequency of defined health system resources. Some measures may monetize the health service by applying a dollar amount such as allowable charges, paid amounts, or standardized prices to each unit of resource use.

Risk Adjustment

Statistical process used to identify and adjust for extraneous variables not associated with care delivery that threaten validity because they affect the outcome being measured outside of the health system’s control. The purpose is a fairer and more accurate comparison of outcomes of care across healthcare organizations or providers.

Sample

A subset of a population. The subset should be chosen in such a way that it accurately represents the whole population with respect to some characteristic of interest. A sampling frame lists all eligible cases in the population of interest (i.e., denominator) and how they are selected.

Scientific Acceptability of the Measure Properties

Extent to which the measure, as specified, produces consistent (i.e., reliable) and credible (i.e., valid) results about the quality of care when implemented.

Score (measure score)

The NQF Glossary of Terms defines measure score as “the numeric result that is computed by applying the measure specifications and scoring algorithm. The computed measure score represents an
aggregation of all the appropriate patient-level data (e.g., proportion of patients who died, average lab value attained) for the entity being measured (e.g., hospital, health plan, home health agency, clinician). The measure specifications designate the entity that is being measured and to whom the measure score applies.” (p. 12)

**Scoring**

Method(s) applied to data to generate results/score. Most quality measures produce rates; however, other scoring methods include categorical value, CV, count, frequency distribution, non-weighted score/composite scale, ratio, and weighted score/composite/scales.

**Semantic Validation**

Method of testing the validity of an eCQM whereby the formal criteria in an eCQM are compared to a manual computation of the measure from the same test database.

**Sensitivity**

As a statistical term, refers to the proportion of actual positives that are correctly identified as such (e.g., percentage of people with diabetes who are correctly identified as having diabetes). Refer to **Specificity**.

**Specifications**

Measure instructions that address data elements, data sources, point of data collection, timing and frequency of data collection and reporting, specific instruments to be used (if appropriate), and implementation strategies.

**Specificity**

As a statistical term, refers to the proportion of negatives that are correctly identified (e.g., percentage of healthy people who are correctly identified as not having the condition). Perfect specificity would mean that the measure recognizes all actual negatives (e.g., all healthy people will be recognized as healthy). Refer to **Sensitivity**.

**Stratification**

Divides a population or resource services into distinct, independent groups of similar data, enabling analysis of the specific subgroups. This type of adjustment can show where disparities exist or where there is a need to expose differences in results.

**Structural (as a measure type)**

Features of a healthcare organization or clinician relevant to the capacity to provide healthcare. This may include, but is not limited to, measures that address health IT infrastructure, provider capacity, systems, and other healthcare infrastructure supports.

**Structural Measure**

Measure that assesses features of a healthcare organization or clinician relevant to its capacity to provide healthcare.
Systematic Literature Review

A review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research. A systematic literature review also collects and analyzes data from studies that are included in the review. Two sources of systematic literature reviews are the AHRQ Evidence-Based Clinical Information Reports and The Cochrane Library.

Target Population

The numerator (i.e., cases) and denominator (i.e., population sample meeting specified criteria) of the measure.

Test-retest Reliability Testing

Assesses the extent to which a survey or measurement instrument elicits the same response from the same respondent across short intervals of time.

Time Interval

Time frame used to determine cases for inclusion in the denominator, numerator, or exclusion. The time interval includes an index event and period of time.

Topped-Out

Sometimes referred to as topped off. A measure has reached a level where rates can no longer increase, so there is no opportunity for performance improvement.

Usability

The NQF defines usability in Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement as the “extent to which potential audiences (e.g., consumers, purchasers, providers, policymakers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations” (p. 25).

Validation

Testing to determine whether the measure accurately represents the concept being evaluated and achieves the purpose for which it is intended (i.e., to measure quality). Validation also is used in reference to statistical risk models where model performance metrics are compared between two different samples of data called the development and validation samples.

Validity (scientific acceptability of measure properties subcriterion)

Measure validity: The measure accurately represents the concept being evaluated and achieves the purpose for which it is intended (i.e., to measure quality). For example, the measure:

- Clearly identifies the concept being evaluated (i.e., face validity)
- Includes all necessary data elements, codes, and tables to detect a positive occurrence when one exists (i.e., construct validity)
- Includes all necessary data sources to detect a positive occurrence when one exists (i.e., construct validity).
**Data element validity**: The extent to which the information represented by the data element or code used in the measure reflects the actual concept or event intended. For example:

- A medication code is used as a proxy for a diagnosis code
- Data element response categories include all values necessary to provide an accurate response.

**Validity Testing**

Empirical analysis of the measure as specified that demonstrates that data are correct and/or conclusions about quality of care based on the computed measure score are correct. Validity testing focuses on systematic errors and bias.

**Validity Threats**

In addition to unreliability, some aspects of measure specifications and data can affect the validity of conclusions about quality. Potential threats include patients excluded from measurement, differences in patient mix for outcome and resource use measures, measure scores generated with multiple data sources/methods, and systematic missing or “incorrect” data (unintentional or intentional).

**Value Set**

Subset of concepts drawn from one or more code systems, where the concepts included in the subset share a common scope of use (e.g., Anticoagulant Therapy).

**Value Set Authority Center (VSAC)**

Central repository for the official versions of value sets that support eCQMs. The NLM maintains the VSAC and provides downloadable access to the value sets and the BPS. The VSAC provides measure developers with tools to search existing value sets, collaborate with other measure developers to harmonize value sets, to create new value sets, and to maintain value set content consistent with current versions of the terminologies they use.
ACRONYMS

AARP  American Association of Retired Persons
ACA  Affordable Care Act
ACE  Angiotensin Converting Enzyme
AHRQ  Agency for Healthcare Research and Quality
AMA  American Medical Association
AMI  Acute Myocardial Infarction
ANOVA  Analysis of Variance
AOE  Acute Otitis Externa
APA  American Psychological Association
API  Application Programming Interface
APM  Alternative Payment Model
ARRA  American Recovery and Reinvestment Act of 2009
ASC  Ambulatory Surgical Center
ASCQR  Ambulatory Surgical Center Quality Reporting
ASPE  Office of the Assistant Secretary for Program and Evaluation
AUC  Area Under the Curve
BPS  Binding Parameter Specification
CABG  Coronary Artery Bypass Graft
CAHPS  Consumer Assessment of Healthcare Providers and Systems
CAU  CQM Annual Update
CBE  Consensus-Based Entity
C-CDA  Consolidated Clinical Document Architecture
CDA  Clinical Document Architecture
CDC  Centers for Disease Control and Prevention
CEHRT  Certified EHR Technology
CHIP  Children’s Health Insurance Program
CHIPRA  Children’s Health Insurance Program Reauthorization Act of 2009
CLABSI  Central Line-Associated Blood Stream Infection
CLD  Content Logical Definition
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMD</td>
<td>Collaborative Measure Development</td>
</tr>
<tr>
<td>CMIT</td>
<td>CMS Measures Inventory Tool</td>
</tr>
<tr>
<td>CMS</td>
<td>Centers for Medicare &amp; Medicaid Services</td>
</tr>
<tr>
<td>COPD</td>
<td>Chronic Obstructive Pulmonary Disease</td>
</tr>
<tr>
<td>COR</td>
<td>Contracting Officer’s Representative</td>
</tr>
<tr>
<td>CPT</td>
<td>Current Procedural Terminology</td>
</tr>
<tr>
<td>CPT4</td>
<td>CPT, 4th Edition</td>
</tr>
<tr>
<td>CQL</td>
<td>Clinical Quality Language</td>
</tr>
<tr>
<td>CQM</td>
<td>Clinical Quality Measure</td>
</tr>
<tr>
<td>CQS</td>
<td>Composite Quality Score</td>
</tr>
<tr>
<td>CRP</td>
<td>Change Review Process</td>
</tr>
<tr>
<td>CSAC</td>
<td>Consensus Standards Approval Committee</td>
</tr>
<tr>
<td>CT</td>
<td>Computerized Tomography</td>
</tr>
<tr>
<td>CV</td>
<td>Continuous Variable</td>
</tr>
<tr>
<td>CY</td>
<td>Calendar Year</td>
</tr>
<tr>
<td>DEL</td>
<td>Data Element Library</td>
</tr>
<tr>
<td>DENEX</td>
<td>Denominator Exclusion</td>
</tr>
<tr>
<td>DENEXCEP</td>
<td>Denominator Exclusion Exception</td>
</tr>
<tr>
<td>DENOM</td>
<td>Denominator</td>
</tr>
<tr>
<td>DERep</td>
<td>Data Element Repository</td>
</tr>
<tr>
<td>DoD</td>
<td>Department of Defense</td>
</tr>
<tr>
<td>DRC</td>
<td>Direct Reference Code</td>
</tr>
<tr>
<td>DTS</td>
<td>Distributed Terminology System</td>
</tr>
<tr>
<td>DUA</td>
<td>Data Use Agreement</td>
</tr>
<tr>
<td>eCQI</td>
<td>electronic Clinical Quality Improvement</td>
</tr>
<tr>
<td>eCQM</td>
<td>electronic Clinical Quality Measure</td>
</tr>
<tr>
<td>ED</td>
<td>Emergency Department</td>
</tr>
<tr>
<td>EHR</td>
<td>Electronic Health Record</td>
</tr>
<tr>
<td>ELM</td>
<td>Expression Logical Model</td>
</tr>
<tr>
<td>ESRD</td>
<td>End-Stage Renal Disease</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>ESST</td>
<td>Environmental Scanning Support Tool</td>
</tr>
<tr>
<td>FACA</td>
<td>Federal Advisory Committee Act</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
</tr>
<tr>
<td>FFS</td>
<td>Fee for Service</td>
</tr>
<tr>
<td>FHIR</td>
<td>Fast Healthcare Interoperability Resources</td>
</tr>
<tr>
<td>FOTO</td>
<td>Focus On Therapeutic Outcomes</td>
</tr>
<tr>
<td>FPL</td>
<td>Federal Poverty Level</td>
</tr>
<tr>
<td>FVC</td>
<td>Forced Vital Capacity</td>
</tr>
<tr>
<td>GPCK</td>
<td>Generic Pack</td>
</tr>
<tr>
<td>GRADE</td>
<td>Grading of Recommendation, Assessment, Development, and Evaluation</td>
</tr>
<tr>
<td>GUID</td>
<td>Globally Unique Identifier</td>
</tr>
<tr>
<td>HAC</td>
<td>Hospital-acquired Condition</td>
</tr>
<tr>
<td>HACRP</td>
<td>Hospital-Acquired Condition Reduction Program</td>
</tr>
<tr>
<td>HAI</td>
<td>Healthcare-Associated Infection</td>
</tr>
<tr>
<td>HCPCS</td>
<td>Healthcare Common Procedure Coding System</td>
</tr>
<tr>
<td>HCQIS</td>
<td>Healthcare Quality Information Systems</td>
</tr>
<tr>
<td>HD</td>
<td>Hemodialysis</td>
</tr>
<tr>
<td>HDL</td>
<td>High Density Lipoprotein</td>
</tr>
<tr>
<td>HHA</td>
<td>Home Health Agency</td>
</tr>
<tr>
<td>HHQR</td>
<td>Home Health Quality Reporting</td>
</tr>
<tr>
<td>HHS</td>
<td>Health and Human Services</td>
</tr>
<tr>
<td>HIPAA</td>
<td>Health Insurance Portability and Accountability Act</td>
</tr>
<tr>
<td>HITECH</td>
<td>Health Information Technology for Economic and Clinical Health Act</td>
</tr>
<tr>
<td>HITEP</td>
<td>Health Information Technology Expert Panel</td>
</tr>
<tr>
<td>HITSC</td>
<td>Health Information Technology Standards Committee</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Viruses</td>
</tr>
<tr>
<td>HL</td>
<td>Hosmer-Lemeshow</td>
</tr>
<tr>
<td>HL7</td>
<td>Health Level Seven International</td>
</tr>
<tr>
<td>HOS</td>
<td>Health Outcomes Survey</td>
</tr>
<tr>
<td>HQI</td>
<td>Hospital Quality Incentive</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>HQMF</td>
<td>Health Quality Measure Format</td>
</tr>
<tr>
<td>HQR</td>
<td>Hospice Quality Reporting</td>
</tr>
<tr>
<td>HRRP</td>
<td>Hospital Readmissions Reduction Program</td>
</tr>
<tr>
<td>HSLOC</td>
<td>Healthcare Service Location Code</td>
</tr>
<tr>
<td>HTML</td>
<td>Hyper Text Markup Language</td>
</tr>
<tr>
<td>HVBP</td>
<td>Hospital Value-Based Purchasing</td>
</tr>
<tr>
<td>ICC</td>
<td>Intraclss Correlation Coefficient</td>
</tr>
<tr>
<td>ICD</td>
<td>International Classification of Diseases</td>
</tr>
<tr>
<td>ICF</td>
<td>International Classification of Functioning, Disability, and Health</td>
</tr>
<tr>
<td>ICU</td>
<td>Intensive Care Unit</td>
</tr>
<tr>
<td>IG</td>
<td>Implementation Guide</td>
</tr>
<tr>
<td>IHI</td>
<td>Institute for Healthcare Improvement</td>
</tr>
<tr>
<td>IMPACT</td>
<td>Improving Medicare Post-Acute Care Transformation</td>
</tr>
<tr>
<td>IN</td>
<td>Specific Ingredient</td>
</tr>
<tr>
<td>IOM</td>
<td>Institute of Medicine</td>
</tr>
<tr>
<td>IP</td>
<td>Initial Population</td>
</tr>
<tr>
<td>IPFQR</td>
<td>Inpatient Psychiatric Facility Quality Reporting</td>
</tr>
<tr>
<td>IQR</td>
<td>Inpatient Quality Reporting</td>
</tr>
<tr>
<td>IRB</td>
<td>Institutional Review Board</td>
</tr>
<tr>
<td>IRF</td>
<td>Inpatient Rehabilitation Facility</td>
</tr>
<tr>
<td>IS-A</td>
<td>Increasing Specificity</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>JSON</td>
<td>JavaScript Object Notation</td>
</tr>
<tr>
<td>LA</td>
<td>LOINC Answers</td>
</tr>
<tr>
<td>LDL</td>
<td>Low-density Lipoprotein</td>
</tr>
<tr>
<td>LOINC</td>
<td>Logical Observation Identifiers Names and Codes</td>
</tr>
<tr>
<td>LTC</td>
<td>Long-term Care</td>
</tr>
<tr>
<td>LTCH</td>
<td>Long-Term Care Hospital</td>
</tr>
<tr>
<td>MACPAC</td>
<td>Medicaid and CHIP Payment and Access Commission</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>MACRA</td>
<td>Medicare Access and CHIP Reauthorization Act</td>
</tr>
<tr>
<td>MAP</td>
<td>Measure Applications Partnership</td>
</tr>
<tr>
<td>MAT</td>
<td>Measure Authoring Tool</td>
</tr>
<tr>
<td>MCC</td>
<td>Multiple Chronic Conditions</td>
</tr>
<tr>
<td>MCCB</td>
<td>MAT Change Control Board</td>
</tr>
<tr>
<td>MDP</td>
<td>Measurement Development Plan</td>
</tr>
<tr>
<td>MDS</td>
<td>Minimum Data Set</td>
</tr>
<tr>
<td>MedPAC</td>
<td>Medicare Payment Advisory Committee</td>
</tr>
<tr>
<td>MIDS</td>
<td>Measure and Instrument Development and Support</td>
</tr>
<tr>
<td>MIF</td>
<td>Measure Information Form</td>
</tr>
<tr>
<td>MIPS</td>
<td>Merit-based Incentive Payment System</td>
</tr>
<tr>
<td>MJF</td>
<td>Measure Justification Form</td>
</tr>
<tr>
<td>MMS</td>
<td>Measures Management System</td>
</tr>
<tr>
<td>MSS</td>
<td>Medicare Shared Savings</td>
</tr>
<tr>
<td>MUC</td>
<td>Measures Under Consideration</td>
</tr>
<tr>
<td>MUD</td>
<td>Measures Under Development</td>
</tr>
<tr>
<td>NCHS</td>
<td>National Center for Health Statistics</td>
</tr>
<tr>
<td>NCQA</td>
<td>National Committee for Quality Assurance</td>
</tr>
<tr>
<td>NDC</td>
<td>National Drug Code</td>
</tr>
<tr>
<td>NIH</td>
<td>National Institutes of Health</td>
</tr>
<tr>
<td>NLM</td>
<td>National Library of Medicine</td>
</tr>
<tr>
<td>NLP</td>
<td>Natural Language Processing</td>
</tr>
<tr>
<td>NQF</td>
<td>National Quality Forum</td>
</tr>
<tr>
<td>NUMER</td>
<td>Numerator</td>
</tr>
<tr>
<td>NUMEX</td>
<td>Numerator exclusion</td>
</tr>
<tr>
<td>O/E</td>
<td>Observed-to-expected</td>
</tr>
<tr>
<td>OASIS</td>
<td>Outcome and Assessment Information Set</td>
</tr>
<tr>
<td>OECD</td>
<td>Organization for Economic Co-operation and Development</td>
</tr>
<tr>
<td>OID</td>
<td>Object Identifier</td>
</tr>
<tr>
<td>OMB</td>
<td>Office of Management and Budget</td>
</tr>
</tbody>
</table>
ONC  Office of the National Coordinator for Health Information Technology
OQR  Outpatient Quality Reporting
P.L.  Public Law
PACE  Program of All-Inclusive Care for the Elderly
PACU  Post-anesthesia Care Unit
PCORI  Patient-Centered Outcomes Research Institute
PCPI  Physician Consortium for Performance Improvement
PCS  Procedure Coding System
PFAC  Patient Family Advisory Council
PHDSC  Public Health Data Standards Consortium
PHIN  Public Health Information Network
PIN  Precise Ingredient
PRA  Paperwork Reduction Act
PRO  Patient-Reported Outcome
PROM  Patient-Reported Outcome Measurement
PROMIS  Patient-Reported Outcomes Measurement Information System
PRO-PM  Patient-Reported Outcome-based Performance Measure
PSI  Patient Safety Indicator
QCDR  Qualified Clinical Data Registry
QDC  Quality Data Code
QDM  Quality Data Model
QIN-QIO  Quality Innovation Network-Quality Improvement Organizations
QIP  Quality Incentive Program
QMTF  Quality Measures Technical Forum
QPP  Quality Payment Program
QPS  Quality Positioning System
QR  Quality Reporting
QRDA  Quality Reporting Document Architecture
QRS  Quality Rating System
RAI  Resident Assessment Instrument
RCT  Randomized Controlled Trial
REST  Representational State Transfer
ROC  Receiver-operating Characteristic
ROI  Return on Investment
RRP  Readmissions Reduction Program
SCD  Semantic Clinical Drug
SCIP  Surgical Care Improvement Project
SDS  Sociodemographic Status
SES  Socioeconomic Status
SIR  Standardized Infection Ratio
SME  Subject Matter Expert
SNF  Skilled Nursing Facility
SNF QRP  Skilled Nursing Facility Quality Reporting Program
SNF VBP  Skilled Nursing Facility Value-Based Purchasing Program
SNOMED CT  Systematized Nomenclature of Medicine—Clinical Terms
SR  Systematic Review
SSA  Social Security Act
STS  Society of Thoracic Surgeons
STU  Standard for Trial Use
TEP  Technical Expert Panel
THA  Total Hip Arthroplasty
TKA  Total Knee Arthroplasty
TRN  Technical Release Note
TTY  RxNorm ingredient type
U.S.  United States
UCUM  Unified Code for Units of Measure
UMLS  Unified Medical Language System
URL  Uniform Resource Locator
USCRS  U.S. SNOMED CT Content Request System
USHIK  United States Healthcare Information Knowledgebase
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>USOW</td>
<td>Umbrella Statement of Work</td>
</tr>
<tr>
<td>USPSTF</td>
<td>United States Preventive Services Task Force</td>
</tr>
<tr>
<td>VADS</td>
<td>Vocabulary Access and Distribution System</td>
</tr>
<tr>
<td>VHA</td>
<td>Veterans Health Administration</td>
</tr>
<tr>
<td>VS</td>
<td>Value Set</td>
</tr>
<tr>
<td>VSAC</td>
<td>Value Set Authority Center</td>
</tr>
<tr>
<td>VSD</td>
<td>Value Set Definition</td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language</td>
</tr>
</tbody>
</table>
Section 6. Appendices
## APPENDIX A: SUMMARY OF CHANGES TO BLUEPRINT

This appendix presents a high-level summary of the changes found in this version of the CMS MMS Blueprint. These changes are arranged by Section and Chapter.

<table>
<thead>
<tr>
<th>Section/Chapter</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>All instances of JIRA changed to Jira.</td>
</tr>
<tr>
<td>Preface</td>
<td>Added statement that Blueprint does not address measures at the portfolio level.</td>
</tr>
<tr>
<td>Section 2 Chapter 7</td>
<td>Added chapter on Tools and Resources in Measure Development.</td>
</tr>
<tr>
<td>Section 3, Chapter 1</td>
<td>Updated to align with current policy.</td>
</tr>
<tr>
<td>Section 3, Chapter 2</td>
<td>Updated to align with current policy.</td>
</tr>
<tr>
<td>Section 3 Chapter 2.2.2</td>
<td>Revised language and added a table with CMS quality programs and the legislation initiating each program.</td>
</tr>
<tr>
<td>Section 3, Chapter 3</td>
<td>Revised chapter including changing name of chapter, added language about the Quality Measure Technical Forum and eCQM Governance Group.</td>
</tr>
<tr>
<td>Section 3, Chapter 5.3</td>
<td>Added graphic depicting relationship between PRO, PROM, and PRO-PM.</td>
</tr>
<tr>
<td>Section 3, Chapter 6</td>
<td>Added language about the CMD Workspace, eCQM Data Element Repository, and eCQM Annual Update.</td>
</tr>
<tr>
<td>Section 3, Chapter 9.9</td>
<td>Added section about tools and resources used in information gathering.</td>
</tr>
<tr>
<td>Section 3, Chapter 16</td>
<td>Numerous updates to the chapter including updating measure examples with measures in CMS programs and providing the NQF number, if applicable.</td>
</tr>
<tr>
<td>Section 3, Chapter 17</td>
<td>Several revisions to chapter to include use of HSLOC terminology for location.</td>
</tr>
<tr>
<td>Section 3, Chapter 19</td>
<td>Extensive revisions to the chapter.</td>
</tr>
<tr>
<td>Section 3, Chapter 24.3.4</td>
<td>Added examples of references for composite methodology.</td>
</tr>
<tr>
<td>Section 3, Chapter 23.4.2.1.2</td>
<td>Added language about data element reliability.</td>
</tr>
<tr>
<td>Section 4</td>
<td>Revised most templates.</td>
</tr>
<tr>
<td>Section 5, Glossary</td>
<td>Added several definitions.</td>
</tr>
<tr>
<td>Section 5, Acronyms</td>
<td>Added several acronyms and deleted a few acronyms.</td>
</tr>
<tr>
<td>Section 6, Appendices C, D, and F</td>
<td>Updated for currency.</td>
</tr>
<tr>
<td>Section 6, Appendix E</td>
<td>Added new Appendix E on Registries.</td>
</tr>
<tr>
<td>Section 6, Appendix G</td>
<td>Updated references.</td>
</tr>
</tbody>
</table>
APPENDIX B: ONC PROJECT TRACKING SYSTEM (JIRA)

Introduction

The ONC Project Tracking System (Jira) is an Atlassian, Inc.-based collaboration platform hosted by the HHS’s ONC used for pre-rulemaking and information sharing for eCQM-related and other projects. CMS uses Jira during most phases of the measure lifecycle, including development, implementation, and maintenance. Jira supports projects such as CMS’s annual MUC process and ONC’s Health IT Certification Program. Jira can produce standard and customized reports to support each project.

Current content for most projects is public facing. Users must create an account to enter new tickets/issues or track existing tickets. Some Jira projects require an account to view tickets and others can be further restricted by the project administrators.

eCQM Specification, Testing, Implementation, and Maintenance Phases

Jira projects use a ticketing process, where content is generated through an issue ticket. Most content is derived from questions submitted through tickets, comments on the issue, and a posted solution to the issue. Measure developers can make use of several Jira projects to support each phase of the measure development lifecycle.

Development Phase – Posting Measures for Public Comment

Measure developers may post eCQM specifications of new and revised eCQMs for public comment in the eCQMs under Development tracker. Measure developers provide information about the measure, including the measure description, numerators, denominators, exceptions and exclusions, the downloadable measure specification packages, and often supplemental documents to provide context and explanation of the measure’s intent. The public may review and provide comments on the measure by posting comments in the measure ticket. Public notice of the posting of new and updated measure specifications in Jira are made via the eCQI Resource Center and the CMS website. Some CMS contracts require development and posting of a document that summarized feedback received and changes to measure specifications based on this feedback.

Implementation Phase – Pre-rulemaking/Measures Under Consideration Process

In the Implementation phase of the measure lifecycle, CMS uses Jira to gather, review, sort, select, and prepare candidate measures for public consideration.

CMS is required to publish an annual list of healthcare quality and efficiency MUCs for public and stakeholder comment, pursuant to Sections 1890A and 1890(b)(7)(B) of the SSA. These sections define the CMS programs that take part in a pre-rulemaking process before measures are adopted for use by CMS programs. CMS makes a Jira project available to authorized users every winter-spring so that measure developers can enter specifications and supporting information concerning candidate measures. Candidate measures may be new measures, existing measures put forward for a new/different CMS program, or existing measures that have undergone substantial changes. CMS also presents a series of informational webinars every spring to orient new users to Jira, to update current users on any changes in processes, and outline CMS measurement priorities. Using defined Jira fields, measure developers enter information such as measure title, description, numerator, denominator, exclusions, testing status, rationale, measure history, and various other specifications. This information
helps CMS, HHS, the MAP multi-stakeholder workgroups, and others decide on appropriate recommendations and actions for measures as they are advanced toward adoption in the field.

Gaining access to the Jira project for the MUC first requires a Jira account, and then emailing a request to MMSSupport@battelle.org with your name, affiliation, and the relevant CMS program. Your request will be verified with the appropriate CMS program staff. Once approved, you will be added to the access list and notified of your access to the Jira MUC project.

For more information and a MUC User Guide, visit the CMS Pre-Rulemaking website.

**Maintenance Phase – Obtaining Feedback from the Field**

Once measures are implemented into CMS quality reporting programs, those responsible for implementing the measures into the different settings or products may have questions related to eCQM specifications. The eCQM Issue Tracker (CQM) Jira project is used for the public to obtain further guidance and clarity on measure specifications. Questions received from the field are triaged to the measure developer or steward for response. Question/answer tickets provide a searchable database for end users.

Measure developers and the public may have questions relating to the standards and tools used in measure development. There are separate Jira projects for QDM, QRDA, CQL, Bonnie, and Cypress. Each of these projects are monitored by SMEs who respond to end user questions.

During the eCQM Annual Update, Jira is used internally to pass eCQMs through the different stages of the review process. This internal review process is restricted from public view. CMS-contracted measure developers should contact their COR regarding participation in eCQM Annual Update and obtain access as needed.
# APPENDIX C: eCQM Metadata

<table>
<thead>
<tr>
<th>Header Data Elements</th>
<th>Definition</th>
<th>Measure Developer Guidance</th>
<th>Preferred Term (Required, None, Not Applicable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>eCQM Title</td>
<td>The title of the quality eCQM.</td>
<td>-</td>
<td>Required</td>
</tr>
<tr>
<td>eCQM Abbreviated Title</td>
<td>The user-entered abbreviated measure title chosen when the measure is created.</td>
<td>The eCQM Abbreviated Title is used when the MAT creates the HQMF and Human-Readable eCQM artifacts.</td>
<td>Required</td>
</tr>
<tr>
<td>eCQM Identifier Measure Authoring Tool</td>
<td>The eCQM identifier generated by the MAT.</td>
<td>Identifiers are generated by selecting the “Generate Identifier” button.</td>
<td>Required</td>
</tr>
<tr>
<td>eCQM Version Number</td>
<td>A value used to indicate the version of the eCQM.</td>
<td>Displays the value provided by the MAT based on a user’s entry. The value has three components: the major version, minor version, and revision number. The revision number provides the number of times the measure version has been packaged. The format follows: major.minor.revision number (e.g., 5.1.001).</td>
<td>Required</td>
</tr>
<tr>
<td>NQF Number</td>
<td>Specifies the NQF number.</td>
<td>“Optional” field in MAT. eCQMs endorsed by NQF should enter this as a 4-digit number (including leading zeros). Only include an NQF number if the eCQM is endorsed.</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>Globally Unique Identifier (GUID)</td>
<td>Represents the globally unique measure identifier for a particular eCQM.</td>
<td>Field is auto-populated by the MAT.</td>
<td>Required</td>
</tr>
<tr>
<td>Finalized Date</td>
<td>This field is left blank until the measure is versioned. When versioned, the date and time of versioning are inserted.</td>
<td>Field is auto-populated by the MAT.</td>
<td>Required</td>
</tr>
<tr>
<td>Measurement Period</td>
<td>The time period for which the eCQM applies.</td>
<td>MM/DD/20xx—MM/DD/20xx.</td>
<td>Required</td>
</tr>
<tr>
<td>Measure Steward</td>
<td>The organization responsible for the continued maintenance of the eCQM.</td>
<td>CMS is the measure steward for measures developed under CMS contracts.</td>
<td>Required</td>
</tr>
<tr>
<td>Measure Developer</td>
<td>The organization that developed the eCQM.</td>
<td>-</td>
<td>Required</td>
</tr>
<tr>
<td>Endorsed By</td>
<td>The organization that has endorsed the eCQM through a consensus-based process.</td>
<td>Users are able to indicate whether a measure is endorsed by NQF by selecting “Yes” or “No” via a dropdown menu.</td>
<td>None</td>
</tr>
<tr>
<td>Description</td>
<td>A general description of the eCQM intent.</td>
<td>A brief narrative description of the eCQM, such as “Ischemic stroke patients with atrial fibrillation/flutter who are prescribed anticoagulation therapy at hospital discharge.”</td>
<td>Required</td>
</tr>
<tr>
<td>Header Data Elements</td>
<td>Definition</td>
<td>Measure Developer Guidance</td>
<td>Preferred Term (Required, None, Not Applicable)</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------</td>
<td>----------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>Copyright</td>
<td>Identifies the organization(s) who own the intellectual property represented by the eCQM.</td>
<td>The owner of the eCQM has the exclusive right to print, distribute, and copy the work. Permission must be obtained by anyone else to reuse the work in these ways. May also include copyright permissions (e.g., “©2010 AMA. All Rights Reserved”).</td>
<td>None</td>
</tr>
<tr>
<td>Disclaimer</td>
<td>Disclaimer information for the eCQM.</td>
<td>This should be brief.</td>
<td>None</td>
</tr>
<tr>
<td>Composite Measure Scoring Method</td>
<td>Indicates how a composite measure is scored.</td>
<td>Select one of three options: • All or Nothing • Opportunity • Patient-level Linear</td>
<td>Required for all CQL-based composite measures</td>
</tr>
<tr>
<td>Measure Scoring</td>
<td>Indicates how the calculation is performed for the eCQM (e.g., cohort, proportion, CV, ratio).</td>
<td>-</td>
<td>Required</td>
</tr>
<tr>
<td>Patient-based Measure</td>
<td>Indicates whether a measure is patient-based or non-patient-based.</td>
<td>The field will be auto-populated upon measure creation for CV measures only.</td>
<td>Required</td>
</tr>
<tr>
<td>Measure Type</td>
<td>Indicates whether the eCQM is used to examine a process or an outcome over time.</td>
<td>This is a checkbox list where a user can select one or more measure types. Values are: • Appropriate use process • Composite* • Cost/Resource use • Efficiency • Patient Engagement/Experience • Outcome • Intermediate clinical outcome • PRO performance • Process • Structure *The ‘Composite’ measure type applies only to composite measures, so this value is not selectable from the measure type list in MAT. Instead, the MAT sets it for the user based on the user creating a CQL-based composite measure.</td>
<td>Required</td>
</tr>
<tr>
<td>Component Measure List</td>
<td>Measures constructed in the MAT may be composite measures. For composite measures, MAT users may add two or more CQL-based component measures. Component measures serve as a reference point for understanding what comprises a composite measure.</td>
<td>Any CQL-based measure may be used as a component measure. If the measure is not a composite measure, this field does not apply and is unavailable. Requirements for Composite and Component eCQMs are spelled out in the most recent version of the CQL-based HQMF IG.</td>
<td>Required for composite measures</td>
</tr>
<tr>
<td>Header Data Elements</td>
<td>Definition</td>
<td>Measure Developer Guidance</td>
<td>Preferred Term (Required, None, Not Applicable)</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td><strong>Stratification</strong></td>
<td>Describes the strata for which the measure is to be evaluated. There are three examples of reasons for stratification based on existing work. These include: (1) Evaluate the measure based on different age groupings within the population described in the measure (e.g., evaluate the whole age group between 14 and 25, and each sub-stratum between 14 and 19, and between 20 and 25). (2) Evaluate the eCQM based on either a specific condition, a specific discharge location, or both (e.g., report ED waiting time results for all patients and for each of two sub-strata: those with a primary mental health diagnosis, and those with a primary diagnosis of sexually transmitted infection) (3) Evaluate the eCQM based on different locations within a facility (e.g., evaluate the overall rate for all intensive care units [ICUs]; some strata may include additional findings such as specific birth weights for neonatal ICUs).</td>
<td>This is a free text field.</td>
<td>None</td>
</tr>
<tr>
<td><strong>Risk Adjustment</strong></td>
<td>The method of adjusting for clinical severity and conditions present at the start of care that can influence patient outcomes, thus impacting valid comparisons of outcome measures across providers. Risk adjustment indicates whether an eCQM is subject to a statistical process for reducing, removing, or clarifying the influences of confounding factors to allow more useful comparisons.</td>
<td>Provide a brief description with instructions where the complete risk adjustment methodology may be obtained. This is a free text field.</td>
<td>None</td>
</tr>
<tr>
<td>Header Data Elements</td>
<td>Definition</td>
<td>Measure Developer Guidance</td>
<td>Preferred Term (Required, None, Not Applicable)</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Rate Aggregation</td>
<td>Rate aggregation describes how to combine information calculated based on logic in each of several populations into one summarized result. It can also be used to describe how to risk-adjust the data based on supplemental data elements described in the eCQM (e.g., a hospital measure for treatment of community-acquired pneumonia may require different antibiotics to be used for patients admitted to the ICU compared with those admitted to non-ICU settings. Rate aggregation provides the method to combine, or aggregate, the two results into one reported rate).</td>
<td>For eCQMs that do not have rate aggregation, enter ‘None.’ This is a free text field.</td>
<td>None</td>
</tr>
<tr>
<td>Rationale</td>
<td>Succinct statement of the need for the measure. Usually includes statements pertaining to <strong>Importance criterion</strong>: impact, gap in care, and evidence.</td>
<td>This is a free text field.</td>
<td>Required</td>
</tr>
<tr>
<td>Clinical Recommendation Statement</td>
<td>Summary of relevant clinical guidelines or other clinical recommendations supporting this eCQM.</td>
<td>This is a free text field.</td>
<td>Required</td>
</tr>
<tr>
<td>Improvement Notation</td>
<td>Information on whether an increase or decrease in score is the preferred result (e.g., a higher score indicates better quality or a lower score indicates better quality, or quality is within a range).</td>
<td>This is a free text field.</td>
<td>None</td>
</tr>
<tr>
<td>Reference(s)</td>
<td>Identifies bibliographic citations or references to clinical practice guidelines, sources of evidence, or other relevant materials supporting the intent and rationale of the eCQM. Users are able to add and remove references by selecting the “+Add Reference” link on the right side of the Reference(s) field.</td>
<td>This is a free text field.</td>
<td>None</td>
</tr>
<tr>
<td>Definition</td>
<td>Description of individual terms, provided as needed.</td>
<td>This is a free text field.</td>
<td>None</td>
</tr>
<tr>
<td>Guidance</td>
<td>Allows measure developers to provide additional guidance for implementers to understand greater specificity than could be provided in the logic for data criteria.</td>
<td>This is a free text field.</td>
<td>None</td>
</tr>
<tr>
<td>Header Data Elements</td>
<td>Definition</td>
<td>Measure Developer Guidance</td>
<td>Preferred Term (Required, None, Not Applicable)</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Transmission Format</td>
<td>URL or hyperlinks for the transmission formats specified for a particular reporting program.</td>
<td>Enter URLs that provide the transmission formats that are specified for a particular reporting program. For measures that do not have Transmission Format information enter ‘None’. This is a free text field.</td>
<td>None</td>
</tr>
<tr>
<td>Initial Population</td>
<td>The Initial Population refers to all events (e.g., patients, episodes) to be evaluated by a specific performance eCQM who share a common set of specified characteristics within a specific measurement set to which a given measure belongs. Details often include information based on specific age groups, diagnoses, diagnostic and procedure codes, and enrollment periods. Some ratio measures will require multiple Initial Populations, one for the Numerator, and one for the Denominator.</td>
<td>Must be consistent with the computer-generated narrative logic in the body of the eCQM. The computer-generated narrative is standardized and concise and can lack the richness of full text that sometimes helps in the understanding of an eCQM. This is especially true for eCQMs that have complex criteria, where the computer-generated text may not be able to express the exact description that a measure developer would like to convey. As part of the quality assurance step, it is important to compare the human-readable description of the measure population (in the header) to the logic representation (in the body) for any discrepancies. This field will be the primary field to fully define the comprehensive eligible population for proportion/ratio eCQMs or the eligible measure population for CV eCQMs. This is a free text field.</td>
<td>Required</td>
</tr>
<tr>
<td>Denominator</td>
<td>This can be the same as the Initial Population or a subset of the Initial Population to further constrain the population for the purpose of the eCQM. Different measures within an eCQM set may have different Denominators. CV eCQMs do not have a Denominator, but instead define a Measure Population.</td>
<td>For proportion/ratio measures, include the text “Equals Initial Population” where applicable. This is a free text field.</td>
<td>Not Applicable (for CV eCQMs)</td>
</tr>
<tr>
<td>Denominator Exclusions</td>
<td>Cases (e.g., patients, episodes) that should be removed from the eCQM Initial Population and Denominator before determining if Numerator criteria are met. Denominator Exclusion is used in proportion and ratio measures to help narrow the Denominator. For example: Patients with bilateral lower extremity amputations would be listed as a Denominator Exclusion for a measure requiring foot exams.</td>
<td>This is a free text field.</td>
<td>None (for proportion or ratio eCQMs) Not Applicable (for CV eCQMs)</td>
</tr>
<tr>
<td>Header Data Elements</td>
<td>Definition</td>
<td>Measure Developer Guidance</td>
<td>Preferred Term (Required, None, Not Applicable)</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------</td>
<td>----------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td><strong>Numerator</strong></td>
<td>Numerators are used in <em>proportion</em> and <em>ratio</em> eCQMs. In proportion measures, the numerator criteria are the processes or outcomes evaluated for each patient, procedure, or other unit of measurement defined in the Denominator. In ratio measures, the Numerator is related to, but not directly derived from the Denominator. For example: A ratio measure numerator listing the number of central line blood stream infections and a denominator indicating the days per thousand of central line usage in a specific time period.</td>
<td>This is a free text field.</td>
<td>Not Applicable (for CV eCQMs)</td>
</tr>
<tr>
<td><strong>Numerator Exclusions</strong></td>
<td>Numerator Exclusions are used only in ratio and proportion eCQMs to define instances that should not be included in the numerator data. For example, in a ratio: If the number of central line blood stream infections per 1,000 catheter days were to exclude infections with a specific bacterium, that bacterium would be listed as a numerator exclusion.</td>
<td>Numerator Exclusions are generally used in proportion measures when the improvement notation is a “lower score indicates better quality.” In proportion measures, numerator exclusion removes instances from the numerator population while retaining them in the denominator. This is a free text field.</td>
<td>None (for ratio and proportion eCQMs) Not Applicable (for CV and cohort eCQMs)</td>
</tr>
<tr>
<td>Header Data Elements</td>
<td>Definition</td>
<td>Measure Developer Guidance</td>
<td>Preferred Term (Required, None, Not Applicable)</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------</td>
<td>----------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Denominator Exceptions</td>
<td>Denominator Exceptions are those conditions that should remove a patient, procedure, or unit of measurement from the denominator of the performance rate only if the Numerator criteria are not met. Denominator Exceptions allow for adjustment of the calculated score for those providers with higher risk populations on individual patient characteristics or preferences. Denominator Exceptions are used only in proportion eCQMs. They are not appropriate for ratio or CV eCQMs. Denominator Exceptions allow for the exercise of clinical judgment and should be specifically defined where capturing the information in a structured manner fits the clinical workflow. Generic Denominator Exception reasons used in proportion eCQMs fall into three general categories: medical reasons, patient reasons, and system reasons.</td>
<td>Be specific for all categories of denominator exception reasons. This is a free text field.</td>
<td>None (for proportion eCQMs) Not Applicable (for ratio or CV eCQMs)</td>
</tr>
<tr>
<td>Measure Population</td>
<td>Measure Population is used only in CV eCQMs. It is a narrative description of the eCQM population. For example, all patients seen in the ED during the measurement period.</td>
<td>For CV eCQMs, include the text “Equals All in Initial Population.” Then add any specific additional criteria if needed. This is a free text field.</td>
<td>Not Applicable (for ratio or proportion eCQMs)</td>
</tr>
<tr>
<td>Measure Population Exclusions</td>
<td>Measure Population Exclusions are those characteristics of patients who meet measure population criteria that should cause them to be removed from the measure calculation. For example, for all patients seen in the ED, exclude those transferred directly to another acute care facility for tertiary treatment.</td>
<td>Measure population exclusions are used only in CV eCQMs. It is a narrative description of the eCQM population to exclude. This is a free text field.</td>
<td>None (for CV eCQMs) Not Applicable (for ratio or proportion eCQMs)</td>
</tr>
<tr>
<td>Measure Observations</td>
<td>Measure Observations is used only in ratio and CV eCQMs. They provide the description of how to evaluate performance. For example, the mean time from arrival to departure for all ED visits during the measurement period.</td>
<td>Measure observations are generally described using a statistical methodology such as: count, etc. This is a free text field.</td>
<td>Not Applicable (for proportion eCQMs)</td>
</tr>
<tr>
<td>Header Data Elements</td>
<td>Definition</td>
<td>Measure Developer Guidance</td>
<td>Preferred Term (Required, None, Not Applicable)</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------</td>
<td>----------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Supplemental Data Elements</td>
<td>CMS defines four required Supplemental Data Elements (i.e., payer, ethnicity, race, and Sex), which are variables used to aggregate data into various subgroups. Comparison of results across strata can be used to show where disparities exist or where there is a need to expose differences in results. Additional Supplemental Data Elements required for risk adjustment or other purposes of data aggregation can be included in the Supplemental Data Element section.</td>
<td>Due to the four CMS-required fields, the measure developer must always populate with payer, ethnicity, race, and sex. For measures used in CMS programs, use this language in the Supplemental Data section: “For every patient evaluated by this measure also identify payer, race, ethnicity, and sex.” Other information may be added for other measures. This is a free text field.</td>
<td>Required</td>
</tr>
</tbody>
</table>
| Measure Set          | A measure set is a unique grouping of measures, that when viewed together, provide a robust picture of the care within a given domain (e.g., cardiovascular care, pregnancy). | For measures that do not have a measure set, enter “none” or “not applicable”. | }
**APPENDIX D: eCQM LOGIC QUALITY ASSURANCE CHECKLIST**

This checklist can be used to review the logic used in eCQMs. Sometimes reviewer’s comments are also provided in a Word document and referenced in the Checklist Comment section.

<table>
<thead>
<tr>
<th>#</th>
<th>Mandatory?</th>
<th>Reviewed?</th>
<th>Passed?</th>
<th>Item</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-1</td>
<td>Y</td>
<td></td>
<td></td>
<td>Is the intent of the measure described in the measure description articulated/captured in the measure logic?</td>
<td></td>
</tr>
<tr>
<td>L-2</td>
<td>Y</td>
<td></td>
<td></td>
<td>Do the logic elements map to definitions in the measure narrative, data dictionary, or supporting reference documentation?</td>
<td></td>
</tr>
<tr>
<td>L-3</td>
<td>Y</td>
<td></td>
<td></td>
<td>Do the populations in the narrative align with the populations defined in the logic?</td>
<td></td>
</tr>
<tr>
<td>L-4</td>
<td>Y</td>
<td></td>
<td></td>
<td>Does the measure adhere to the CQL Style Guide?</td>
<td></td>
</tr>
<tr>
<td>L-5</td>
<td>Y</td>
<td></td>
<td></td>
<td>Has the logic been represented using the most concise language and logic operators without changing the original intent of the measure? (This can be accomplished by creating definitions for reused logic.) Is a shared CQL library used if available rather than duplicate logic defined in the shared library?</td>
<td></td>
</tr>
<tr>
<td>L-6</td>
<td>Y</td>
<td></td>
<td></td>
<td>Are queries expressed correctly? (For example, filter criteria are being executed at the correct level.) As defined in the CQL Author’s Guide Section 3 Queries.</td>
<td></td>
</tr>
<tr>
<td>L-7</td>
<td>Y</td>
<td></td>
<td></td>
<td>Are all QDM elements time-bound (either directly or indirectly)? This includes properly “sorting” queries and lists. (As defined in the CQL Author’s Guide Section 3.3 Sorting.)</td>
<td></td>
</tr>
<tr>
<td>L-8</td>
<td>Y</td>
<td></td>
<td></td>
<td>Do mathematic inequalities reflect the measure intent and represent the intended populations? (For example, when intended the inequality represents less than rather than less than and/or equal to.) Does the logic properly utilize Precision-Based Timing constructs? (As defined in the CQL Author’s Guide Section 5 Precision-Based Timing.)</td>
<td></td>
</tr>
<tr>
<td>L-9</td>
<td>Y</td>
<td></td>
<td></td>
<td>Are operator precedence rules followed as specified as defined in the CQL Developer’s Guide Section 1.7 Operator Precedence?</td>
<td></td>
</tr>
<tr>
<td>L-10</td>
<td>Y</td>
<td></td>
<td></td>
<td>Have “Specific Occurrences” been properly defined following the guidance in the CQL Formatting and Usage Wiki? Are episode-of-care measures returning counts?</td>
<td></td>
</tr>
<tr>
<td>L-11</td>
<td>Y</td>
<td></td>
<td></td>
<td>Are time intervals represented in similar units (e.g., hours)? Are all timing comparisons using the correct precision? Are interval beginning and ending appropriately marked as exclusive (i.e., open, indicated by a parentheses) or</td>
<td></td>
</tr>
<tr>
<td>#</td>
<td>Mandatory?</td>
<td>Reviewed?</td>
<td>Passed?</td>
<td>Item</td>
<td>Comment</td>
</tr>
<tr>
<td>----</td>
<td>------------</td>
<td>-----------</td>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>L-12</td>
<td>Y</td>
<td></td>
<td></td>
<td>Are annotations included in the logic for sections of the measure that have been updated/changed?</td>
<td></td>
</tr>
<tr>
<td>L-13</td>
<td>N</td>
<td></td>
<td></td>
<td>Does the measure demonstrate at least 100% coverage of test patients with at least one positive and one negative patient for each population?</td>
<td></td>
</tr>
<tr>
<td>L-14</td>
<td>N</td>
<td></td>
<td></td>
<td>Has the measure been tested with Bonnie? Has the measure author provided the Bonnie account (email address) where the measure was tested?</td>
<td></td>
</tr>
<tr>
<td>L-15</td>
<td>N</td>
<td></td>
<td></td>
<td>Additional comments / issues / suggestions?</td>
<td></td>
</tr>
</tbody>
</table>
Appendix E: Registries

1 Qualified Clinical Data Registry

A QCDR is a CMS-approved vendor that is in the business of improving healthcare quality. These organizations may include specialty societies, regional health collaboratives, large health systems or software vendors working in collaboration with one of these medical entities. One of the ways QCDRs can help to improve the quality of care patients receive is by collecting clinical data from clinicians and reporting this data to CMS on their behalf for purposes of MIPS. QCDR submission differs from QR submission in that QCDRs can submit QCDR measures, as well as MIPS quality measures. They may also submit data for the Promoting Interoperability and Improvement Activity performance categories on behalf of clinician(s). QCDRs must submit data through secure submission methods – in either the QRDA Category III or an approved QPP JSON or XML format.

QCDRs can develop and/or submit QCDR measures to CMS for CMS review and approval. A QCDR may submit no more than 30 measures of which a portion may be approved, provisionally approved, or rejected for a given program year.

- Provisionally approved measures may require revisions, harmonization, or performance data submission for the next year’s approval. CMS provides a rationale for the provisional status and may require performance data on QCDR measures approved in a previous year to show a continued performance gap.
- Rejected measures will not be reconsidered unless the measure has been revised or there is evidence of a performance gap.

QCDRs may submit QCDR measures that are:

- Not contained in the annual list of MIPS quality measures
- Have substantive differences in the population covered by an existing MIPS measure
- Have a different manner of submission of an existing MIPS measure
- Developed by the QCDR, specialty societies, or regional quality collaboratives
- An NQF-endorsed measure that is not part of MIPS.

QCDRs are also permitted to customize the Clinician and Group CAHPS survey and should consider including non-Medicare beneficiaries.

QCDR measures should be developed in accordance with the Blueprint, but do not need to go through the pre-rulemaking process. QCDR measures are submitted as part of the self-nomination process and must be reviewed and approved annually regardless of their previous approval status. CMS is willing to meet with QCDRs to review measure concepts prior to the start of self-nomination period. Figure 52 depicts the development, review, and posting process.

65 Beginning with the MIPS 2020 performance period a QCDR will be defined as an entity with clinical expertise in medicine and in quality measurement development that collects medical or clinical data on behalf of a MIPS eligible clinician for the purpose of patient and disease tracking to foster improvement in the quality of care provided to patients.
Annually, CMS opens up the self-nomination period beginning on July 1 and close it by September 1 in the year prior to the performance period. A Self-Nomination Toolkit for QCDRs and Qualified Registries is published in the QPP Resource Library prior to the start of the self-nomination period. The Toolkit contains the following items:

- MIPS Self-Nomination User Guide for QCDRs and Qualified Registries
- MIPS QCDR Self-Nomination Fact Sheet
- MIPS QR Self-Nomination Fact Sheet
- QCDR Measure Development Handbook
- Self-Nomination QCDR Measures Submission Template.

QCDRs must self-nominate annually, regardless of whether they were previously approved. Beginning with the 2019 self-nomination, CMS offers a simplified self-nomination process for existing QCDRs in good standing. The self-nomination application will be available through the QPP portal for the 2020 performance period. Applicants are required to provide all information at the time of self-nomination.

There are several benefits of using a QCDR. A few examples are:

- QCDRs streamline data collection and manage the submission of three MIPS categories to CMS.
- QCDRs are required to provide quarterly feedback reports to their participating clinicians, which provides the clinicians the opportunity to make more rapid changes to improve quality of care.
- QCDR measures are clinically relevant measures that address gaps in care for specialties, preventive care, and/or disease management.
- QCDR measures are measures that are not contained in the annual list of MIPS CQMs for the applicable performance period of MIPS.
- Publicly reporting QCDR data on Physician Compare expands the quality measure data available for eligible clinicians and group practices regardless of specialty and provides more quality data to consumers to help them make informed decisions.
- Since many QCDRs are specialty-based, QCDR measures may be more meaningful and applicable to the eligible clinician.
- Reporting through a QCDR reduces burden for MIPS reporting.
2  QUALIFIED REGISTRY

A Qualified Registry is a CMS-approved vendor that collects clinical data from an individual MIPS eligible clinician, group, or virtual group and submits it to CMS on their behalf. The CY2017 Quality Payment Program final rule codified the definition a QR at 42 C.F.R. §414.1305 to be “a medical registry, a maintenance of certification program operated by a specialty body of the American Board of Medical Specialties or other data intermediary that, with respect to a particular performance period, has self-nominated and successfully completed a vetting process (as specified by CMS) to demonstrate its compliance with the MIPS qualification criteria specified by CMS for that performance period” (81 Fed. Reg. 77382). Clinicians work directly with the Qualified Registry of their choosing to submit data on the selected MIPS quality measures or a specialty set of measures. Qualified Registries can help to improve the quality of care patients receive by collecting clinical data from clinicians and reporting this data to CMS for purposes of MIPS. Qualified Registry submissions differ from QCDR submissions in that Qualified Registries may not submit QCDR measures. Qualified Registries are required to submit data on MIPS quality measures and may choose to support data submission for the Promoting Interoperability and Improvement Activity performance categories.
APPENDIX F: MEASURE CALCULATIONS

1  **PROPORTION MEASURES**

A proportion is a measure in which the *numerator* is a subset (or part) of the *denominator* and can be written as \( \frac{a}{a+b} \) (Figure 53). A proportion measure differs from a *ratio* measure (refer to Section 16.3, Develop Specifications and Definitions) because in a ratio measure the numerator is not a subset of the denominator population. Although the numerator and denominator populations may be related or may overlap for ratio measures, these populations do not have a superset/subset relationship like proportion measures.

![Figure 53. Proportion Measure Populations](image)

1.1 **ELEMENTS OF A PROPORTION MEASURE**

From these relationships and definitions, these sequential steps are used to determine whether a patient falls into a given population:

1. **Initial population** (IP): Identify those patients who meet the IP criteria.
2. Denominator (DENOM): Identify the subset of the IP that meet the DENOM criteria.
3. **Denominator exclusion** (DENEX): Identify the subset of the DENOM that meet the DENEX criteria. These are patients who should be removed from the denominator as exclusions. Once these patients are removed, the remaining subset would reflect the group of patients for which the numerator criteria will be evaluated.
4. Numerator (NUMER): Identify those in the DENOM and not in the DENEX that meet the NUMER criteria. In proportion measures, the numerator criteria are the processes or outcomes expected for each patient, procedure, or other unit of measurement defined in the denominator.
5. **Numerator exclusion** (NUMEX): Identify that subset of the NUMER that meets the NUMEX criteria. NUMEX is used to define instances that should not be included in the numerator data.

6. **Denominator exception** (DENEXCEP): Identify those in the DENOM and not in the DENEX and NOT in the NUMER that meet the DENEXCEP criteria. These patients are removed from the IP and/or DENOM.

### 1.2 Performance Rates

Specific programs may require reporting of performance rates. The performance rate is the number of patients meeting NUMER criteria (accounting for exceptions), divided by patients in the DENOM (accounting for exclusion and exception). Performance rate can be calculated using this formula:

\[
\text{Performance rate} = \frac{\text{NUMER} - \text{NUMEX}}{\text{DENOM} - \text{DENEX} - \text{DENEXCEP}}
\]

Exception: 0% Performance Rates

- **Inverse measures**: For inverse measures, a lower rate indicates better performance and a 0% performance rate will be counted as satisfactorily reporting (e.g., 30-day mortality rate for AMI).
- **Null Scores**: If the measure is not applicable for all patients within the sample, the performance rate would be 0/0 (null) and would be considered satisfactorily reporting. Performance exclusion QDCs are not counted in the performance denominator.

### 1.3 Proportion Measure Examples

Examples of the mathematical relationships between populations in proportion measures:

**Example #1: Proportion Measure**

A fictitious proportion measure defines the population criteria:

- **IP**: All patients aged 65 years and older with an active diagnosis of diabetes mellitus
- **DENOM**: Equals IP
- **DENEX**: Bilateral blindness
- **NUMER**: Dilated eye exam for diabetic retinopathy
- **NUMEX**: None
- **DENEXCEP**: Bed confinement status in a community where mobile eye exam imaging is unavailable.

Mr. Jones is 75 years old and has an active diagnosis of diabetes. There is no mention of blindness in his chart. He has a documented dilated eye exam for diabetic retinopathy.

- (IP = YES) Mr. Jones meets the IP criteria.
- (DENOM = YES) Mr. Jones meets the DENOM criteria.
- (DENEX = NO) By the positive evidence principle, Mr. Jones does not meet the DENEX criteria.
- (NUMER = YES) Mr. Jones meets the NUMER criteria.
- (NUMEX = NO).
- (DENEXCEP = NO) DENEXCEP was not evaluated because the patient met the NUMER criteria.

---

66 Also referred to as the raw rate.
Mr. Smith is 75 years old and has an active diagnosis of diabetes. There is no mention of blindness in his chart. There is no mention of a dilated eye exam in his chart. There is no mention in his chart that he is bed bound.

- (IP = YES) Mr. Smith meets the IP criteria.
- (DENOM = YES) Mr. Smith meets the DENOM criteria.
- (DENEX = NO) By the positive evidence principle, Mr. Smith does not meet the DENEX criteria.
- (NUMER = NO) By the positive evidence principle, Mr. Smith does not meet the NUMER criteria.
- (NUMEX = NO).
- (DENEXCEP = NO) By the positive evidence principle, Mr. Smith does not meet the DENEXCEP criteria.

Mr. Johnson is 85 years old and has an active diagnosis of diabetes. There is no mention of blindness in his chart. He has a documented dilated eye exam for diabetic retinopathy. He is known to be confined to bed in a community where mobile eye exam imaging is unavailable.

- (IP = YES) Mr. Johnson meets the IP criteria.
- (DENOM = YES) Mr. Johnson meets the DENOM criteria.
- (DENEX = NO) By the positive evidence principle, Mr. Johnson does not meet the DENEX criteria.
- (NUMER = YES) Mr. Johnson meets the NUMER criteria.
- (NUMEX = NO).
- (DENEXCEP = NO) By definition, Mr. Johnson does not meet the DENEXCEP criteria because DENEXCEP criteria are not applicable to those meeting the NUMER criteria.

**Example #2: Inverse Proportion Measure**

A fictitious inverse proportion measure (i.e., where improvement is a decrease in the rate) defines the population criteria:

- IP: Patients aged 8 to 65 who gave birth.
- DENOM: Equals IP.
- DENEX: Patients with gestational age < 37 weeks.
- NUMER: Patients with medical induction of labor or C-section.
- NUMEX: Patients in active labor or with spontaneous rupture of membranes before induction of labor or C-section.
- DENEXCEP: None.

Mrs. Jones is a 31-year-old woman who gave birth at 37 weeks’ gestation. She has a medical induction of labor and no evidence of active labor or spontaneous rupture of membranes before being induced.

- (IP = YES) Mrs. Jones meets the IP criteria.
- (DENOM = YES) Mrs. Jones meets the DENOM criteria.
- (DENEX = NO) Mrs. Jones does not meet the DENEX criteria.
- (NUMER = YES) Mrs. Jones meets the NUMER criteria.
- (NUMEX = NO) By the positive evidence principle, Mrs. Jones does not meet the NUMEX criteria.
- (DENEXCEP = NO).
Mrs. Thompson is 31 years old and had a C-section at 38 weeks after a spontaneous rupture of membranes.

- (IP = YES) Mrs. Thompson meets the IP criteria.
- (DENOM = YES) Mrs. Thompson meets the DENOM criteria.
- (DENEX = NO) Mrs. Thompson does not meet the DENEX criteria.
- (NUMER = YES) Mrs. Thompson meets the NUMER criteria.
- (NUMEX = YES) Mrs. Thompson meets the NUMEX criteria.
- (DENEXCEP = NO).

Mrs. Hill is 31 years old and gave birth at 36 weeks, being induced after a spontaneous rupture of membranes.

- (IP = YES) Mrs. Hill meets the IP criteria.
- (DENOM = YES) Mrs. Hill meets the DENOM criteria.
- (DENEX = YES) Mrs. Hill meets the DENEX criteria.
- (NUMER = NO) By definition, Mrs. Hill does not meet the NUMER criteria because NUMER criteria are not applicable to those meeting the DENEX criteria.
- (NUMEX = NO) By definition, Mrs. Hill does not meet the NUMEX criteria because NUMEX criteria are only applicable to those meeting the NUMER criteria.
- (DENEXCEP = NO).

**Example #3: Measure Aggregate Calculations**

Aggregate scores are simply the counts of individuals in each population. Thus, the aggregate IP is the count of individuals meeting the IP criteria.

Building on Example #1, counting all individuals within the population, the aggregate counts are determined:

- Initial population: N=150 (e.g., 150 patients meet the IP criteria).
- Denominator: N=150.
- Denominator exclusion: N=20 (meet DENOM and meet DENEX).
- Numerator: N=75 (meet DENOM, not in DENEX, and meet NUMER criteria).
- NUMEX: N=0.
- Denominator exception: N=5 (meet DENOM, not in DENEX, not in NUMER, and meet the DENEXCEP criteria).

Performance rate = \((\text{NUMER} - \text{NUMEX}) / (\text{DENOM} - \text{DENEX} - \text{DENEXCEP})\) = \((75-0) / (150-20-5)\) = 0.6.
2 CONTINUOUS VARIABLE CALCULATIONS

This section provides further guidance on the precise mathematical relationships between populations in a CV measure, and the process to be used to determine individual and aggregate scores.

In a CV measure, each individual value for the measure can fall anywhere along a continuous scale, for example, mean number of minutes between presentation of chest pain to the time of administration of thrombolytics.

2.1 ELEMENTS OF CONTINUOUS VARIABLE MEASURES

There is a fixed mathematical relationship between the populations in CV measure, as shown in Figure 54.

![Figure 54. CV Measure Populations](image)

The CV measure query process is defined as:

- IP: Identify those cases that meet the IP criteria.
- Measure population (MSRPOPL): Identify the subset of the IP that meet the MSRPOPL criteria.
- Measure population exclusion (MSRPOPLEX): Identify the subset of the MSRPOPL that meet the MSRPOPLEX criteria.

For a CV measure, an individual observation is determined for each case in the MSRPOPL and not in the MSRPOPLEX.

 pena. eCQM CQL logic should not explicitly remove the MSRPOPLEX cases because the execution environment for the eCQM is responsible for excluding the MSRPOPLEX cases and calling the Measure Observation function for each of the remaining cases.
Per Requirement 13 of the HL7 Version 3 Implementation Guide: CQL-based HQMF Release 1, STU 3, aggregation rules are mandatory: “methodCode shall be populated to indicate the aggregation method for the measure.”

2.2 **Continuous Variable Aggregate Calculations**

Aggregate scores for CV measures are more complex than for proportion measures in that they are more than just the counts of cases in each population. In addition to the identification of measure population(s), CV measures define observations that are to be made on cases falling into various populations. These individual observations are then aggregated according to aggregation rules specific to each measure, but not in the measure population exclusion.

In CV measures, an individual observation is made for each case falling into the measure population and not in the measure population exclusion. These individual observations are then used to calculate the aggregate CV, which constitutes the measure “score.” In other words, the individual observations made for all cases in the MSRPOPL and not in the MSRPOPLEX are used to calculate the aggregate score.

🔍 It is important to note that observations on the individual can be sent in QRDA Category I, whereas aggregate calculations can be sent in QRDA Category III.

2.3 **CV Measure Example**

A fictitious CV measure defines the population criteria:

- IP: All inpatient encounters ending during the measurement period with length of stay fewer than or equal to 120 days
- MSRPOPL: Inpatient encounters preceded by an ED visit
- MSRPOPLEX: None
- Individual observations: Time in minutes from ED admission to ED discharge for patients admitted to the facility from the ED
- MSRPOPL observation: Number of minutes in the ED

**Aggregation rules:**

🔍 For eCQMs, HQMF methodCode specifies median. Mr. Jones is 75 years old and was admitted to the hospital from the ED. He spent 90 minutes in the ED.

- (IP = YES) Mr. Jones has inpatient encounters that meet the IP criteria.
- (MSRPOPL = YES) Mr. Jones meets the MSRPOPL criteria.
- (MSRPOPLEX = NO).
- Individual observations for Mr. Jones: 90 minutes.
- MSRPOPL observation: 90 minutes.

Mr. James is 75 years old, admitted directly to the hospital from an outside facility.

- (IP = YES) Mr. James has inpatient encounters that meet the IP criteria.
- (MSRPOPL = NO) By the positive evidence principle, Mr. James does not meet the MSRPOPL criteria.
- (MSRPOPLEX = NO).
- Individual observations for Mr. James: None.
MSRPOPL observation: N/A (Mr. James has no inpatient encounters that meet the MSRPOPL criteria.)

Measure aggregate calculations:

Building upon the examples above, but now considering all encounters within the population, the aggregate counts are determined:

- IP: N = 150 (i.e., 150 encounters meet the IP criteria)
- MSRPOPL: N = 120
- MSRPOPLEX: N = 0

CV Score: 96 minutes (median of all individual Measure Population observations of time spent in the ED)
3 **RATIO MEASURE CALCULATIONS**

This section provides further guidance on the precise mathematical relationships between populations in a ratio measure, and the process to be used to determine individual and aggregate scores.

A ratio measure is derived by dividing a count of one type of data by a count of another type of data (e.g., the number of patients with central lines who develop infection divided by the number of central line days). A ratio measure differs from a proportion measure because in a ratio measure the numerator is not a subset of the denominator population. Although the numerator and denominator populations may be related or may overlap, these populations do not have a superset/subset relationship.

3.1 **ELEMENTS OF A RATIO MEASURE**

There is a fixed mathematical relationship between the populations in ratio measure, as shown in Figure 55. Ratio measures are often two CV calculations for related populations (e.g., median ED waiting time for the index hospital, median ED waiting time for the region in which the hospital is located as the numerator).

![Figure 55. Ratio Measure Populations](image)

From these relationships and definitions, the ratio measure query process is defined as:

- **IP**: Identify those cases that meet the IP criteria. (Some ratio measures will require multiple IPs: one for the numerator and one for the denominator.)
- **DENOM**: Identify the subset of the IP that meet the DENOM criteria.
- **DENEX**: Identify the subset of the DENOM that meet the DENEX criteria.
- **NUMER**: Identify the subset of the IP that meet the NUMER criteria.
- **NUMEX**: Identify the subset of the NUMER that meet the NUMEX criteria.
Individual observations for ratio measures are computed both for the denominator and numerator populations:

- For each case in the DENOM and not in the DENEX, determine the individual DENOM observations.
- For each case in the NUMER and not in the NUMEX, determine the individual NUMER observations.

### 3.2 Ratio Measure Aggregate Calculations

As with CV measures, aggregate scores for ratio measures are more than just the counts of cases in each population. In addition to the identification of measure population(s), ratio measures define observations that are to be made on cases falling into various populations. These individual observations are then aggregated according to aggregation rules specific to each measure.

In ratio measures, for each population, individual observations are made for denominator and numerator cases, which are then used to calculate the aggregate ratio:

\[
\text{Ratio} = \frac{\text{aggregate NUMER}}{\text{aggregate DENOM}}
\]

where the aggregate DENOM is calculated using individual observations for all cases in the DENOM and not in the DENEX, and the aggregate NUMER is calculated using individual observations for all cases in the NUMER and not in the NUMEX, calculate the aggregate NUMER.

### 3.3 Ratio Measure Example

A fictitious ratio measure looking at the rate of central line bloodstream infections (i.e., the ratio of number of central line infections per total central line days) defines the population criteria:

- IP: All hospitalized patients with a laboratory confirmed bloodstream infection
- DENOM: Presence of central line or umbilical catheter for >2 calendar days
- DENEX: Patient is immunosuppressed
- NUMER: Central line bloodstream infection
- NUMEX: Infection is deemed to be a contaminant

**Individual Observations:**

- DENOM observation: Number of hospital days with central line
- NUMER observation: The observation in this case is fully defined by the population criteria (i.e., the observation is simply whether the case met the population criteria or not).

**Aggregation Rules:**

- DENOM aggregation: Sum of number of hospital days with central line – DENEX
- NUMER aggregation: Count of NUMER – NUMEX

Mr. Jones is 75 years old and was hospitalized for 7 days. He had a central line in place for 5 days. There is no mention of immunosuppression in his chart. There is no mention of a central line bloodstream infection in his chart.

- (IP = YES) Mr. Jones meets the IP criteria.
- (DENOM = YES) Mr. Jones meets the DENOM criteria.
• (DENEX = NO) By the positive evidence principle, Mr. Jones does not meet the DENEX criteria.
• (NUMER = NO) Mr. Jones does not meet the NUMER criteria.
• (NUMEX = NO) By definition, Mr. Jones does not meet the NUMEX criteria because NUMEX criteria are only applicable to those meeting the NUMER criteria.

Individual observations for Mr. Jones:
• DENOM observation: 5
• NUMER observation: N/A (Mr. Jones does not meet the NUMER criteria.)

Mr. James is 75 years old and was hospitalized for 24 days. He had a central line in place for 17 days. There is no mention of immunosuppression in his chart. He has a documented central line blood stream infection while the central line was in place. There is no mention that the infection is a contaminant.

• (IP = YES) Mr. James meets the IP criteria.
• (DENOM = YES) Mr. James meets the DENOM criteria.
• (DENEX = NO) By the positive evidence principle, Mr. James does not meet the DENEX criteria.
• (NUMER = YES) Mr. James does meet the NUMER criteria.
• (NUMEX = NO) By the positive evidence principle, Mr. James does not meet the NUMEX criteria.

Individual observations for Mr. James:
• DENOM observation: 17
• NUMER observation: Mr. James meets the NUMER criteria.

Measure aggregate calculations:

Building from the examples above, considering all cases within the population, the aggregate counts are determined:
• IP: N = 150 (i.e., 150 cases meet the IP criteria)
• DENOM: N = 20 of the 150 had laboratory confirmed blood stream infection with central lines
• DENEX: N = 2
• NUMER: N = 6 (Mr. James was one of the six numerator cases)
• NUMEX: N = 1

Aggregate DENOM: 108 (In this example, 108 is an assumed sum of central line days across all cases in the DENOM and not in the DENEX)

Aggregate NUMER: 5 (total number of central line blood stream infections, excluding those deemed to be contaminants)

Ratio = aggregate NUMER / aggregate DENOM = 5 / 108 = 0.046

Additional information about eCQMs and proportion, ratio, and CV measure calculation can be found at the eCQI Resource Center.
APPENDIX G: REFERENCES


