Blueprint for the CMS Measures Management System

Version 13.0 | May 2017

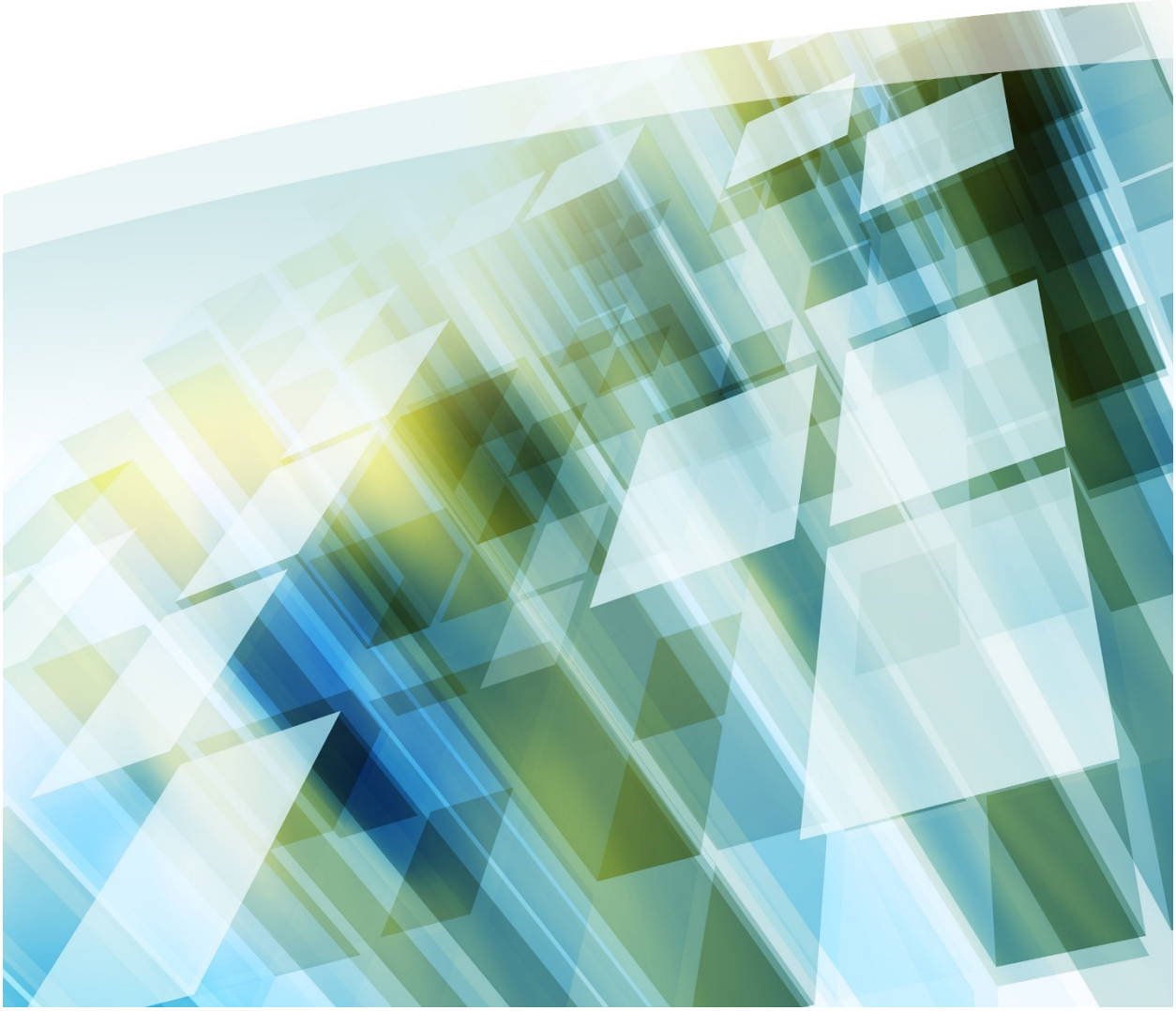

Table of Contents

PREFACE XII
Using the Blueprint xiii
Changing the Blueprint xiv

SECTION 1. INTRODUCTION 1

1 CMS Quality Measure Development 2
1.1 CMS Quality Strategy 3
1.2 Successes to Date 3
1.3 Critical Challenges 4
1.4 General Principles for Measure Development 4
1.5 Technical Principles for Measure Development 5

2 The Measures Management System 6
2.1 Role of the Measure Developer 6
2.2 Role of the Contracting Officer's Representative (COR) for the Measure Development Contract 7
2.3 Role of the Measures Manager 7

SECTION 2. THE MEASURE LIFECYCLE 9

1 Measure Conceptualization 12
1.1 Information Gathering 13
1.2 Business Case Development 14
1.3 Stakeholder Engagement 16
1.3.1 Technical Expert Panel (TEP) 16
1.3.2 Person and Family Engagement 17
1.3.3 Public Comment 17

2 Measure Specification 19
2.1 Technical Specification 19
2.2 Harmonization 21
2.3 Stakeholder Engagement 22

3 Measure Testing 23
3.1 The Measure Testing Process 23
3.2 Stakeholder Engagement 25

4 Measure Implementation 26
4.1 Pre-Rulemaking 26
4.2 Other Submissions 26
4.3 Stakeholder Engagement 27
4.4 The Implementation Process 27

5 Measure Use, Continuing Evaluation, and Maintenance 29
5.1 Continuing Evaluation 29
5.2 Measure Maintenance 29
5.2.1 Measure Production and Monitoring 30
5.2.2 Measure Maintenance Reviews 31
5.2.3 Stakeholder Engagement 31

SECTION 3. IN-DEPTH TOPICS 32

1 Health Care Quality Strategies 33
1.1 National Quality Strategy 33
1.2 CMS Quality Strategy 34
1.3 Crosswalk Between the NQS and the CMS Quality Strategy 35

2 Priorities Planning 36
2.1 Measure Prioritization 36
2.2 CMS Measure Planning Inputs 37
2.2.1 NQS and CMS Quality Strategy 37
2.2.2 Legislative Mandates 38
2.2.3 Quality Measure Development Plan 39
2.2.4 Patients, Public, and Other Stakeholders 39
2.2.5 Impact Assessment and Other Reports 39
2.3 Role of the Measure Developer in Priorities Planning 40
2.4 Role of the Measures Manager in Priorities Planning 41

3 Measure Governance 43
3.1 Developers 43
3.2 Stewards 43

4 Measure Classification 44

5 Selected Measure Types 47
5.1 Cost and Resource Use Measures 47
5.2 Composite Performance Measures 48
5.2.1 Purpose of Composite Measures 49
5.2.2 Component Performance Measures 49
5.3 Patient-Reported Outcome (PRO) Measures 50
5.3.1 Patient-reported Outcomes 50
5.3.2 Patient-reported Outcome Measurement Tools 50
5.3.3 Patient-reported Outcome-based Performance Measures 51
5.3.4 Approaches to Developing Patient-reported Outcome-based Performance Measures 51
5.4 Multiple Chronic Conditions Measures 53
5.4.1 MCC Definition 53
5.4.2 Need for Measure Development 54
5.4.3 Considerations for Measure Development Targeting Persons with MCC 54

6 Introduction to Electronic Clinical Quality Measures (eCQMs) 57
6.1 eCQM Components 58
6.2 Special Considerations for eCQMs 59

7 Electronic Clinical Quality Measure (eCQM) Standards-Based Guidance and Tools 60
7.1 ONC JIRA Tracking System 60
7.2 Health Quality Measure Format (HQMF) 60
7.3 Quality Data Model (QDM) 61
7.3.1 Stating Logic Using the Quality Data Model 62
7.4 Measure Authoring Tool (MAT) 62
7.4.1 Define HQMF Metadata 62
7.4.2 MAT Package 63
7.5 Quality Reporting Document Architecture (QRDA) 63
7.6 Tools for Testing eCQMs 63
7.6.1 Bonnie Tool 63
7.6.2 Cypress Tool 64

8 Master List of Deliverables 65

9 Information Gathering 67
9.1 Conduct an Environmental Scan 67
9.2 Conduct an Empirical Data Analysis, as Appropriate 68
9.3 Evaluate Information Collected during Environmental Scan and Empirical Data Analysis 69
9.4 Conduct a Measurement Gap Analysis to Identify Areas for New Measure Development 70
9.5 Determine the Appropriate Basis for Creation of New Measures 70
9.6 Apply Measure Evaluation Criteria 71
9.7 Submit the Information Gathering Report 72
9.7.1 Summary of Literature Review (annotated bibliography) 72
9.7.2 Summary of Clinical Practice Guidelines Review 73
9.7.3 Review of Existing Measures, Related Measures, and Gap Analysis Summary 73
9.7.4 Empirical Data Analysis Summary 73
9.7.5 Summary of Solicited and Structured Interviews, if applicable 73
9.8 Prepare an Initial List of Measures or Measure Topics 73

10 Environmental Scan 75
10.1 Literature Review 75
10.2 Quality of the Body of Evidence 76
10.3 Clinical Practice Guidelines 77
10.4 Existing and Related Measures 77
10.5 Stakeholder Input to Identify Measures and Important Measure Topics 78
10.6 Call for Measures 79

11 Business Case 80
11.1 MMS Business Case Best Practices 84
11.2 Business Case Template 84

12 Technical Expert Panel (TEP) ... 85
12.1 Timing of TEP input ... 85
12.2 TEP Structures ... 86
12.2.1 Traditional TEP Structure ... 86
12.2.2 Standing TEP Structure ... 87
12.3 Steps of the TEP ... 87
12.3.1 Draft the TEP Charter and Consider Potential TEP Members for Recruitment ... 88
12.3.2 Complete the Call for TEP Web Page Posting Form ... 88
12.3.3 Notify Relevant Stakeholder Organizations ... 89
12.3.4 Post the Call for Nominations following COR Review ... 89
12.3.5 Select TEP and Notify the COR of the Membership List ... 90
12.3.6 Select Chair or Meeting Facilitator ... 91
12.3.7 Post the TEP Composition Documentation (Membership) List and Projected Meeting Dates ... 91
12.3.8 Arrange TEP Meetings ... 91
12.3.9 Send Materials to the TEP ... 91
12.3.10 Conduct the TEP Meetings ... 92
12.3.11 Prepare TEP Summary Report and Propose Recommended Set of Candidate Measures ... 93
12.3.12 Post the TEP Summary Report ... 94

13 Person and Family Engagement ... 95
13.1 Background and Definition ... 95
13.2 Options for Engagement and Selected Best Practices ... 95
13.3 Engagement Activities: Virtual vs. In-Person ... 99
13.4 Recruitment ... 99
13.5 Options for Engagement, by Measure Lifecycle Stage and Selected Best Practices ... 101
13.5.1 Measure Conceptualization ... 101
13.5.2 Measure Specification ... 101
13.5.3 Measure Testing ... 102
13.5.4 Measure Implementation ... 103
13.5.5 Measure Use, Continuing Evaluation, and Maintenance ... 104
13.6 Other Considerations ... 104

14 Public Comment ... 106
14.1 Timing of Public Comment ... 106
14.2 Federal Rulemaking ... 106
14.3 Steps for Public Comment ... 107
14.3.1 Prepare the Call for Public Comment ... 107
14.3.2 Notify Relevant Stakeholder Organizations ... 107
14.3.3 Post the Measures Following COR Approval ... 108
14.3.4 Collect Information ... 108
14.3.5 Summarize Comments and Produce Report ... 108
14.3.6 Send Comments to the TEP for Consideration ... 109
14.3.7 Finalize the Public Comment Report, including Verbatim Comments ... 109
14.3.8 Arrange for the Final Public Comment Summary Report to be Posted on the Website ... 109

15 MMS Website Posting ... 110
15.1 Posting Timeframe ... 111
15.2 Posting Format ... 111
15.3 Posting Template ... 111
15.4 Documents to Include ... 111

16 Measure Technical Specification ... 113
16.1 Develop the Candidate Measure List ... 113
16.1.1 New Measures ... 114
16.1.2 Adopted Measures ... 114
16.1.3 Adapted Measures ... 114
16.1.4 Respecified Measures ... 115
16.1.5 Composite Performance Measures ... 115
16.2 Develop Precise Technical Specifications and Update the Measure Information Form ... 115
16.2.1 Develop Measure Name and Description ... 115
16.2.2 Define the Initial Population ... 116
16.2.3 Define the Denominator ... 117
16.2.4 Define the Numerator ... 117
16.2.5 Determine if Denominator Exception or Denominator Exclusion Is Needed ... 118
16.2.6 Mathematical Relationships between Populations in a Proportion Measure ... 121
16.2.7 Continuous Variable and Ratio Measure Calculations ... 125
16.3 Measure Stratification ... 130
16.4 Define the Data Source ... 131
16.4.1 Administrative Data ... 131
16.4.2 Electronic Clinical Data ... 132
16.4.3 Patient Medical Records (Paper-based or Electronic) ... 132
16.4.4 Registry ... 132
16.4.5 Patient Assessments ... 132
16.4.6 Patient Reported Data/Surveys ... 133
16.5 Specify the Code Systems ... 133
16.6 Construct Data Protocol ... 133
16.6.1 Describe the Level of Measurement/Analysis ... 134
16.6.2 Describe Sampling ... 135
16.6.3 Determine Risk Adjustment ... 135
16.6.4 Clearly Define Any Time Intervals ... 135
16.6.5 Describe How the Measure Results are Scored and Reported ... 136
16.6.6 Develop the Calculation Algorithm ... 137
16.7 Document the Measures and Obtain COR Approval ... 137

17 Codes, Code Systems, and Datasets ... 139
17.1 Code Systems/Vocabularies/Terminologies ... 139
17.1.1 Encoding Clinical Information ... 139
17.1.2 Health Information Technology Standards Committee Recommendations ... 140
17.1.3 Quality Data Model (QDM) Categories with Recommended Code Systems ... 141
17.1.4 Use of Specific Code Systems ... 145

17.2 Value Sets ... 151
17.2.1 Use of Value Sets ... 151
17.2.2 Constructing Value Sets ... 151
17.2.3 Representing the Codes to Be Included ... 151
17.2.4 Determining the Value Set Code System ... 152
17.3 Value Set Reuse ... 154
17.4 Value Set Authority Center (VSAC) ... 154
17.5 VSAC Collaboration ... 155
17.6 Value Set Use in eCQMs ... 155
17.6.1 Use of Value Sets to Address Negation Rationale in eCQMs ... 155
17.6.2 Value Set Package Review ... 156
17.6.3 Pre-MAT Value Set Review ... 157
17.6.4 Post-MAT Value Set Review ... 158
17.6.5 NLM Value Set Review ... 158

18 Measure Harmonization ... 159
18.1 Adapted or Respecified Measures ... 160
18.2 Adopted Measures ... 161
18.3 New Measures ... 162
18.4 Harmonization during Measure Maintenance ... 162
18.4.1 Decide whether Harmonization is Indicated ... 162
18.4.2 Implement Harmonization Decisions ... 163
18.4.3 Test Scientific Acceptability of Measure Properties ... 163
18.4.4 NQF Evaluates for Harmonization during Measure Maintenance ... 163

19 Risk Adjustment ... 164
19.1 Risk Adjustment Strategies ... 164
19.2 Attributes of Risk Adjustment Models ... 165
19.2.1 Sample Definition ... 166
19.2.2 Appropriate Time Frames ... 166
19.2.3 High Data Quality ... 167
19.2.4 Appropriate Variable Selection ... 167
19.2.5 Appropriate Analytic Approach ... 168
19.2.6 Complete Documentation ... 168
19.3 Risk Adjustment Procedure ... 168
19.3.1 Choose and Define an Outcome ... 169
19.3.2 Define the Conceptual Model ... 169
19.3.3 Identify the Risk Factors and Timing ... 170
19.3.4 Acquire Data (Sample, if Necessary) ... 171
19.3.5 Model the Data ... 171
19.3.6 Assess the Model ... 174
19.3.7 Document the Model ... 176

20 Cost and Resource Use Measure Specification ... 178
20.1 Measure Clinical Logic ... 178
20.2 Measure Construction Logic ... 178
20.2.1 Time Frames ... 178
20.2.2 Assigning and Triaging Claims ... 178

20.3 Adjusting for Comparability ... 179
20.3.1 Define Risk Adjustment Approach ... 179
20.3.2 Define Stratification Approach ... 179
20.3.3 Define Costing Methodology ... 179

20.4 Measure Reporting ... 179
20.4.1 Attributing Resource Use Measures ... 179
20.4.2 Peer Group Identification and Assignment ... 179
20.4.3 Calculating Comparisons ... 180
20.4.4 Setting Thresholds ... 180
20.4.5 Providing Detailed Feedback ... 180
20.4.6 Reporting with Descriptive Statistics ... 180

21 Composite Measure Technical Specifications ... 181

22 Measure Testing ... 186
22.1 Develop the Testing Work Plan ... 186
22.2 Submit the Plan and Obtain CMS Approval ... 187
22.3 Implement the Plan ... 187
22.4 Analyze the Test Results ... 187
22.5 Refine the Measure ... 188
22.6 Retest the Refined Measure ... 188
22.7 Compile and Submit Deliverables to CMS ... 188
22.7.1 Measure Testing Summary Report ... 188
22.8 Support CMS during NQF Endorsement Process ... 190

23 Alpha and Beta Testing ... 191
23.1 Alpha Testing ... 192
23.2 Beta Testing ... 192
23.3 Sampling ... 192

24 Measure Evaluation ... 194
24.1 Measure Evaluation Criteria and Subcriteria ... 194
24.2 Applying Measure Evaluation Criteria ... 195
24.3 Timing of Measure Evaluation ... 197
24.4 Testing and Measure Evaluation Criteria ... 198
24.4.1 Importance ... 199
24.4.2 Scientific Acceptability ... 199
24.4.3 Usability ... 204
24.4.4 Feasibility ... 205
24.5 Evaluation during Measure Maintenance ... 205
24.5.1 Apply Measure Evaluation Criteria ... 206
24.5.2 Report Results of Evaluation ... 206

25 Testing for Special Types of Measures 207
25.1 Adapted or Respecified Measures 207
25.2 eCQM Testing 207
25.2.1 Types of eCQM Testing 207
25.2.2 Feasibility 208
25.2.3 Validity 209
25.2.4 Reliability 211
25.2.5 Testing Multiple Sites and Multiple EHRs 211
25.3 Composite Measures 211
25.3.1 Component Reliability and Validity Testing 211
25.3.2 Component Coherence 211
25.3.3 Composite Reliability and Validity Testing 211
25.3.4 Appropriateness of Aggregation Methods 212
25.3.5 Feasibility and Usability of Composite Components 213

26 Evaluation for Special Types of Measures 214
26.1 Evaluating Composite Measures 214
26.1.1 Considerations for Evaluating Composite Measures 214
26.2 Evaluating Cost and Resource Use Measures 215
26.2.1 Considerations for Evaluating Resource Use Measures 215
26.3 Evaluating eCQMs 216
26.4 Evaluating Patient-reported, Outcome-based Performance Measures 216

27 NQF Endorsement and Maintenance 217
27.1 Measure Submission to NQF 218
27.1.1 NQF Issues a Call for Measures 218
27.1.2 The COR Decides to Submit a Measure to NQF 218
27.1.3 The Measure Developer Completes the NQF Measure Submission 218
27.1.4 The COR Approves the NQF Measure Submission 220
27.1.5 The Measure Developer or the COR Submits the Measure to NQF According to NQF Processes 220
27.2 NQF Endorsement Process 220
27.2.1 Initial Review by NQF Staff 220
27.2.2 Standing Committee and Technical Advisory Panel Review 221
27.2.3 NQF Makes Endorsement Decision 221
27.2.4 Review Endorsement Results 223
27.3 Measure Developer's Role during NQF Evaluation 223
27.4 Trial Use Approved Measures 223
27.5 Measure Maintenance for NQF 223

28 Measure Selection 225
28.1 Pre-Rulemaking Process 225
28.1.1 Measures Under Consideration (MUC) 225
28.1.2 CMS Measure Selection Criteria: 226
28.1.3 MAP Recommendations 226
28.1.4 CMS Considers MAP Input for Final Selection 226
28.2 CMS Rulemaking Processes 227
28.3 Rollout, Production, and Monitoring of Measures 227

28.4 Measure Maintenance 227
28.5 Impact Assessment of Medicare Quality Measures 228

29 Measure Rollout 229
29.1 Measures are Selected by CMS 230
29.2 Implement the Rollout Plan 231
29.3 Implement the Data Management Processes 231
29.4 Develop the Auditing and Validation Plan 232
29.5 Develop an Appeals Process 232
29.6 Implement Education Processes 232
29.7 Conduct the Dry Run 233
29.8 Submit Reports 233

30 Measure Production and Monitoring 234
30.1 Conduct Data Collection and Ongoing Surveillance 235
30.2 Respond to Questions about the Measure 235
30.3 Produce Preliminary Reports 236
30.4 Report Measure Results 236
30.5 Monitor and Analyze the Measure Rates and Audit Findings 236
30.6 Perform Measure Maintenance or Ad Hoc Review, when Appropriate 237
30.7 Provide Information that CMS can Use in Measure Priorities Planning 237

31 Measure Maintenance Reviews 239
31.1 Annual Update 239
31.1.1 Potential for Harmonization 240
31.1.2 Procedure 240
31.2 Comprehensive Reevaluation 243
31.2.1 For Measures in CMS Programs Not Owned by CMS or Maintained by a CMS Measure Developer 244
31.2.2 Harmonization 244
31.2.3 Procedure 244
31.3 CMS Ad Hoc Review 248
31.3.1 Trigger for an Ad Hoc Review 248
31.3.2 Deferring an Ad Hoc Review 249
31.3.3 Procedure 249
31.4 NQF Ad Hoc Reviews 252
31.5 Possible Outcomes of Maintenance Reviews 253

SECTION 4. FORMS AND TEMPLATES 256

1 Environmental Scan Outline 257

2 Business Case Form Instructions 260

3 Business Case Template 263

4 Measure Information Form Instructions 267

5 Blank Measure Information Form Template 273

6 Measure Justification Form Instructions 275

7 Blank Measure Justification Form Template 291

8 Measure Evaluation Criteria and Instructions 296

9 Blank Measure Evaluation Report Template 310

10 CMS Measures Management System WebPage Posting Instructions 314

11 Public Comment Call Web Posting 315

12 Public Comment Summary Web Posting 316

13 Public Comment Summary Report Template 317

14 Technical Expert Panel (TEP) Call for TEP Web Posting 319

15 Technical Expert Panel (TEP) Nomination Form Template 321

16 Technical Expert Panel (TEP) Composition (Membership) List Template 324

17 Technical Expert Panel (TEP) Composition (Membership) List Web Posting 325

18 Technical Expert Panel (TEP) Charter Template 326

19 Technical Expert Panel (TEP) Summary Web Posting 327

SECTION 5. GLOSSARY AND ACRONYMS 328

Glossary 329
Acronyms 347

SECTION 6. APPENDICES 353

Appendix A: eCQM Logic Quality Assurance Checklist 354
Appendix B: eCQM Metadata 355
Appendix C: Summary of Changes to Blueprint 363

List of Figures

Figure 1. Timeline of Measure Lifecycle 10
Figure 2. Interaction among Measure Lifecycle Phases 10
Figure 3. Flow of the Measure Lifecycle—Measure Conceptualization 12
Figure 4. Information Gathering Deliverables 13
Figure 5. Business Case Deliverables 15
Figure 6. Technical Expert Panel Deliverables 16
Figure 7. Public Comment Deliverables 17
Figure 8. Flow of the Measure Lifecycle—Measure Specification 19
Figure 9. Measure Specification Deliverables 20
Figure 10. Flow of the Measure Lifecycle—Measure Testing 23
Figure 11. Measure Testing Deliverables 24
Figure 12. Overview of Measure Testing 25
Figure 13. Flow of the Measure Lifecycle—Measure Implementation 27
Figure 14. Measure Implementation Deliverables 28
Figure 15. Measure Maintenance Deliverables 29
Figure 16. Flow of the Measure Lifecycle—Measure Use, Continuing Evaluation, and Maintenance 30
Figure 17. CMS Priorities Planning and Measure Selection 37
Figure 18. Relationship Between Resource Use, Value, and Efficiency 48
Figure 19. Environmental Scan Data Sources 68
Figure 20. Potential Costs, Benefits, and Savings 80
Figure 21. Inputs and Uses for the Business Case 82
Figure 22. Flow of Inputs into Business Case Development 83
Figure 23. Person and Family Engagement Concepts 95
Figure 24. TEP and Working Group Best Practices 98
Figure 25. Recruitment Practices 100
Figure 26. Measure Conceptualization Best Practice 101
Figure 27. Measure Conceptualization and Specification Best Practice 102
Figure 28. The Web Posting Process 110
Figure 29. Proportion Measure Populations 121
Figure 30. Ratio Measure Populations 126
Figure 31. CV Measure Populations 126
Figure 32. VSAC Value Set Lifecycle 152
Figure 33. Structure Process Outcome Relationship 164
Figure 34. Risk Adjustment Deliverables 169
Figure 35. Example of ROC Curves 175
Figure 36. Applying Measure Evaluation Criteria 197
Figure 37. Measure Submission 217
Figure 38. Measure Review Cycle Timeline 224
Figure 39. Pre-rulemaking Process Timeline 225
Figure 40. Overview of the Measure Rollout Process 229
Figure 41. Overview of the Measure Monitoring Process 234
Figure 42. Annual Update Deliverables 239
Figure 43. Comprehensive Reevaluation Deliverables 243
Figure 44. Ad Hoc Review Deliverables 249
Figure 45. CMS Criteria for Measure Disposition: Revise or Retain 254

Figure 46. CMS Criteria for Measure Disposition: Retire 254
Figure 47. CMS Criteria for Measure Disposition: Remove 255

List of Tables

Table 1. Framework for Progression of Payment to Clinicians and Organizations in Payment Reform 2
Table 2. NQS and CMS Quality Strategy 35
Table 3. National Quality Measures Clearinghouse (NQMC) Clinical Quality Measure (CQM) Domains 44
Table 4. Examples of Measures Addressing Each of the NQS Priorities 45
Table 5. Measurement Settings 46
Table 6. The Business Case is an Ongoing Process that Occurs Throughout the Measure Lifecycle 81
Table 7. Best Practices for Implementing Person/Family Engagement Activities, by Phase of Engagement 96
Table 8. Measure Populations Based on Type of Measure Scoring 120
Table 9. ONC HITSC Recommended Vocabulary Summary 140
Table 10. ONC HITSC Transition Vocabulary Standards Summary and Plan 141
Table 11. Quality Data Model Categories with ONC HITSC Recommended Vocabularies 141
Table 12. Value Set Review Areas and Remedial Actions 157
Table 13. Harmonization Decisions during Measure Development 160
Table 14. Harmonization Decisions during Measure Maintenance 163
Table 15. Attributes of Risk Adjustment Models 166
Table 16. Types of Composite Measure Scoring 183
Table 17. Features of Alpha and Beta Testing 191
Table 18. Extent of Measure Evaluation as a Function of Prior Comprehensive Evaluation and Measure Use 247
Table 19. Logic Review Checklist 354

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 seen through the measure lifecycle. Its five chapters correspond to the five phases of the lifecycle: Measure Conceptualization; Measure Specification; Measure Testing; Measure Implementation; and 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 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 every aspect of measure development. These article topics range from Centers for Medicare & Medicaid Services (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 of the forms and templates required for completion of the measure development process and the delivery of CMS contract deliverables.

Section 5 contains the Glossary.

Section 6 contains Appendices.

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 detailed list of significant changes this latest version, see Appendix C.

Recommendations for changes to the content, structure, or organization of the Blueprint are welcome. Please submit all suggestions to the MMS support mailbox (MMSSupport@battelle.org). Please include specifics about the recommended change, including:

- Version of the Blueprint being referenced
- Relevant section, chapter number, and title
- Page number
- Relevant text to modify, if applicable
- New text to add, if applicable
- Rationale for change
- Point of contact information.

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 underway in our healthcare system. It is a transformation fundamentally driven by performance measurement. In nearly every setting of care, the Centers for Medicare & Medicaid Services (CMS) is moving from paying for volume to paying for value. 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 Affordable Care Act (ACA) and more recently driven by the Medicare Access and CHIP[1] 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[2]

	Category 1: Fee for Service—No Link to Quality	Category 2: Fee for Service—Link to Quality	Category 3: Alternative Payment Models on Fee—for Service Architecture	Category 4: Population-Based Payment
Description	Payments are based on volume of services and not linked to quality or efficiency	At least a portion of payments vary based on the quality or efficiency of healthcare delivery	• Some payment is linked to the effective management of a population or an episode of care • Payments still triggered by delivery of services, but opportunities for shared savings or 2-sided risk	• Payment is not directly triggered by service delivery, so volume is not linked to payment • Clinicians and organizations are paid and responsible for the care of a beneficiary for a long period (e.g., >1yr)
Medicare	• Limited in Medicare FFS • Majority of Medicare payments now are linked to quality	• Hospital value-based purchasing • Physician value-based modifier • Hospital Readmissions/ Hospital-Acquired Condition Reduction Programs	• Accountable Care Organizations • Medical Homes • Bundled Payments	• Eligible Pioneer accountable care organizations in Years 3 through 5 • Some Medicare Advantage plan payments to clinicians and organizations
Medicaid	Varies by state	• Primary Care Case Management • Some managed care models	• Integrated care models under FFS • Managed fee-for-service models for Medicare–Medicaid beneficiaries • Medicaid Health Homes • Medicaid shared savings models • Medicaid waivers for delivery reform incentive payments • Episodic-based payments	• Some Medicaid managed care plan payments to clinicians and organizations • Some Medicare–Medicaid (duals) plan payments to clinicians and organizations

With this payment model transition, the stakes are higher than ever for patients and providers, and the onus is on CMS and other payers to ensure that meaningful robust clinical quality measures (CQMs) are available for determination of quality and value of clinical care across all settings. In striving to achieve the goals of the CMS Quality Strategy, developed measures must be meaningful to patients and the

[1] Children's Health Insurance Program (CHIP)
[2] Rajkumar R, Conway PH, Tavenner M. The CMS—Engaging Multiple Payers in Risk-Sharing Models. JAMA. Doi:10.1001/jama.2014.3703.

providers who serve them, represent opportunities for improvement in care quality, and differentiate quality in a meaningful and valid way. In order to accomplish this, the following strategies for measure development that must be kept at the forefront.

1.1 CMS Quality Strategy

CMS released its Quality Strategy in late 2013. Building off the framework of the National Quality Strategy (NQS), our Quality Strategy articulates six goals to improve the quality of care in our healthcare system:

1) Make Care Safer
2) Strengthen Person and Family Engagement
3) Promote Effective Communication and Coordination of Care
4) Promote Effective Prevention and Treatment
5) Work with Communities to Promote Best Practices of Healthy Living
6) Make Care Affordable.

Each goal has a set of objectives and desired outcomes. The Quality Strategy also identifies ongoing and future initiatives and activities that CMS and front line providers can engage in to achieve each goal and objective. In the context of CMS programs, measure developers should familiarize themselves with the CMS Quality Strategy, and should explicitly link proposed measure concepts to the goals and objectives while taking into consideration the foundational principles described in this document.

For in-depth information on healthcare quality strategies, see Section 3, Chapter 1, Health Care Quality Strategies.

1.2 Successes to Date

For the first time in many years, a number of critically important metrics at the national level have 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, the NQF endorsement process has become more challenging and CMS continually seeks innovations and process improvement to meet these challenges. CMS has also started to remove measures from our programs that have little room for improvement, are no longer supported by evidence, or are of low value from the patient or clinical workflow perspective.

CMS is rebalancing our portfolio of measures to contain more outcome and fewer process measures, with the goal of better addressing performance gaps in the six domains[3] of the NQS and the CMS quality strategy. As part of this effort, the Measures Manager maintains the CMS Measures Inventory. The Inventory is inclusive of measures under development, measures under consideration, and measures

[3] Note that CMS and other HHS agencies define and use the term "domain" differently from one another. Therefore, within the Blueprint, you will see the term "domain" defined differently in different contexts, depending on the relevant agency within the discussion.

that have entered the rulemaking process for dozens of CMS programs and initiatives. The Inventory includes measure title, description, NQF endorsement status, measure type, NQS domain, etc., 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)[4] and cannot all be enumerated here. However, some of the key challenges 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 Tools (PROMs) and appropriate use measures
- Developing measures that promote shared accountability across settings and providers.

Now is the time to address these challenges head-on using Lean techniques in all phases of measure development. CMS also 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

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

- Explicitly align with the CMS Quality Strategy and its goals and objectives.
- Align with other players, including Medicaid, other federal partners, and private payers.
- Address a performance gap where there is known variation in performance, not just a measure gap.
- Focus on what is best for patients.
- Be developed in a rapid-cycle fashion, in accordance with Lean principles.
- Be based on collaboration among measure developers, and share best practices/new learnings freely.
- Reorient and align around patient-centered outcomes that span across clinical settings–this 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.
- Focus on outcomes (including patient-reported outcomes, such as functional status after knee replacement), safety, patient experience, care coordination, and appropriate use.
- Be meaningful to patients, caregivers, and providers.

[4] CMS Quality Measure Development Plan. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/Final-MDP.pdf. Accessed on: March 10, 2017.

- Be focused on outcomes (including patient-reported outcomes), 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. [5]

1.5 Technical Principles for Measure Development

The following principles govern the technical execution of a measure:

- 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) over data from claims and chart abstraction, where appropriate.
- Maintain a focus on iterative testing using both synthetic and real data.
- Consider approaches to aggregate multiple data sources 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 is 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. For example, 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. Strive to develop each measure with sufficient statistical power to detect and report statistically significant differences in provider performance.
- Consider strategies to allow clinicians with smaller practices and low-volume facilities to reliably report a measure.
- Strive to develop measures that can progress multi-payer applicability using all-payer databases where available.[6]

[5] Ibid.
[6] Ibid.

2 THE MEASURES MANAGEMENT SYSTEM

The CMS Measures Management System (MMS)[7] is a standardized system for developing and maintaining the quality measures used in CMS' 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 manual, the CMS Measures Management System Blueprint (the 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. See also Measure Governance under the In-Depth Topics section 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. Because this Blueprint is designed as a guide for entities holding a measure development and maintenance contract with CMS, it most often will 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 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."[8] 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.[9]

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).
- 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,

[7] Centers for Medicare & Medicaid Services. *Measures Management System.* Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/. Accessed on: May 6, 2015.
[8] National Quality Forum. *Phrasebook: A Plain Language Guide to NQF Jargon.* Washington, DC: National Quality Forum; 2013. Available at: http://public.qualityforum.org/NQFDocuments/Phrasebook.pdf. Accessed on: May 25, 2015.
[9] http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: May 25, 2015.

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 CMS COR on all communications with the Measures Manager.

2.2 Role of the Contracting Officer's Representative (COR) 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 owner and measure steward. This means that CMS holds and retains ultimate responsibility for measures developed by its measure developers. Within the context of this Blueprint, the measure developer's 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 points 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. This 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 between multiple internal Department of Health and Human Services (HHS), CMS, 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 CMS CORs on all Measures Manager to measure developer communications.
- Maintaining listserv of measure development stakeholders and communicating with measure development public about topics of interest.

Section 2. The Measure Lifecycle

The end product of measure development is a precisely specified, valid, reliable, and clinically significant measure that is directly linked to the 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. Though the figure depicts the five phases of the measure lifecycle in a linear, sequential fashion, it should be understood that 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. Figure 2 depicts the way in which the measure lifecycle phases may overlap and interact in a nonlinear fashion.

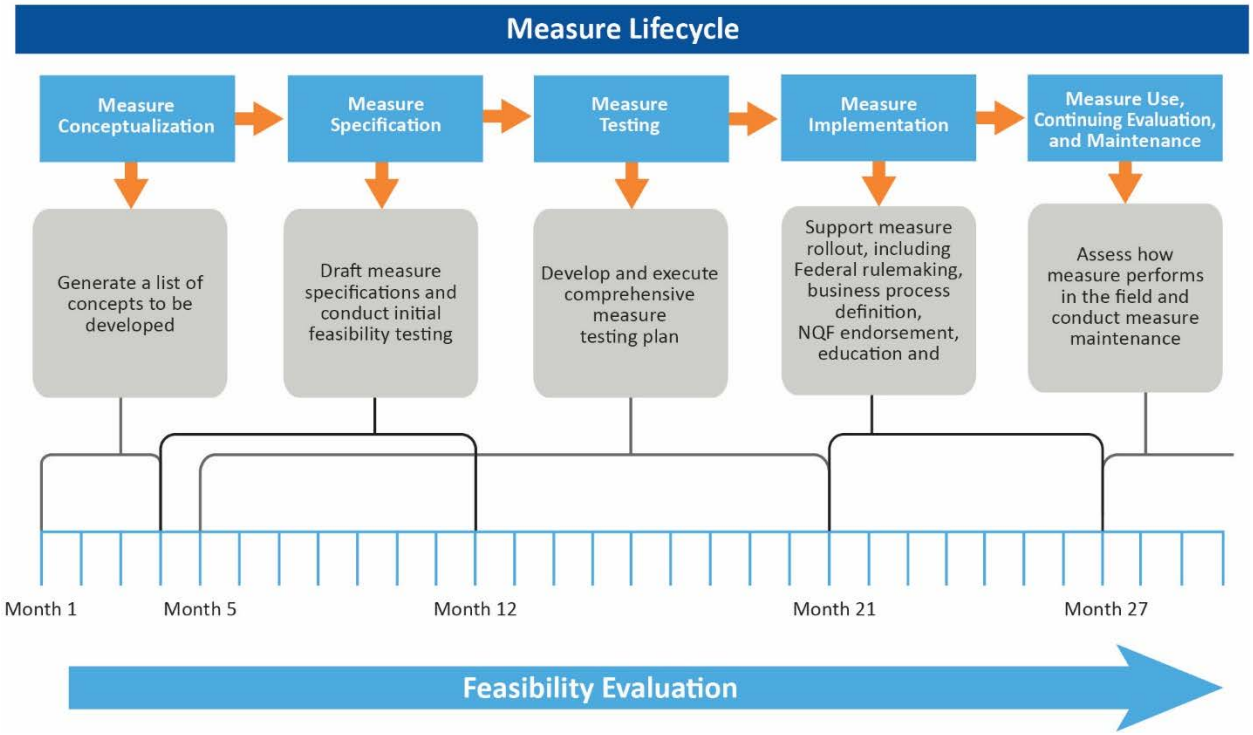

Figure 1. Timeline of Measure Lifecycle

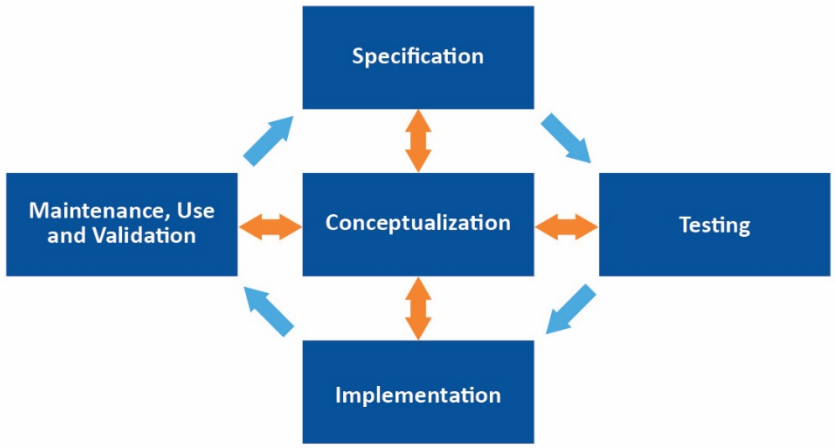

Figure 2. Interaction among Measure Lifecycle Phases

Note that the discussions in this document reflect both traditional measures and electronic clinical quality measures (eCQMs). Information pertaining uniquely to eCQMs is designated with a computer icon.

For more background on eCQMs, see Section 3, Chapter 6, Introduction to Electronic Clinical Quality Measures (eCQMs).

1 MEASURE CONCEPTUALIZATION

In the first phase of the measure lifecycle, measure developers compile the evidence base for the concept and the basic elements of the measures are compiled. Figure 3 depicts measure conceptualization in the context of the entire measure lifecycle. In the measure conceptualization stage, the measure developer identifies whether existing measures may be adopted, adapted, 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 CMS COR for approval. Upon approval from the COR, the measure developer proceeds with the development of draft specifications for the measures.

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) 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 avoid rework later in the process.

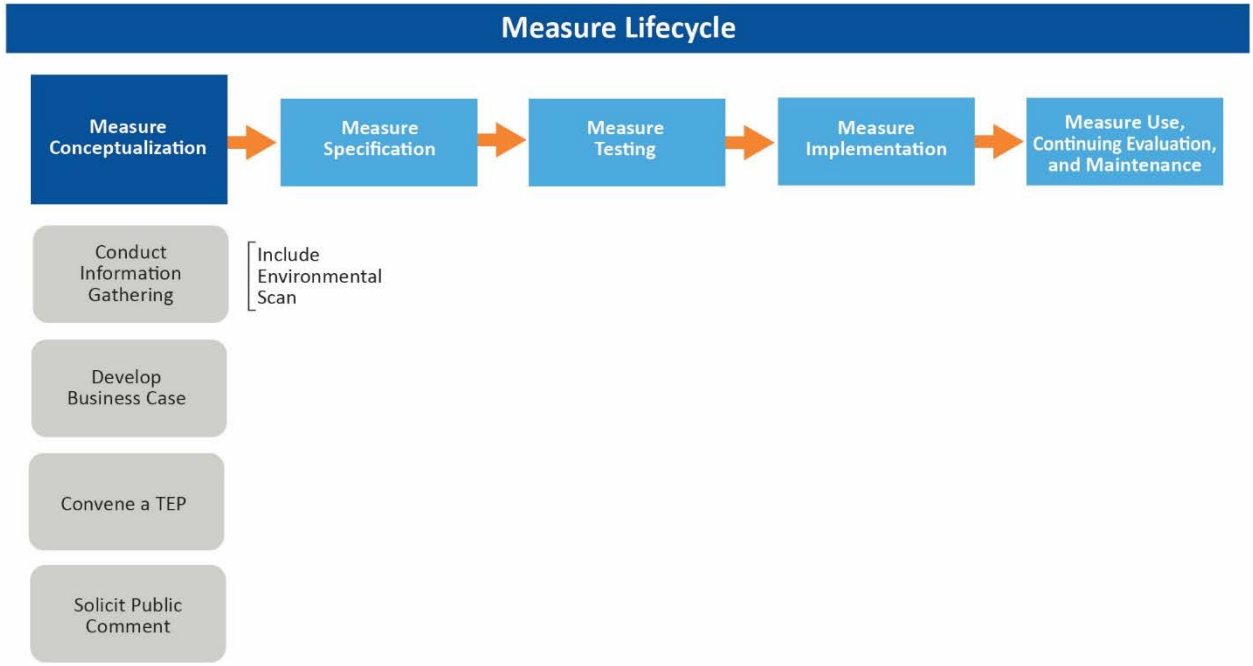

Figure 3. Flow of the Measure Lifecycle—Measure Conceptualization

The main components of Measure Conceptualization are:

- Information gathering
- Business case development
- Stakeholder input
  - TEP
  - Person/family engagement
  - Public comment.

1.1 Information Gathering

Information gathering is a broad term that includes an environmental scan (literature review, clinical practice guidelines search, interviews, 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.[10]

Information Gathering Deliverables

1. Information gathering report
2. List of potential candidate measures
3. Measure documentation, including:
   - Measure Information Form (MIF)
   - Measure Justification Form (MJF)
   - Business case
   - Expert input report (if applicable).

Figure 4. Information Gathering Deliverables

Good information gathering will provide a significant knowledge base that includes the quality goals, the 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.

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 Measure Information Form (MIF) and Measure Justification Form (MJF).[11]

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

- Conduct an 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

[10] A master list of all deliverables associated with the measure development process is in Section 3, Chapter 8.
[11] Completed NQF measure submission forms may be used for contract deliverables in lieu of the MIF and MJF. This is permissible only if the contract allows for it.

- Submit the information gathering report
- Prepare an initial list of measures or measure topics.

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

1.2 Business Case Development

The CMS Measure and Instrument Development and Support (MIDS) master contract requires a business case to be developed for each candidate measure. The business case provides CMS with the information needed to assess the anticipated benefits of a measure against the resources and costs required to implement a measure. It should include enough information to demonstrate the strategic fit of the measure in CMS' measure portfolio, addressing the strategic goals and objectives of the CMS quality strategy, 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 both 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 categories below align with the National Quality Strategy's three aims, the following types of information should be systematically evaluated to build the business case:[12]

Better Care:

- Improve the overall quality, by making health care more patient-centered, reliable, accessible, and safe.
- Improve processes of care delivery where every patient gets the right care every time (e.g., preventive screening services, immunizations, pharmacotherapy, counseling)
- Improve intermediate outcomes (e.g., lower blood pressure, lower HbA1c values among diabetics)
- 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)
- Improve outcomes of care (e.g., fewer heart attacks, decubitus ulcers, etc.).

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
- 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
- Savings from preventing complications and adverse events and overall improved health.

[12] Leatherman, S, Berwick, D, et al. The business case for quality: case studies and an analysis. *Health Affairs*. 2003; 22, (number 2): p. 18. Available at: http://content.healthaffairs.org/content/22/2/17.full.pdf. Accessed on: March 14, 2016.

Continual Improvement:

- Monitor the magnitude and timeframe of the expected benefits 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.

While other models may be used, a cost savings model is the most prevalent for evaluating the potential quality measures' business case, that is 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 claims 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, 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.[13]

Business Case Deliverables

1. Initial business case
2. Final business case

Figure 5. Business Case Deliverables

[13] A master list of all deliverables associated with the measure development process is in Section 3, Chapter 8.

1.3 Stakeholder Engagement

1.3.1 Technical Expert Panel (TEP)

A TEP is a group of stakeholders and experts 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 (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 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:

- Technical Expert Panel (Call for TEP) nominations
- The TEP Composition Documentation with meeting dates
- The 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 (TEP), and should be performed when convening the TEP and conducting the TEP meetings. TEP deliverables are outlined in Figure 6.[14]

Technical Expert Panel Deliverables

1. Call for TEP
2. TEP nomination forms
3. List of stakeholders
4. TEP Charter
5. TEP Composition Documentation (TEP Membership List)
6. TEP meeting schedule
7. Meeting minutes
8. Potential measures presented to the TEP
9. Measure Evaluation Report
10. Update MIF and MJF

Figure 6. Technical Expert Panel Deliverables

[14] A master list of all deliverables associated with the measure development process is in Section 3, Chapter 8.

1.3.2 Person and Family Engagement

Involving persons[15] and family representatives in the measure development process (e.g., on TEPs, in focus groups, and 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.[16] In this context, a *person* is a non-healthcare professional representing those who receive healthcare.[17] Family representatives are other non-healthcare professionals, such as caregivers, supporting those who receive health care. Guidance for obtaining input from persons and family member stakeholders is provided in Section 3, Chapter 13, Person and Family Engagement.

1.3.3 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 below will apply whenever public comment is obtained.

The Call for Public Comment involves several postings to the dedicated CMS MMS website.[18] 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 5 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 Deliverables

1. Public Comment Call Web Posting template form
2. List of stakeholders for notification
3. Measure Justification Form (for candidate measures)
4. eCQM specifications (as appropriate)
5. Public Comment Summary report, including verbatim comments
6. Responses to Public Comment

Figure 7. Public Comment Deliverables

The following eight steps are essential to successfully soliciting public comment. Deviation from the following procedure requires COR approval. Public comment deliverables are outlined in Figure 7.[19]

- Prepare the Call for Public Comment
- Notify relevant stakeholder organizations
- Post the measures following COR approval
- Collect information

[15] The term "persons" is used in lieu of "patients" throughout the *Blueprint* in order to reflect an individual's identity as more than a patient, to recognize his or her participation in prevention and wellness. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/Person-and-Family-Engagement-Strategy-Summary.pdf

[16] CMS 2016 Quality Strategy. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/CMS-Quality-Strategy.pdf. Accessed 2/25/2016

[17] The term "patients" is not used here because the term "patient" refers to individuals receiving inpatient care in hospitals.

[18] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Measures Management System: Public Comment.* Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/CallforPublicComment.html. Accessed on: March 14, 2016.

[19] A master list of all deliverables associated with the measure development process is in Section 3, Chapter 8.

- Summarize comments and produce report
- Send comments to the TEP for consideration
- Finalize the Public Comment Report, including verbatim comments
- 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.

2 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. The fields within the MIF are completed and updated 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.

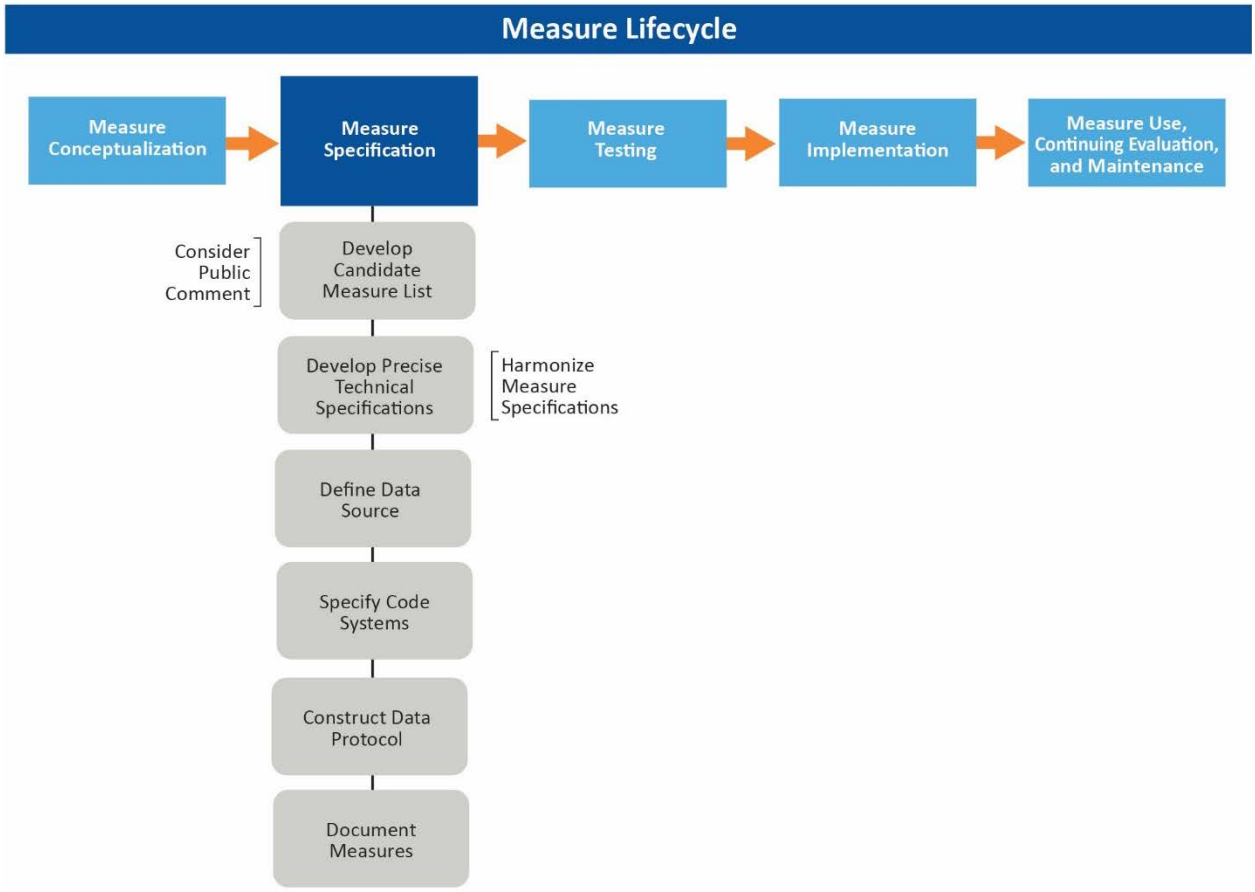

Figure 8. Flow of the Measure Lifecycle—Measure Specification

2.1 TECHNICAL SPECIFICATION

The MIF is used to document the technical specifications of the measures.[20] 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 from the measure specification process are outlined in Figure 9.[21]

[20] Completed NQF measure submission forms may be used for contract deliverables in lieu of the MIF and MJF.
[21] A master list of all deliverables associated with the measure development process is in Section 3, Chapter 8.

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 3, 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.

Measure Specification Deliverables

1. List of potential measures to be developed and timelines
2. Appropriate measure specification forms by case:
   a. MIF and MJF, or equivalent, for candidate measures
   b. MJF, or equivalent, to document the measure evaluation for new or adapted measures, or measures that are developed and in use by another organization but are not NQF-endorsed
   c. For eCQMs: Health Quality Measure Format (HQMF), SimpleXML, human-readable HTML, and value sets
3. For risk-adjustment measures:
   a. Risk Adjustment Methodology Report
   b. MIF with completed risk adjustment sections for each measure
   c. For eCQMs, the eCQM human-readable rendition (HTML) file includes instructions where the complete risk adjustment methodology may be obtained

Figure 9. Measure Specification Deliverables

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)
- Exclusion and exception
- Data sources
- Key terms, data elements, codes, and code systems
- Unit of measurement or analysis
- Sampling and stratification
- Risk adjustment (Section 3, Chapter 19, Risk Adjustment)
- Timing and time intervals
- Calculation algorithm.

eCQMs must also have eCQM XML file (HQMF), SimpleXML file, an eCQM human-readable rendition (HTML) file, and value sets. Most of the components listed above are in these files.

The following 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 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. See 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. See Section 3, Chapter 20, Cost and Resource Use Measure Specification, for details on developing technical specifications for those measure types.

2.2 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, etc.) 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. 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 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: code systems, code lists, descriptions
- Allowable conditions for inclusion in the denominator: code systems, code lists, descriptions
- Exclusion categories, whether the exclusion is from the denominator or numerator, 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 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 behind the new measure does not support using the same variable(s) found in the existing measure
- CMS's intent for the measure 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 18–75. A new process-of-care measure is based on new clinical practice guidelines that recommend a particular treatment only for individuals aged 65 years and older.
- An existing diabetes measure includes individuals aged 18–75. CMS has requested measures for beneficiaries aged 75 years and older.

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

2.3 Stakeholder Engagement

Though it is advisable to obtain public comments at several points during measure development, a key time to get 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, see Section 3, Chapter 14, Public Comment.

3 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 measure evaluation criteria.[22] 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.

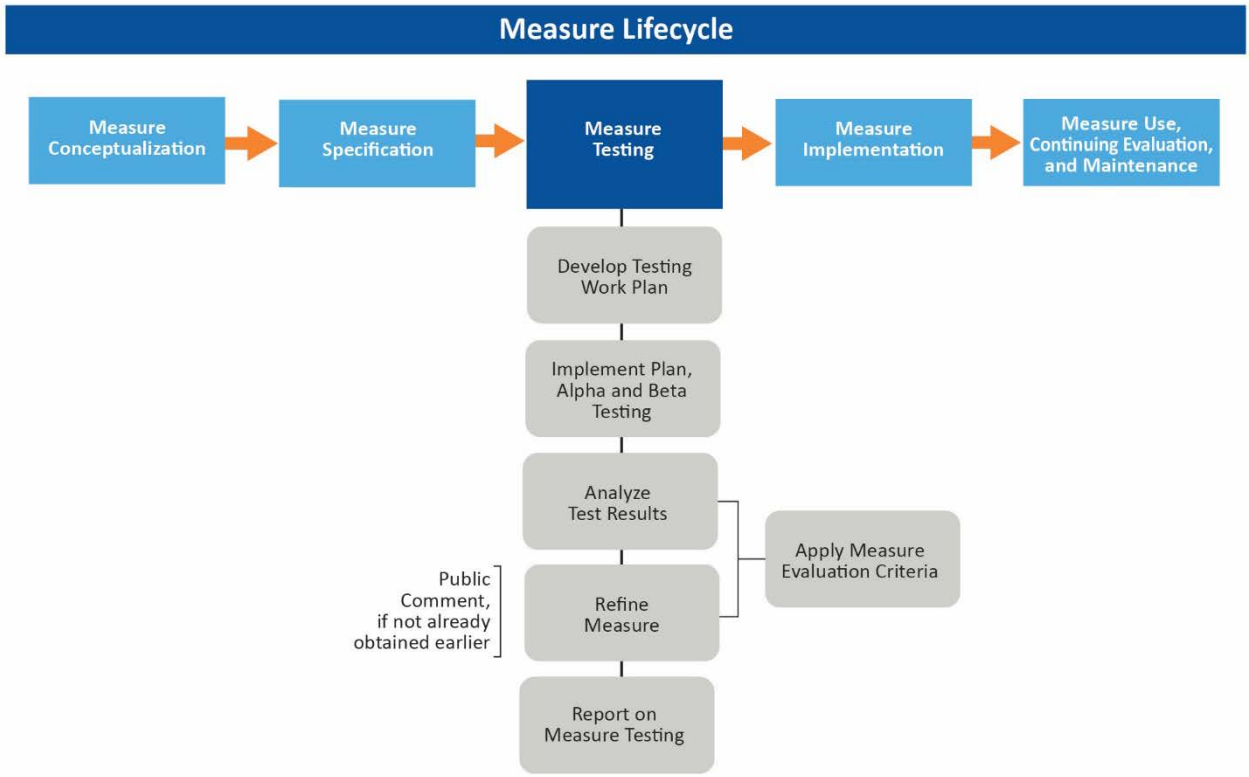

Figure 10. Flow of the Measure Lifecycle—Measure Testing

3.1 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 CMS COR, and key considerations when analyzing and documenting results of testing and analysis, including incorporation of stakeholder inputs after testing is complete.

[22] NQF Measure Evaluation Criteria. Available at: https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=83123. Accessed on: December 2, 2016.

When testing a measure (or set of measures) for CMS, a measure developer is required to submit specific reports. Though 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 measure testing deliverables.[23]

Measure Testing Deliverables

1. Measure Testing Plan
2. Measure Testing Summary Report
3. Updated Measure Information Form
4. Updated Measure Justification Form
5. Updated Measure Evaluation Report

Figure 11. Measure Testing Deliverables

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:

- Develop the testing work plan
- Submit the plan and obtain CMS approval
- Implement the plan
- Analyze the test results
- Refine the measure, including incorporation of stakeholder inputs
- Retest the refined measure
- Compile and submit deliverables to CMS
- Support CMS during NQF endorsement process.

[23] A master list of all deliverables associated with the measure development process is in Section 3, Chapter 7.

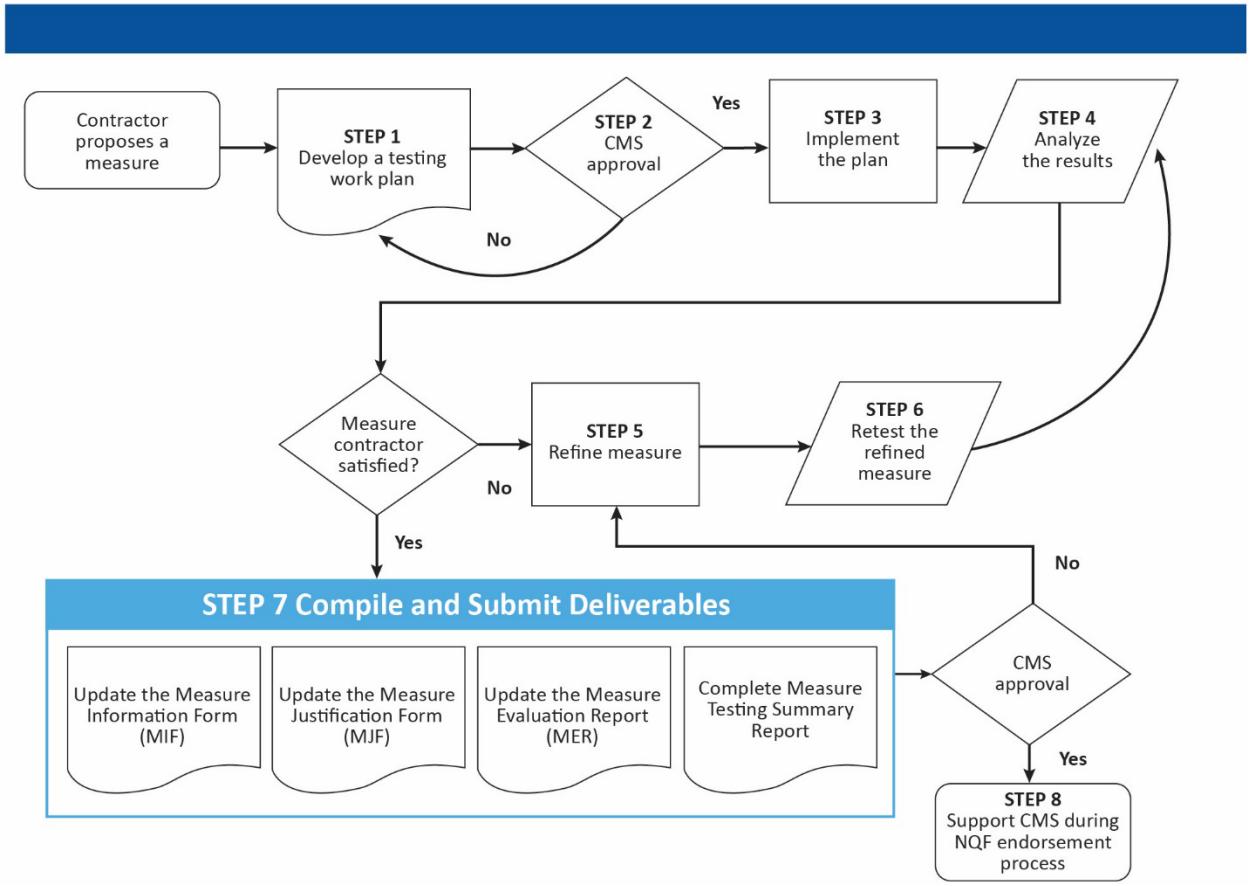

Figure 12. Overview of Measure Testing

3.2 Stakeholder Engagement

It is often appropriate to obtain stakeholder inputs at several points during the testing process. This includes obtaining face validity inputs at alpha testing, feasibility and burden inputs at beta testing, and other inputs bases 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 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 operationalized the inputs in a way consistent with the stakeholders' needs. It also improves the likelihood that stakeholders will remain engaged for ongoing support on current or future measures.

4 MEASURE IMPLEMENTATION

Quality measure implementation includes all activities associated with taking a measure from a development state to an active, in-use state. This includes but is not limited to consensus endorsement processes, measure selection processes, and measure rollout. CMS identifies and selects measures it is considering through a transparent process that is open to stakeholders and public comment. CMS adopted a set of criteria to ensure a consistent approach. When considering a measure for a topic already measured in another program, CMS prefers to use the same measure or a harmonized measure.

4.1 PRE-RULEMAKING

Depending on the CMS program, there are different paths that a measure can take for implementation. One path is through the pre-rulemaking and rulemaking process. The CMS programs involved in pre-rulemaking include those identified under ACA 3014. Measures for these programs are submitted to the Measures Under Consideration (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 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,[24] 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 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.

4.2 OTHER SUBMISSIONS

Some measures or measure programs go through neither the pre-rulemaking nor the 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:

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

Some measure programs have their own submission processes. Measure developers should check the relevant program's requirements for additional guidance.

[24] National Quality Forum. *Measuring Performance: Measure Applications Partnership.* Available at: https://qualityforum.org/map/. Accessed on: March 14, 2016.

4.3 Stakeholder Engagement

Measures under consideration 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.

4.4 The Implementation Process

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

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

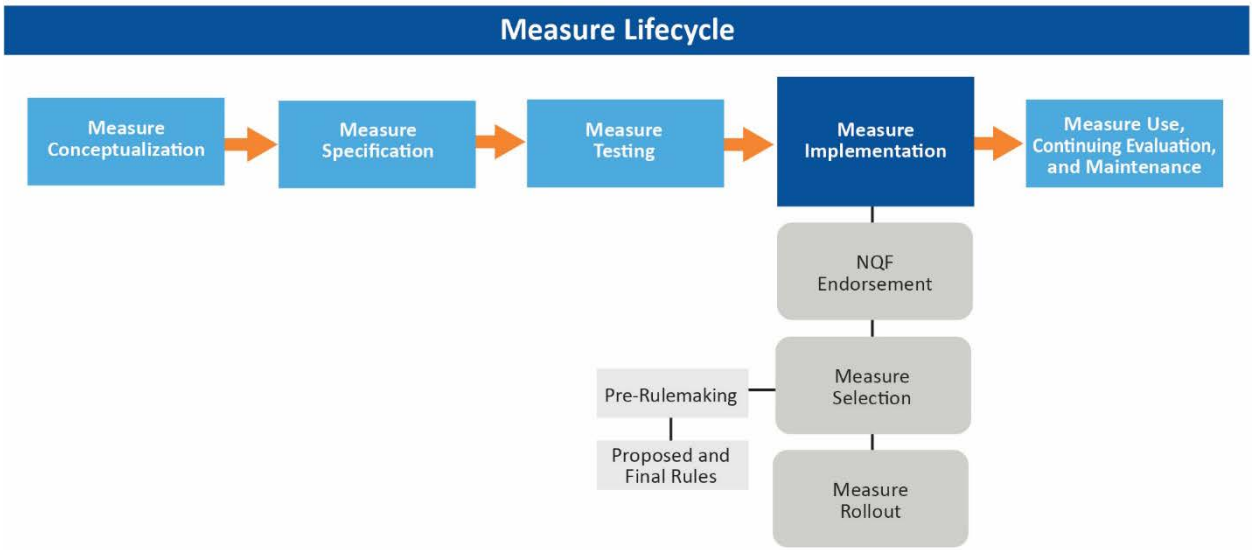

Figure 13. Flow of the Measure Lifecycle—Measure Implementation

The process of implementing measures 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
  - 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 measure implementation deliverables.[25] For detailed information on measure implementation phases, see Section 3, Chapters 27, 28, and 29 (NQF Endorsement and Maintenance, Measure Selection, and Measure Rollout, respectively).

Measure Implementation Deliverables

1. Public Description of Quality Measures
2. Timeline for Data Item and/or Quality Measure Implementation
3. Implementation Stakeholder Meetings
4. Questions and Answers Support
5. Implementation Process Roadmap
6. Measure Calculations/Results
8. Compare Sites Files and Measures as applicable)
7. Pre-Posting Preview Results
9. Implementation Algorithm (also called Calculation Algorithm/Measure Logic)
10. Data Use Agreement

Figure 14. Measure Implementation Deliverables

[25] A master list of all deliverables associated with the measure development process is in Section 3, Chapter 8.

5 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.

5.1 Continuing Evaluation

The measure developer uses the continuing evaluation process to update the Measure Justification Form 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 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 particular 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.

5.2 Measure Maintenance

As depicted in Figure 15, there are multiple steps to measure maintenance. These steps, known collectively as measure production and monitoring, 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. Deliverables associated with measure maintenance are listed in Figure 16.[26]

Measure Maintenance Deliverables

- Audit and validation reports
- Audit and validation appeals reports
- Preview reports, if required by the CMS program using the measure
- Periodic measure rate trend reports
- Analysis of the measure results
- Ad hoc analyses, as requested by CMS
- Questions and answers support
- Periodic environmental scans
- Data files of the measure rates and/or demographic information suitable for posting on CMS website
- Updated Measure Justification Form, as necessary

Figure 15. Measure Maintenance Deliverables

[26] A master list of all deliverables associated with the measure development process is in Section 3, Chapter 7.

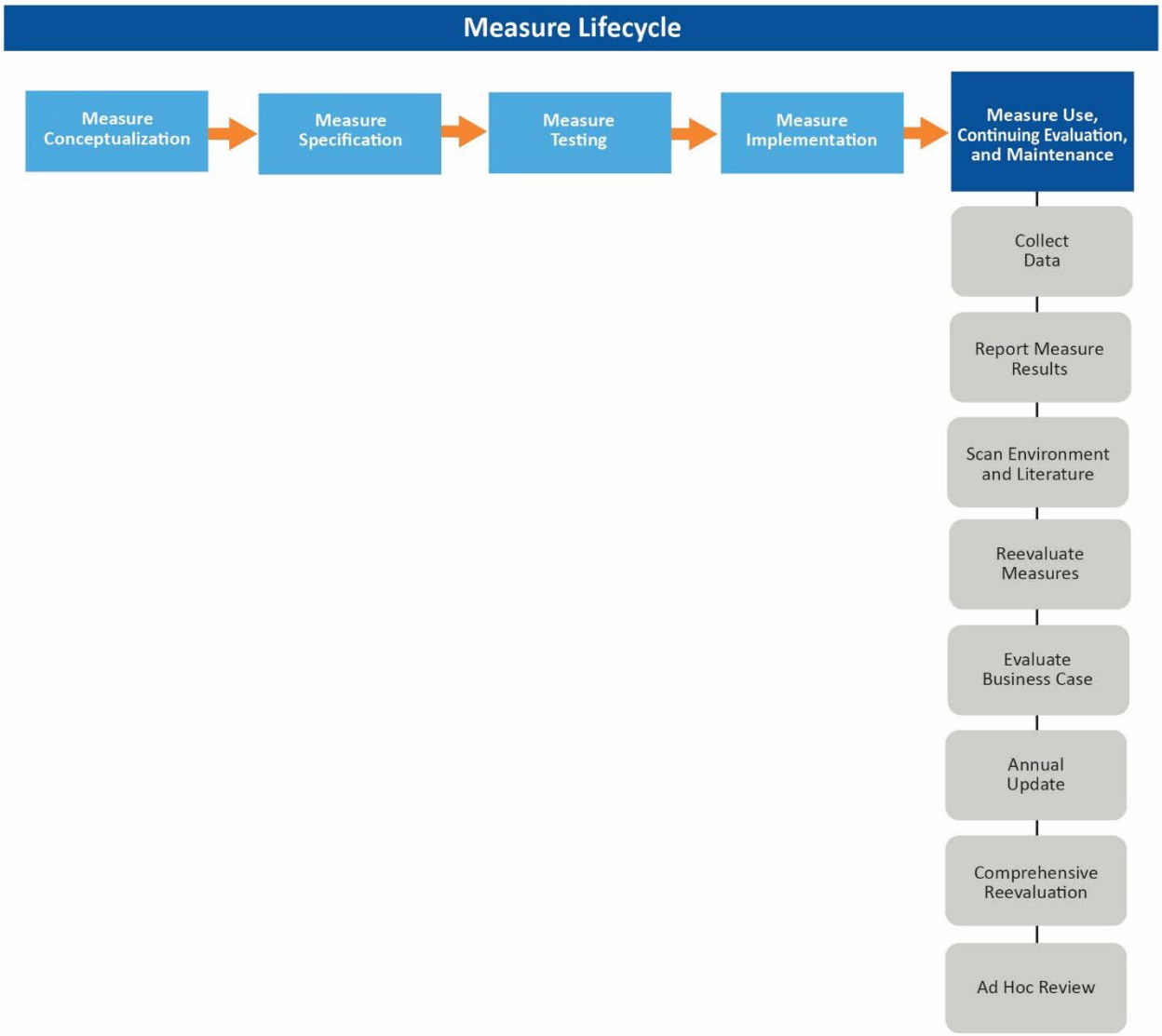

Figure 16. Flow of the Measure Lifecycle—Measure Use, Continuing Evaluation, and Maintenance

5.2.1 Measure Production and Monitoring

The following steps are involved in the continuous production and monitoring of implemented measures:

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

Details on these steps can be found in Section 3, Chapter 30, Measure Production and Monitoring.

5.2.2 Measure Maintenance Reviews

The following three types of maintenance reviews are described in Section 3, Chapter 31, 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, see Section 3, Chapter 18, Measure Harmonization, and Section 3, Chapter 24, Measure Evaluation.

5.2.3 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 of these 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, 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, for various reasons, these schedules may not align completely.

Section 3. In-Depth Topics

1 HEALTH CARE QUALITY STRATEGIES

1.1 NATIONAL QUALITY STRATEGY

CMS selects measures to develop and implement based on several key inputs, most important of which is the National Strategy for Quality Improvement in Health Care (NQS). The Patient Protection and Affordable Care Act of 2010, commonly called the Affordable Care Act (ACA), seeks to increase access to high-quality, affordable healthcare for all Americans. Section 3011 of the ACA requires the Secretary of the Department of HHS to establish a NQS that sets priorities to guide this effort and includes a strategic plan for how to achieve it. The initial NQS established three aims and six priorities for quality improvement.

The three aims are:

- Better care
- Healthy people/healthy communities
- Smarter spending.

The six priorities are:

- Making care safer by reducing harm caused in the delivery of care
- Ensuring that each person and his or her family (caregivers) are engaged as partners in their care
- Promoting effective communication and coordination of care
- Promoting the most effective prevention and treatment practices for the leading causes of mortality, starting with cardiovascular disease
- Working with communities to promote wide use of best practices to enable healthy living
- Making quality care more affordable for individuals, families, employers, and governments by developing and spreading new healthcare delivery models.[27]

The NQS has produced several reports since its first in April 2012:

- The first annual progress report to Congress elaborated on the six priorities and established long-term goals and national tracking measures to monitor quality improvement progress.
- The second annual progress report to Congress, published in July 2013, updated results of public and private payers' collaborative efforts to align their quality measures' progress against the national tracking measures.
- The third annual progress report to Congress, published in September 2014, featured the NQS Priorities in Action, which highlighted some of the promising and transformative quality improvement programs at the federal, state, and local levels. The report also analyzed the current state of quality measurement, with consideration for the continued need for harmonization and alignment.
- The fourth annual progress report to Congress, published in October 2015, reported significant progress on the NQS priorities, backed by data published annually by the National Healthcare Quality and Disparities Report, an Agency for Healthcare Research and Quality (AHRQ) publication. The report noted that such progress was supported through alignment to the NQS aims and priorities.

[27] Department of Health and Human Services, Agency for Health Research and Quality, National Quality Strategy. *Working For Quality*. Available at: http://www.ahrq.gov/workingforquality/. Accessed on: May 6, 2015.

- The fifth annual progress report to Congress, published in April 2016, integrated into one document the NQS with the National Health Care Quality and Disparities Report. The key findings from this combined report included a dramatic rise in access to care due to greater availability of health insurance and an increased number of Americans with a usual source of healthcare. While disparities persist, the report asserted that quality of healthcare continues to improve overall, most notably in effective treatment, while care coordination continues to lag.

1.2 CMS Quality Strategy

The CMS Quality Strategy pursues and aligns with the three broad aims of the NQS as well as Delivery System Reform payment goals. Like the NQS, the CMS Quality Strategy was developed through a participatory, transparent, and collaborative process that included the input of a wide array of stakeholders. CMS's vision is to optimize health outcomes by improving clinical quality and transforming the health system.

Each of the six NQS priorities has become a goal in the CMS Quality Strategy. In addition, CMS has developed four foundational principles: eliminate racial and ethnic disparities, strengthen infrastructure and data systems, enable local innovations, and foster learning organizations.

CMS supports these 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 in 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."[28]

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 include organizations such as Quality Innovation Network-Quality Improvement Organizations (QIN-QIOs),[29] university researchers, health services research organizations, and consulting groups. The Measures Manager supports the CMS CORs and their various measure developers in their work implementing the MMS.

[28] Conway PH, Mostashari F, Clancy C. The future of quality measurement for improvement and accountability. *JAMA.* 2013;309(21):2215-2216. doi:10.1001/jama.2013.4929. Accessed on: March 15, 2017.
[29] Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityImprovementOrgs/index.html?redirect=/QualityImprovementOrgs. Accessed on: November 7, 2015.

1.3 Crosswalk Between the NQS and the CMS Quality Strategy

Table 2 compares the NQS with the goals and objectives of the CMS Quality Strategy.[30]

Table 2. NQS and CMS Quality Strategy

NQS Priorities	CMS Quality Strategy Goals and Objectives
1. Making care safer by reducing the harm caused in the delivery of care.	Goal 1: Make care safer by reducing harm caused in the delivery of care. • Improve support for a culture of safety. • Reduce inappropriate and unnecessary care. • Prevent or minimize harm in all settings.
2. Ensuring that each person and family (caregiver) are engaged as partners in their care.	Goal 2: Strengthen person and family (caregiver) engagement as partners in their care. • Ensure all care delivery incorporates patient and caregiver preferences. • Improve experience of care for patients, caregivers, and families. • Promote patient self-management.
3. Promoting effective communication and coordination of care.	Goal 3: Promote effective communication and coordination of care. • Reduce admissions and readmissions. • Embed best practices to manage transitions to all practice settings. • Enable effective healthcare system navigation.
4. Promoting the most effective prevention and treatment practices for the leading causes of mortality, starting with cardiovascular disease.	Goal 4: Promote effective prevention and treatment of chronic disease. • Increase appropriate use of screening and prevention services. • Strengthen interventions to prevent heart attacks and strokes. • Improve quality of care for patients with multiple chronic conditions (MCC). • Improve behavioral health access and quality care. • Improve perinatal outcomes.
5. Working with communities to promote wide use of best practices to enable healthy living.	Goal 5: Work with communities to promote best practices of healthy living. • Partner with and support federal, state, and local public health improvement efforts. • Improve access within communities to best practices of healthy living. • Promote evidence-based community interventions to prevent and treat chronic disease. • Increase use of community-based social services support.
6. Making quality care more affordable for individuals, families, employers, and governments by developing and spreading new healthcare delivery models.	Goal 6: Make care affordable. • Develop and implement payment systems that reward value over volume. • Use cost analysis data to inform payment policies.

[30] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *CMS Quality Strategy 2013 — Beyond*. Nov 11, 2013. Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy.html. Accessed on: November 7, 2015.

2 PRIORITIES PLANNING

CMS responds to a variety of inputs to develop and implement its quality measurement agenda for the next 5–10 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 CMS Quality Strategy. CMS places emphasis on electronically specified measures for implementation in quality initiatives. These 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 CMS Quality Strategy and its alignment with the NQS aims and priorities,[31] with an emphasis on process, access, outcome, structure, and patient experience measures.[32] These focus areas drive measure development, selection, and implementation activities. In addition, CMS intends to "optimize health outcomes by leading clinical quality improvement and health system transformation."[33] CMS also sets priorities based on inputs from the National Impact Assessment reports. Though 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.

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, producing "Families of Measures" in areas such as safety, care coordination, cardiovascular conditions, diabetes, and dual eligible beneficiaries. In addition, the CMS commitment to the NQS and CMS Quality Strategy in measure development has contributed to significant improvement in this area—closing gaps and generating streamlined measure sets. The recent Institute of Medicine (IOM) Vital Signs report[34] and the 2015 Impact Report,[35] "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 17 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, 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

[31] Department of Health and Human Services, Agency for Health Research and Quality, National Quality Strategy. Working for Quality. Available at: http://www.ahrq.gov/workingforquality/. Accessed on: November 7, 2016.
[32] Available at https://www.qualitymeasures.ahrq.gov/about/domain-definitions.aspx.
[33] Department of Health and Human Services, Centers for Medicare & Medicaid Services. CMS Quality Strategy 2013–Beyond. Nov 11, 2013. Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy.html. Accessed on: November 7, 2016.
[34] Institute of Medicine of the National Academies. *Vital Signs Core Metrics for Health and Health Care Progress.* Available at: http://iom.nationalacademies.org/Reports/2015/Vital-Signs-Core-Metrics.aspx. Accessed on March 1, 2016.
[35] Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/National-Impact-Assessment-of-the-Centers-for-Medicare-and-Medicaid-Services-CMS-Quality-Measures-Reports.html. Accessed on November 7, 2016.

Medicare programs. The MUC list is made available to the public no later than December 1st of every year.

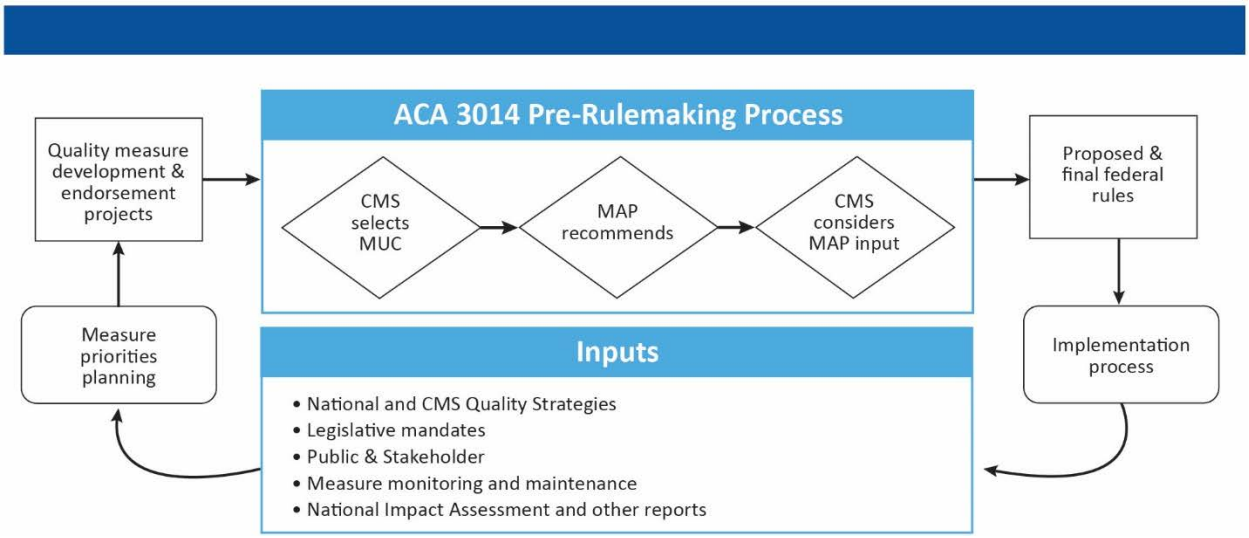

Figure 17. CMS Priorities Planning and Measure Selection

Around the second quarter of each federal year, through a call for quality and efficiency measures, CMS begins the annual pre-rulemaking cycle of collecting and compiling the MUC list. In late April or early May, 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. The NQF, the consensus-based entity with a current contract under Section 1890 of the Act, 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 the following four workgroups: Clinicians, Post-Acute Care/Long-Term Care, Hospitals, and Dual Eligible Beneficiaries. Annually, the MAP workgroups and the Coordinating Committee meet to provide program-specific recommendations to HHS by February 1st. Measure developers are strongly encouraged to attend the MAP. Participation by measure developers leads to the MAP making decisions using complete and accurate information on MUC. United States Code provides general requirements for informal rulemaking.[36]

2.2 CMS Measure Planning Inputs

2.2.1 NQS and CMS Quality Strategy

The NQS 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 helps prioritize quality improvement efforts, share lessons, and measure collective successes.

[36] https://www2.gwu.edu/~clai/recent_events/2010/Summer_Regulatory_Program/June_2010_Reg_Program_Presentations/June_2010_IFC_Reg_Map.pdf

Section 3011 of the ACA requires this strategic plan to include:

- Coordination among HHS agencies and other federal partners
- HHS agency-specific strategic plans
- A process for reporting measure performance and activities
- Benchmarks for measure results
- Strategies to align public and private payers
- Incorporation of quality measurement and improvement into Health Information Technology (HIT).

The CMS Quality Strategy[37] is built on the foundation of the NQS, as described in the previous chapter. Based on this foundation, CMS makes strategic choices in measure development and maintenance contracts as well as targeted measure selection for programs.

2.2.2 Legislative Mandates

CMS uses priorities mandated under several laws to define its measure domains.[38] Recently, Congress passed and the President of the United States signed the Medicare Access and CHIP Reauthorization Act (MACRA) of 2015 (P.L. 114-10). This Act 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 patient reported outcome and functional status measures; (b) patient experience measures; (c) care coordination measures; and (d) measures of appropriate use of services, including measures of overuse.

Specifically, MACRA, ACA, and the American Recovery and Reinvestment Act of 2009 (ARRA) have the largest influence on CMS's quality measurement priorities, which have led to the broad payment reform and quality-based payment models.[39] With MACRA in 2015, Congress mandated that several quality reporting incentive programs will phase out in 2018, while the Merit-based Incentive Payment System (MIPS) will continue well beyond 2019. Under MACRA, CMS will develop 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 of 2010 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 Systems-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 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.

[37] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *CMS Quality Strategy*. Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/CMS-Quality-Strategy.pdf. Accessed on: February 26, 2016.

[38] Note that CMS and other HHS agencies define and use the term "domain" differently from one another. Therefore, within the Blueprint, you will see the term "domain" defined differently in different contexts, depending on the relevant agency within the discussion.

[39] Available at: http://www.hhs.gov/news/press/2015pres/01/20150126a.html

2.2.3 Quality Measure Development Plan

On May 2, 2016, CMS finalized the Quality MDP, mandated under the MACRA, to support the new MIPS and advanced alternative payment models (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 more information, please view the CMS Quality MDP and Annual Report.

2.2.4 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 1, 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, procedures, etc., 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 very important stakeholders in the quality measurement enterprise, and CMS has committed to gather their input during priorities planning. More detail about ways the patient's voice can be included is found in Section 2, Chapter 1, Measure Conceptualization.

2.2.5 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. CMS uses this input to decide whether to remove, retire, or retain measures in use.

In addition to measure maintenance, CMS conducts various evaluations and assessments of its measures and programs. CMS conducts evaluations 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.

A variety of 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 are:

- Medicare Payment Advisory Committee (MedPAC) and Medicaid and CHIP Payment and Access Commission (MacPAC) quality reports
- AHRQ—National Healthcare Quality and Disparities Reports
- CMS Center for Strategic Planning—Chronic Conditions among Medicare Beneficiaries
- Universities, researchers, and healthcare facilities—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 the 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 measures under development. Section 2, Chapter 1, 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. This 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 CMS Quality Strategy 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 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 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.

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 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. This may include:

- Review HHS and CMS strategic plans, goals, and initiatives.
- Monitor the progress of CMS measure development and maintenance projects against the CMS Quality Strategy and identify 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 workgroups 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 six categories:[40]
    - Proposed – A measure proposed for use within a CMS Program via a Federal Rule.
    - Rescinded – The proposal to incorporate a measure into a program has been rescinded via Federal Rule. The measure will not be finalized or implemented.
    - Finalized – The proposal to incorporate a measure into a CMS program has been finalized per Federal Rule. The measure will be implemented within a designated timeframe.
    - Implemented – A measure that is finalized and currently used within a CMS program.
    - Suspended – A finalized measure that has been removed from current use within a program. The measure is no longer implemented.
    - Removed – A measure that has been removed from a CMS program via Federal Rule. The measure is no longer implemented.

[40] Measures can be categorized many ways, such as by National Quality Strategy Priority, by type (structure, process, outcome, etc.) by data source, by setting of care, or by level of analysis.

3 MEASURE GOVERNANCE

Measure developers have distinct roles and responsibilities of governance 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 DEVELOPERS

Measure developers create, edit, and submit measures to a designated steward, in this case, CMS. Stewards approve, reject, and publish submitted measures from developers. Developers submit measures to their assigned stewards for approval, and stewards withdraw measures from 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.

3.2 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 specific purpose. Stewards are responsible for approving measure content.

4 MEASURE CLASSIFICATION

Measures may be classified according to a variety of schemes, including measurement domain,[41] by the NQS priority(ies) addressed, and by measurement setting. Elements of these classification schemes and examples are provided in Tables 3 and 4. A list of measurement settings is shown in Table 5.

Table 3. National Quality Measures Clearinghouse (NQMC) Clinical Quality Measure (CQM) Domains[42]

Measurement Domain	Definition	Example
Process	A process of care is a healthcare-related activity performed for, on behalf of, or by a patient. Process measures are supported by evidence that the clinical process—that is the focus of the measure—has led to improved outcomes. These measures are generally calculated using patients eligible for a service in the denominator, and the patients who either do or do not receive the service in the numerator.	The percentage of patients with chronic stable coronary artery disease (CAD) who were prescribed lipid-lowering therapy.
Access	Access to care is the attainment of timely and appropriate healthcare by patients or enrollees of a healthcare organization or clinician. Access measures are supported by evidence that an association exists between the measure and the outcomes of or satisfaction with care.	The percentage of members 12 months to 19 years of age who had a visit with a primary care practitioner in the past year (based on evidence that annual visits lead to better health outcomes for children and youth).
Outcome	An outcome of care is a health state of a patient resulting from healthcare. Outcome measures are supported by evidence that the measure has been used to detect the impact of one or more clinical interventions. Measures in this domain are attributable to antecedent healthcare and should include provisions for risk adjustment.	The risk-adjusted rate of in-hospital hip fracture among acute care inpatients aged 65 years and over, per 1,000 discharges.
Structure	Structure of care is a feature of a healthcare organization or clinician related to the capacity to provide high-quality healthcare. Structure measures are supported by evidence that an association exists between the measure and one of the other CQM domains.	Does the healthcare organization use Computerized Physician Order Entry (CPOE) (based on evidence that the presence of CPOE is associated with better performance and lower rates of medication error)?
Patient Experience	Experience of care is a patient's or enrollee's report of observations of and participation in healthcare, or assessment of any resulting change in their health. Patient experience measures are supported by evidence that an association exists between the measure and patients' values and preferences, or one of the other clinical quality domains. These measures may consist of rates or mean scores from patient surveys.	The percentage of adult inpatients that reported how often their doctors communicated well.

[41] Note that CMS and other HHS agencies define and use the term "domain" differently from one another. Therefore, within this Blueprint, you will see the term "domain" defined differently in different contexts, depending on the relevant agency within the discussion.
[42] National Quality Measures Clearinghouse. http://www.qualitymeasures.ahrq.gov/about/domain-framework.aspx. Website accessed 10/21/2015.

Table 4. Examples of Measures Addressing Each of the NQS Priorities[43]

NQS Priority	Examples of Measures Addressing Each NQS Priority
Making care safer by reducing harm caused in the delivery of care	Acute care prevention of falls: rate of inpatient falls per 1,000 patient days NQF 2726 Prevention of Central Venous Catheter (CVC)-Related Bloodstream Infections Hospital acquired conditions: standardized infection ratio of CLABSI, Catheter-associated Urinary Tract Infections (CAUTI), and Surgical Site Infection (SSI)
Ensuring that each person and family is engaged as partners in their care	NQF 1919 Cultural Competency Implementation Measure NQF 0310 Back Pain: Shared Decision Making Behavioral healthcare patients' experiences: percentage of adult patients who reported whether they were provided information about treatment options
Promoting effective communication and coordination of care	Adult depression in primary care: percentage of patients who have a follow-up contact within 3 months of diagnosis or initiating treatment Care for older adults: percentage of adults 66 years and older who had a medication review during the measurement year NQF 0243 Stroke and Stroke Rehabilitation: Screening for Dysphagia
Promoting the most effective prevention and treatment practices for the leading causes of mortality, starting with cardiovascular disease	Acute myocardial infarction (AMI)/chest pain: percentage of emergency department (ED) patients with AMI or chest pain who received aspirin within 24 hours before ED arrival or prior to transfer Adult trauma care: percentage of patients age 18 years and older admitted to hospital with an injury diagnosis and Deep Venous Thrombosis prophylaxis prescribed within 24 hours of hospital admission
Working with communities to promote wide use of best practices to enable healthy living	Annual dental visit: percentage of members 2 to 21 years of age who had at least one dental visit during the measurement year NQF 0039 Flu Vaccinations for Adults Ages 18 and Older
Making quality care more affordable for individuals, families, employers and governments by developing and spreading new healthcare delivery models	NQF 1604 Total Cost of Care Population-based Per-member Per-month Index Imaging efficiency: percentage of brain CT studies with a simultaneous sinus CT

[43] https://www.nap.edu/catalog/13059/finding-what-works-in-health-care-standards-for-systematic-reviews

Table 5. Measurement Settings[44]

Measurement Settings
• Accountable Care Organizations • Ambulatory Procedure/Imaging Center • Ambulatory/Office-based Care • Ancillary Services • Assisted Living Facilities • Behavioral Health Care • Community Health Care • Emergency Department • Emergency Medical Services • Home Care • Hospices • Hospital – Other • Hospital Inpatient • Hospital Outpatient • Intensive Care Units • Long-term Care Facilities • Managed Care Plans • National Public Health Programs • Patient-centered Medical Homes • Regional, County, or City Public Health Programs • Rehabilitation Centers • Residential Care Facilities • Rural Health Care • Skilled Nursing Facilities/Nursing Homes • State/Provincial Public Health Programs • Substance Abuse Treatment Programs/Centers • Transitions of Care

[44] http://www.qualitymeasures.ahrq.gov/browse/by-national-quality-strategy.aspx; Accessed 10/23/2015

5 SELECTED MEASURE TYPES

5.1 COST AND RESOURCE USE MEASURES

The CMS strategic goals for 2013–2017 include providing better care and lower costs of care for all Americans.[45] 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—These measures are 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—These are measures of the 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:[46] safety, timeliness, effectiveness, efficiency, equity, and patient centeredness.
- Efficiency—This term is associated with measuring cost of care associated with a specified level of quality of care.
- Value of care—This type of measure 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.

As a country with high healthcare costs, but poorer than expected health outcomes relative to many parts of the world,[47] the challenge for CMS 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 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-

[45] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *CMS Strategy: The Road Forward, 2013-2017.*Mar 2013. Available at: http://www.cms.gov/About-CMS/Agency-Information/CMS-Strategy/Downloads/CMS-Strategy.pdf. Accessed on: March 14, 2016.

[46] Crossing the Quality Chasm: A New Health System for the 21st Century. Committee on Quality of Health Care in America, Institute of Medicine, Washington, DC, USA: National Academies Press; 2001.

[45]OECD report - OECD (2015), Health at a Glance 2015: OECD Indicators, OECD Publishing, Paris.
DOI: http://dx.doi.org/10.1787/health_glance-2015-en

centeredness."[48] NQF uses Figure 18 (adapted) to illustrate the relationship between resource use, efficiency, and value.

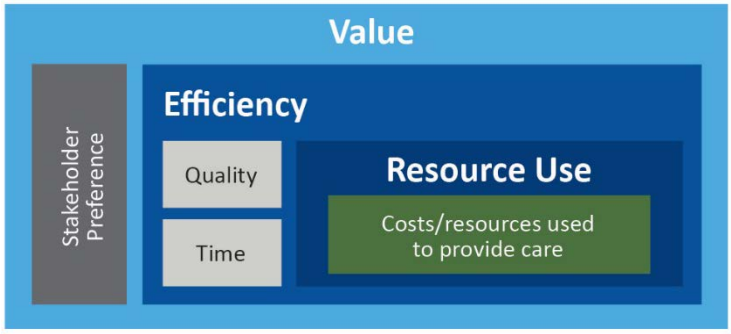

Figure 18. Relationship Between Resource Use, Value, and Efficiency

Cost and resource use measures must be linked to quality outcomes as well as to the processes that are required to achieve those outcomes. Consider ways those types of 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

NQF defines a composite performance measure as "a combination of two or more individual performance measures in a single performance measure that results in a single score."[49] 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.

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.

[48] National Quality Forum. *Measurement Framework: Evaluating Efficiency Across Patient-Focused Episodes of Care*. Washington, DC: Available at http://www.qualityforum.org/Publications/2010/01/Measurement_Framework__Evaluating_Efficiency_Across_Patient-Focused_Episodes_of_Care.aspx. National Quality Forum; Washington, DC; Jan 2010. Accessed on: March 14, 2016.

[49] National Quality Forum. *Composite Performance Measure Evaluation Guidance*. Washington, DC: National Quality Forum; Apr 2013; Contract No. HHSM-500-2009-00010C. Available at: http://www.qualityforum.org/Publications/2013/04/Composite_Performance_Measure_Evaluation_Guidance.aspx. Accessed on: March 14, 2016.

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).[50]

Composite performance measure development should follow these principles:[51]

- 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.[52]

5.2.2 Component Performance Measures

The following are some considerations for selecting measures to be included in a composite:[53]

- 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

[50] Ibid.
[51] Peterson ED, Delong ER, Masoudi FA, et al. ACCF/AHA 2010 Position Statement on Composite Measures for Healthcare Performance Assessment: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures (Writing Committee to develop a position statement on composite measures). *Circulation.* 2010;121(15):1780-1791.
[52] The Physician Consortium for Performance Improvement® Convened by the American Medical Association Measures Development, Methodology, and Oversight Advisory Committee: Recommendations to PCPI Work Groups on Composite Measures Approved by the PCPI in Dec 2010.
[53] National Quality Forum. *Composite Performance Measure Evaluation Guidance*. Washington, DC: National Quality Forum; Apr 2013; Contract No. HHSM-500-2009-00010C. Available at: http://www.qualityforum.org/Publications/2013/04/Composite_Performance_Measure_Evaluation_Guidance.aspx. 2013. Accessed on: March 14, 2016.

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.

Composite measures are not yet available as electronic clinical quality measures (eCQMs).

5.3 Patient-Reported Outcome (PRO) Measures

PRO measures are quality measures that are derived from outcomes reported by patients. These measures present some design challenges that are described below, with some approaches to those challenges.

Ensuring that patients and families are engaged as partners in their care, one of the NQS priorities, can also be an effective way to measure the quality of patient care. Though patient reports of their health and experience with care are not the only outcomes that should be measured, they certainly are an important component. Patient experience and satisfaction with care has been measured historically, but the infrastructure to collect other PROs and use them in quality and accountability programs is still under construction. Tools to collect these data (e.g., PROMIS tools) have mostly been used in academic settings and are being tested for clinical application.

5.3.1 Patient-reported Outcomes

PROs are "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."[54] Self-reported patient data provide a rich data source for outcomes. This definition reflects the key domains listed in the NQF report on Patient-Reported Outcome-Based Performance Measurements (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)[55]

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 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

[54] United States Food and Drug Administration. *Guidance for Industry, Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. Federal Register. 2009;74(35):65132-133. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed on: March 14, 2016.
[55] National Quality Forum. *Patient-Reported Outcomes (PROs) in Performance Measurement*. Jan 2013. Available at: https://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx. Accessed on: March 14, 2016.

reporting, and health improvement. All managed care plans with Medicare Advantage contracts must participate.

- Focus On Therapeutic Outcomes, Inc. (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.[56]

5.3.4 Approaches to Developing Patient-reported Outcome-based Performance Measures

Though PROs are a special type of outcome measure, the principles for development are the same. Patient-reported outcome-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 also must 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

[56] National Quality Forum. Fast Forward Briefs on New Work by NQF: *Creating Valid and Reliable Patient-Reported Outcome Measures.* National Quality Forum. 2013; Issue No. 1. 2013. Available at: http://www.qualityforum.org/Publications/2013/04/Fast_Forward__Creating_Valid_and_Reliable_Patient-Reported_Outcome_Measures.aspx. Accessed on: March 14, 2016.

validity. With further testing in clinical settings, they can be used for PROMs. 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 measures under development must be evaluated. Detailed specifications must be submitted, using the MIF.[57]

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, *Patient-Reported Outcomes (PROs) in Performance Measurement*.[58] Documentation of these items should be submitted to the COR as specified in the contract.

[57] The NQF submission may be acceptable for this deliverable. National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: March 14, 2016.
[58] National Quality Forum. *Patient-Reported Outcomes (PROs) in Performance Measurement*. Jan 2013. Available at: https://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx. Accessed on: March 14, 2016.

5.4 Multiple Chronic Conditions Measures

In an article published in 2013, the Centers for Disease Control and Prevention (CDC) stated that 68.4 percent of Medicare beneficiaries had two or more chronic conditions.[59] Though the numbers vary by counting methods, the CDC also found the prevalence of MCC in all adults had increased from 21.8 percent in 2001 to 26.0 percent in 2010.[60] These individuals constitute a particular challenge to the healthcare system because their conditions complicate each other, are ongoing, and are very 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.[61] In 2011, CMS found that Medicare beneficiaries with MCC were the heaviest users of healthcare services.[62] For those with six or more chronic conditions, two-thirds were hospitalized during 2010 and they accounted for about half of Medicare spending on hospitalizations.[63] However, very few measures exist that are specifically designed to evaluate the quality of care provided to these people.[64]

5.4.1 MCC Definition

HHS recently contracted with NQF to develop a measurement framework for persons with MCC. The NQF MCC Measurement Framework defined MCC as follows:

Persons with MCC are defined 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.[65]

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.

Importantly, 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.[66]

[59] Lochner KA, Cox CS. Prevalence of multiple chronic conditions among Medicare beneficiaries, United States, 2010. *Preventing Chronic Disease;* Centers for Disease Control.2013;10:120137. Available at: http://dx.doi.org/10.5888/pcd10.120137. Accessed on: March 14, 2016.
[60] Ward BW, Schiller JS. Prevalence of multiple chronic conditions among US adults: estimates from the National Health Interview Survey, 2010. *Preventing Chronic Disease;* Centers for Disease Control. 2013;10:120203. Available at: http://dx.doi.org/10.5888/pcd10.120203. Accessed on: March 14, 2016.
[61] Tinetti ME, McAvay GJ, Chang SS, et al. Contribution of multiple chronic conditions to universal health outcomes. *Journal of the American Geriatric Society.* 2011; 59(9):1686–91. doi: 10.1111/j.1532-5415.2011.03573.x.
[62] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Chronic Conditions among Medicare Beneficiaries, Chart Book: 2012 Edition.* Centers for Medicare & Medicaid Services; Baltimore, MD. 2012. Available at: http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/2012ChartBook.html. Accessed on: March 14, 2016.
[63] Ibid.
[64] National Quality Forum. *Multiple Chronic Conditions Measurement Framework.* National Quality Forum; May 2012. Available at: http://www.qualityforum.org/Projects/Multiple_Chronic_Conditions_Measurement_Framework.aspx. Accessed on: March 14, 2016.
[65] Ibid.
[66] Ibid.

5.4.2 Need for Measure Development

Though 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 their needs.[67] Current quality measures are largely based on performance standards derived from clinical practice guidelines for management of specific diseases.[68] 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.[69] Few measures exist to evaluate inappropriate care in these situations.

5.4.3 Considerations for Measure Development Targeting Persons with MCC

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 do exist that are especially pertinent to people with MCC. The following measurement concepts were identified as having potential for high-leverage in quality improvement for patients with MCC:[70]

- 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

The NQF Framework identified that quality measures for persons with MCC should be guided by several principles. Quality measures should:[71]

- 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

[67] Ibid.
[68] Tinetti ME, Bogardus ST, Agostini JV. Potential pitfalls of disease-specific guidelines for patients with multiple conditions. *New England Journal of Medicine.* 2004;351:2870–4.
[69] Ibid.
[70] Ibid.
[71] Ibid.

- 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 very nature of chronic conditions requires observation over time. Especially in the case of outcome measures for patients with MCCs, it is very 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, 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 also mean 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 see 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 MCC. Risk adjustment should be used with caution in the situation of MCC. 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 difficult 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 difficult 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. It is important for measure developers to standardize data collection methods.

5.4.3.3 When Testing and Evaluating Measures for Persons with MCC

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. 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. Other situations when a measure developer may require additional guidance may exist that are not covered here. For those situations, contact the Measures Manager and the appropriate COR.

6 INTRODUCTION TO ELECTRONIC CLINICAL QUALITY MEASURES (ECQMS)

Collecting and reporting accurate healthcare performance data has historically been a complex and time-consuming manual process. To limit the need for extensive record reviews, 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 collected through EHRs an essential tool for implementing the CMS Quality Strategy by transitioning the measuring and public reporting of providers' quality performance using EHR data.

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 Health Quality Measure Format (HQMF) provides for quality measure consistency and unambiguous interpretation. HQMF is a component of a larger quality end-to-end framework evolving to a time when providers ideally will be able to provide the most appropriate care using clinical decision support and evaluate their own performance using the eCQMs imported into their EHRs. The HQMF Release 1 (R1) was published as a standard for trial use (STU) in 2009[72] and is the underlying structured representation used by the CMS Measure Authoring Tool (MAT) for eCQMs developed through June 2014. HQMF was updated to STU Release 2.1 in 2015 and updated to a normative standard in January 2017.

The representation of an eCQM is simplified and standardized when measure developers author their eCQM in the MAT. The HQMF was 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 HIT 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 following:

- The Blueprint
- The MAT User Guide. The MAT is a web-based tool that allows measure developers to author eCQMs in HQMF using the Quality Data Model (QDM) and healthcare industry standard vocabularies. Measure developers should consult with their COR if they have questions about using the MAT software.
- QDM. The QDM is an information model that defines clinical concepts in a standardized format to enable electronic quality performance measurement.
- The Value Set Authority Center. The Value Set Authority Center (VSAC) is provided by the National Library of Medicine (NLM), in collaboration with the Office of the National Coordinator for Health Information Technology (ONC) and CMS. Requiring a free Unified Medical Language System (UMLS) license, the VSAC provides searchable and downloadable access to all official

[72] Announcement press release available at: http://www.hl7.org/documentcenter/public_temp_116FD0C6-1C23-BA17-0C50779726E5BD29/pressreleases/HL7_PRESS_20090827.pdf. Accessed December 7, 2016.

versions of value sets used by each of the eCQM releases used in CMS and other quality reporting programs (e.g., The Joint Commission).

- JIRA. JIRA is an issue tracking system licensed by ONC. It is a collaboration platform that supports the implementation of HIT 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 on the MAT, QDM, VSAC, and JIRA is found in Section 3, Chapter 7, eCQM Standards-Based Guidance and Tools.

6.1 eCQM Components

eCQMs are written to conform to the Health Level Seven (HL7) HQMF standard for representing a health quality measure as an electronic eXtensible Markup Language (XML) document. eCQMs are specified using patient-level information coded in a format intended for extraction from EHRs and other electronic clinical systems. 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, ratio). However, eCQMs require additional steps to map measure data elements to corresponding QDM components and standard terminologies to assemble the data criteria. An eCQM has three components:

- XML—Contains important details about the measure, how the data elements are defined, and the underlying logic of the measure calculation. The file uses HQMF XML syntax. The XML includes a Header and a Body. The Header identifies and classifies the document and provides important metadata about the measure. Appendix B 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., data criteria, population criteria, supplemental data elements).
- Hyper Text Markup Language (HTML) file (.html)—Displays the eCQM content in a human-readable format directly in a web browser. This file does not include the underlying HQMF syntax.
- Value sets—Conveys specific coded values that are allowed for the data elements within the eCQM. Value sets are identified by an object identifier (OID), and include several metadata elements. The value set includes a list of codes (i.e., the value set “expansion code set”), descriptors of those codes, the code system from which the codes are derived, and the version of that code system.

The MAT produces the following items for an eCQM:

- eCQM XML
- SimpleXML file
- eCQM HTML file
- Standalone spreadsheet that contains all the value sets referenced by the measure. Note, this spreadsheet is not the authoritative list of value set content.

6.2 Special Considerations for eCQMs

For the most part, the measure development process for eCQMs is not that much different than non-eCQMs. The measure conceptualization process is the same for eCQM as for measures developed using other data sources. However, eCQMs are based on information that should exist in a structured format in electronic clinical systems such as EHRs.[73] 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 (validity and reliability) of eCQMs is based on some unique assumptions and special considerations:

- 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 sources of error as non-eCQMs that could result in low evaluation ratings for the reliability and validity of data elements and measure scores. Careful analysis is required to avoid 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.[74] However, using these rarely updated fields for quality measurement purposes may encourage enhanced reporting, resulting in improved data quality and usefulness. Other examples of potential sources of error include:
    - Incorrect or incomplete measure specifications, including value sets, logic, or computer-readable programming language
    - 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 entries made into the wrong EHR field
    - Inaccurate interpretation of data by natural language processing 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.
- Initial data element validity for an eCQM can be evaluated based on complete agreement between data elements and computed measure scores obtained by applying the eCQM specifications to a simulated test EHR data set with known values for the critical data elements and computed measure score.

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

[73] However, data not in a structured field may be used in conjunction with natural language processing software or similar tools.
[74] eCQM specifications do not designate where (e.g., problem list) in the EHR the data should be extracted.

7 Electronic Clinical Quality Measure (eCQM) Standards-Based Guidance and Tools

eCQM specification development and maintenance has evolved into a complex process that requires input from multiple stakeholders (e.g., CMS, NLM, measure steward) as well as use of multiple standards-based guidance and tools at various stages during development and maintenance. The systems used during eCQM development and maintenance include measure analysis and information gathering tools, MAT, VSAC, testing tools (i.e., Bonnie), as well as JIRA. The standards-based guidance and tools described here apply to de novo eCQMs, respecified eCQMs, and eCQM maintenance. Aligning an eCQM with the transmission format (e.g., the interoperability specification that governs how the individual or aggregate patient data are to be communicated to CMS), known as QRDA, maximizes consistency and may lead to successful data submission.

7.1 ONC JIRA Tracking System

JIRA is a commercial product licensed by the ONC. The ONC JIRA project system is a platform that enables teams to create, capture, and organize comments and issues; develop solutions; and follow team activities for multiple HIT projects. In addition, issue tracking in JIRA enables users to search for issues that have been resolved, those currently pending, discussions between users, answers to questions, and real-time feedback to federal agencies that relate to development and release of quality standards. Therefore, JIRA serves as a collaborative environment to support the development, implementation, and maintenance of eCQMs.

eCQM developers are responsible for responding to the JIRA tickets pertaining to the eCQMs they have developed or are maintaining. Some tickets can be closed by providing simple answers, whereas some may require discussions with other developers or stakeholders. JIRA tickets that cannot be resolved with such discussions should be addressed through discussion and consensus of proposed solutions.

The function of JIRA has evolved and expanded from communication among various stakeholders on numerous HIT projects to a source of information as a central file-sharing repository for eCQM maintenance, submission of MUC for CMS, and aggregate public comments on a variety of issues. More information on JIRA can be found in the Measure Logic Document on the eligible hospital (EH) and eligible professional (EP)/eligible clinician pages of the Electronic Clinical Quality Improvement (eCQI) Resource Center and on the JIRA website.

7.2 Health Quality Measure Format (HQMF)

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. HQMF is an HL7 standard for representing a health quality measure as an electronic XML document.[75] Through standardization of a measure's structure, metadata, definitions, and logic, the HQMF provides quality measure consistency and unambiguous interpretation. HQMF is a component of a larger quality end-to-end framework evolving to a time when providers ideally will be able to provide the most appropriate care using clinical decision support and evaluate their own performance using the eCQMs imported into their electronic clinical systems. The eCQMs ideally can be turned in to queries that automatically gather data from the EHR

[75] Refer to eCQI Resource Center pages for eligible hospital and clinician measures for up to date examples of how eCQMs will appear when rendered in XML and HTML.

data repositories and generate reports for quality reporting. From there, individual and/or aggregate patient quality data can be transmitted to the appropriate agency.

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, measure stewards, measure type and measure scoring, as well as information about whether the measure is NQF-endorsed. The HQMF Body contains eCQM chapters (e.g., population criteria, data criteria, supplemental data, and risk adjustment variables). Population Criteria can contain narrative descriptions and all chapters contain formally encoded HQMF entries.

Any eCQM intended to be submitted for NQF endorsement must be submitted in HQMF. This process is assured when measure developers author their eCQMs in the MAT. The MAT, see Section 3, Chapter 7.4, MAT, was developed under contract with CMS to aid in the creation of eCQMs.[76]

7.3 Quality Data Model (QDM)

The QDM was initially established by the Health Information Technology Expert Panel (HITEP) convened by NQF in 2009. 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 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.

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

The QDM allows definition of an eCQM data element,[77] which is an atomic (smallest possible) unit of information that has precise meaning to communicate the data required within a quality measure. Each eCQM 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 eCQM data element
- Code system recommendations—standard terminology, taxonomy, or classification system used for an eCQM data element's category (e.g., SNOMED CT, Logical Observation Identifiers Names and Codes [LOINC], RxNorm)
- Values —a direct referenced code[78] or list of codes (value set) used to define the specific eCQM data element
- Attribute—specific detail about an eCQM data element that further constrains the concept.

An eCQM 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. For example, defining a value set for pneumonia and applying the category (diagnosis) and the QDM datatype "diagnosis" forms the eCQM data element, "diagnosis: pneumonia" as a specific instance for use in a measure.

The process for changes to the QDM are outlined in the QDM User Group Charter.

[76] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Measure Authoring Tool.* Available at: https://www.emeasuretool.cms.gov/. Accessed on: November 9, 2016.
[77] Currently referred to as a QDM data element or QDM element in the QDM document.
[78] Direct referenced codes are a future enhancement.

7.3.1 Stating Logic Using the Quality Data Model

Measure developers should develop eCQMs using eCQM data elements and expression logic.[79] This approach, built into the MAT, takes each datatype (e.g., Diagnosis) in the QDM and its respective direct referenced code or value set and attributes with logic. The MAT allows the measure developer to create eCQM data elements by assigning direct referenced codes **(future enhancement)** or value sets to QDM datatypes and add attributes as appropriate. The MAT further allows assembly of data elements into data criteria using expression logic.

Examples:

"Diagnosis: Pneumonia" *overlaps* "Encounter, Performed: Encounter Inpatient"

|QDM Datatype| Value Set | Expression Logic| QDM Datatype | Value Set |

"Medication, Administered: Antibiotics" *starts during* "Encounter, Performed: Encounter Inpatient"

| QDM Datatype | Value Set | Expression Logic | QDM Datatype | Value Set |

See Appendix C: eCQM Logic Quality Assurance Checklist to help with review of eCQM logic.

7.4 Measure Authoring Tool (MAT)

The MAT is a publicly available, web-based tool that is used by measure developers to create eCQMs.[80] The tool enables measure developers to create their eCQMs in HQMF, a structured document format, without extensive knowledge of HQMF standards. Measure developers use the MAT to express measure criteria using QDM. The MAT is continuously maintained, supported, and updated as needed by modifications to the QDM, HQMF, and other standards to meet future measure authoring requirements. If a change to QDM is needed, there is a process to update the QDM and subsequently the MAT. All changes to the MAT are evaluated and prioritized by a MAT Change Control Board (MCCB) coordinated by CMS and other federal agencies.

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.

7.4.1 Define HQMF Metadata

The HQMF Header of an eCQM identifies and classifies the document and provides important metadata about the measure. The Header contains the information needed for the MIF for the eCQM. Appendix B: eCQM Metadata are summarized in listing elements in the order in which they are conventionally displayed as generated from the MAT.

The eCQM Header should include the appropriate information for each element as described in the Definition column. The default for each element in the MAT is a blank field. However, all Header fields require an entry. Appendix B: eCQM Metadata lists all Header (metadata) fields, their definitions, and guidance for measure developers to describe the field and conventions to use (e.g., "none" or "not applicable") if the field is optional. "Required" in the preferred term column indicates that the measure developer must populate the metadata field as defined in the second column. All eCQM Header fields must have information completed OR placement of a "None" or "Not Applicable" in the Header field. Conventions for when to use "None" versus "Not Applicable" are described in Appendix B: eCQM

[79] At time of publication, QDM expression logic is used for eCQMs. Clinical Quality Language (CQL) expression logic is in testing and may replace QDM logic in the future.
[80] At the time of publication, the MAT uses the term "eMeasure." The MAT will use eCQM in a future version.

Metadata for each metadata field and should be entered per the Preferred Term column instructions (e.g., for measures not endorsed by NQF, the metadata element NQF Number should be populated with "Not Applicable").

Note that for risk-adjusted measures, the Risk Adjustment metadata field points to an external risk model. The risk model itself is not part of the eCQM.

7.4.2 MAT Package

The MAT output includes the HQMF, a SimpleXML document,[81] a human-readable HTML, and a spreadsheet of the value sets. As previously noted, the spreadsheet of value sets is not the authoritative source for value set information. The VSAC is the authoritative source for value sets.

7.5 Quality Reporting Document Architecture (QRDA)

Once eCQMs are specified, tested, and implemented, the EHR systems vendors turn the eCQM into queries that retrieve the necessary information from the EHR's data repositories and generate quality data reports. eCQM reporting (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 (individual patient data) and Category III (aggregate patient data) reports, respectively. Both QRDA Category I and Category III are HL7 STU 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 Implementation Guides[82] 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 the 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 (e.g., a total count of patients who meet the denominator population criteria of a measure within a health system over a specific period of time).

7.6 Tools for Testing eCQMs

7.6.1 Bonnie Tool

Bonnie was released on April 3, 2014, and designed for testing eCQMs. With Bonnie, measure developers can evaluate logic for a measure created in the MAT by creating test cases with expected results. Test cases are first defined to cover each logic branch and scenario. The measure developer then enters these test cases into Bonnie. Bonnie executes these test cases and provides immediate feedback

[81] The SimpleXML is anticipated to be deleted as an output in a future version of the MAT.
[82] QRDA I - http://www.hl7.org/implement/standards/product_brief.cfm?product_id=35, QRDA III - http://www.hl7.org/implement/standards/product_brief.cfm?product_id=286

on whether the logic behaved as expected. This tool enables the user to readily identify criteria each test case (patient) must satisfy to qualify for a specific population (e.g., denominator and numerator). The ability to evaluate such test cases reduces the manual quality assurance efforts evaluating measure logic written in plain text. The tool facilitates building of test patients and identifies errors missed with manual review. It is especially helpful identifying recursive logic. Measure implementers also use Bonnie to validate their implementation of eCQMs and troubleshoot.

Refer to the Bonnie User Guide for more information.

7.6.2 Cypress Tool

Cypress is an open source testing tool used by vendors to certify eCQMs in their EHRs and EHR modules for calculating eCQMs. Cypress is the official testing tool for the EHR Certification program supported by ONC. Testing involves importing an eCQM into Cypress and applying QRDA Category I and Category III test patient records to test the measure logic and process calculations against the eCQM. Cypress can also be used in the measure development process. Like Bonnie, Cypress supports efficiency in measure development when integrated early in the process to identify additional measure logic discrepancies and calculation errors that potentially could occur during adoption and implementation.

8 MASTER LIST OF DELIVERABLES

Phase/Topic	Deliverables
Information Gathering	1. Information gathering report 2. List of potential candidate measures 3. Measure documentation, including: • MIF • Measure Justification Form (MJF) • Business case • Expert input report (if applicable)
Business Case	1. Initial business case 2. Final business case
Technical Expert Panel	1. Call for TEP 2. TEP nomination forms 3. List of stakeholders 4. TEP Charter 5. TEP Composition Documentation (TEP Membership List) 6. TEP meeting schedule 7. Meeting minutes 8. Potential measures presented to the TEP 9. Measure Evaluation Report 10. Updated MIF and MJF
Public Comment	1. Public Comment Call Web Posting template form 2. List of stakeholders for notification 3. MIF and MJF (for candidate measures) 4. eCQM specifications (as appropriate) 5. Public Comment Summary report, including verbatim comments
Measure Specification	List of potential measures to be developed and timelines Appropriate measure specification forms, by case: MIF and MJF, or equivalent, for candidate measures MJF, or equivalent, to document the measure evaluation for new or adapted measures, or measures that are developed and in use by another organization but are not NQF-endorsed For risk-adjusted measures: Risk Adjustment Methodology Report MIF with completed risk adjustment sections for each measure For eCQMs: eCQM XML file (HQMF), SimpleXML file, an eCQM human-readable rendition (HTML) file, and value sets
Measure Testing	1. Measure Testing Plan 2. Measure Testing Summary Report 3. Updated MIF 4. Updated MJF 5. Updated Measure Evaluation Report
Measure Implementation	1. Public Description of Quality Measures 2. Timeline for Data Item and/or Quality Measure Implementation 3. Implementation Stakeholder Meetings 4. Questions and Answers Support 5. Implementation Process Roadmap 6. Measure Calculations/Results 7. Pre-Posting Preview Results 8. Compare Sites Files and Measures (as applicable) 9. Implementation Algorithm (also called the Calculation Algorithm/Measure Logic) 10. Data Use Agreement

Phase/Topic	Deliverables
Measure Maintenance	1. Audit and validation reports 2. Audit and validation appeals reports 3. Preview reports, if required by the CMS program using the measure 4. Periodic measure rate trend reports 5. Analysis of the measure results 6. Ad hoc analyses, as requested by CMS 7. Questions and Answers Support 8. Periodic environmental scans 9. Data files of the measure rates and/or demographic information suitable for posting on CMS website
Risk Adjustment	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 2. MIF with completed risk adjustment sections for each measure 3. For eCQMs, the eCQM human-readable rendition (HTML) file includes instructions where the complete risk adjustment methodology may be obtained
Measure Update	An updated MIF showing all recommended changes to the measure. If there are changes relevant to the Measure Justification, that form should be updated as well. For measure developers maintaining eCQMs, the revised eCQM XML file (HQMF), SimpleXML file, an eCQM human-readable rendition (HTML) file, and value sets must be submitted detailing the changes to the measure. A document summarizing changes made, such as Release Notes, if not included in the updated MIF. NQF Annual Update online submission regardless if any change was made to the measure. NQF submission documentation for any material changes to the measure.
Comprehensive Reevaluation	An updated MIF detailing all recommended changes to the measure. For measure developers maintaining eCQMs, the revised eCQM XML file (HTML), SimpleXML file, eCQM human-readable rendition [HTML] file, and value sets must be submitted detailing the changes to the measure. A document summarizing changes made, such as Release Notes, if not included with the updated MIF. An updated MJF documenting the environmental scan results, any new controversies about the measure, and any new data supporting the measure's justification. An updated Measure Evaluation Report describing measure performance compared to the measure evaluation criteria and the performance of the measure. An updated business case that reports on the measure performance trend and trajectory as compared to the projections made during measure development, including recommendations. NQF endorsement maintenance online submission documentation (at the scheduled 3-year endorsement maintenance). If it is time for 3-year maintenance review (comprehensive reevaluation) but the NQF project is not ready, an annual update report may be submitted online.
Ad Hoc Review	Updated MIF, if the ad hoc review results in changes to the measure specifications. For measure developers maintaining eCQMs, the revised eCQM XML file (HTML), SimpleXML file, eMeasure human-readable rendition [HTML] file, and value sets, if the ad hoc review results in changes to the measure specifications. Updated MJF, 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. Updated Measure Evaluation Report, if the review resulted in a change to the measure's strengths and/or weaknesses.

9 INFORMATION GATHERING

Information gathering is conducted via the following eight steps, which may not occur sequentially:

- Conduct an environmental scan
- Conduct an empirical data analysis, as appropriate
- Evaluate information collected during 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
- Submit the information gathering report
- Prepare an initial list of measures or measure topics.

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 Umbrella Statement of Work, contractors 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 contractor must identify any applicable measures in current use that might be appropriate for the specific Task Order. This would occur through analysis of resources, including employers, commercial plans, managed care plans, Tricare, NQF, MedPac, IOM, Institute for Healthcare Improvement (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 CMS Quality Measurement Technical Forum Goals, 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 contractor 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 contractor to evaluate existing quality measures to support development of outcome and process measures that have established histories of quality or process improvement. The Task Order might require the contractor to evaluate measures that address safety issues, adverse events, healthcare acquired conditions (e.g., pressure ulcers), patient-centered care (e.g., symptom management), patient engagement and experience, care coordination, readmissions, and population health.

Among the many important areas to scan, contractors must consider the IOM's Six Aims of Care, which include safety, timeliness, efficiency, effectiveness, equitability, and patient centeredness. Contractors 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, expert input (including the TEP and other experts), and stakeholder input–inclusive of all relevant stakeholders, including patients (Figure 19).

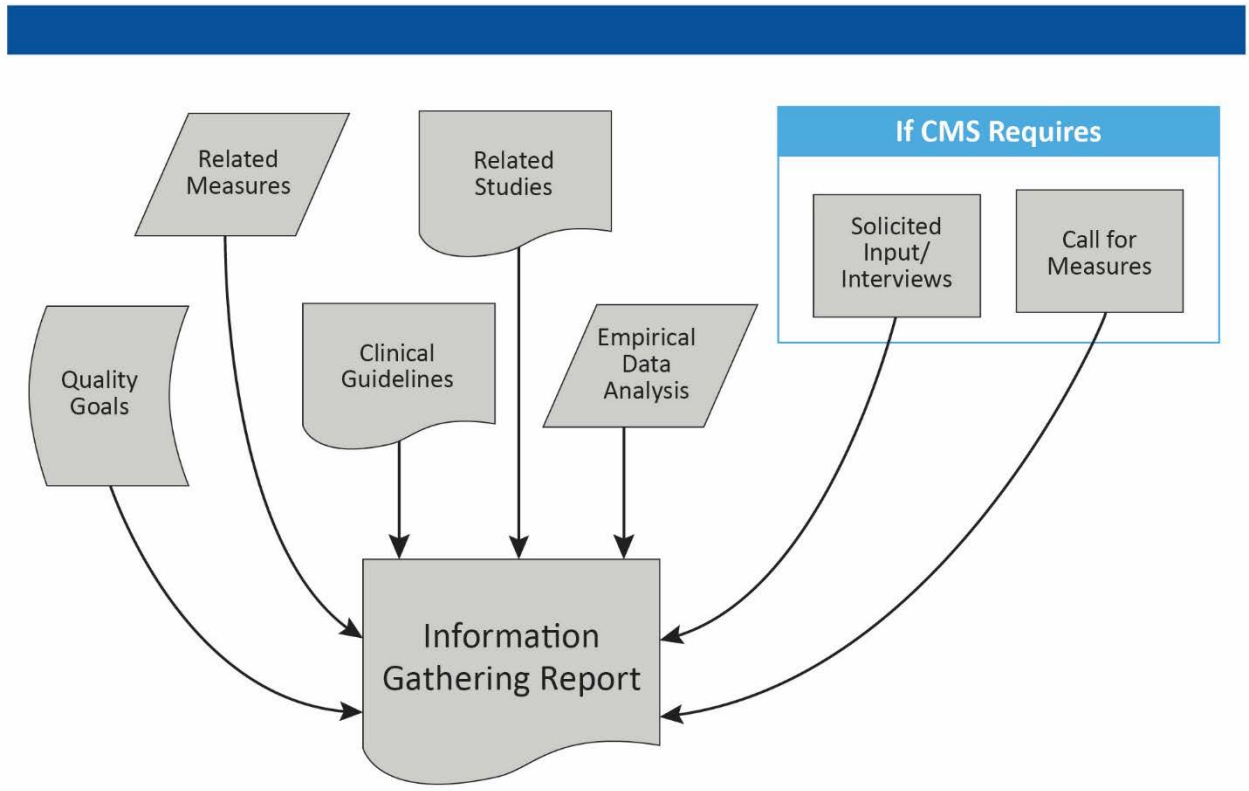

Figure 19. Environmental Scan Data Sources

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,[83] and in the business case.

[83] The NQF submission may be acceptable for this deliverable. National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: March 14, 2016.

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. This will enable measure developers sufficient time to: (1) replace or revise data elements, (2) consider an alternative measure type, (3) assess implementation burden versus value of measure, or (4) 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:

- 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.

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

A respecified measure is a type of adapted measure that is moved into HQMF format (eCQM) for use with EHR data. Hence, a respecified measure will often include subset of the original measure's data elements.

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, it may be possible for the measure developer to adapt the measure for the new use.

Example of an adapted 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 owner of the adult depression screening measure to expand the population in the measure to the adolescent population.

Begin evaluating whether to adapt 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 adapted 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. The Section 3, Chapter 22, Measure Testing provides further details of the process.

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

- 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)
- Change in the importance of the original measure in a new setting (i.e., 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)
- 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 a measure is copyright protected, there may be issues relating to its ownership or to proper referencing of the parent measure. In either case, contact the measure owner for permission or clarification. Upon receiving approval from the original developer to use the existing measure, include the detailed specifications for the measure.

9.4 Conduct a Measurement Gap Analysis to Identify Areas for New Measure Development

Develop a framework to organize the measures gathered. The purpose of this gap analysis is to identify measure types or concepts that may be missing for the measure topic or focus. 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, adapted, or identify new measures that need to be developed.

9.5 Determine the Appropriate Basis for Creation of New Measures

If no existing measures are suitable for adoption, adaptation, 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 the following 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 clinical 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. This 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 Medicare population; effects on Medicare 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.[84] 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

[84] The NQF submission may be acceptable for this deliverable. National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: March 14, 2016.

measure submission[85]. 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.[86]

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 adapted measures and 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 to the COR that summarizes the information obtained from the previous steps. This report should include, but not be limited to the following items.

9.7.1 Summary of Literature Review (annotated bibliography)

Provide the following information (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
- Characteristics of the study (i.e., population, study size, data sources, study type, and method)
- Which measure evaluation criteria (i.e., importance, scientific acceptability, usability, and feasibility) 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 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 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.

[85] Either the MIF and MJF or the NQF measure submission forms may be submitted as contract deliverables.
[86] National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: March 14, 2016.

9.7.2 Summary of Clinical Practice Guidelines Review

Provide 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.

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

Provide a summary of findings and measurement gaps:

- 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 Form:

- 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

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, adapted, respecified, or 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.

10 ENVIRONMENTAL SCAN

The following six steps are fundamental to creating an environmental scan:

1. 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.
2. Determine the frame for relevant work, including literature databases and search engines, keywords and phrases, inclusion and exclusion criteria, and domain experts.
3. 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, seminality/originality, and quality of writing.
4. 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.
5. 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.
6. Refine research questions and develop hypotheses. Generate a general analysis plan, including data sources and estimation procedures.

In addition, measure developers will want to:

- Be strategic in planning and managing the scan
- Formalize their scanning process
- Design the scan in collaboration with domain experts
- Manage the information obtained.[87]

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 24, Measure Evaluation 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[88] associated with the measure topic. This is especially true if:

- Clinical practice guidelines are unavailable.
- The 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

[87] Choo, C.W. (1999). *Bulletin of the American Society for Information Science. "The Art of Scanning the Environment". Available at:* http://bit.ly/1KfZfvm. Accessed 10/29/2015.
[88] http://www.ahrq.gov/workingforquality/about.htm

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

Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence 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; and in some cases, it may not be appropriate. In these cases, non-RCT studies may be relied on including quasi-experimental studies, observational studies (e.g., cohort, case-control, cross-sectional, epidemiological), and qualitative studies. Review the:

- Quantity—Five or more RCT studies are preferred, but are a general guideline.[89] This count refers to actual studies, not papers or journal articles written about the study.
- Consistency of results across studies—Summarize the consistency of direction and magnitude of clinically/practically meaningful benefits over harms to the 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. The measure developers are not required to grade the evidence; rather, the goal is to assess whether the evidence was graded, and if so, what the process entailed.
- Summary of controversy and contradictory evidence, if applicable.
- Information related to healthcare disparities—Review these across patient demographics, in clinical care and in outcomes. This 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
- Published in journals from respected organizations
- Written recently (within the past 5 years)
- Based on data collected within the past 10 years
- Unpublished studies or reports such as those described as grey literature. Governmental agencies such as the AHRQ, CMS, and the CDC produce studies and reports that are publicly available but not peer-reviewed.

[89] National Quality Forum. *Review and Update of Guidance for Evaluating Evidence and Measure Testing*. Oct 2013. Available at: http://www.qualityforum.org/Publications/2013/10/Review_and_Update_of_Guidance_for_Evaluating_Evidence_and_Measure_Testing_-_Technical_Report.aspx. Accessed on: March 14, 2016.

- 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.
- IOM report: Finding What Works in Health Care Standards for Systematic Reviews[90]
- NQF report: Guidance for Evaluating the Evidence Related to the Focus of Quality Measurement and Importance to Measure and Report.[91]

10.3 Clinical Practice Guidelines

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 physician organizations or federal agencies are preferred. However, guidelines and other evidence documents developed by non-American organizations, as well as non-physician organizations, may also be acceptable and should be assessed to determine if they are a sufficient basis for measure development.

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 recommendation is noted, document if the guideline recommendations are valid, useful, and applicable.

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 National Guideline Clearinghouse 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. Determine what measurement gaps exist for the topic area, as well as existing measures that may be adopted, adapted, or respecified for the project. For example, if a contract objective is the development of immunization measures for use in the home

[90] Institute of Medicine. *Finding What Works in Health Care Standards for Systematic Reviews.* Washington, DC: The National Academies Press, 2011. Available at: https://www.nap.edu/catalog/13059/finding-what-works-in-health-care-standards-for-systematic-reviews. Accessed on: March 14, 2016.

[91] National Quality Forum. *Guidance for Evaluating the Evidence Related to the Focus of Quality Measurement and Importance to Measure and Report.* Jan 2011. Available at: http://www.qualityforum.org/Measuring_Performance/Improving_NQF_Process/Evidence_Task_Force_Final_Report.aspx. Accessed on: March 14, 2016.

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 in the same setting, but for a different topic
- Measures in a different setting, but for the same topic
- Measures that are 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. Following are links to a few readily available sources:

- NQMC
- HHS Measures Inventory
- CMS Measures Inventory and Pipeline
- NQF's Quality Positioning System
- American Medical Association-Physician Consortium for Performance Improvement.

Search for other sources of information such as performance indicators, accreditation standards, or preferred practices that may pertain to the contract topic. Though 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 National Committee for Quality Assurance. 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 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 by having patients participate as part of the TEP, the TEP 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 vendors 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, TEP.

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 if 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 who 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 owner of those measures or measure concepts is willing to expand the measures for use by CMS. A 14-day call period is recommended.

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 4, Measure Implementation covers these types of calls.

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 owner of the original measure to discuss issues related to ownership, maintenance, and testing.

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.

Compile a list of the initial measures received during the Call for Measures and evaluate these measures using the measure evaluation criteria.

11 BUSINESS CASE

The business case documents all the 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 20). 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 sectors 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.

Potential Costs, Benefits, and Savings

In making the business case, it is important to qualify and quantify the pros and cons of implementing the measure, including hard and soft benefits. Possible items to consider include:

Patients: health outcomes, length of stay, readmissions, patient satisfaction, adverse events, medical errors, trust of the healthcare system.

Employee and organizational: workplace safety, staff time, staff turnover, sick time, training, turnover, hiring costs, staff supervision costs.

Liability: worker's compensation claims, liability insurance premiums, litigation and judgment costs, fines.

Materials: product purchase, maintenance, storage, and disposal.

Figure 20. Potential Costs, Benefits, and Savings

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. 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 6.

Table 6. The Business Case is an Ongoing Process that Occurs Throughout the Measure Lifecycle

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

The importance criteria in the Measure Evaluation Report and in the NQF Measure Evaluation Criteria[92] contain requirements for information that will be used to begin a business case. The guidance provided here and in the Business Case Form Instructions will help measure developers identify additional information to collect and to construct a case to meet CMS requirements. To the extent possible, CMS has aligned its Business Case Form with NQF's Measure Submission Forms. 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 21 diagrams business case inputs and the impact of the business case throughout the measure lifecycle and

Figure 22 shows the overall flow of inputs to business case development.

[92] National Quality Forum. *Measure Evaluation Criteria.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards/Measure_Evaluation_Criteria.aspx. Accessed on: March 14, 2016.

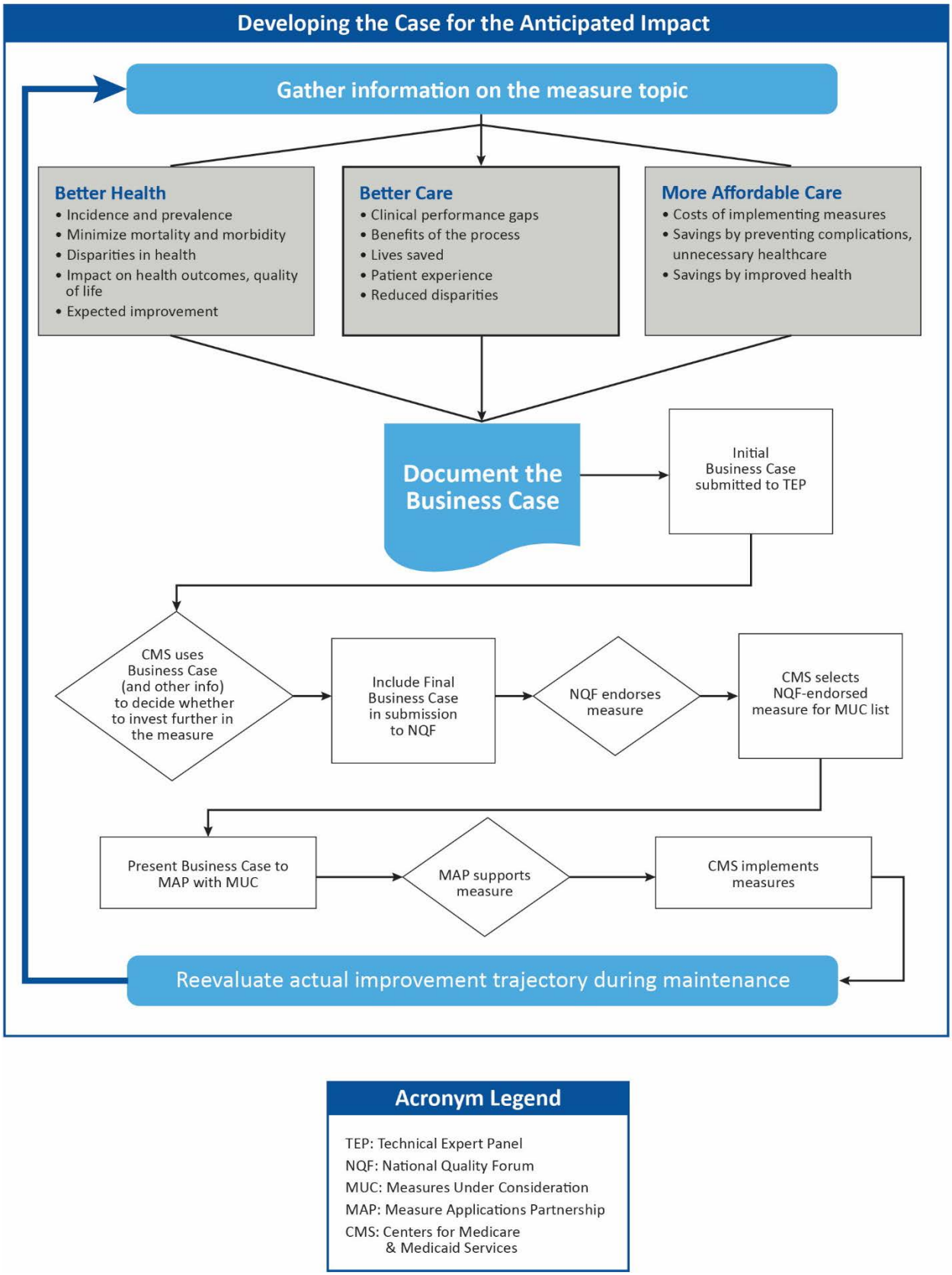

Figure 21. Inputs and Uses for the Business Case

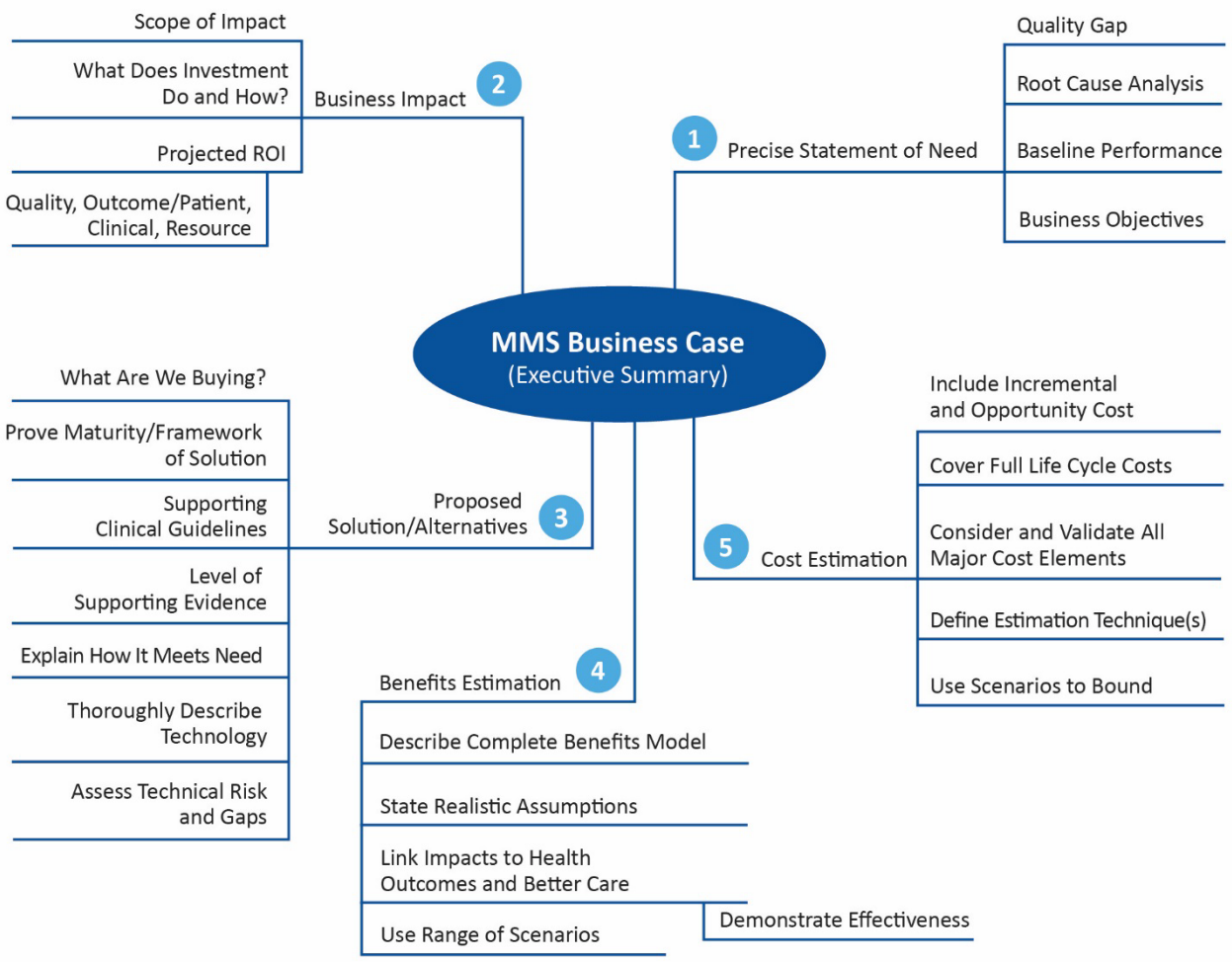

Figure 22. Flow of Inputs into Business Case Development

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 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 the reasons the improvements are not happening. The business case should be adjusted for any changes in the environment or if the assumptions used initially 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. It is recognized that 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). Following are the key elements that may be included in the executive summary:

- 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 informing measure developers when fields request information that is also required for a NQF business case submissions, 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 measures under development and maintenance. Consult with the COR on the types of information and final format of the business case that will be required.

12 TECHNICAL EXPERT PANEL (TEP)

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/Information Technology (IT) vendors—preferably at least two different 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 contractor 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 the following 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 (the TEP should assess the feasibility of alternative methods to address the measurement opportunity such as 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). See 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 contractors will convene several TEP meetings, either by teleconference 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 give 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. See Section 2, Chapter 5, 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 contractor 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 contractor should be able to show how the recommended measures relate to overall HHS goals, including the NQS priorities, CMS Quality Strategy, measurement priorities, and relevant program 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 a lengthy and resource-intensive nomination and review process being undertaken by measure developers 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 concurrently 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.

12.2.2 Standing TEP Structure

Some measure developers have migrated toward the development of a standing TEP structure to alleviate some of these issues and concerns. Under this model, the measure developer nominates and gathers a standing TEP with a 2- to 3-year term of membership. This TEP has a diversity of membership with broad-based expertise (e.g., policy and program, measure development, clinicians, patients/advocates, 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 broad-based policy surrounding each of the measures under development or considered for future development. This cross-cutting focus enables the standing TEP to view and help problem solve across the portfolio of measures under development.

In concert with the standing TEP, the measure developer also convenes a series of expert workgroups through targeted outreach. These workgroups 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 particular 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 deep dive into the technical aspects of a measure. The expert workgroups 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 efficient—the TEP nomination process not used for every measure
- Continuity, perspective, and programmatic knowledge within the standing TEP membership
- Trust building among TEP members who meet regularly and get to know each other
- Less alienation and confusion for patient and caregiver representatives, since technicalities are tackled separately in the expert workgroup
- 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 of 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
- 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 for measure development, 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 a patient's 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 an actual 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. Call for TEP documents should be written in language that lay participants can clearly understand. The following 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 where TEP nominations and any questions are to be sent.

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.[93]

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 may submit their TEP nomination packages to MMS Support (MMSSupport@battelle.org) for posting on the CMS TEP webpage. The CMS Events Calendar also lists dates when TEP nomination periods open and close. The calendar can be found on the CMS Resource Materials webpage.

[93] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Measures Management System: Technical Expert Panels.* Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/TechnicalExpertPanels.html. Accessed on: March 14, 2016.

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 8–15 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 should incorporate the patient perspective in measure development through patient and/or caregiver input into quality issues that are important to patients.[94] The measure developer then proposes the list of TEP members to CMS.

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

- Geography—Include representatives from multiple areas of the country and other characteristics such as rural and urban settings.
- Diversity of experience—Consider individuals with diverse backgrounds and experience in different types of organizations and organizational structures.
- 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.

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

[94] Requirement quoted from MIDS contract 2013 language.

also be sent to the COR. Notify the COR about the TEP membership list within one 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 achieve the following tasks:

- Convene and conduct the meeting in a professional and timely manner
- Conduct the meeting according to the agenda
- Recognize speakers
- Call for votes.

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 Technical Expert Panel Composition (Membership) List Template. Use the Technical Expert Panel Composition (Membership) List Web Page Posting form that includes the meeting schedule. Patients included on the TEP who indicated that they wanted their name to remain confidential on the TEP Nomination will be identified as "Patient" 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, via telephone conferencing, 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 one week prior to 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 the following materials:

- 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.[95]
    - 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 at any time while serving on the TEP, a member's status changes and a potential conflict of interest arises, 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, as well as overarching goals for improvement. Provide emphasis on presenting materials in a manner that lay members of the TEP will be able to understand in order to highlight the important role of the patient/caregiver voice in measure development.

[95] The NQF submission may be acceptable for this deliverable. National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: March 14, 2016.

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 24, Measure Evaluation provides a description 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 the following:

- 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 delivered to CMS by the measure developer at this time. The Measure Evaluation Report includes information on how each measure met or did not meet each subcriterion. Additionally, it provides CMS with information regarding the feasibility of strengthening the rating of any subcriterion that was rated "low." At this time, 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 Technical Expert Panel 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 regarding whether the technical specifications need to be revised.

It is important to note that the TEP may be consulted for its advice during any stage of the 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 CORs discretion) be appropriate to summarize discussions held during multiple meetings.

13 Person and Family Engagement

13.1 Background and Definition

Person and family engagement is the process of involving persons and/or family representatives in a meaningful way throughout the measure lifecycle. As used here, the term *person* refers to a non-healthcare professional representing those who receive healthcare. In this context, *family representatives* are other non-healthcare professionals supporting those who receive healthcare (e.g., caregivers).

Strengthening persons and families as partners in their care is one of the goals of CMS Quality Strategy[96] 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 7. 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) person-family engagement framework when engaging consumers (Figure 23) and observe best practices for conducting qualitative research, survey and interview construction, and testing, as applicable and with approval from CMS.

Concepts highlighted by PCORI that are applicable to person/family member engagement in the measure development process include the following:

- Reciprocal Relationships: Roles and decision-making authority of all involved are defined collaboratively and clearly stated.
- Co-Learning: It is important to ensure that all participants understand the measure development process, person and family engagement, and person-centeredness.
- Partnership: 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.
- Trust, Transparency, Honesty: 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.

Figure 23. Person and Family Engagement Concepts

[96]CMS Quality Strategy. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy.html. Accessed on February 20, 2016

Table 7. Best Practices for Implementing Person/Family Engagement Activities, by Phase of Engagement

Phase	Best Practices for Implementing Person/Family Engagement Activities
Preparing for Person/Family Engagement Activities[97]	• 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 1–2 days prior to 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 get 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.

[97] Institute for Patient and Family Care. *Partnering with Patients and Families to Enhance Safety and Quality: A Mini Toolkit*. 2013. Available at http://www.ipfcc.org/. Accessed on: May 31, 2015

Prior to measure conceptualization, developers should put together 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 in the following paragraphs.

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.[98,99] In some instances, developers have found appointing a patient as the leader of the TEP an effective strategy.
    - Ask patients 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 is 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.

[98] Institute for Patient and Family Care. *Partnering with Patients and Families to Enhance Safety and Quality: A Mini Toolkit*. 2013. Available at http://www.ipfcc.org/. Accessed on: October 15, 2015.

[99] Noted by Kyle Campbell. MIDS C3 Forum Contract Overview Presentation. October 21, 2015

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 5–6 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 5–6 individuals, including patients, families, 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 24 contains a list of best practices for TEPs and working groups.

Selected Best Practices: TEPs and Working Groups

- Schedule meetings at times that are convenient for participants
- Ensure participants are well prepared for the meeting
- Provide read-ahead materials that are easy to understand
- Communicate with participants between meetings.

Figure 24. TEP and Working Group Best Practices

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 3, 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.[100]

[100]Additional information about plain language testing can be found through resources such as http://www.plainlanguage.gov/ and http://centerforplainlanguage.org/.

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 timeframe. 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 social 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 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).

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 ensure 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 Technical Expert Panel (Call for TEP) Web Page Posting form may be used, but other sources and methods may also be required. The following 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).

The following websites are examples 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:

- AARP
- The Empowered Patient Coalition
- WebMD
- Patient Voice Institute
- AgingCare.com
- Caring.com
- Connecticut Center for Patient Safety
- Daily Strength
- Informed Medical Decisions Foundation (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. The websites listed above 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 very 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 25 contains an example of a featured best practice for recruitment.

Featured Practice: Recruitment

A measure developer will be holding a TEP meeting in Washington, D.C., to discuss new measures being considered for the Readmissions Reduction Program (RRP). To facilitate person participation, the measure developer made the following options available:

- Option to be picked up at home by a ride service and driven to the meeting for those living within 50 miles of the meeting venue.
- Option to dial-in via a toll-free conference line, and/or participate virtually via web-based meeting software.

Figure 25. Recruitment Practices

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

As discussed in Section 2, 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 stages of the measure lifecycle when particular engagement techniques are most useful are described below.

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. See also Figure 26 for a best practice on measure conceptualization.

Featured Practice: Measure Conceptualization

In support of measure development efforts for the Inpatient Psychiatric Facility Reporting Program, an interviewer asked patients and family members to describe their experience during each step in the care experience: admission, care during the inpatient stay, discharge, and after care in the community. Next, the interviewer asked patients to say what went well and what could have gone better. In addition, the interviewer also prompted patients to prioritize a list of topics identified through evidence review and expert input. This process made it possible to identify areas of agreement about priority topics from patients, family members, and other experts.

Figure 26. Measure Conceptualization Best Practice

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?*"

- *One-on-one interviews* with a skilled interviewer using a planned study 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 this 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, patient reported outcome 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. See Figure 27 for a best practice on measure conceptualization and specification.

Techniques. The following mechanisms that allow for 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 will help focus the recommendations they provide to the developer.

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 27. Measure Conceptualization and Specification Best Practice

13.5.3 Measure Testing

During the measure testing phase, the developer tests the measure to make sure 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 the following 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 measure work best at this stage. One-on-one data collection methods are often useful.

- *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 surfing 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 implementation stage, the 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, 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 allow for 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, 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 these points 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 make sure 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. This 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 (1) at least every 2–3 years to ensure the concepts are fresh and relevant, and (2) 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 Office of Management and Budget (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.[101] Measure developers working under contract with CMS should consult with the COR to determine if their project is eligible for an exemption.

[101] https://www.gpo.gov/fdsys/pkg/USCODE-2008-title44/pdf/USCODE-2008-title44-chap35.pdf

Developers working with CMS programs that are not PRA-exempt should factor time—on average, 6–8 months—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 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 measures under development 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 allow 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.

- Measure conceptualization, information gathering: Aggregate comments on the summary of the TEP meetings.
- Measure specification: Draft technical specifications can be posted with summaries of subsequent TEP meetings.
- Measure testing: If a TEP reviews testing results and updated specifications, those summaries can be posted for further public comment.
- Measure implementation:
    - The MUC list is posted for public comment as part of the pre-rulemaking process.
    - The MAP [102] 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.
- 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

[102] National Quality Forum. *MAP 2014 Recommendations on Measures for More Than 20 Federal Programs Final Report.* Available at: http://www.qualityforum.org/Publications/2014/01/MAP_Pre-Rulemaking_Report__2014_Recommendations_on_Measures_for_More_than_20_Federal_Programs.aspx. Accessed on: May 31, 2015.

after the measure is developed and is being proposed for implementation; therefore, the measure developer could miss the opportunity to address the 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 following eight steps are essential to successfully soliciting public comment. Deviation from the following 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 site that lets readers know that a measure is up for comment on another website.

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

For eCQMs, the ONC has licensed JIRA, a generic software tracking tool, for collecting and monitoring feedback on different stages of the measure development process, including posting calls for public comment. JIRA is also used to collect measures for the MUC List.

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., American 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 group.

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 workgroup 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."[103]

As a rule, the call should be posted on the website for at least 2 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; for example:
    - Identifying important quality goals related to Medicare services
    - 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)
- Posting for public comment
- 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
- List of the TEP members (use the TEP Panel Roster form), including any potential conflicts of interest disclosed by the members
- Information about the measure developer and subcontractors developing this measure set.

14.3.4 Collect Information

Commenters submit their comments via email or other tool 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

[103] The NQF submission may be acceptable for this deliverable. National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: March 14, 2016.

the TEP. This report should be submitted to the COR and the TEP within 2 weeks following the end of the public comment period.

The report should include:

- 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 may be done 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 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 within 3 weeks (or as directed by the COR) 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 (Call for Measures, TEPs and Call for Public Comment) linked through the MMS Overview site.

Figure 28 demonstrates the steps in the process for posting to the CMS MMS website.

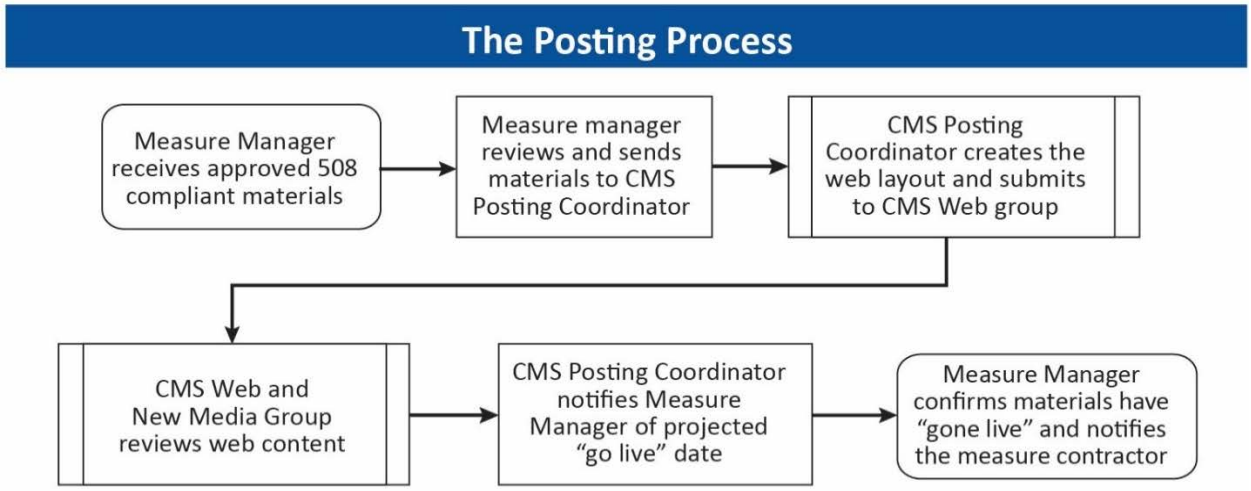

Figure 28. The Web Posting Process

1. 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.

2. Submissions must be sent to MMS Support inbox (MMSSupport@battelle.org) for posts on:

- Call for Measures web page
- Public Comment web page
- Technical Expert Panel (TEP) web page

3. Upon receipt by the Measure Manager, the material to be posted is reviewed again for Section 508 compliance, completeness of information, confirmed COR approval, and compliance with formatting requirements.

4. The materials are then sent to the CMS Website Posting Coordinator.

5. CMS Website Posting Coordinator creates the updated web page layout and submits it to the CMS Web group for posting.

    a. 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."

6. The CMS Website Posting Coordinator sends confirmation that the approved materials have been moved into the production environment.
7. The Measure Manager verifies that the materials are available on the site and notifies the measure developer and their COR.

15.1 Posting Timeframe

Allow **at least 5 business days** for processing your post; however, posts *may* be posted prior to this timeframe. If your post needs to be published on a specific date, please note this in your email and CMS will work to accomplish 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.** Please 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[104] provides guidance for the measure developer to ensure measures developed for CMS have complete technical specifications that are detailed and precise.

The following factors influence the development of technical specifications for a new measure:

- Literature review
- Existing measures
- TEP input
- Public comment
- Alpha testing
- Beta testing.

All these factors 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 across providers.

Most quality measures are expressed as a rate. Usually, the basic construct of a measure begins with the numerator, denominator, exclusions, exceptions, and measure logic. Then, the measure concept is more precisely specified with increasing amounts of detail, including the appropriate values or value sets and/or detailed and precisely defined data elements.

The following 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 code systems
- Construct data protocol
- Document the measures and obtain COR approval.

16.1 DEVELOP THE CANDIDATE MEASURE LIST

Use the information collected from the environmental scan, measure gap analysis, and other information gathering activities to determine if there are existing or related measures before deciding to develop new measures. Use the information obtained from the information gathering process to identify if there are existing measures for the project within a specific topic or condition. If there are no existing or related measures that can be adapted, respecified, or adopted, then it is appropriate to develop a new measure.

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, adapted existing measures, or measures adopted from an existing set).

Avoid selecting or constructing measures that can be met primarily through documentation without evaluating the quality of the activity (i.e., often satisfied with a checkbox, date, or code).

[104] Some of the direction provided in this chapter is based on guidance from the NQF, and in some instances the wording remains unchanged to preserve the intent of the original documents.

Examples of such measures include:

- A completed assessment
- A completed care plan
- An instruction such as teaching or counseling that is simply delivered.

More important than whether a patient received teaching (or any of the other examples) is whether a patient understands how to manage their care, which is best measured from the patient's perspective.[105] Although it is relatively easy to document having counseled or educated a patient about a specific issue, it is important, although more difficult, to document that the patient understood the counseling or came away with a gained self-care competency through the encounter. The latter is closer to an outcome measure, where the former (counseling) is simply a process measure that may not have the intended effect in measuring a patient's understanding or competence. For this reason, it is important to base performance measures on the patient's perspective as much as possible.

16.1.1 New Measures

Begin work on a new measure if it has been determined through the information gathering process and input from the TEP that no existing or related measures are applicable for the topic. Determine the appropriate basis for the new measures in consultation with the TEP, keeping in mind the measure evaluation criteria as a framework. The appropriate basis will vary by type of measure. The following criteria for measure evaluation are discussed in Section 3, Chapter 24, Measure Evaluation:

- Draft the measure statement with high-level numerator and denominator statements.
- With input from the TEP, consider the populations to be included in both the numerator and denominator. In addition, develop a high-level algorithm describing the overall logic that will be used to calculate the measure. Alpha (or formative) testing may be used at this stage to reinforce development of the conceptual measure.
- For measures that are developed using administrative data, data analysis may be conducted to determine strategies for obtaining the desired populations.
- For measures using medical record information, interviews with clinicians or small-scale tests may assess the feasibility, usability and validity of the measure or portions of the measure. EHR data experts and informaticists would be valuable resources for conducting early feasibility testing. Section 3, Chapter 22, Measure Testing includes more details of this process.
- After determining any areas for potential harmonization (described in Section 3, Chapter 18, Measure Harmonization), the measure developer develops the detailed specifications.

16.1.2 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 particular to the measure's implementation use (e.g., data submission instructions).

16.1.3 Adapted Measures

In adapting a measure to a different setting, the measure developer needs to consider accountability, attribution, data source, and reporting tools of the new setting. Measures that are being adapted 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, particularly

[105] National Quality Forum. *CSAC Guidance on Quality Performance Measure Construction*. May 2011. Available at: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=68999. Accessed on: January 25, 2017.

where the measure is being adapted for use in a new setting with a new data source, this aspect of the adapted measure will need to be evaluated, respecified and tested. Section 3, Chapter 18, Measure Harmonization describes adapted measures. Before the decision is made to adapt a measure in existence, the following issues should be considered:

- 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 owner 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 owner. NQF endorsement status may need to be discussed with NQF. After making any changes to fit the use, the detailed specifications will be developed.

16.1.4 Respecified Measures

Respecified measures are eCQMs developed from current claims/registry/paper measures that are placed into HQMF format. Hence, a respecified measure will often include different data elements than the original CQM created based on the same evidence.

16.1.5 Composite Performance Measures

Composite measures are comprised of multiple component measures. Select the component measures to be combined in the composite performance measure. Concepts within Section 3, Chapter 5, Selected Measure Types describe the way component measures are used to construct composite performance measures.

16.2 Develop Precise Technical Specifications and Update the Measure Information Form

Development of the complete technical specifications is an iterative process. Alpha or formative testing should be conducted, as needed, concurrently with the development of the technical specifications. The timing and types of tests performed may vary depending on variables such as data source; complexity of measures; and whether the measure is new, adapted, respecified, or adopted. At a minimum, measures should be specified with the broadest applicability (e.g., target population, setting, level of measurement/analysis) as supported by the evidence.[106]

16.2.1 Develop 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.

Format—[target population] who received/had [measure focus]

Examples:

- Patients with diabetes who received an eye exam

[106] National Quality Forum. *CSAC Guidance on Quality Performance Measure Construction*. May 2011. Available at: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=68999. Accessed on: January 25, 2017.

- Long-stay residents with a urinary tract infection
- Adults who received a Body Mass Index assessment.

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 (i.e., to demonstrate which clinical activity is being captured in the numerator).

Examples:

- Percentage of patients admitted to a PACU, regardless of age, who are under the care of an anesthesia practitioner in which a formal post-anesthetic transfer of care protocol or checklist is used that includes the key transfer of care elements.
- Percentage of residents with a valid target assessment and a valid prior assessment whose need for help with daily activities has increased.
- Median time from emergency department arrival to administration of fibrinolytic therapy in emergency department (ED) patients with ST-segment elevation or left bundle branch block (LBBB) on the electrocardiogram (ECG) performed closest to ED arrival and prior to transfer.
- Percentage of diabetics, aged 18 years and older, appropriately adhering to chronic medication regimes.

16.2.2 Define the Initial Population

The initial population refers to all patients to be evaluated by a specific performance measure 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.

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 cohort from which the denominator population is selected must be specified. Details often include information based on specific age groups, diagnoses, diagnostic and

procedure codes, and enrollment periods. 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.

16.2.3 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 is a subset of the initial population to further constrain the population for the purpose of 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. The denominator statement should be precisely defined and include parameters such as:

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

Format—The number of patients, aged [age or age range], with [condition] in [setting] during [time frame]

Examples:

- The number of patients, aged age 18–75, with diabetes in ambulatory care during a measurement year.
- The number of female patients, aged 65 and older, who responded to the survey indicating they had a urinary incontinence problem in the past 6 months.
- The number of patients, aged 18 and older, who received at least a 180-day supply of digoxin, including any combination products, in any care setting during the measurement year.
- The number of patients, aged 18 and older, with a diagnosis of chronic obstructive pulmonary disease (COPD) who have a forced expiratory volume in 1 second/forced vital capacity (FEV1/FVC) of less than 70 percent and have symptoms.
- The number of patients on maintenance hemodialysis during the last hemodialysis treatment of the month, including patients on home hemodialysis.
- The number of patients, aged 65 and older, discharged from any inpatient facility (e.g., hospital, skilled nursing facility, rehabilitation facility) and seen within 60 days following discharge in the office by the physician providing ongoing care.

16.2.4 Define the Numerator

The numerator statement describes the process, condition, event, or outcome that satisfies the measure focus or intent. Numerators are used in proportion and ratio measures only. 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. They should be precisely defined and include 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— The number of denominator-eligible patients who received/had [measure focus] {during [time frame] if different than for target population}

Examples:

- The number of denominator-eligible patients that received a foot exam, including visual inspection, sensory exam with monofilament, or pulse exam
- The number of denominator-eligible patients that had documentation of receiving aspirin within 24 hours before ED arrival or during their ED stay
- The number of denominator-eligible nursing home residents who had an up-to-date pneumococcal vaccination within the six-month target period as indicated on the selected Minimum Data Set target record (assessment or discharge).

16.2.5 Determine if Denominator Exception or Denominator Exclusion Is Needed

Identify patients who are in the denominator (target population) but who should not receive the process or are not eligible for the outcome for some other reason, particularly where their inclusion may bias results. The intent of measure specification, therefore, 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 goal of the denominator inclusion and exclusion criteria is to have a population or sample with a similar risk profile in terms of meeting the numerator criteria. Though not all scenarios can be accounted for through measure specifications, measure developers consider the most appropriate care and clinical scenario through the measure development process. Particularly in the case of people with MCC, exceptions and exclusions will determine if the care for this potentially vulnerable group is examined.

Exception permits the exercise of clinical judgment and implies that the treatment was at least considered for, or offered to, each potentially eligible patient. They are most appropriate when contraindications to drugs or procedures being measured are relative, and patients who qualify for exclusion may still receive the intervention after the physician has carefully considered the entire clinical picture.[107] For this reason, most measures apply exception only to cases where the numerator is not met. Denominator exception is only used in proportion measures. It is not appropriate for ratio or continuous variable measures.

Following is an example of an exception allowing for clinical judgment in the case of two chronic conditions:

- 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 fails the measure.

[107] Spertus JA, Bonow RO, Chaun P, et al. ACCF/AHA New Insights Into the Methodology of Performance Measurement: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures. *Circulation.* 2010; 122:2091–2106.

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 reasons, patient reasons, and system reasons.

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. A patient's reasons for not receiving the service specified may be an exception to allow for patient preferences. System reasons are generally rare. They should be limited to identifiable situations that are known to occur (e.g., temporarily running out of a vaccine).

Examples:

- Medical reason: The medication specified in 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.
- Patient reason: The patient has a religious conviction that precludes the patient from receiving the specified treatment. The physician explained the benefits of the treatment and documented the patient's refusal in the record.
- System reason: 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 particular minority population. This may indicate a need for a more targeted, culturally appropriate patient education.

Exceptions may sometimes be reported as numerator positives rather than being removed from the denominator. This is sometimes done to preserve denominator size when there is an issue of small numbers, and out of respect for the clinical judgment and autonomy of the clinician. To ensure transparency, allowable exception—either included as numerator positives or removed from the denominator—must be captured in a way that they could be reported separately, in addition to the overall measure rate.

Denominator exclusion refers to criteria that result in removal from the denominator before determining if numerator criteria are met. Exclusion is absolute, meaning that the numerator event is not applicable and would not be considered for a population. Missing data should not be specified as an exclusion. Missing data may indicate a quality problem, so excluding those missing cases may present an inaccurate picture of quality. Systematic missing data (e.g., when poor performance is selectively not

reported) also reduces the validity of conclusions that can be made about quality.[108] One example of an exclusion would be patients with bilateral lower extremity amputations from a measure of foot exams.

An allowable exclusion or exception must be supported by:

- Evidence of sufficient frequency of occurrence such that the measure results will be distorted without the exclusion and/or exception
- Evidence that the exception is clinically appropriate to the eligible population for the measure
- Evidence that the exclusion significantly improves the measure validity.

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

Definitions of each population are provided in Table 8 and in the glossary. Although no single agreed-upon approach exists, there seems to be consensus that exception provide valuable information for clinical decision-making. Measure developers that build exception into measure logic should be cautioned that—once implemented—exception rates may be subject to reporting, auditing, and validation of appropriateness, and these factors need to be factored into the measure design and development. The difficulty in capturing exception as a part of clinical workflow makes the incorporation of exclusion more desirable in an EHR environment.

Table 8. Measure Populations Based on Type of Measure Scoring

	Initial Population	Denominator	Denominator Exclusion	Denominator Exception	Numerator	Numerator Exclusion	Measure Population	Measure Population Exclusion
Proportion	R	R	O	O	R	O	NP	NP
Ratio	R*	R	O	NP	R	O	NP	NP
Continuous Variable	R	NP	NP	NP	NP	NP	R	O

In the table above, R=Required, O=Optional, and NP=Not Permitted.

* Some ratio measures will require multiple Initial Populations: one for the numerator and one for the denominator.

16.2.5.1 Activities That Were "Not Done"

A negation rationale attribute may be used to identify situations where an action did not occur or was not observed for a documented reason. Negation rationale is used for the medical/patient/system reason for not doing something.

Example:

- Assessment, performed not done: "Medical Reason" for "Tobacco Use Screening" less than or equal to 24 month(s) starts before end of "Measurement Period"

See Section 3, Chapter 17, Codes, Code Systems and Datasets and Section 3, Chapter 17.2, Value Sets for a more complete discussion of negation rationale.

[108] National Quality Forum. *CSAC Guidance on Quality Performance Measure Construction*. May 2011. Available at: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=68999. Accessed on: January 25, 2017.

16.2.6 Mathematical Relationships between Populations in a Proportion Measure

This section provides further guidance on the precise mathematical relationships between populations in a proportion measure, and the process used to determine individual and aggregate scores. This ensures that all implementers arrive at the same scores, given the same data and same measures.

A proportion is a ratio in which the numerator is a subset (or part) of the denominator and can be written as a/(a+b). A proportion measure differs from a ratio 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 for ratio measures, these populations do not have a superset/subset relationship like proportion measures.

Figure 29 shows a fixed mathematical relationship between the populations in a proportion measure.

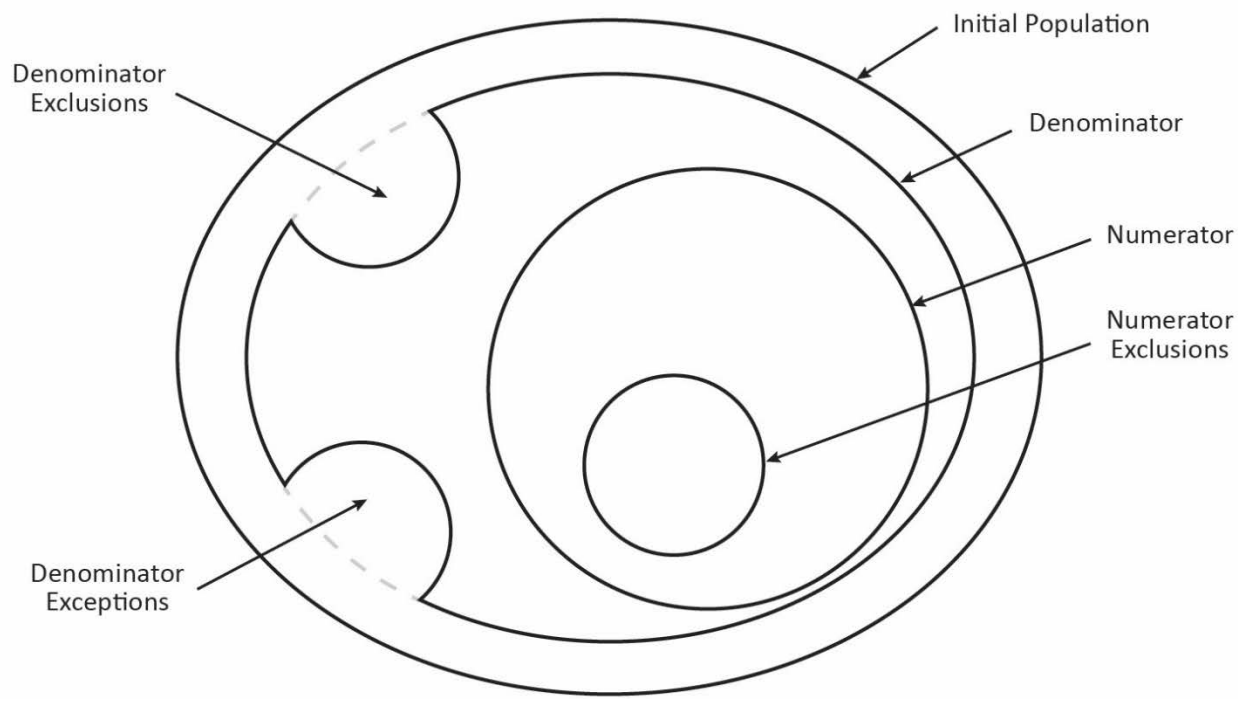

Figure 29. Proportion Measure Populations

From these relationships and definitions, the following sequential steps are used to determine whether a patient falls into a given population:

1. Initial population (IPOP): Identify those patients who meet the IPOP criteria.
2. Denominator (DENOM): Identify the subset of the IPOP 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 meet 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.

Queries 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 NUMER 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 DENOM 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.[109]

Specific programs may require reporting of performance rates. The performance rate of a proportion measure is defined as:

Performance Rate = (NUMER – NUMEX)/(DENOM – DENEX – 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 (100% performance rate would not be considered satisfactorily reporting for an inverse measure).
- 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 quality data codes are not counted in the performance denominator. If the provider submits all performance exclusion quality data codes, the performance rate would be 0/0 (null) and would be considered satisfactorily reporting.

16.2.6.1 Proportion Measure Examples

Following are some examples of the mathematical relationships between populations in proportion measures and the process to determine individual and aggregate scores.

[109] 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."

Example #1

A fictitious proportion measure defines the following population criteria:

- IPOP: All patients aged 65 years and older with an active diagnosis of diabetes mellitus
- DENOM: Equals IPOP
- 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.

Measure individual determination:

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.

- (IPOP = YES) Mr. Jones meets the IPOP 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) By definition, Mr. Jones does not meet the DENEXCEP criteria, because DENEXCEP criteria are not applicable to those meeting the NUMER criteria.

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.

- (IPOP = YES) Mr. Smith meets the IPOP 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.

- (IPOP = YES) Mr. Johnson meets the IPOP 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

A fictitious inverse proportion measure (where improvement is a decrease in the rate) defines the following population criteria:

- IPOP: Patients aged 8 to 65 who gave birth.
- DENOM: Equals IPOP.
- 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.

Measure individual determination:

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.

- (IPOP = YES) Mrs. Jones meets the IPOP 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.

- (IPOP = YES) Mrs. Thompson meets the IPOP 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.

- (IPOP = YES) Mrs. Hill meets the IPOP 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)

Measure Aggregate Calculations

Aggregate scores are simply the counts of individuals in each population. Thus, the aggregate IPOP is the count of individuals meeting the IPOP criteria.

The performance rate is a ratio of patients meeting NUMER criteria, divided by patients in the DENOM (accounting for exclusion and exception). Performance rate can be calculated using this formula:

Performance rate = (NUMER - NUMEX) / (DENOM – DENEX – DENEXCEP)

Building on Example #1, counting all individuals within the population, the following aggregate counts are determined:

- Initial population: N=150 (e.g., 150 patients meet the IPOP 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 = (NUMER – NUMEX) / (DENOM – DENEX - DENEXCEP) = (75-0) / (150-20-5) = 0.6.

16.2.7 Continuous Variable and Ratio Measure Calculations

This section provides further guidance on the precise mathematical relationships between populations in a continuous variable measure and a ratio measure, and the process to be used to determine individual and aggregate scores. This ensures that all implementers arrive at the same scores, given the same data and same measures.

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.

Aggregate scores for ratio and continuous variable (CV) measures are more complex than for proportion measures in that they are more than just the counts of individuals in each population. In addition to the identification of measure population(s), ratio and 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.

Measure Populations

There is a fixed mathematical relationship between the populations in a ratio or CV measure, as shown in Figures 30 and 31. 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).

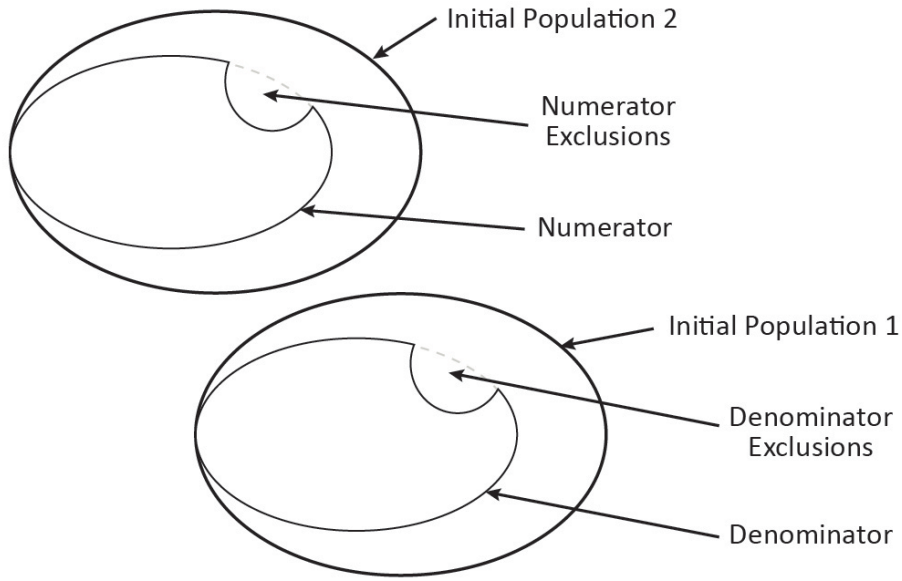

Figure 30. Ratio Measure Populations

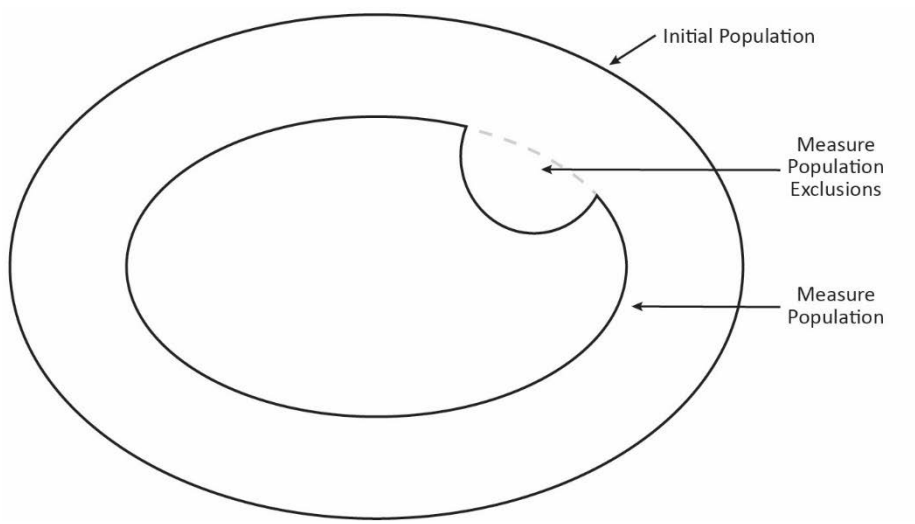

Figure 31. CV Measure Populations

From these relationships and definitions, the ratio measure query process is defined as:

- IPOP: Identify those cases that meet the IPOP criteria. (Some ratio measures will require multiple initial populations: one for the numerator and one for the denominator.)
- DENOM: Identify the subset of the IPOP that meet the DENOM criteria.
- DENEX: Identify the subset of the DENOM that meet the DENEX criteria.
- NUMER: Identify the subset of the IPOP that meet the NUMER criteria.
- NUMEX: Identify the subset of the NUMER that meet the NUMEX criteria.

The CV measure query process is defined as:

- IPOP: Identify those cases that meet the IPOP criteria.
- Measure population (MSRPOPL): Identify the subset of the IPOP that meet the MSRPOPL criteria.
- Measure population exclusion (MSRPOPLEX): Identify the subset of the MSRPOPL that meet the MSRPOPLEX criteria.

Queries 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 NUMER criterion is "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 DENOM 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.[110]

Individual Observations

CV measures, and frequently ratio measures, use quantities of interest other than population counts. The individual observations made for each case meeting population criteria are aggregated to produce the measure result. For instance, a CV measure might require that time spent waiting in the ED is measured for all patients meeting the measure population criteria, and might require average wait time be calculated for the measure population as a whole.

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.
- For a CV measure, an individual observation is determined for each case in the MSRPOPL and not in the MSRPOPLEX.

Measure Aggregate Calculations

Aggregate scores for ratio and CV measures are more complex than for proportion measures in that they are more than just the counts of cases in each population.

In ratio measures, for each population, individual observations are made for denominator and numerator cases, which are then used to calculate the aggregate ratio:

- Using individual observations for all cases in the DENOM and not in the DENEX, calculate the aggregate DENOM.
- Using individual observations for all cases in the NUMER and not in the NUMEX, calculate the aggregate NUMER.

Ratio = aggregate NUMER / aggregate DENOM

[110] 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".

In CV measures, an individual observation is made for each case falling into the measure population. 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.

Ratio Measure Example

Note: This is an example of the mathematical relationships between populations in a ratio measure and the process to be used to determine individual and aggregate scores.

A fictitious ratio measure looking at the rate of central line blood stream infections (i.e., the ratio of number of central line infections per total central line days) defines the following population criteria:

- IPOP: 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 blood stream 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

Determination of Measure Individual Observations

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 blood stream infection in his chart.

- (IPOP = YES) Mr. Jones meets the IPOP 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.

- (IPOP = YES) Mr. James meets the IPOP 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 following aggregate counts are determined:

- IPOP: N = 150 (i.e., 150 cases meet the IPOP 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

CV Measure Example

This is an example of the mathematical relationships between populations in a CV measure and the process to be used to determine individual and aggregate scores.

A fictitious CV measure defines the following population criteria:

- IPOP: 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:

- Aggregate MSRPOPL: Median minutes in ED

Determination of measure individual observations:

Mr. Jones is 75 years old and was admitted to the hospital from the ED. He spent 90 minutes in the ED.

- (IPOP = YES) Mr. Jones meets the IPOP 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.

- (IPOP = YES) Mr. James meets the IPOP 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 does not meet the MSRPOPL criteria.)

Measure aggregate calculations:

Building upon the examples above, but now considering all individuals within the population, the following aggregate counts are determined:

- IPOP: N = 150 (i.e., 150 cases meet the IPOP criteria)
- MSRPOPL: N =120
- MSRPOPLEX: N = 0

Aggregate MSRPOPL: 96 minutes (median of all individual Measure Population observations of time spent in the ED)

Additional information about eCQMs and proportion, ratio, and CV measure calculation can be found at the eCQI Resource Center.

16.3 Measure Stratification

Measure developers may define reporting strata, which are variables on which the measure is designed to report inherently (e.g., report different rates by type of Intensive Care Unit (ICU) in a facility; stratify and report separately by age group [14–19, 20–25, and total 14–25]). Measure stratification is described in a measure's specification document such as the "Specifications Manual for National Hospital Inpatient Quality Measures."[111]

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 three numerators, and the

[111] The Joint Commission. *Specifications Manual for National Hospital Inpatient Quality Measures.* Available at: https://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx. Accessed on: January 25, 2017.

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. When the measure definition includes stratification, each population in the measure definition should be reported both without stratification, and stratified by each stratification criteria.

For eCQMs, the Reporting Stratification section is included in the human-readable rendition. 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 is listed separately under the population criteria section.

16.4 Define the Data Source

The data source used to calculate a measure can determine reliability, usability, validity, and feasibility of the measure. Measure specifications should include the data sources and the method of data collection that are acceptable. This may be defined by the contract or be determined by the measure developer. If the measure is calculated from more than one data source, develop detailed specifications for each data source. Collect evidence that the results calculated from the different data sources are comparable.

16.4.1 Administrative Data

Electronic data often include transactional data that have been created for billing. This information can come from claims that have been submitted and adjudicated or from the provider's billing system.

Benefit programs categorize Medicare claims as follows:

- Part A is hospital insurance provided by Medicare. Part A covers inpatient care in skilled nursing facilities, critical access hospitals, and hospitals. Hospice and home healthcare are also covered by Part A.
- Part B covers outpatient care, physician services, physical or occupational therapists, and additional home healthcare.
- Part D is a standalone prescription drug coverage insurance administered by companies offering prescription drug plans.

Claims from each of these Medicare benefits have specific types of information and are unique sources of data containing data elements that can be used in the development of a quality measure. Claims data can be used if CMS or its measure developers will calculate the measure results.

Similar data elements may exist in the provider's billing system that can be used to produce claims. This information may be appropriate if the provider is to calculate the measure or identify cases for the denominator.

Other types of administrative data include patient demographics obtained from eligibility or enrollment information, physician office practice management systems, and census information. Payroll data and other databases containing information about providers can also be a source for some types of measures.

16.4.2 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. Information that is captured by an EHR but is not coded in a structured field, may require special processing by measure implementers. EHRs are one form of electronic clinical data, and these systems often include laboratory, imaging, and pharmacy data that can be queried and extracted for the measure.

16.4.3 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. Information manually abstracted from an EHR, which may include data from unstructured fields, clinical laboratory data, imaging services data, personal health records, and pharmacy data, may be used in a quality measure and should be considered the same or like a paper patient record. Manual extraction is labor-intensive.

16.4.4 Registry

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 system records all clinically relevant information about each patient, as well as the population of patients as a whole.

Registries may be components of an EHR of an individual clinician practice, or they may be part of a larger regional or national system that may operate across multiple clinicians and institutions. An example of a registry that is part of an EHR of an individual physician or practice is a diabetes registry. This type of registry identifies all the patients in the practice who have diabetes and tracks the clinical information on this set of patients for this condition.

Examples of national registries include the ACTION Registry-GWTG (from the American College of Cardiology Foundation and American Heart Association), the Society of Thoracic Surgeons Database, and the Paul Coverdell National Acute Stroke Registry. These registries generally collect data at the facility level that can be reported to the facility for local quality improvement or aggregated and reported at a regional or national level.

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.

16.4.5 Patient Assessments

CMS uses data items or elements from 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), the Minimum Data Set (MDS), and others.

16.4.6 Patient Reported Data/Surveys

Survey data may be collected directly from a patient (e.g., Consumer Assessment of Healthcare Providers and Systems (CAHPS) surveys that collect information on beneficiaries' experiences of care).[112] Surveys can provide the following advantages:

- Survey data (e.g., CAHPS) are readily available.
- Surveys ask about concepts such as satisfaction that are not available elsewhere.
- Surveys provide a unique window into patients' feelings.
- Surveys can collect patient reported outcomes.

16.5 Specify the Code Systems

Most CMS measures rely at least in part on the use of various standardized 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 may require copyright statements to accompany their use

See Section 3, Chapter 17, Codes, Code Systems, and Datasets for guidance on the section of concepts for inclusion in measures. Some measures with look-back periods that allow for patient population identification using historical data will need to consider the need to include code system content (codes) that are not active in the current code system version.

16.6 Construct Data Protocol

Explicitly identify the types of data and aggregate or link these data so that the measure can be calculated reliably and validly. Merging data from different sources or systems must be done 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
- 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.

Define Key Terms, Data Elements, and Code Systems

Terms used in the numerator or denominator statement, or in allowable exclusion/exception, 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 "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."

[112] CAHPS® is a registered trademark of the Agency for Healthcare Research and Quality (AHRQ).

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 is 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 BP < 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 ICD-10-CM code I21.xx, excluding those with I22.xx (AMI, subsequent episode of care).

Include enough detailed information in the denominator, numerator, exclusion, and exception 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.

16.6.1 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.

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.”
- Medication measures developed for use in populations (state or national level), Medicare Advantage plans, prescription drug plans, and individual physicians and physician groups.

16.6.2 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.

16.6.3 Determine Risk Adjustment

Risk adjustment is the statistical process used to identify and adjust for differences in patient characteristics (or risk factors) before examining outcomes of care. The purpose of risk adjustment is to facilitate a fairer and more accurate comparison of outcomes of care across healthcare organizations. Statistical risk models should not include factors associated with disparities of care as these factors will obscure quality problems related to disparities.[113]

All measure specifications, including the risk adjustment methodology, are to be fully disclosed. The risk adjustment method, data elements, and algorithm are to be fully described in the MIF[114] and the Risk Adjustment Methodology Report. 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 value assigned need not be disclosed because it varies over time. Section 3, Chapter 19, Risk Adjustment provides details of the procedure.

16.6.4 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 should be 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.

ISO 8601:2004 defines data elements and interchange formats for the representation of dates and times, including time intervals. The eCQI Resource Center provides a summary of important terms defined in the standard that are of particular importance and can be drawn upon to 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 days is the time interval and the hospital discharge date is the index event. If the minimum sensitivity and level of granularity desired was one month instead of 30 days, then the measure

[113] National Quality Forum. *CSAC Guidance on Quality Performance Measure Construction*. May 2011. Available at: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=68999. Accessed on: January 26, 2017.

[114] The NQF submission may be acceptable for this deliverable. National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: March 14, 2016.

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.

Appendix B: Time Interval Calculation Conventions in the QDM provides conventions that are intended to standardize the time calculation units for durations (e.g., difference between two date/time elements, typically with time relationships defined in the QDM, such as “starts after start of” and “ends before start of”). As with the Time Units and Time Interval Definitions, the Time Interval Calculation Conventions can be used for any type of CQM, not just eCQMs.

16.6.5 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.

- Categorical variable: A categorical variable groups items into predefined, discrete, non-continuous classes (male, female), (board certified, not board certified). Categories may reflect a natural order, in which case they are called ordinal (e.g., cancer stage: I, II, III, or IV); hospitals rankings: good, better, best).
- CV: A measure score in which each individual value for the measure can fall anywhere along a continuous scale (e.g., mean time to thrombolytics, which aggregates the time in minutes from a case presenting with chest pain to the time of administration of thrombolytics).
- Frequency distribution: A display of cases divided into mutually exclusive and contiguous groups according to a quality-related criterion.
- Non-weighted score/composite/scale: A combination of the values of several items into a single summary value for each case.
- Rate and proportion: 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).
- Ratio: A score that may have a value of zero or greater that is derived by dividing a count of one type of data by a count of another type of data. The key to the definition of a ratio is that the numerator is not in the denominator (e.g., the number of patients with central lines who develop infection divided by the number of central line days).
- Weighted score/composite/scale: A combination of the values of several items into a single summary value for each case where each item is differentially weighted (i.e., multiplied by an item-specific constant).

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

- Better quality = higher score
- Better quality = lower score
- Better quality = score within a defined interval
- Passing score defines better quality.

Avoid measures where improvement decreases the denominator population, unless they are based on Appropriate Use criteria (e.g., denominator—patients who received a diagnostic test; numerator—patients who inappropriately received the diagnostic test). With improvement, fewer will receive the diagnostic test).[115]

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. Also consider stratifying by population characteristics as CMS has a continued interest in identifying and mitigating disparities in clinical care areas/outcomes across patient demographics. Therefore, stratification may effectively detect potential disparities in care/outcomes among populations related to the measure focus. If results are to be stratified by population characteristics, describe the variables used.

Examples

- A vaccination measure numerator that includes the following: (1) the patient received the vaccine, (2) the patient was offered the vaccine and declined, or (3) the patient has an allergy to vaccine.
- Overall rate includes all three numerator conditions in the calculation of the rate.
- Overall rate is reported along with the percentage of the population in each of the three categories.
- Overall rate is reported with the vaccination rate. The vaccination rate would include only the first condition—that the patient received the vaccine—in the numerator.
- A measure is to be stratified by population type (e.g., race, ethnicity, age, social risk factors, income, region, gender, primary language, disability).

16.6.6 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 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.

16.7 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.[116]

[115] National Quality Forum. *CSAC Guidance on Quality Performance Measure Construction*. May 2011. Available at: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=68999. Accessed on: January 26, 2017.

[116] The NQF submission may be acceptable for this deliverable. National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: March 14, 2016.

Information from measure testing, the public comment period, 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.

The MIF and MJF have been aligned with the NQF measure submission 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 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 (LOINC, 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 the International Health Terminology Standards Development Organisation for SNOMED CT, that 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. This 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 is done 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. However, encoding content is critical to computable interoperability because it allows computer-based systems to find and operate upon data without human intervention. Encoding also benefits clinical interoperability, allowing 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 9) 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 Office of the National Coordinator for Health Information Technology (ONC) provides the latest Task Force's transitional vocabulary recommendations.

It is always important to use the appropriate standardized terminology for the task. When billing code systems, considered transitional for eCQMs are the focus of the analysis, then the measure should utilize the billing code system(s).

Table 9. ONC HITSC Recommended Vocabulary Summary

Clinical Terminology Standards	Primary Use
SNOMED CT	Clinical conditions, procedures, and general clinical information
Logical Observation Identifiers Names and Codes (LOINC)	Observables including laboratory test ordered or performed
RxNorm	Non-vaccine medications
CVX—Vaccines Administered	Vaccine medications
Centers for Disease Control and Prevention (CDC) Race and Ethnicity	Race and ethnicity
Unified Code for Units of Measure (UCUM)	Units of measure
ISO-639	Language
Public Health Data Standards Consortium (PHDSC) Source of Payment Typology	Source of payment

The transition vocabulary standards summary and plan are listed in Table 10.

Table 10. ONC HITSC Transition Vocabulary Standards Summary and Plan[117]

Transition Vocabulary	Transition period: Acceptable for reporting eCQM results	Final date for reporting eCQM results
International Classification of Diseases-9th Revision-Clinical Modification (ICD-9-CM)	With dates of service before the implementation of ICD-10	Not acceptable for reporting eCQM results for services provided after the implementation of ICD-10[118]
ICD-10-CM	With dates of service on or after the implementation of ICD-10	Final Date: Initially proposed one 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.
ICD-10-Procedure Coding System (PCS)	With dates of service on or after the implementation of ICD-10	Final Date: Initially proposed one 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.
Common Procedural Terminology (CPT)	During Stage 1, 2, 3 if unable to report using clinical vocabulary standards	Final Date: Initially proposed one 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.
Healthcare Common Procedure Coding System (HCPCS)	During Stage 1, 2, 3 if unable to report using clinical vocabulary standards	Final Date: Initially proposed one 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.

17.1.3 Quality Data Model (QDM) Categories with Recommended Code Systems

Table 11 provides guidance on the recommended code system to be used when a noted clinical concept is required for an eCQM; this includes general guidance for clinical concepts in any CQM as well as transitional vocabularies where specified.

Table 11. Quality Data Model Categories with ONC HITSC Recommended Vocabularies[119]

General Clinical Concept	Quality Data Model Datatypes	Quality Data Model Attribute	Clinical Vocabulary Standards	Transition Vocabulary
Adverse Effect/ Allergy/Intolerance	Device, Adverse Event Device, Allergy Device, Intolerance	Reaction	SNOMED CT	N/A
	Diagnostic Study, Adverse Event Diagnostic Study, Intolerance	Reaction	SNOMED CT	
	Immunization, Allergy Immunization, Intolerance	Reaction	SNOMED CT	
	Intervention, Adverse Event Intervention, Intolerance	Reaction	SNOMED CT	

[117] 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.
[118] ICD-9-CM can be used for lookback and historical data.
[119] These recommendations were made by HITSC in 2012 and 2015 using program information and language current at the time. With the adoption of the Quality Payment Program and other changes to quality reporting programs, these recommendations may be updated in the future.

General Clinical Concept	Quality Data Model Datatypes	Quality Data Model Attribute	Clinical Vocabulary Standards	Transition Vocabulary
	Laboratory Test, Adverse Event Laboratory Test, Intolerance	Reaction	SNOMED CT	
	Medication, Adverse Effects Medication, Allergy Medication, Intolerance	Reaction	SNOMED CT	
	Procedure, Adverse Event Procedure, Intolerance	Reaction	SNOMED CT	
	Substance, Adverse Event Substance, Allergy Substance, Intolerance	Reaction	SNOMED CT	
Patient Care Experience	Patient Care Experience	N/A	SNOMED CT (or LOINC if part of an Evaluation Tool)	NA
Substance	Substance, Administered Substance, Order Substance, Recommended	N/A	SNOMED CT	N/A
Condition/ Diagnosis/Problem	Diagnosis	N/A	SNOMED CT	ICD-9-CM, ICD-10-CM
Symptom	Symptom	N/A	SNOMED CT	N/A
Encounter (any patient-provider interaction, e.g., telephone call, email regardless of reimbursement status, status—includes traditional face-to-face encounters)	Encounter, Active Encounter, Order Encounter, Performed Encounter, Recommended	N/A	SNOMED CT	CPT, HCPCS, ICD-9-CM Procedures, ICD-10-PCS
Device	Device, Applied Device, Order Device, Recommended	N/A	SNOMED CT	N/A
Physical exam (definition of the components of the physical exam performed)	Physical Exam, Order Physical Exam, Performed Physical Exam, Recommended	N/A	LOINC	N/A
Physical exam (expression of the answers/responses for the physical exam component)	Physical Exam, Performed	Result	SNOMED CT or LOINC Normative Responses	N/A

General Clinical Concept	Quality Data Model Datatypes	Quality Data Model Attribute	Clinical Vocabulary Standards	Transition Vocabulary
Laboratory test (names)	Laboratory Test, Order Laboratory Test, Performed Laboratory Test, Recommended	N/A	LOINC	N/A
Laboratory test (results)	Laboratory Test, Performed	Result	SNOMED CT or LOINC Normative Responses	N/A
Diagnostic study test names	Diagnostic Study, Order Diagnostic Study, Performed Diagnostic Study, Recommended	N/A	LOINC	HCPCS
Diagnostic study test results	Diagnostic Study, Performed	Result	SNOMED CT or LOINC Normative Responses	N/A
Units of measure for results	N/A	N/A	UCUM – Unified Code for Units of Measure	N/A
Intervention	Intervention, Order Intervention, Performed Intervention, Recommended	N/A	SNOMED CT	CPT, HCPCS, ICD-9-CM Procedures, ICD-10-PCS
Procedure	Procedure, Order Procedure, Performed Procedure, Recommended	N/A	SNOMED CT	CPT, HCPCS, ICD-9-CM Procedures, ICD-10-PCS
Assessment instrument answers/responses (e.g., responses to questions for assessing patient status used as part of clinical workflow, clinical outcome evaluation, social functional and emotional status, patient preference, experience, characteristics)	Assessment, Performed	Result	SNOMED CT, or LOINC Normative Responses	N/A
Categories of function	Assessment, Performed Assessment, Recommended	N/A	ICF – International Classification of Functioning, Disability, and Health	N/A
Communication	Communication: from Patient to Provider Communication: from Provider to Patient Communication: from Provider to Provider	N/A	SNOMED CT	CPT, HCPCS

General Clinical Concept	Quality Data Model Datatypes	Quality Data Model Attribute	Clinical Vocabulary Standards	Transition Vocabulary
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)	Assessment, Performed Assessment, Recommended	N/A	LOINC	N/A
Medications (administered, excluding vaccines)	Medication, Active Medication, Administered Medication, Discharge Medication, Dispensed Medication, Order	N/A	RxNorm	N/A
Vaccines (administered)	Immunization, Administered Immunization, Order	N/A	CVX—Vaccines Administered	N/A
Patient characteristic, date of birth	Patient Characteristic Birthdate	N/A	Fixed to LOINC code 21112-8 (Birth date) and therefore cannot be further qualified with a value set	N/A
Patient characteristic, expired	Patient Characteristic Expired	N/A	Fixed to SNOMED-CT code 419099009 (Dead) and therefore cannot be further qualified with a value set.	
Patient characteristic, sex	Patient Characteristic Sex	N/A	ONC Administrative Sex – VSAC OID 2.16.840.1.113762.1.4.1 Definition Version: 20150331	
Patient characteristic, ethnicity	Patient Characteristic Ethnicity	N/A	Centers for Disease Control National Center for Health Statistics – VSAC OID 2.16.840.1.114222.4.11.837 Definition Version: 20121025 Detailed Ethnicity: HL7 Terminology – VSAC OID 2.16.840.1.114222.4.11.877 Definition Version: 20160213	N/A

General Clinical Concept	Quality Data Model Datatypes	Quality Data Model Attribute	Clinical Vocabulary Standards	Transition Vocabulary
Patient characteristic, race	Patient Characteristic Race		Centers for Disease Control National Center for Health Statistics – VSAC OID 2.16.840.1.114222.4.11.836□Definition Version: 20121025 Detailed Race: HL7 Terminology – VSAC OID 2.16.840.1.113883.1.11.14914 Definition Version: 20150424	
Patient characteristic, preferred language	Individual Characteristic	N/A	ISO-639-1:2002	N/A
Patient characteristic, payer	Patient Characteristic Payer	N/A	Payer Typology (PHDSC Payer Typology) (scope) VSAC OID 2.16.840.1.114222.4.11.3591 Definition Version: 20121025 http://www.phdsc.org/standards/pdfs/SourceofPaymentTypologyVersion5.0.pdf In PHIN VADS (vocabulary) http://phinvads.cdc.gov/vads/ViewCodeSystemConcept.action?oid=2.16.840.1.113883.221	N/A
Patient characteristic, unspecified	Patient Characteristic (unspecified)	N/A	SNOMED CT	N/A

17.1.4 Use of Specific Code Systems

17.1.4.1 International Classification of Diseases (ICD)

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 has transitioned from ICD-9-CM to ICD-10-CM/PCS beginning October 1, 2015. It is generally a good practice to not 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 Centers for Disease Control and Prevention (CDC) National Center for Health Statistics (NCHS) for use in all U.S. 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.[120]

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) identifying 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% to 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% to 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
- A09, Infectious gastroenteritis and colitis, unspecified—This is a standalone code and does not require any additional digits to be valid.
- A08, Viral and other specified intestinal infections—This is 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 that additional requirements also be met.[121]

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.

17.1.4.2 Current Procedural Terminology, Fourth Edition (CPT4)

CPT is a registered trademark of the American Medical Association (AMA) for the Current Procedural Terminology, 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.

---

[120] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *ICD-10 CM/PCS – The Next Generation of Coding*. Medicare Learning Network. ICN 901044; August 2016. Available at: http://www.cms.gov/ICD10/Downloads/ICD-10Overview.pdf. Accessed on: November 27, 2016.

[121] National Quality Forum. Inclusion of ICD-10 Codes in Measures Using ICD-9 Codes. http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=72793. Accessed on November 27, 2016.

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.[122]

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 via the Internet 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 over 357,000 healthcare concepts with unique meanings and formal logic-based definitions organized into hierarchies. The fully populated table 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. This includes conditions, diagnoses, symptoms, and signs, all of which are a type of "finding." SNOMED CT also represents procedures, observations including 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 above. 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, the core content architecture, and the SNOMED CT Core content. SNOMED CT Core 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 the Technical Implementation Guide.

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 name 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. Measure developers must sign up for a UMLS Terminology Services account to log into the USCRS. Measure developers should

[122] American Medical Association/Centers for Medicare & Medicaid Services. License for Use of Current Procedural Terminology, Fourth Edition ("CPT®). https://www.cms.gov/apps/physician-fee-schedule/license-agreement.aspx. Accessed on November 27, 2016.

consider contractual timelines when considering applying for new concepts. For information on obtaining the standard, contact: http://www.snomed.org/ or https://www.nlm.nih.gov/healthit/snomedct/index.html.

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 also 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:

- 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.

Radiopharmaceuticals, bulk powders, contrast media, food, dietary supplements, and medical devices such as bandages and crutches are all out of scope for RxNorm.[123]

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.

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.

[123] https://www.nlm.nih.gov/research/umls/rxnorm/overview.html

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. This reflects that, 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, currently, measures are only updated on an annual basis. Given that RxNorm (and all drug information vendor products) can be used to map from the more stable general identifier to a branded identifier, the branded RxNorm TTYs were often not included under the assumption that if an implementer had a branded ID, they could map to the included general RxNorm ID.

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 RxNorm should include only the ingredient (IN) or precise ingredient (PIN) TTY.

Measure Developer Guidance:

- Always consider including a measure clause 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, that a patient with an allergy/intolerance to any one of the substances will likely remove that 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," for example, "Antithrombotic Therapy Allergen" and "Beta Blocker Therapy Allergen."

17.1.4.4.6 Medication Value Sets

Medication value sets, when drawn from RxNorm, should include RxNorm concepts having TTYs:

- Generic Pack (GPCK)
- Semantic Clinical Drug (SCD).

17.1.4.5 Logical Observation Identifier Names and Codes (LOINC)

LOINC[124] is a common language (set of identifiers, names, and codes) for clinical and laboratory observations. 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, because of this, 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

[124] LOINC codes are copyrighted by Regenstrief Institute and the Logical Observation Identifier Names and Codes Consortium.

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 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 if the individual observations are copyright protected and if they can be used as independent observations outside of the complete survey instrument as a complete questionnaire. Many survey instruments include summary final scores (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 including data in a measure that is not already identified in the Quality Data Model (QDM) Categories with Recommended Code Systems section, determine if a specific authoritative code system is in widespread use and consider including that code system into the measure.

Examples of this include the following:

- CVX (for vaccines)
- 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 the following:

- 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 is a direct referenced code[125]
- 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*[126] (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 (codes) to be included in the value set *expansion*. A value set expansion is the actual list of codes that are to be used when the value set is implemented. Many value set definition CLDs are constructed by enumerating each specific code that is desired. This has traditionally been called an enumerated or *extensional* value set definition. All current VSAC value sets use this type of CLD. Some value sets would be best defined using the structure of the code system used, for example "All the codes that are descendants of the condition INSULIN DEPENDENT DIABETES MELLITUS." This is an example of an *intensional* value set definition. VSAC intends to provide tooling to support this type of definition in the near future.

Value sets have a lifecycle similar to many persistent objects. The VSAC is a tool suite developed by NLM to support the creation and maintenance of value sets. In addition to the life cycle noted in Figure 32[127], ongoing maintenance of value sets needs occur where value set authors modify the content to address improvements in clinical understanding, address changes in available coded concepts that occur with updates to the code system, and to address errors.

[125] Direct referenced codes will be introduced in the future.
[126] http://wiki.hl7.org/index.php?title=Value_Set_Definition_Standard_Project
[127] https://www.nlm.nih.gov/vsac/support/authorguidelines/authoringworkflow.html

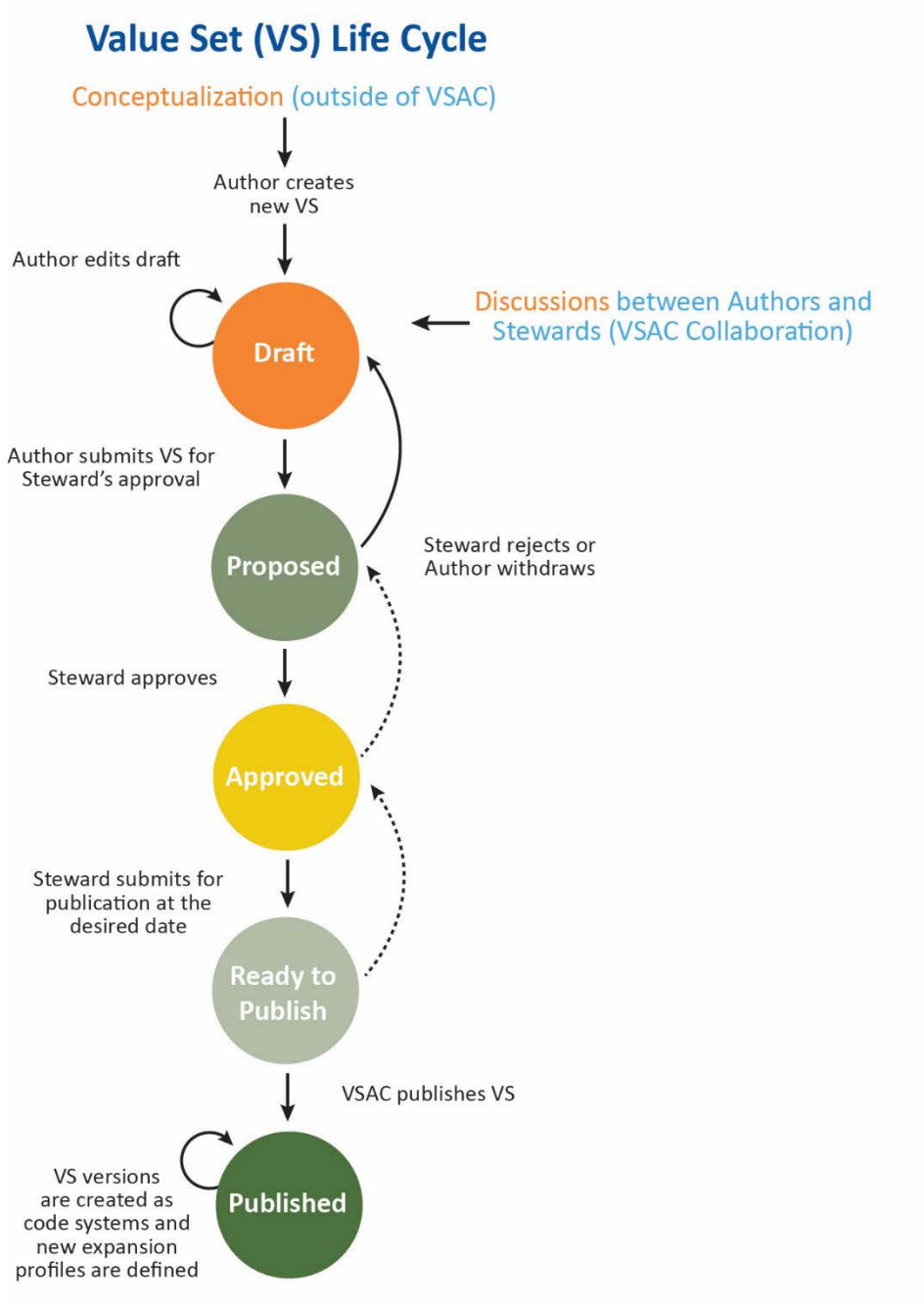

Figure 32. VSAC Value Set Lifecycle

17.2.4 Determining the Value Set Code System

Using the guidance noted in the Quality Data Model (QDM) Categories with Recommended Code Systems section, a value set may need concepts from more than one code system. This 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 child value set and a Diabetes Mellitus ICD-10-CM child value set. They then define a Diabetes Mellitus Grouping value set that combines the two subsets and associate the value set grouping with the measure data element. The VSAC allows only one level of grouping. Therefore, the MAT measure logic clause is needed to combine two value set groupings. 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 HIV infection might be similarly grouped.

17.2.4.1 Value Set Metadata

When creating a value set, the author must specify value set metadata that describe what it represents so that it can be used properly in measures and so that others can find it and reuse it in different measures where appropriate. to 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, the value set purpose (its scope is required and is intended to describe, in text for other humans to read), and the breadth of concepts that are to be included.

The required metadata for each value set includes:

- Name—developed based on guidelines presented on the VSAC Authoring Best Practices website. Include 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. This can be the information about clinical relevance or a statement about the general focus of the value set, such as a description of types of messages, payment options, geographic locations, etc.
- Data element scope—a free text statement describing how the Data Element in the intended information model (for example the 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. This 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. This is because the 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 value set definition. Measure developers need to notify NLM the version of the code system for the retired code and NLM will load that version of the code system into VSAC if that version is not available in VSAC.

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 should be used, rather than creating a new value set. This 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. Measure developers should define a value set as an enumerated list of codes. For example, a diabetes mellitus value set may include an enumerated list of fully specified ICD-10-CM codes, such as E10.1, E10.2, E10.3 and so on, and SNOMED CT codes as well. VSAC is the only authoritative tool to author value sets for eCQMs. 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. Some of these tools can take value set criteria (e.g., all ICD-10-CM codes beginning with E10) and expand them into an enumerated list. Where such value set criteria exist, they should be included as part of the value set definition.

When multiple value sets appear to represent the same thing, 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 with the VSAC Support Center.

Note that in the future, direct referenced codes may also be reused. VSAC will provide access to direct referenced codes in current use to allow such reuse and collaboration.

17.4 Value Set Authority Center (VSAC)

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 eCQM releases used in CMS and other quality reporting programs (e.g., The Joint Commission). Each value set consists of the codes (concept identifiers from specified code systems) and human-readable names (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 to about VSAC access, value set lifecycles and work flow, governance, Author and Steward roles, and best practices for value set development. In addition, the VSAC Support Center offers monthly users forums, catalogs release notes, and provides links to VSAC publications.

The key benefits of using VSAC include the following; for details please refer to the VSAC website:

- VSAC serves as the authority and central repository for the official versions of value sets that support the 2014 Edition eCQMs.
- VSAC provides search, retrieval, and download capabilities through a Web interface and APIs.
- VSAC provides authoring and validation tools for creating new and revising published value sets.
- VSAC hosts the 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.
- VSAC requires a 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.

- VSAC offers complete value set authoring guidance.

17.5 VSAC Collaboration

VSAC Collaboration[128] is a companion tool to the VSAC value set authoring and maintenance environment. This tool provides a central site where value set authors can post value sets for collaborative discussion. The site allows value set authors to collaborate, communicate, and provide workflow and document management by with stewards and their invited external collaborators. Value set authors and stewards can post value sets from the VSAC Authoring Tool into VSAC Collaboration to be viewed, discussed, or voted by collaboration group members. Like the VSAC Authoring Environment, the VSAC Collaboration tool requires a UMLS License Agreement for each user.

It is expected that VSAC authors will use VSAC Collaboration for most value sets to obtain input into the value set content from participating users and clinical experts. VSAC Collaboration also provides quality test and analysis of concept changes in defined value sets to support authors in the maintenance activity required to keep value set content aligned with code system version changes and clinical knowledge enhancements. It is important to note that any use of value sets can benefit from utilization of VSAC Collaboration; this is not focused solely on eCQM development.

As the VSAC Collaboration site utilization grows, more functionality will be added.

17.6 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: (a) 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*); (b) 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 (c) 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 this third definition.

QDM addresses negation rationale as an attribute of 73 QDM datatypes. By specifying negation rationale in a measure clause, the measure developer provides logic criteria requiring assertion that an action was not performed, most often allowable only for 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 Electronic Clinical Quality Improvement Resource Center describes how a measure developer can use negation rationale. Negation rationale is part of the current production

[128] National Institutes of Health, National Library of Medicine. VSAC Collaboration. Available at: https://vsaccollab.nlm.nih.gov/collab/page/user/sharkez/dashboard. Accessed November 27, 2016.

version of QDM and draft versions for use with Clinical Quality Language (CQL) logic. A discussion about negation and methods for expressing negation with CQL can be found on the CQL Formatting Wiki.

Negation rationale is presented in measure logic as *not done.* Using negation requires two value sets or, in the future, direct referenced codes, one to indicate the action that was not performed and the other to indicate the acceptable reason(s) for avoiding the expected action.

Example:

> Medication, Administered not done: *Angiotensin Converting Enzyme (ACE) Inhibitor* for *Medical Reason for avoiding ACE Inhibitors*
>
> ACE Inhibitor references the value set of the expected medications for which documentation asserts was not administered. Medical Reason references another 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.

In most of the 73 QDM datatypes, the same value set is used to indicate the action taken as the action not taken. Initially, eCQM reporting required clinicians to select a specific code from the referenced value set to indicate which one of the codes was not performed. This issue was problematic because the intent of negation is that none of the items of the value set was performed. Thus, the eCQMs required confusing documentation challenges. To avoid long lists of codes for medications, measure developers initially used the RxNorm ingredient codes to create value sets for medication actions with negation rationale (i.e., "not done"). Ingredient codes were more consistent with clinical activity avoidance, but that practice did not resolve the issue for other QDM datatypes. Indicating that a device was not applied caused similar issues since the HL7 standard on which reporting is based, QRDA could not directly indicate the device was negated. Some work in the HL7 standards space resolved the issue by allowing the clinician to indicate that an action was not performed (avoided) due to a reason. This new change allows the clinician to indicate "I did not perform *any* of the items in the value set" rather than having to choose a specific code. With that change, negation rationale with any applicable QDM datatype can use the same value set as the clause that seeks to determine if the action actually occurred. The new negation representation basically allows the HQMF to indicate that negation applies to "any" of the value set contents.

17.6.2 Value Set Package 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. 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. See Section 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 an Excel spreadsheet for value set review is that it is simple, easy to browse 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 an 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.

To overcome the disadvantages of an MS Excel spreadsheet, links to the value sets directly within the NLM's VSAC system is 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, although access to the VSAC requires that all users apply for 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. Because of the UMLS license, value set access and creation is available to all VSAC users. Implementers must make sure 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 the 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 (QA) team, internal or external terminologist, and steward as the NLM is not involved in this review. The steward will make the final decision of the selection of codes based on the existing data and feedback from clinicians. The QA team and terminologist will focus on the areas listed in Table 12 and take the appropriate remedial action. It is acceptable to use an MS 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 12. Value Set Review Areas and Remedial Actions

Area	Remedial Action
Value Set Duplication	Duplicate value sets should be replaced by normalized value sets.
Clinical Validity	Value sets must correspond to the intent and purpose of the clinical perspective.
Code List Completeness	A value set should contain all the relevant codes for a particular data element.
Metadata Completeness	Apply a common desirable pattern, implementing NLM VSAC guidance for extensional and grouping value set metadata.
Alignment of Code System to the Standards	Value sets must use recommended terminology systems for an extensional value set. Update code sets from transition vocabularies to those ideally desired.
Terminological Correctness	Only root codes and their descendants should be present in the value set. Combining of terminologies should use a grouping value set approach.
Single and Multiple Concepts	Extensional value sets should not combine more than one concept and terminology. Extensional value sets based on one concept may be reused in conjunction with other values sets to create a grouping value set for representing a combination of concepts.
Impact to Measure Logic Size	Sections of measure logic that deal with identical logic may be replaced by using grouping value sets which combine appropriate extensional value sets.

17.6.4 Post-MAT Value Set Review

The spreadsheet in the MAT output package is the main source for the Post-MAT value set review. 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 and subsequently shared with 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

To improve value set authorship, curation, and delivery, NLM performs QA 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 QA parameters. Measure developers should take proper actions as specified by NLM based on the analysis outcome. If the VSAC or NLM QA teams identify value set deficiencies, measure developers should correct the value sets using the VSAC Authoring Tool.

18 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 (numerator criteria)
- Have the same target population (denominator criteria)
- Apply to many measures (such as 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 depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.[129]

Measure alignment is defined as, "Encouraging the use of similar standardized performance measures across and within public and private sector efforts."[130] 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.[131] A core set of measures increases "signal" for public and private recognition and payment programs.[132] 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:

[129] National Quality Forum. *Changes to NQF's Harmonization and Competing Measures Process: Information for Measure Developers.* National Quality Forum; Washington, DC, 2013. Available at: https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=72716. Accessed on: May 31, 2015.
[130] Ibid.
[131] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *CMS Quality Strategy 2013 — Beyond*. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy.html. Accessed on: May 31, 2015.
[132] Conway PH, Mostashari F. and Clancy, C. The future of quality measurement for improvement and accountability. *Journal of the American Medical Association.* 2013; 309(21):2215–2216.

- 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 adapt an existing measure, respecify an existing measure, adopt an existing measure, or develop a new measure.

Harmonization should be considered when:

- Developing measure concepts, by:
    - Conducting a thorough environmental scan to determine if 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 13 summarizes ways to identify whether measures are related, competing, or new, and indicates the appropriate action based on the type of harmonization issue.

Table 13. Harmonization Decisions during Measure Development

Measure	Harmonization Issue	Action
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 (adapted) • Justify differences • Adapt 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 Adapted or 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 adapted. This process includes changing the numerator or denominator specifications or revising a measure to meet the needs of a different care setting, data source, or population. Alternatively, it may simply require adding new specifications to fit the new use. An example of these types of adaptations would be adapting the pressure ulcer quality measures used in nursing homes for use in other post-acute settings such as long-term acute care hospitals or inpatient rehabilitation facilities.

In adapting 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 adapted 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 adapting a measure for use in a new setting, a new population, or with a new data source, the newly adapted measure must be evaluated and possibly respecified and tested. Before deciding to adapt or respecify a measure already in existence, consider the following issues:

- 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 owner 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 owner. NQF endorsement status may need to be discussed with NQF. After making any changes to the numerator and denominator statement to fit the particular use, new detailed specifications will be required.

The first step in evaluating whether to adapt or 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 soundness 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 the following:

- Changes in the relative frequency of critical conditions used in the original measure specifications when applied to a new setting or population such as when there is a dramatic increase in the occurrence of exclusionary conditions.
- Change in the 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.
- Changes in the applicability of the original measure; for example, when the original measure composite contains preventive care components that are not appropriate in a new setting such as hospice care.
- Change in the feasibility of the data when changing the data source to an EHR.

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 particular to the measure's implementation use (such as data submission instructions). 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 a 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[133] (in development or planned for development). If there are no existing or related measures that can be adapted 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 developer or 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. The broader topic of measure harmonization is also discussed in Section 3, Chapter 18, Measure Harmonization. The following four steps taken 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 other 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 14 describes harmonization issues and actions based on the numerator and denominator specifications.

[133] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *CMS Measures Inventory.* Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/CMS-Measures-Inventory.html. Accessed on: May 31, 2015.

Table 14. Harmonization Decisions during Measure Maintenance

	Harmonization Issue	Action
Numerator: Same measure focus Denominator: Same target population	Competing measures	Replace the measure under reevaluation with the competing measure to promote alignment. Justify continuation of the additional measure.
Numerator: Same measure focus Denominator: Different target population	Related measures	Harmonize on measure focus (revise) Justify differences Adapt one of the existing measures by expanding the target population and retire the duplicative measure.
Numerator: Different measure focus Denominator: Same target population	Related measures	Harmonize on target population (revise) Justify differences
Numerator: Different measure focus Denominator: Different target population	Unique measures	Maintain existing measure

18.4.2 Implement Harmonization Decisions

After evaluating for harmonization, the following are possible outcomes:

- 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 where 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 then request that the 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

Performance measures include three basic types of measures used to assess the quality of healthcare: structure, process, and outcome.[134] Outcome measures assess results of healthcare experienced by patients: patients' clinical events, patients' recovery and health status, patients' experiences in the health system, and efficiency/cost. Outcomes depend on process of care because they are by definition the results of the actions of the healthcare system. Figure 33 depicts the relationship between structures, processes, and outcomes that measures should evaluate.[135]

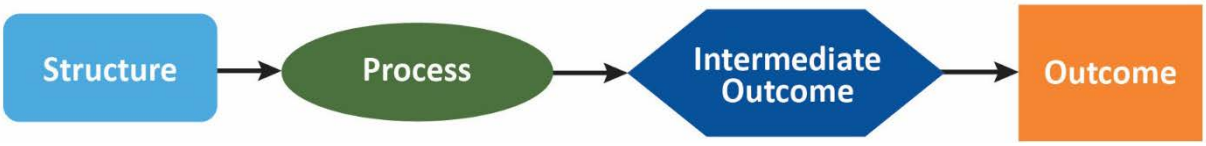

Figure 33. Structure Process Outcome Relationship

Multiple provisions in the ACA support the development of quality measures, and Section 10303 specifically calls for outcome measures to be developed.[136] Outcomes capture what patients and society care most about—the results of care. However, when constructing performance measures using outcomes, the outcomes must link to processes that are within the influence of providers or other entities being held accountable. The challenge when developing outcome measures that can be used for accountability is to ensure that the outcomes are indeed within the influence of the provider and not caused by intrinsic patient factors or other extraneous variables. For that reason, outcome measures are generally risk adjusted.

When outcomes are used as performance measures for assessing healthcare services and providers, there often needs to be a process of controlling for factors outside the influence of the providers included in those measures, which is risk adjustment. Many terms describe the concept of risk adjustment. Risk adjustment, severity adjustment, and case-mix adjustment are all often used to describe similar methods. All such methods are used either separately or in combination to "level the playing field" when comparing healthcare outcomes achieved by healthcare services and providers.

19.1 RISK ADJUSTMENT STRATEGIES

Information in this section references evidence-based risk adjustment strategies that encompass both statistical risk models and risk stratification, using terms employed by the NQF. 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. In this chapter, the term risk adjustment refers to the statistical process used to adjust for differences in population characteristics (i.e., risk factors) before comparing outcomes of care. Risk-adjusted outcome, risk factor, and risk adjustor are expressions related to a risk adjustment model. In contrast, the term risk stratification, as used in this chapter, refers to reporting outcomes separately for different groups, unadjusted by a risk model.

---

[134] Donabedian A. Explorations in quality assessment and monitoring. Vol. I. The definition of quality and approaches to its assessment, 1980; Vol. II. The criteria and standards of quality, 1982; Vol. III. The methods and findings of quality assessment and monitoring: an illustrated analysis. Ann Arbor: Health Administration Press, 1985.

[135] National Quality Forum. Patient-Reported Outcomes (PROs) in Performance Measurement. Jan 2013 Available at: http://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx. Accessed on: March 14, 2016.

[136] 111th Congress of the United States. *Patient Protection and Affordable Care Act*, 42 U.S.C & 18001 (2010). United States Government Printing Office. Available at: http://www.gpo.gov/fdsys/pkg/BILLS-111hr3590enr/pdf/BILLS-111hr3590enr.pdf Accessed on: March 14, 2016.

Within this framework of a risk adjustment strategy, the purpose of any measure risk adjustment model is to facilitate fair and accurate comparisons of outcomes across healthcare organizations, providers, or other groups.[137] Risk adjustment of healthcare outcome measures is encouraged because the existence of risk factors before or during healthcare encounters may contribute to different outcomes independently of the quality of care received. Accounting for social risk factors should be considered where appropriate. Strategies such as adjusting for social risk factors can help to avoid misleading comparisons. However, developers should also consider whether risk adjustment for these factors could obscure disparities in care for patients with social risk factors. Developers should also consider what method would be most appropriate for accounting for social risk factors (risk adjustment, stratification by groups within a measure, stratification at the measure level, etc). The exploration of a risk adjustment strategy (i.e., the use of a statistical risk adjustment model and, if necessary, risk stratification for selected populations)[138] is required for measures developed using the Blueprint. For a measure to be accepted by CMS and endorsed by NQF, the measure developer must demonstrate the appropriate use of a risk adjustment strategy and risk stratification, as needed. 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 if variation in factors intrinsic to the patient should be accounted for before outcomes can be compared and how to best apply these factors in the measure specifications. It is important to remember that risk adjustment does not itself provide the answers to study questions about measures, but instead provides a method for determining the most accurate answers.[139] 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 cannot currently 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 Attributes 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 attributes. Some of these attributes are listed in Table 15, which was partially derived from a description of preferred attributes of

[137] An example of different methods to adjust within and across groups is found in The American Academy of Actuaries May 2010 Issue Brief titled "Risk Assessment and Risk Adjustment" that discusses risk adjustment in the context of the Patient Protection and Affordable Care Act (PPACA) and issues needing attention and accommodation prior to the 2014 inclusion of small group markets.

[138] The National Quality Forum (NQF) has undertaken a 2-year trial period in which new measures and measures undergoing maintenance review have been assessed to determine if risk adjustment for social risk factors is appropriate for these measures. This trial entails temporarily allowing inclusion of social risk factors in the risk-adjustment approach for some performance measures. At the conclusion of the trial, NQF will issue recommendations on future inclusion of social risk factors in risk adjustment for quality measures.

[139] Management Decision and Research Center; Washington, DC: VA Health Services Research and Development Service in collaboration with Association for Health Services Research 1997. Risk Adjustment: A Tool for Leveling the Playing Field. W84.AA1 R595 1997.

models used for publicly reported outcomes.[140] Each of the attributes listed in the table are described in detail in the sections below.

Table 15. Attributes of Risk Adjustment Models

Attribute	Description
Sample definition	Sample(s) should be clearly defined, clinically appropriate for the measure's risk adjustment, and large enough for sufficient statistical power and precision
Appropriate time frames	Time frames for model variables should be clearly defined, sufficiently long to observe an outcome, and recent enough to retain clinical credibility
High data quality	The data should be reliable, valid, complete, comprehensive, and rely on as few proxy measures as can be accomplished with due diligence
Appropriate variable selection	Selected adjustment or stratification variables should be clinically meaningful
Appropriate analytic approach	Analytic approach must be scientifically rigorous and defensible, and take into account multilevel or clustered organization of data (if necessary)
Complete documentation	Risk adjustment and/or stratification details and the model's performance must be fully documented and all known issues disclosed

19.2.1 Sample Definition

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. Samples are intended to be microcosms such that the distributions of characteristics and their interactions should mimic those in the overall population. Researchers need to explain their rationale for using selected samples and offer justification of the sample's appropriateness.

19.2.2 Appropriate Time Frames

All of the criteria used to formulate decisions regarding the selection of the time frame should be clearly stated and explained in the measure documentation. Criteria used to identify risk factors for the stated outcomes should be clinically appropriate and clearly stated. 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).

[140] Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group. *Circulation* 2006;113, 456–62.

19.2.3 High Data Quality

The measure developer must ensure that the data used for risk adjustment are of high quality. Considerations in determining the quality of data include the following:

- The data are collected in a reliable way. That is, the method of collection must be reproducible with very little variation between one collection and another if the same population is the source.
- Data must be 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 must be 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.
- The data collected are as recent as possible. If the measure developer were using 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.
- The data collected are as complete as possible. The data should contain as few missing values as possible. Missing values are difficult to interpret and lower the validity of the model.
- Documentation of the data sources including when the data were collected, if and how the data were cleaned and manipulated, and the data's assumed quality should be 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 the 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.

Occasionally, less obvious 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; but, in certain instances, the same data may be used to develop a linear regression model when key statistical assumptions are not violated.[141] Selecting the correct statistical model is absolutely imperative, because an incorrect model can lead to entirely erroneous results. The analytic approach should also take into account 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 developer 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 all of the steps that were used to create a risk adjustment model, others cannot understand or fully evaluate the model. HHS policies emphasize transparency.[142] NQF policy on the endorsement of proprietary measures promotes the full disclosure of all aspects of a risk adjustment model used in measure development.[143]

The risk adjustment method used; performance of the risk adjustment model, its components, and its algorithms; and the sources of the data and methods used to clean or manipulate the data should be fully described. Documentation should be sufficient to allow others to reproduce the findings. The 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.

19.3 Risk Adjustment Procedure

The following seven steps are recommended in the development of a risk adjustment model:

- Choose and define an outcome
- Define the conceptual model
- Identify the risk factors and timing
- Acquire data (sample, if necessary)
- Model the data
- Assess the model
- Document the model.

[141] 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 but where the available data are dichotomous, such as the probability of receiving a specified healthcare service.

[142] Department of Health and Human Services. *Grants Management Information: Grants Policy Topics.* Available at: http://www.hhs.gov/grants/grants/grants-policies-regulations/index.html. Accessed on: March 14, 2016.

[143] National Quality Forum. . *http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx*. Accessed on: December 4, 2016..

Some models may not lend themselves appropriately to all of these steps, and an experienced statistician and clinical expert can determine the need for each step. A list of risk adjustment deliverables can be found in Figure 34.

19.3.1 Choose and Define an Outcome

Though risk adjustment should not be applied to structure and process measures that are entirely within the measured provider's control, risk adjustment may be necessary for outcome measures that are not fully within the measured providers' control (e.g., readmission rates, mortality, and length of stay).[144] When selecting outcomes that are appropriate for risk adjustment, the time frame 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 allow 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 3, 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.

Risk Adjustment Deliverables

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.
2. Measure Information Form (MIF) with completed risk adjustment sections for each measure.
3. For electronic clinical quality measures (eCQMs), HQMF, SimpleXML file, and human-readable rendition (HTML) file that include instructions where the complete risk adjustment methodology may be obtained.

Figure 34. Risk Adjustment Deliverables

Risk variables in risk-adjusted outcome eCQMsare 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 the 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 modern computing power allow for

[144] National Quality Forum. *National Voluntary Consensus Standards for Ambulatory Care—Measuring Healthcare Disparities: A Consensus Report.* Washington, DC; National Quality Forum; Mar 2008; Report No. ISBN 1-9338 75-14-3. Available at: http://www.qualityforum.org/Publications/2008/03/National_Voluntary_Consensus_Standards_for_Ambulatory_Care%E2%80%94Measuring_Healthcare_Disparities.aspx. Accessed on: March 14, 2016.

statistical routines (e.g., jackknifing) to explore the data for relationships between outcomes and potential adjustment factors that might not yet be clinically identified, but empirically exist.

The first step in developing or selecting the conceptual model is identifying the relationship among variables. This process should include the following:

- Conducting a review of clinical literature and canvassing expert opinion to establish variable relationships and resolvable confounds.
- Obtaining expert opinion. The experts consulted should include 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 (TEP) covers the standardized process used for convening a TEP.
- When appropriate data are available, automated computer routines can be used 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 attributes:

- Clinically relevant
- 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. CMS also generally precludes the use of risk factors that obscure disparities in care associated with race, social risk factors, and gender factors. Below examples of factors that generally should not be used in risk adjustment models for quality measures, even if their inclusion improves the predictive ability of the model. When such factors exist, it may be more appropriate to develop measures that stratify the population rather than include the factors in a risk adjustment model.

19.3.3.1 Race or Ethnicity

Populations representing certain races/ethnicities are at different levels of risk for disease and mortality, and they have different level-of-care needs. Risk models should not obscure disparities in care for populations by including factors that are associated with differences or inequalities in care such as race or ethnicity. It is preferable to stratify by this factor, rather than use it in a risk adjustment model.

19.3.3.2 Medicaid Status

Though there have been CMS programs that warrant the use of Medicaid status in a risk adjustment model (e.g., Medicare HOS or comparisons between groups with diverse proportions of dual-eligible beneficiaries), CMS prohibits the use of risk factors that obscure differences associated with social risk factors. Consequently, it is preferable to stratify by this factor or to consult with CMS prior to including it in any competing or alternate risk adjustment model.

19.3.3.3 Gender

Males and females often show differences relative to treatment, utilization, risk levels for mortality and disease, and necessary level of care. Though restricting a measure to a single gender using exclusion criteria may be reasonable, inclusion of gender in a risk model should be avoided to prevent obscuring disparities in care. It is preferable to stratify by gender, rather than use it in a risk adjustment model.

19.3.4 Acquire Data (Sample, if Necessary)

Healthcare data can be acquired from many sources, but the three most frequently used are administrative data, patient record data, and survey data. Of these, the most common source of data for developing risk adjustment models is administrative data reported by the provider. Once the 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 2, Chapter 3, Measure Testing provides more details of the processes.

19.3.5 Model the Data

In addition to the clinical judgment used to define the conceptual model and candidate variables, empirical modeling should also be conducted to help determine the risk factors to include or exclude. A number of concerns exist in data modeling and the following 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 not particularly rare 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.[145] 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 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 (overfitting), and can also result in reduced stability of parameters due to increased multicollinearity among a larger number of predictors. In contrast, a model with fewer predictors may reduce the amount of explained variance possible for the measure (underfitting).

[145] Lezzoni LI. Risk Adjustment for Measures Health Care Outcomes—Third Edition. Chicago, IL: Health Administration Press; 2002.

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 multicollinearity 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.[146] 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 little improvement in predictive validity or overall model fit, and by combining clinically similar conditions to improve performance of the model across time and populations.

19.3.5.3 Methods to Retain/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/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.[147] 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 jackknife analysis and assessment of the model in multiple/diverse samples (refer also to the Generalizability section below). Consultation of clinical expertise, ideally used during candidate variable selection, is also strongly recommended when examining the performance of candidate variables in the 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 (or calibration) dataset and a validation dataset. The development (or calibration) dataset is used to develop the model (or calibrate the coefficients), and the validation dataset is used to determine the extent to which the model can be appropriately applied to the 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 allows 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 the stability of the model over time.

---

[146] In situations with high visibility or potentially wide-spread fiscal repercussions, CMS has employed some of the most sophisticated models available, such as Hierarchical Generalized Linear Models (Statistical Issues in Assessing Hospital Performance, Commissioned by the Committee of Presidents of Statistical Societies, November 28, 2011).

[147] Harrell, Frank E. Regression Modeling Strategies, with Applications to Linear Models, Survival Analysis and Logistic Regression. New York, NY: Springer; 2001.

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, who may in turn be nested under hospitals). The risk adjustment model should account for these multilevel relationships, when present, and risk adjustment development should investigate theoretical and empirical evidence for potential patterns of correlation in this multilevel data. For example, patients in the same inpatient rehabilitation facility (IRF) may tend to have similar outcomes based on a variety of factors, and this should be addressed by the risk adjustment model.

Such multilevel relationships are often examined by building models designed to account for relationships between observations within larger groups. Terms for these types of models include multilevel 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/characteristic being measured is the same across samples (e.g., male versus female, nonprofit versus 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 (such as 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.

Multilevel 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 take into account variation at different grouping levels and allow modeling of hypothesized factors at these different levels. For example, a multilevel 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 available. 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.[148]

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 is 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 the 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), and Hosmer-Lemeshow test. However, several other statistical techniques exist that allow 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 the values of the included risk factors.

The R^2 value for a model can vary, and no firm standard exists for what is the optimal expected value. Past experience or previously developed models may inform what 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, AUC, and C-statistic

A 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 35. 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

[148]Committee of Presidents of Statistical Societies-Centers for Medicare & Medicaid Services. (2011). *Statistical issues in assessing hospital performance [White Paper]*. Revised Jan 27, 2012. Available at http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf. Accessed on: March 14, 2016.

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 area under the curve (AUC). The AUC, sometimes referred to as the c-statistic, is a value that varies from 0.5 (discriminating power not better than chance) to 1.0 (perfect discriminating power). It 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 (the null hypothesis). Both non-parametric and parametric methods exist for calculating the AUC, and this varies by statistical software.

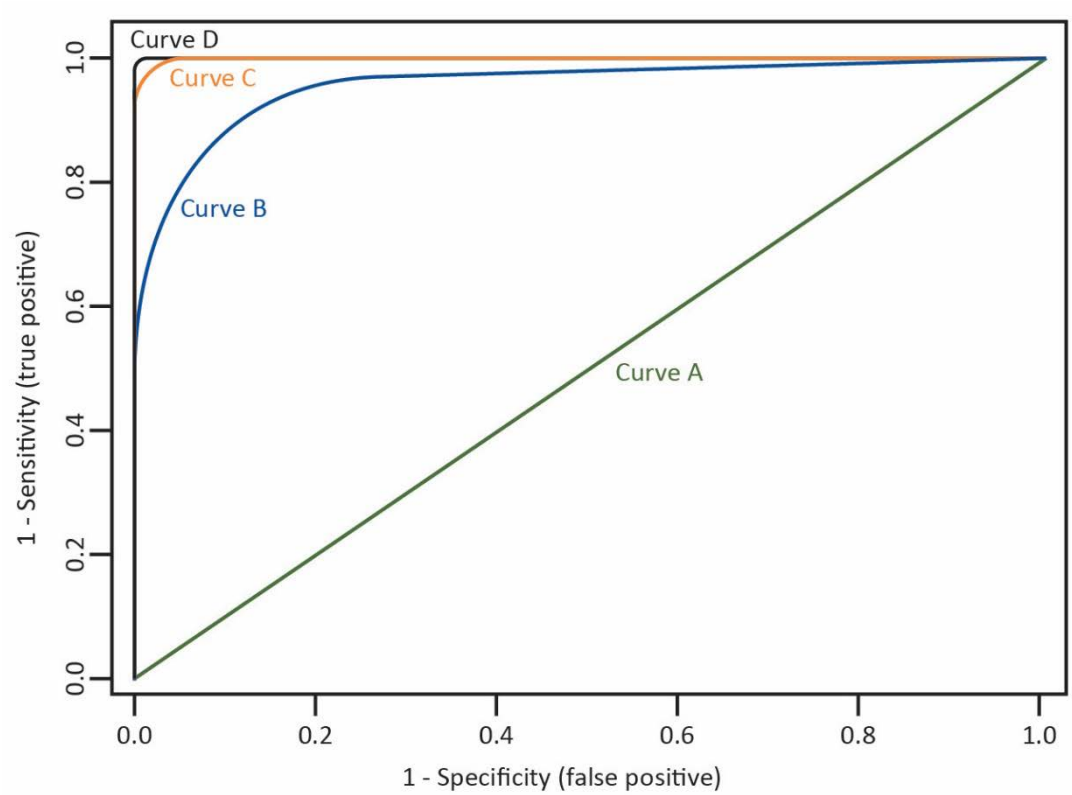

Figure 35. Example of ROC Curves

19.3.6.3 The Hosmer-Lemeshow Test

Though 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 Hosmer-Lemeshow 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 Hosmer-Lemeshow 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 that should be addressed by an experienced statistician.

A statistician with experience in such methodology can determine the adequacy of any model. It is expected that the measure developer 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 what is cited here. For example, a risk adjustment model may discriminate very well based on the c-statistic but 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 percent 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 percent 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, the 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 at hand are also expected to be consulted (or employed) 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 considered a required deliverable and is expected at the conclusion of a measure development project. This report ensures that relevant information about the development and limitations of the risk adjustment model are available for review by consumers, purchasers, and providers. It also allows 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 the following:

- Identification or review of the need for risk adjustment of the measures
- 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. For example, classifying a high-outcome provider as a low one or the reverse[149]

[149] Austin PC. Bayes rules for optimally using Bayesian hierarchical regression models in provider profiling to identify high-mortality hospitals. *BMC Medical Research Methodology.* 2008;8:30.

- Enough summary information about the comparison between unadjusted and adjusted outcomes to evaluate if 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 later updated as needed.

All measure specifications, including the risk adjustment methodology, must be fully disclosed. The risk adjustment method, data elements, and algorithm are to 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

It is important to submit instructions and analytic steps for aggregating data when designing cost and resource use measures.[150] This 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, the reliability and validity of the data 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 episode
- 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. The analytic steps are designed to create appropriately homogeneous units for measurement.

20.2 Measure Construction Logic

20.2.1 Time Frames

Decisions about when to start or end a measurement period must be specified for each measure. These time frames 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 which 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.

[150] National Quality Forum. *National Voluntary Consensus Standards for Cost and Resource Use.* 2012. Available at: http://www.qualityforum.org/Publications/2012/04/National_Voluntary_Consensus_Standards_for_Cost_and_Resource_Use.aspx. Accessed on: March 14, 2016.

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 where known disparities exist or where 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:

- The count of services
- The 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, the care of a population, or event to a provider (e.g., physician, physician groups) or other entity (e.g., health plan) and 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 particular 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 is designed to identify the decision maker, perhaps the primary care physician, and hold 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 (higher or lower performance is better), preferred resource use amounts often are not standardized; and it is not always clear if 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

Observed-to-expected (O/E) ratio compares the value for each resource use measure attributed to a physician or entity (observed amount) and divides it by the average resource use within the identified peer group (expected amount—the amount of resource use expected if the entity measured were performing at the mean).

More sophisticated statistical approaches such as 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, 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. All of these actions should be documented so users can understand the full context.

20.4.5 Providing Detailed Feedback

After all of 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 simply 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

Though technical specifications of all components of the composite may already be documented, they should also be completed for the composite. The Measure Information Form and Measure Justification Form are aligned with the requirements of the NQF measure submission and guide the measure developer to ensure that the 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 all of the component measures may not meet all of 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 24, Measure Evaluation.

The methodology and considerations for scoring includes ensuring that the weighting and scoring of the components support the goal that is articulated for the measure. Then, using a specified method, combine the component scores into one composite.

Descriptions of five common types of composite performance measure scoring are provided in

Table 16. 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 as follows:

- All-or-none
- Any-or-none
- Linear combinations
- Regression-based composite performance measures
- Opportunity scoring.

Table 16. Types of Composite Measure Scoring

Type of Scoring	Advantages	Disadvantages	Examples/Evidence
All-or-None (Defect-free Scoring) Process Measures The patient is the unit of analysis. Only those patients who received all indicated processes of care are counted as successes. Performance is defined by the proportion of patients receiving all of the specified care processes for which they were eligible. No credit is given for patients who receive some but not all required items.	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 in which achieving a desired clinical outcome empirically requires reliable completion of a full set of tasks (that is, when partial completion does not gain partial benefit).	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 (one unreliable component measure will contaminate the whole score), so it is essential that each of the component measures be well designed.	Minnesota Community Measurement Optimal Diabetes Care measure. Institute for Healthcare Improvement (IHI) Bundles: ventilator, central line. Society of Thoracic Surgeons (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) 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.[151]
Any-or-None Outcome Measures Similar to all-or-none, but is used for events that should not occur. The patient is the unit of analysis. A patient is counted as failing if he or she experiences at least 1 adverse outcome from a list of 2 or more adverse outcomes.	Promotes a high standard of excellence. Useful when component measures are rare events.	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.	STS Postoperative Risk-Adjusted Major Morbidity, 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.

[151] Stulberg JJ, Delaney CP, Neuhauser DV, Aron DC, Fu P, Koroukian SM. Adherence to Surgical Care Improvement Project Measures and the Association with Postoperative Infections. *The Journal of the American Medical Association.* 2010;303(24):2479-2485. doi:10.1001/jama.2010.841.

Type of Scoring	Advantages	Disadvantages	Examples/Evidence
Linear Combinations Can be simple average or weighted average of individual measure scores.	Has the advantage of simplicity and transparency.	Does not account for potential differences in the validity, reliability, and importance of the different individual measures. Equal weighting may be undesirable if there is a considerable imbalance in the numbers of measures from different domains. Different stakeholders have different priorities; one weighting method may not meet the needs of all potential users. 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. 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 allow 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.	Premier/CMS Hospital Quality Incentive Demonstration uses a composite of process and outcome to measure quality for coronary artery bypass graft (CABG). The composite quality score (CQS) was based on an equally weighted combination of seven measures (four process measures and three outcome measures). The actual 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. The US News & 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 3 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. The AHRQ Patient Safety Indicators (PSI) composite performance measure uses a weighted average of various individual component measures. The weighting was determined by an expert panel.
Regression-based Composite Performance Measures If a certain outcome is regarded as a gold standard, the weighting of individual items may be determined empirically by optimizing the predictability of the gold standard end point.	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.

Type of Scoring	Advantages	Disadvantages	Examples/Evidence
Opportunity Scoring Opportunity scoring counts the number of times a given care process was actually 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.	Provides an alternative to simple averaging often used for aggregating individual process measures. Has the advantage of increasing the number of observations per unit of measurement, consequently potentially increasing the stability of a composite estimate, particularly when the sample size for individual measures is not adequate.	Rate is influenced by the most common care processes, regardless of whether they are the most important methods.	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 Hospital Quality Incentive (HQI) Demonstration project uses the opportunity scoring method for the process composite rate for each of 5 clinical areas. The sum of all the numerators is divided by the sum of all the denominators in each clinical area.

Composite eCQMs are not yet possible. HQMF does support expressing component measure metadata in the header. Future adoption of CQL will support composite eCQMs. More information may be found on the eCQI Resource Center and the CQL Formatting and Usage Wiki.

22 MEASURE TESTING

The 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.

See also Section 2, Chapter 3, Measure Specification for details on the interrelationships and chronology of the following measure testing steps:

- Develop the Measure Testing Plan
- Submit the plan and obtain CMS approval
- Implement the plan
- Analyze the test results
- Refine the measure
- Retest the refined measure
- Compile and submit deliverables to CMS
- Support CMS during NQF endorsement process.

For details on alpha and beta testing as well as testing considerations for selected measure types, see Section 3, Chapter 23, Alpha and Beta Testing. For more information about special considerations pertaining to eCQM testing, see Section 3, Chapter 25, Testing for Special Types of Measures.

22.1 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 the measure specifications have been developed, and it 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 the following:

- Name(s) of measure(s)
- Type of testing (alpha or beta; Section 3, Chapter 23, Alpha and Beta Testing)
- Study objective(s)
- The timeline for the testing and report completion
- Data collection methodology
- Description of test population; include number and distribution of test sites/data sets, when available
- Description of the data elements that will 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 a description of test statistics that will 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 such as means and detectable differences, dispersion of comparison groups, etc. More formal testing, if requested by CMS, may require assessment via structured surveys or focus groups to evaluate the 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[152]
- Training and qualification of staff. For example, identifying those who:
    - Manage the project (and their qualifications)
    - Conduct the testing (and their qualifications)
    - Conduct or oversee data abstraction
    - Conduct or oversee data processing
    - Conduct or oversee data analysis.

22.2 Submit the Plan and Obtain CMS Approval

Submit the work plan to the COR with any necessary supporting documents. Revise as necessary to meet CMS approval.

22.3 Implement the Plan

Following COR review and approval, execute the approved work plan.

22.4 Analyze the Test Results

Once all the 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.

[152] 104th Congress of the United States. *Paperwork Reduction Act. United States Government Printing Office.* Available at: http://www.gpo.gov/fdsys/pkg/PLAW-104publ13/pdf/PLAW-104publ13.pdf. Accessed on: March 14, 2016.

22.5 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 the testing results.

Example:

- Following alpha testing, measure respecification or efforts to overcome implementation barriers are often undertaken.
- Following beta testing, changes in the definition of the population or adjustments to the comparison group definition may occur.
- If changes to the measure are made, consultation with the TEP is recommended prior to retesting the measure.

22.6 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.

22.7 Compile and Submit Deliverables to CMS

Communicate findings of the measure testing with revised measure specifications to CMS for review. Update the Measure Information Form with revised specifications, and update the Measure Justification Form 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.[153]

Based on the results from beta testing, prepare a Measure Evaluation Report for each measure to summarize how well the 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.

22.7.1 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 summarize the measure testing findings. The goal of these summaries is to document sufficient evidence to support approval by CMS 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. This 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 also summarize expert findings/consensus with respect to measure:

[153] The NQF submission may be acceptable for this deliverable. National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: March 14, 2016.

- Importance
- Acceptability
- Usability
- Feasibility.

The following are recommendations for the content of the measure Testing Summary Report. 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 the characteristics of the measure.

The summary of testing may include the following:

- Name of measure or measure set.
- An executive summary of the tests and resulting recommendations.
- Type of testing conducted (alpha or beta), and an overview of the 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, and 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 the 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.
- A detailed discussion of testing results compared to NQF requirements, including whether the NQF requirements are sufficiently met or if additional testing is required.
- Examples of limitations of the alpha or beta testing:
    - The sample limited to two sites or three EHR applications.
    - The sample used registry data from one state, and registry data are known to vary across state.
    - Testing was formative alpha test only and was not intended to address validity and reliability.

22.8 Support CMS during NQF Endorsement Process

If the measure(s) will be submitted to NQF for endorsement, the measure developer helps the COR, as directed, by completing the measure submission including results of the measure testing. Information documented in the Measure Information Form should be used to complete the NQF submission. Measure developers also provide additional information as needed and are available to discuss testing results with NQF throughout the endorsement process.

23 ALPHA AND BETA TESTING

Testing provides an opportunity to refine the draft specifications before they are finalized; augment or reevaluate earlier judgments about the measure's importance; and assess the feasibility, usability, and scientific acceptability of the measure (For more information on measure testing as it relates to evaluation criteria (see Section 3, Chapter 24, Measure Evaluation).

Initial testing during development (sometimes referred to as pilot testing) is generally conducted within the framework of alpha and beta tests. Though 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 the development of measure specifications.

Attributes of each phase of testing are shown in Table 17, and these may be used as considerations when developing a work plan for alpha or beta tests.

Table 17. Features of Alpha and Beta Testing

	Alpha Testing	Beta Testing
Timing	• Usually carried out prior to the completion of technical specifications • May be carried out multiple times in quick succession	• After the measure developer's detailed and precise technical specifications are developed
Scale	• Typically smaller scale • Only enough records to ensure data set contains all elements needed for the measure • Only enough records to identify common occurrences or variation in the data	• Strives to achieve representative sample sizes • Requires appropriate sample selection protocols • May require evaluation of multiple sites in a variety of settings depending on the data source (e.g., administrative, medical chart)
Sampling	• Convenience sampling	• Sufficient to allow adequate testing of the measure's scientific acceptability • Representative of the target population • Representative of the people, places, times, events, and conditions important to the measure • If based on administrative data, use the entire eligible population
Specification Refinement	• Permits the early detection of problems in the technical specifications (e.g., identification of additional inclusion and exclusion criteria)	• Used to assess or revise the complexity of computations required to calculate the measure
Importance	• Designed to look at the volume, frequency, or costs related to a measure topic (cost of treating the condition, costs related to procedures measured, etc.) • Establishes on a preliminary basis that the measure can identify low levels of care quality • Provides support for further development of the measure	• Allows for enhanced evaluation of a measure's importance including evaluation of performance thresholds and outcome variation • 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)
Scientific Acceptability	• Limited in scope if conducted during the formative stage. Usually occurs later in development	• Assesses measure reliability and validity • Reports results of analysis of exclusion (if any used) • Tests results of risk adjustment model, quantifying relationships between and among factors

	Alpha Testing	Beta Testing
Feasibility	• Provides initial information about the feasibility of collecting the required data and calculating the measures using the technical specifications • Identifies barriers to implementation • Offers an initial estimate of the costs or burden of data collection and analysis	• Provides enhanced information regarding feasibility including greater determination of the barriers to implementation and costs associated with measurement • Evaluates the feasibility of stratification factors based on occurrences of target events in the sample • Identifies unintended consequences, including susceptibility to inaccuracies and errors • Reports strategies to ameliorate unintended consequences
Usability	• No formal analytic testing at this stage. The TEP may be used to assess the potential usability of the measure	• May consist of focus groups or similar means of assessing usefulness of the measure by consumers. This type of testing is often not in the scope of measure development contracts • The TEP may also be used to assess the potential usability

23.1 Alpha Testing

Alpha tests (also called formative tests) are of limited scope since they usually occur before detailed specifications are fully developed. Alpha testing, particularly regarding the feasibility of the concept in the context of the data source, may be conducted as part of the information gathering empirical analysis. Alpha testing may also be performed concurrently with the development of the technical specifications as part of an iterative process. The alpha tests include methods to determine if individual data elements are available and if 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 very little alpha testing. In contrast, measures that address areas for which specifications have never been developed may require multiple iterations of an alpha test. 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. The results can impact decisions about what is included in a measure.

23.2 Beta Testing

Beta testing (also called field testing) generally occurs after the 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. They 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.

23.3 Sampling

The need for sampling often varies depending on the type of test (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. 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 particular 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.

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 general public. However, with the passage of the MACRA, data collection for quality measure development is now exempt from PRA requirements. [154] Measure developers should consult with their COR about the ramifications of the PRA and MACRA exemption before requesting information from the public.

[154] Chapter 35, Title 44 of United States Code. Available at: http://www.gpo.gov/fdsys/search/pagedetails.action;jsessionid=p6ThV2Hd4MhNPP8ZJYKpF77RLxvvmM6wv122Kg8n9N1f7cbrTnqF!-1506635365!930041035?collectionCode=USCODE&searchPath=Title+44%2FCHAPTER+35&granuleId=USCODE-2008-title44-chap35&packageId=USCODE-2008-title44&oldPath=Title+44%2FChapter+35%2FSubchapter+I%2FSec.+3510&fromPageDetails=true&collapse=true&ycord=900. Accessed on July 27, 2015.

24 MEASURE EVALUATION

CMS aims to develop quality measures of the highest caliber that will drive significant healthcare quality improvement and inform consumer choices. To gain CMS approval for measure implementation, the measure developer must first provide strong evidence that the measure adds value to existing measurement programs and that it is constructed in a sound manner. CMS gives preference to measures that are already endorsed by the NQF or are likely to become endorsed for implementation in its programs. Therefore, measure developers should develop measures that meet NQF evaluation criteria and are likely to be endorsed if they are submitted for endorsement.

Each proposed measure should undergo rigorous evaluation during the development process to determine its value and soundness based on a set of standardized criteria and subcriteria including the importance to measure and report on the topic, scientific acceptability of measure properties, feasibility, usability and use, and harmonization. Each criterion is composed of a set of subcriteria that are evaluated to determine if the criteria are met. Measure evaluation is an iterative process to build, then strengthen justification that the measures will impact an important healthcare quality need, are scientifically sound, can be implemented without undue burden, and are useful for accountability and performance improvement.

This chapter provides an overview of the measure evaluation criteria and guidance for rating measures according to those criteria.[155] The measure evaluation criteria should be considered throughout the measure lifecycle from information gathering in measure development through measure maintenance.

The measure developer should self-evaluate measures using these criteria and report results and improvements as indicated. This Measure Evaluation Report documents for CMS the extent to which the measure meets the criteria. It also documents any plans the measure developer has to improve the rating when a measure is rated low or moderate on any subcriterion. The measure developer will use this report to document the pros and cons, cost benefit, and any risks associated with not further refining the measure. To facilitate efficient and effective development of high-caliber measures, the materials in this chapter have been revised to reflect changes implemented by NQF as of August 2016.[156] A Measure Evaluation Report template is available in Forms and Templates.

Though measure evaluation should be conducted throughout the measure development and measure maintenance process, a formal Measure Evaluation Report for each measure is submitted to CMS (if required by the measure development contract) when:

- Recommending approval of candidate measures for further development
- Recommending approval of fully tested and refined measures for implementation
- Conducting comprehensive reevaluation.

24.1 MEASURE EVALUATION CRITERIA AND SUBCRITERIA

Measure developers should apply the standardized evaluation criteria to their measures throughout the development process. The more effectively the measure properties meet the 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 Measure

[155] The criteria and evaluation guidance given in the Blueprint align with the NQF criteria and guidance.
[156] National Quality Forum. *Measure Evaluation Criteria.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards/Measure_Evaluation_Criteria.aspx. Accessed on: May 31, 2015.

Justification Form is intended to provide information demonstrating that the evaluation criteria have been met. The form should be updated continuously with any information demonstrating the strength of the measure. CMS and NQF use the following criteria when evaluating measures:

- Evidence, performance gap, and priority (impact)—importance to measure and report
- Reliability and validity—scientific acceptability of measure properties
- Feasibility
- Usability and use
- Comparison to related or competing measures—harmonization.

Measure evaluation does not end when the measure is fully developed. Measures must also be continuously reevaluated during maintenance, with reports submitted at specified periods. Though there may be differing evaluation details for the specific reevaluations, the general principles are the same.

The Measure Evaluation criteria descriptions in the Measure Evaluation Criteria and Instructions, the guidance from NQF on applying the criteria,[157] and the Measure Evaluation Report form facilitate a systematic approach for applying the measure evaluation criteria, rating the strength of the measure and tracking the results. The 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, allowing measure developers to anticipate the evaluation the measure may receive when submitted. 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 (pros/cons, costs/benefits) to increase the measure's evaluation rating versus the risks if it is left unchanged.

24.2 Applying Measure Evaluation Criteria

Throughout measure development, measures are evaluated to determine the degree to which each 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.

[157] National Quality Forum. 2016 *Measure Evaluation Criteria and Guidance* . National Quality Forum; Revised Oct 2013. Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx Accessed on: December 4, 2016

Figure 36 diagrams the process of applying the measure evaluation criteria.

Measure evaluation criteria are applied as follows:

- During information gathering to guide the search for appropriate measures and measure concepts
- During the TEP meetings to inform the TEP members and contribute to meaningful deliberation
- As specifications are refined and tested to strengthen the measures
- When developing a testing plan
- When preparing the following deliverables: Measure Evaluation Report, MIF, and MJF.[158]

[158] The NQF submission may be acceptable for this deliverable. National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: March 14, 2016.

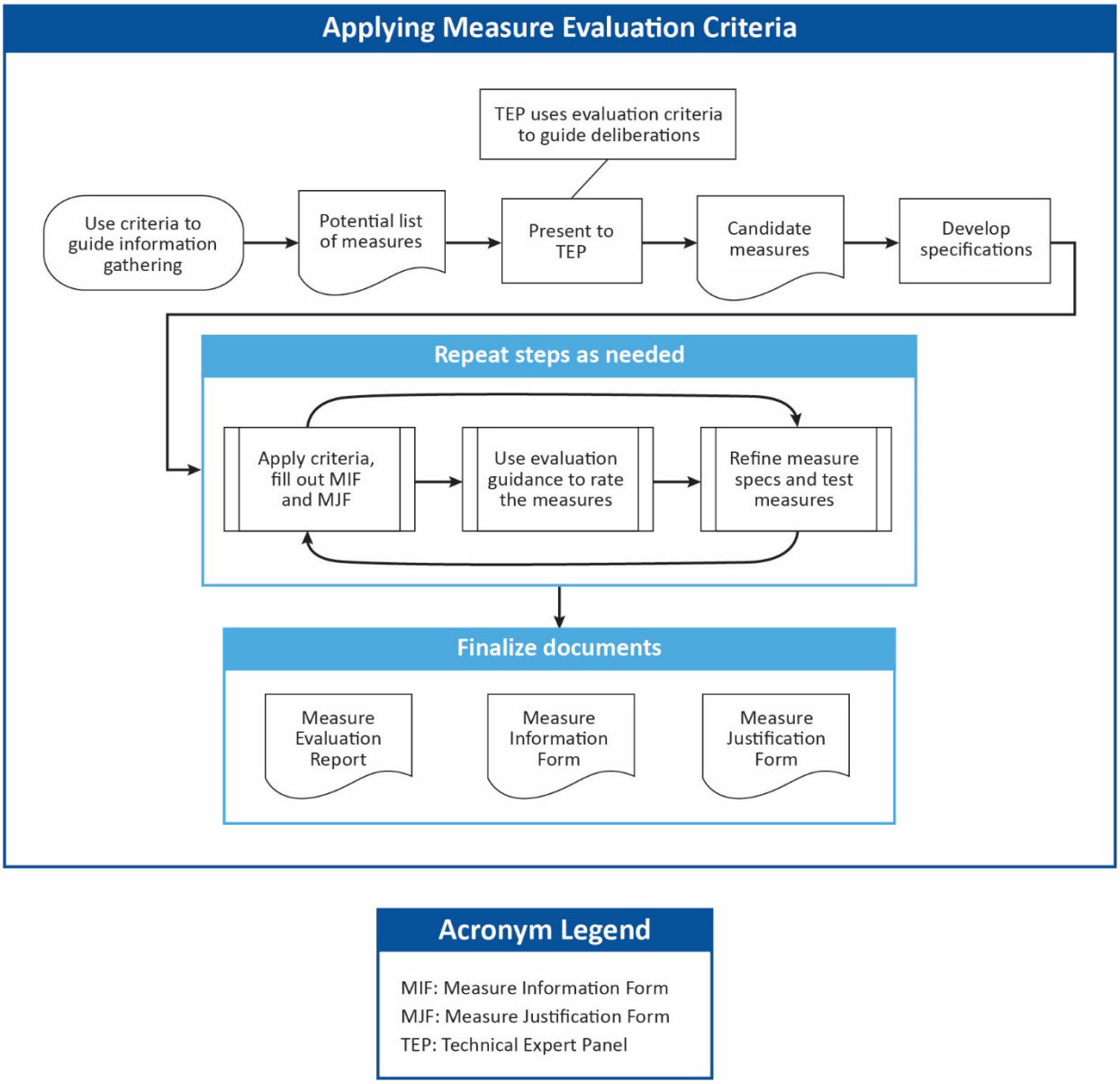

Figure 36. Applying Measure Evaluation Criteria

24.3 Timing of Measure Evaluation

While a formal Measure Evaluation Report is required in only three of the measure lifecycle phases, evaluating measures 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 corrections can be made or weaknesses strengthened at each point rather than waiting for the formal reporting time.

Measure Conceptualization (Section 2, Chapter 1)

- Provide the TEP with an analysis of how the measures might perform by applying the measure evaluation criteria to candidate measures.
- Use the criteria when refining the candidate measure list (***formal report required***).

Measure Specification (Section 2, Chapter 2)

- Report how the measure's proposed technical specifications function.
- Evaluate how the risk model works for outcome measures.

Measure Testing (Section 2, Chapter 3)

- 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 4)

- 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 5)

- 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 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. The testing results are reported in the Measure Evaluation Report. It is important for CMS to be fully informed of the pros and cons; costs and benefits for improving the rating; and the risks if the weaknesses cannot be corrected.

The COR will work with the measure developer to identify the 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.

24.4 Testing and Measure Evaluation Criteria

The results of measure testing are used to demonstrate a measure's alignment with the measure evaluation criteria. Because testing is often an iterative process, both alpha and beta testing findings may provide information that addresses 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.

See Section 3, Chapter 22, Measure Testing; Section 3, Chapter 23, Alpha and Beta Testing; and Section 3, Chapter 25, 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 below.

24.4.1 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.

24.4.2 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, lab values, medication, or surgical procedures) that are aggregated at the comparison group level (e.g., hospital, nursing home, or 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. Some examples of common measure testing and reporting errors are shown here.

- **Reporting is limited to descriptive statistics.** Lack of evidence for empirical testing or appropriate methods for reliability or validity testing.[159] Descriptive statistics demonstrates that data are available and can be analyzed but does not provide evidence of reliability or validity.
- **Lack of testing of adapted measures**. When adapting a measure (e.g., using similar process criteria for a different population or denominator), the newly adapted measures still require testing to obtain empirical evidence of reliability and validity.

[159] National Quality Forum. Review and Update of Guidance for Evaluating Evidence and Measure Testing, Technical Report, Approved by CSAC on October 8 2013: 11

- **Inadequate evidence of scientific acceptability for commonly used measure elements.** Measure's 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 exclusion**. Analyses and/or clinical guidelines justifying 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 the 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.

Though not replacing the expert judgment of the measure development team, the following subsections describe the general considerations for evaluating reliability and validity of both a measure score and its component elements.

24.4.2.1 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.

24.4.2.1.1 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 shown below:

Inter-rater (inter-abstractor) reliability. 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. 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 also is sometimes called parallel-forms reliability. 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 or across automated data extraction from different data sources. 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 also is sometimes called temporal reliability. 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.[160] 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 the 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 the 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, estimation of variance components within a hierarchical mixed (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.

24.4.2.1.2 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.

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.

[160] Cronbach's alpha has been used to evaluate internal consistency reliability for several decades. Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika. 1951;*16:297–334.

24.4.2.2 Validity

In measure development, the term validity has a particular application known as test validity. Test validity refers to the degree to which evidence, clinical judgment, and theory support the 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 assess) and process (i.e., how it is assessed).

24.4.2.2.1 Types of Validity

Validity testing of a measure score can be assessed in many different ways. Though some view all types of validity as a special case of construct validity, researchers commonly reference the following types of validity separately: construct validity, discriminant validity, predictive validity, convergent validity, criterion validity, and face validity.[161]

Construct validity. This refers to 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. 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 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 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 refers to verification of data elements against some reference criterion determined to be valid (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.

Face validity. Face validity 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 that a formal consensus process be used for the review of face validity such as a

[161] Messick, S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*. 1995;50:741–749.

modified Delphi approach where participants systematically rate their agreement, and formal aggregating and consensus failure processes are followed.[162] NQF criteria allow the use of face validity in lieu of empirical testing 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. [163]This type of formal process can also be used when addressing whether specifications of the measure are consistent with medical evidence.

24.4.2.2.2 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. Though the patient level data elements are important, it is the computed measure scores that are used to draw conclusions about the targeted aspect of care. Therefore, data element testing alone is not sufficient.[164]

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:

- Administrative data—Claims data where codes that are used to represent the primary clinical data (e.g., International Classification of Disease [ICD], CPT) can be compared to manual abstraction from a sample of patient medical records.
- Standardized patient assessment instrument—Standardized information (e.g., 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—EHR information extracted using automated processes based on measure technical specifications can be compared to manual abstraction of the entire. For measures that rely on many data elements, testing may not necessarily be conducted for every single data element. Rather, testing may involve only critical data elements that contribute most to the computed measure score.

24.4.2.3 Prior Evidence of Reliability and Validity for Measure Elements

When prior evidence of reliability or validity of the data elements comprising the measure exists, it can sometimes be used in place of testing of the measure's data elements. In contrast to a measure's data elements, though prior evidence can augment findings for a calculated measure score under development, 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. Prior evidence of either 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.[165] Prior evidence of reliability or validity testing may include published or unpublished testing results of same data elements, same data type, and representative sample of sufficient size.

[162] National Quality Forum. *Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties*. The National Quality Forum—Task Force on Measure Testing. 2011:50.
[163] National Quality Forum. Review and Update of Guidance for Evaluating Evidence and Measure Testing, Technical Report, Approved by CSAC on October 8 2013: 12.
[164] Ibid.
[165] National Quality Forum. *Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties*. The National Quality Forum—Task Force on Measure Testing. 2011:13.

NQF guidance states:[166]

- Validity—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. Data elements that represent an existing standardized scale are also often excluded when a judgment is made that the validity of the scale has already been confirmed.
- Reliability—Separate reliability testing of the data elements is not required if validity testing was conducted on the data elements. If validity testing was not conducted, prior evidence of reliability of data elements can be used.

24.4.2.4 Testing of Exclusion/Exception

Review of measure exclusion and exception should be based on the testing data. The review should include at a minimum:

- Evidence of sufficient frequency of occurrence of the exclusion/exception
- Evidence that measure elements (e.g. codes) used to identify exclusion/exception are valid.

Review may also include evidence that measure results are distorted without the exclusion/exception. For example, evidence that exclusion distorts a measure may include variability of 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 if 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 both with and without the exception, and the proportion of exception should be included for any group-level tabulations.

24.4.2.5 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 is not needed for process measures.

Empirical evidence for the adequacy of risk adjustment or rationale that risk adjustment is not necessary to ensure fair comparisons must be provided.

Information should include 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.

24.4.3 Usability

Formal usability[167] testing is often not 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

[166] National Quality Forum. *Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties*. The National Quality Forum—Task Force on Measure Testing. 2011: APPENDIX A, tables A-2, A-4.

[167] NQF defines usability 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." *Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement*, August 2016, p.20.

respect to intended audiences (e.g., consumers, purchasers, providers, and 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.

24.4.4 Feasibility

Testing can be used to assess measure feasibility to determine the extent to which the required data are available, 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 through the use of 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 the following:

- The availability of data (e.g., evidence that required data, including any exclusion criteria, are routinely generated and used in care delivery)
- The extent of missing data, measure susceptibility to inaccuracies, and the ability to audit data to detect problems
- An estimate of the costs or burden of data collection and analysis
- Any barriers encountered in implementing performance measure specifications, data abstraction, measure calculation, or performance reporting
- The 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
- The identification of unintended consequences.

eCQMs have a separate feasibility scorecard assessing data availability, data accuracy, data standards, and workflow.

24.5 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. The following two steps help CMS ensure that its measures retain NQF endorsement.

24.5.1 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 same set of standardized measure evaluation criteria used in measure development.

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. Through the measure evaluation process, developers update justification for the measure and any changes to the technical specifications to demonstrate the following:

- That 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
- That data elements, codes, and parameters included in the specifications are the best ones to use to quantify the particular measure, and that data collection still does not cause undue burden on resources
- That 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.

24.5.2 Report Results of Evaluation

Measure Evaluation Criteria and subcriteria are detailed in the Measure Evaluation Criteria and Instructions with a link to the 2016 NQF measure evaluation guidance document. 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:

- 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.

25 TESTING FOR SPECIAL TYPES OF MEASURES

25.1 ADAPTED OR RESPECIFIED MEASURES

When adapting a measure for use in a new domain (e.g., new setting or population) or respecifying a measure 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 the following:

- 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.

If eCQMs are adapted for use in different settings or with different populations, the adapted measures must also be tested and evaluated.

25.2 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 the 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.

25.2.1 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 clinical quality measure (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 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 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.

25.2.2 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. 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 (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 the following:

- Data availability (including standardization)
- Accuracy of the information in the data
- Maturity of standards
- Standard terminologies
- Extent to which the 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 2, Chapter 3, 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 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 when determining the threshold for feasibility. Refer to the NQF Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement[168] and the NQF eCQM Feasibility Scorecard for more information and guidance.

[168] National Quality Forum. *Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement.* Effective August 2016. Available at: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=83123. Accessed on: November 9, 2016.

25.2.3 Validity

Testing the different measures from the eCQM package validate different aspects of the measure: measure as a whole, measure logic, data element in the measure, and measure score.

Validity testing for the eCQM confirms the intent of the measure, ensures the eCQM logic is not ambiguous and expected test patients fall in the correct populations, data elements are aligned with national standards, and checks calculated scores from automated extraction for accuracy.

Ideally, certified EHR technology (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 EHR (i.e., chart abstraction or administrative 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 reliability and validity.

A subjective evaluation of the human-readable document of the eCQM should be conducted to confirm 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 workgroup or TEP that eCQM concepts reflect the intent. Measure level (face) validity testing may involve iterative discussions with the measure steward or clinical workgroup/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.

25.2.3.1 Measure Logic Validity

An objective evaluation of measure logic should be performed 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 (see Section 3, Chapter 7, eCQM Standards-Based Guidance and Tools) is a measure testing tool that provides feedback on the behavior of eCQM logic. Measure developers use Bonnie as an introductory testing tool to ensure the measure algorithm performs as intended. Bonnie may be used at multiple points in the measure development process as measure logic is revised. Bonnie is sponsored by CMS and the ONC.

25.2.3.2 Data Element Validity

An objective evaluation of whether data elements automatically extracted from an EHR are comparable to similar data elements visually abstracted by the reviewers should be conducted. 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 automatic extraction and compared to a gold standard EHR extract to assess the validity of the automatic 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 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 would therefore 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 of the 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 research; or
- 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 reliability testing of measure data elements may be supplanted by evidence of measure data element validity.

Calculate at the score level, and confirm the calculation and score are consistent with the expected results.

25.2.3.3 Standards Conformance Validation

To help ensure the accuracy of data elements, measure developers are expected to validate the content of the eXtensible Markup Language (XML). This is often achieved using three methods:

- Syntactic validation—This method of accuracy validation ensures that the XML content follows (i.e., conforms to) specific constraints required by the Health Level Seven (HL7) HQMF Standard and the XML patterns based on the QDM. These quality-checking processes are built into the MAT application.
- The HL7 ISO-based Schematron is a possible mechanism for validating XML that is written outside the MAT; however, it may not include all the components that are now built into the MAT. Additional resources for information—including technical specifications—on the ISO Schematron may be found at the ISO website.[169]
- Narrative validation—A fully constructed eCQM is an XML document that can be viewed in a standard Web browser in HTML. When rendered in a Web browser, the eCQM is in a human-readable format which allows the measure author to assess the extent to which the machine-

[169] http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html. Accessed on: March 14, 2016.

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.

25.2.4 Reliability

Testing for reliability involves experts assessing the human-readable format (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.

25.2.5 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 consistently be entered into the fields expected by the vendor. Variances in results from such testing at multiple sites should be evaluated to determine if changes are needed in the measure logic or definition. Testing must encompass at least two EHRs.

25.3 Composite Measures

A composite measure is a combination of two or more individual measures into a single measure that results in a single score. The use of composite measures creates unique issues associated with measure testing. To meet the NQF criteria for endorsement of composite measures, testing the measure composite score must be augmented by testing the individual components of the composite. However, this does not apply to measure components previously endorsed by the NQF or for components of a scale/instrument that cannot be used independently of the total scale. Below are recommendations for testing a composite measure in support of submission to CMS for approval and NQF for endorsement.

25.3.1 Component Reliability and Validity Testing

Demonstration of reliability and validity is recommended for both the composite and the components of the composite. Composite components must individually demonstrate adequate reliability and validity, but the composite measure as a whole must also meet these criteria.

25.3.2 Component Coherence

Testing is recommended to determine if components of a composite measure adequately support the goals articulated in the constructs for the measure. The reliability of the components can be tested using correlation analyses or confirmatory factor analysis methods. If the components are coherent, the component items meet the intent of the measure construct.[170]

25.3.3 Composite Reliability and Validity Testing

The 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

[170] National Quality Forum, Composite Measure Evaluation Framework and National Voluntary Consensus Standards for Mortality and Safety—Composite Measures, A Consensus Report, 2009: 12.

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 intraclass correlation coefficient (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 dependent variable
- Clinical justification, demonstrating correlation of the individual component measures to a common outcome measure.

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.

25.3.4 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.

25.3.4.1 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.

25.3.4.2 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, the 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 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 emergency department (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 see 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 (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.

25.3.5 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, and demonstrate that the measure can be deconstructed into its components at the group/organization level to facilitate transparency and can be understood by the intended measure audience.

26 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.

26.1 EVALUATING COMPOSITE MEASURES

A composite performance measure is 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.[171] There are two primary types of composite performance measures:

1. Measures of two or more individual performance areas scored using an algorithm that produces as single score as its only output. This type of composite cannot produce individual scores for component elements.
2. 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.

26.1.1 Considerations for Evaluating Composite Measures

The following information from the NQF 2013 report, *Composite Performance Measure Evaluation Guidance*, describes NQF's approach to evaluation.

A coherent quality construct and rationale for the composite performance measure are essential for determining the following:

- What components are included in a composite performance measure
- How the components are aggregated and weighted
- What analyses should be used to support the components and demonstrate reliability and validity
- Added value over that of individual measures alone.

[171] National Quality Forum. *Composite Performance Measure Evaluation Guidance*. Washington, DC: National Quality Forum; Apr 2013; Contract No. HHSM-500-2009-00010C. Available at: http://www.qualityforum.org/Publications/2013/04/Composite_Performance_Measure_Evaluation_Guidance.aspx. Accessed on: March 14, 2016.

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.

- 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.
- The 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.
- The 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.[172]

26.2 Evaluating Cost and Resource Use Measures

The resource use measure evaluation criteria are grounded in the standard NQF evaluation criteria (version 1.2), 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 (in terms of units or dollars) applied to a population or population sample. Resource use measures count the frequency of specific resources, and 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.

26.2.1 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 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

[172] Ibid.

a COPD resource use measure because cancer is considered the dominant medical condition with known high costs.

- Testing for resource use measure exclusion 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.

26.3 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.

26.4 EVALUATING PATIENT-REPORTED, OUTCOME-BASED PERFORMANCE MEASURES

Evaluation for patient-reported outcome-based performance measures is a special case of overall outcome measure evaluation. In its January 2013 report, *Patient-Reported Outcomes in Performance Management*,[173] NQF outlined criteria specific to patient-reported outcomes-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 burden to patients of responding should be considered.
- Person-centered—These 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—These measures should capture impact on health-related quality of life, symptom burden, experience with care, and achievement of personal goals.
- Amenable to change—The 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.

[173] National Quality Forum. *Patient-Reported Outcomes In Performance Management.* National Quality Forum; Dec 2012. Available at: http://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx. Accessed on: March 14, 2016.

27 NQF Endorsement and Maintenance

The NQF currently serves as the consensus-based entity 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.

NQF's measure endorsement process, or consensus development project, is cyclical; measures in various topic areas are considered on a fairly routine basis (approximately every 3-4 years or possibly more often, depending on the topic). NQF-endorsed measures are due for routine maintenance review every 3 years and will be reevaluated in relevant topic areas such as Cardiovascular, Patient Safety, Person and Family (caregiver)-Centered Care, and Cardiovascular Care, pending funding. NQF endorsement and maintenance projects follow a 7-month timeline from measure submission to the appeals period.

The measure submission process is summarized in Figure 37. These steps, as well as those involved in the endorsement process that follows, are discussed below.

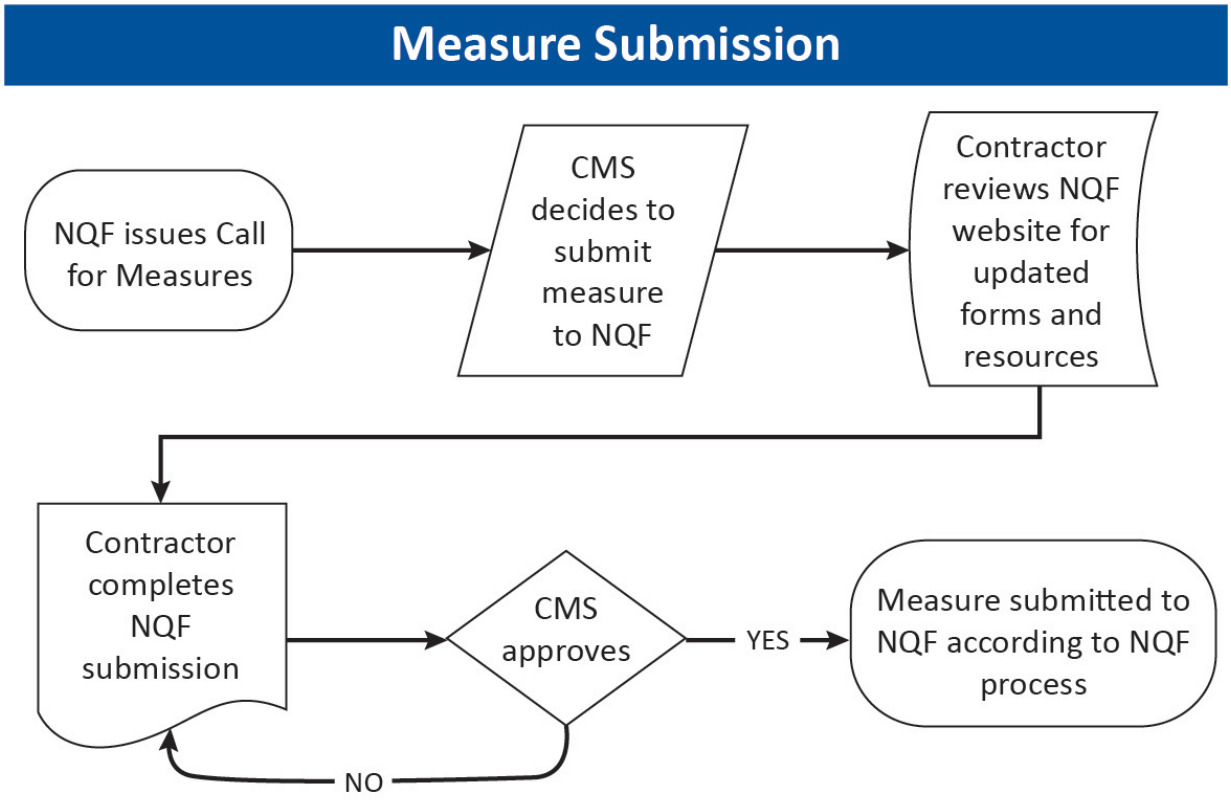

Figure 37. Measure Submission

27.1 Measure Submission to NQF

27.1.1 NQF Issues a Call for Measures

No less than 2 months before the start of a project, NQF issues a formal Call for Measures requesting measure developers to submit their measures relevant to the project. NQF also allows measure developers to submit measures at any time prior to an official Call for Measures. Measure developers should obtain approval from their COR before initiating online submission of a measure.

Beginning the online submission prior to a Call for Measures allows the measure developers additional time to prepare and thoroughly review the submission form. This may be useful when many measures are being considered for submission by a single measure developer. Though measure developers may submit their measure at any time, the measure will only be reviewed by NQF when there is a Call for Measures for that topic.

27.1.2 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, and contact appropriate NQF project staff for technical assistance.

With the introduction of the EHR incentive programs, there is a movement toward the development and/or respecifying of measures specified for use with the EHR (eMeasures or eCQMs). The COR will provide guidance as to 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[174] prior to submission and to contact NQF staff for technical assistance. Note that while many measures can be calculated from EHR data, only some of them are considered eCQMs

27.1.3 The Measure Developer Completes the NQF Measure Submission

Measure developers must submit their measures via an online measure submission.[175] The online form is available on the NQF website and allows users to do the following:

- 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. This allows 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. Measure developers should inform their COR of this option. However, users must coordinate the timing at which they save their respective edits or their edits could be overwritten.

[174] Available at: http://www.qualityforum.org/Electronic_Quality_Measures.aspx. Accessed on: March 14, 2017.
[175] http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx

The CMS Measure Information Form (MIF) and Measure Justification Form (MJF) have been aligned 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 to gather the information needed for a successful NQF submission and organize it to minimize rework. The MIF and MJF are CMS forms designed to present measures in a standardized way. Every effort has been made to ensure that the MIF and MJF are aligned with the NQF measure submission requirements to facilitate online information entry.[176]

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 as follows:

- Contact NQF project staff for measuremaintenance@qualityforum.org 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 pilot 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 that no questions are left unanswered (i.e., no fields should be left blank and all questions should have a response).

Measure developers are free to contact the Measures Manager for content questions while completing the online submission. For technical questions about the online submission, please 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.

[176] The NQF submission may be acceptable for these deliverables. National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: March 14, 2016.

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 is the standard by which NQF evaluates harmonization. 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 then select a superior measure or request that both responsible measure developers create a harmonization plan addressing 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.

27.1.4 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.

27.1.5 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 NQF using the online process.

NQF follows a standardized measure review process to consider granting endorsement to the measure. NQF's current endorsement process is described below.

27.2 NQF Endorsement Process

27.2.1 Initial Review by NQF Staff

After NQF receives the measure submission, in addition to checking for completeness of the document, NQF staff also performs an initial review to ensure that the measures submitted meet all of the following conditions for consideration:

- The measure is in the public domain or a measure steward agreement is signed.
- The measure owner/steward verifies that there is an identified responsible entity and a process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years.
- The intended use of the measure includes both accountability applications (including public reporting) and performance improvement to achieve high-quality, efficient health care.
- The measure is fully specified and tested for reliability and validity and is submitted for trial use.
- The measure developer/steward attests that harmonization with related measures and issues with competing measures have been considered and addressed, as appropriate.

- The requested measure submission information is complete and responsive to the questions so that all the information needed to evaluate all criteria is provided.[177]

NQF staff will inform the developer of any items that need additional attention, and arrange for the submission to be updated. There is usually a very quick turn-around time for these updates (typically 48 hours). Once this completeness check is done, the measures are accepted by NQF. NQF staff then prepare a summary of the submission for the Standing Committee (this is known as the NQF "preliminary analysis"). As part of this process, NQF staff may have clarifying questions for the developer. Once the NQF preliminary analysis is complete, the measure is considered ready for evaluation by the project Standing Committee.

27.2.2 Standing Committee and Technical Advisory Panel Review

After the measures are accepted, the measures are reviewed by a topic-specific Standing Committee that has been selected from nominations submitted during a public call for nominations. This Standing Committee oversees the portfolio of NQF-endorsed measures for that topic area, offers expert advice, ensures that input is obtained from relevant stakeholders, and makes recommendations to the NQF membership and Consensus Standards Approval Committee about measures that are proposed for endorsement.

A technical advisory panel may also be used depending on the technical expertise of the Standing Committee. The technical advisory panel members are experts in their field and provide guidance to the Standing Committee around specific technical issues related to some or all of the measures under review. The Standing Committee evaluates each submitted measure using the NQF Measure Evaluation Criteria.

After evaluation of the measure against the criteria, the Standing Committee recommends that a measure be endorsed or not. The measure developer should be available to provide an overview and respond to Standing Committee questions by attending the meeting in person or by teleconference. The COR and a member from the TEP involved in measure development may also attend the meetings to answer any questions that the committee members may have on the measures.

Measures that are recommended for endorsement by the Standing Committee are then posted on the NQF website site for public and NQF member comment. After the comment period, NQF members vote on the measures, and are then sent to the Consensus Standards Approval Committee (CSAC). The CSAC is the governing body that makes endorsement decisions regarding national voluntary consensus standards. It has the most direct responsibility for overseeing the implementation of NQF's consensus development process.[178]

27.2.3 NQF Makes Endorsement Decision

The CSAC reviews the recommendations of the Standing Committee and the results of the NQF member voting period. As during the Standing Committee evaluation, the measure developer and COR (and possibly a member from the TEP) involved in measure development should attend the CSAC meeting in person or by teleconference to answer any questions that the CSAC members may have about the measures.

---

[177] National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: March 14, 2016.
[178] National Quality Forum. *Consensus Development Process*. Available at: http://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process.aspx. Accessed on: March 3, 2016.

After review of the measure, the CSAC will make the final endorsement decision, without ratification by another body. Specifically, the CSAC may do the following:

- Grant full endorsement. This requires no further documentation. According to NQF, full endorsement is given to measures meeting NQF endorsement requirements.
- Defer endorsement decision. This designation is generally used when the CSAC requires additional information from the measure steward or requests specific changes to a measure, until such information can be submitted and reviewed. The CSAC will propose a timeline; within 14 days, the developer will confirm the feasibility of the recommended changes and agree to the proposed timeline. If an endorsement decision is deferred, the measure retains its endorsement during the deferral timeframe.
- Decline to endorse the measure altogether. According to NQF, measures are not endorsed when they do not fully meet NQF criteria for endorsement.

Standing Committees and the CSAC may evaluate selected eCQMs for Approval for Trial Use rather than for endorsement. NQF's eMeasure and Maintenance teams screen candidate eCQMs and determine whether they meet requirements for Approval for Trial Use. Measures deemed by NQF staff to be candidates for the Trial Use Program are ready for implementation, but cannot yet be adequately tested to meet NQF endorsement criteria. Approval for Trial Use expires after three years; however, developers can submit a Trial Use measure for potential endorsement once empirical testing of the measure has been completed.

Once the CSAC's endorsement decision has been made public on the NQF website, a 30-day appeals period begins. Any interested party may file an appeal with the Appeals Board during the appeals period. All appeals will be adjudicated by a five-member Appeals Board, which will meet within 45 days of the end of the appeals period. The only measure endorsement decisions eligible for appeal are as follows:

- The CSAC endorses a measure that a standing committee recommends for endorsement
- The CSAC declines to endorse a measure that a standing committee recommends for endorsement.

Eligible measure endorsement decisions may only be appealed on the following grounds:

- Procedural errors reasonably likely to affect the outcome of the original endorsement decision, such as a failure to follow NQF's Consensus Development Process
- New information or evidence, unavailable at the time the CSAC made its endorsement decision, that is reasonably likely to affect the outcome of the original endorsement decision.

The Appeals Board may render the following decisions:

- Uphold the CSAC endorsement decision
- Overturn the CSAC endorsement decision
- Dismiss the appeal because it does not state sufficient grounds for appeal or it does not raise decisions that are eligible for appeal.

All decisions made by the NQF Appeals Board will be final.

27.2.4 Review Endorsement Results

After NQF has completed its consensus development process, the measure developer meets with the measure developer's COR and the Measures Manager to discuss the results of the consensus development process, discuss why NQF came to its decision, identify lessons learned about both the NQF process and the CMS MMS processes, and discuss potential next steps.

27.3 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 measures are 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 the 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 measures. The measure developer proposes responses to the questions, which the COR reviews and approves, before the measure is submitted to NQF.

During the NQF 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 makes the changes and submits the revised measure to NQF.

27.4 Trial Use Approved Measures

On a very limited basis, NQF may grant an eCQM Approval for Trial Use. 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.[179]

27.5 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 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 the CMS MMS maintenance reviews, but

[179] http://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process/CSAC_Decision.aspx

Figure 38 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. More information on the CMS scheduled maintenance reviews can be found in Section 3, Chapter 31, Measure Maintenance Reviews.

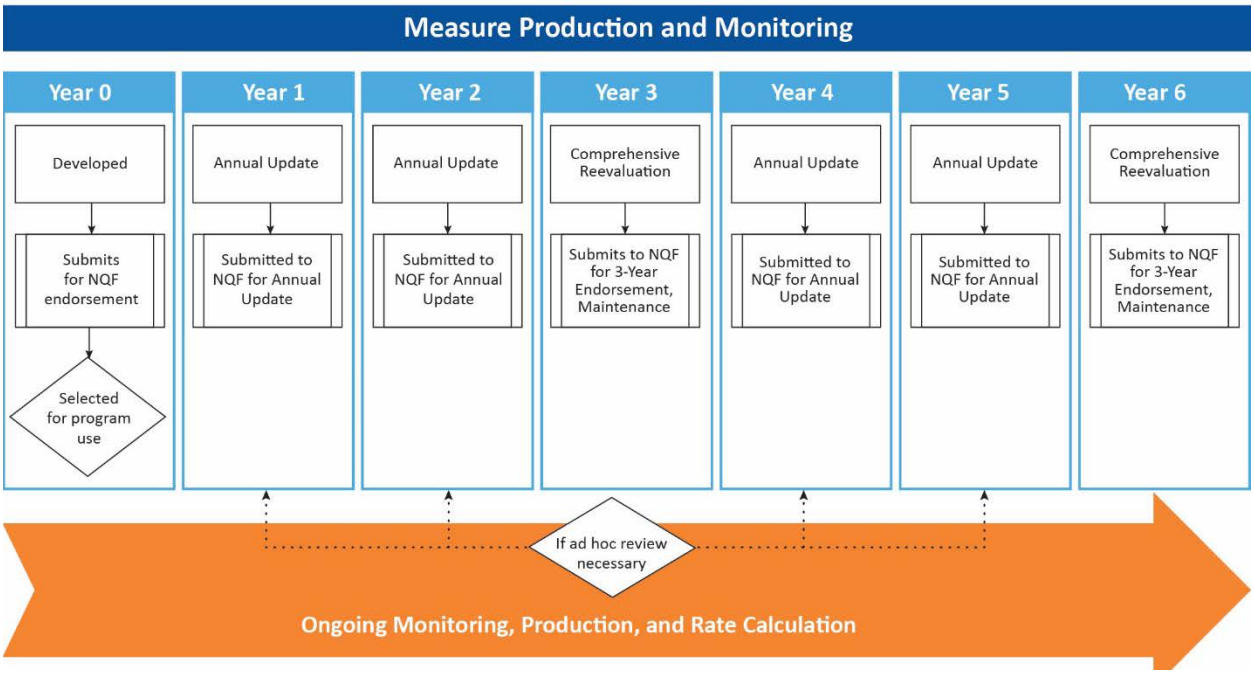

Figure 38. Measure Review Cycle Timeline

28 MEASURE SELECTION

28.1 PRE-RULEMAKING PROCESS

Section 3014 of the ACA[180] 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 (in proposed or final rules) that is not NQF endorsed.

28.1.1 Measures Under Consideration (MUC)

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 if 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 HHS can develop the MUC list for qualifying programs and make it publicly available by December 1 annually. 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 39 contains a timeline for the pre-rulemaking process.

Figure 39. Pre-rulemaking Process Timeline

[180] 111th Congress of the United States. *Patient Protection and Affordable Care Act*, 42 U.S.C & 18001 (2010). United States Government Printing Office. Available at: http://www.gpo.gov/fdsys/pkg/BILLS-111hr3590enr/pdf/BILLS-111hr3590enr.pdf. Accessed on: March 14, 2016.

28.1.2 CMS Measure Selection Criteria:

- 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 or more of the six NQS priorities.
- 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 little opportunity for improvement—in other words, measures that are topped out.
- Potential use of the measure in a program does not result in negative unintended consequences such as reduced lengths of stay, overuse or inappropriate use of treatment, and limiting access to care.
- Measures should not duplicate other measures currently implemented in programs.
- eCQMs must be fully developed and tested.
- eCQMs must be created using the MAT, and expressed in the HQMF.

Applying the measure selection criteria listed above, CMS develops the MUC list.[181] 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.

28.1.3 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.[182] To be fully enfranchised in this process, measure developers are strongly encouraged to attend the MAP.

28.1.4 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 the 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 was not previously endorsed by NQF.

[181] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *2014 Measures under Consideration List Program Specific Measure Priorities and Needs.* April 21, 2014.
[182] http://www.qualityforum.org/Setting_Priorities/Partnership/Measure_Applications_Partnership.aspx

28.2 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 rule and publishes it 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.

28.3 Rollout, Production, and Monitoring of Measures

When the 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.

28.4 Measure Maintenance

After the measures are implemented, the measure developers monitor the performance of the measures, respond to ongoing feedback, and continuously scan the environment regarding the measures. For example, for eCQMs, the ONC's instance of 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 5, Measure Use, Continuing Evaluation, and Maintenance. Five different outcomes are possible following maintenance review of CMS measures:

- Retire—Cease to collect or report the measure indefinitely. This applies only to measures owned by CMS. CMS will not continue to maintain these measures. 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.
- Retain—Keep the measure active with its current specifications and minor changes.
- Revise—Update the measure's current specifications to reflect new information.
- Suspend—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.)
- Remove—A measure is no longer included in a particular 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 particular 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.

28.5 Impact Assessment of Medicare Quality Measures

Also mandated by Section 3014 of the ACA, once every 3 years, 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). This triennial report assesses how well CMS, through the use of quality measures, has achieved the NQS's three aims and six priorities. It 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. This report can be accessed at the CMS website. The most recent report was published in March 2015 and greatly expanded from the trend data reported in 2012. The next report is expected to be published in spring 2018.

29 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 conducted in this chapter, as with all parts of the Blueprint, will comply with the requirements of the Data Quality Act. The Data Quality Act 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". [183] The measure development and maintenance procedures detailed in this Blueprint also comply with HHS' Guidelines for Ensuring the Quality of Information Disseminated to the Public.

Figure 40 Overview of the Measure Rollout Process depicts the process of measure implementation as well as associated responsibilities for related tasks when rolling out measures approved by the COR.

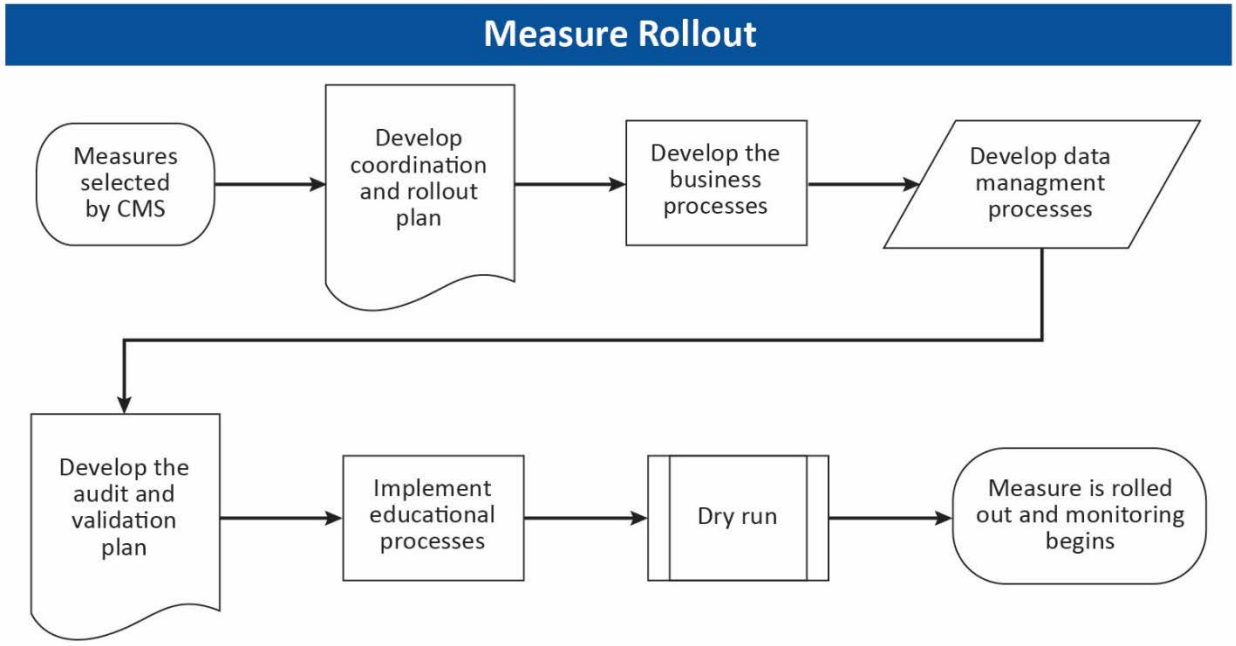

Figure 40. Overview of the Measure Rollout Process[184]

Once the COR has approved a measure for use in a particular program, several tasks have to be completed for rollout. Perform the following nine steps simultaneously whenever possible to achieve an efficient timeline.

[183] https://www.ftc.gov/site-information/website-policy/data-quality-act
[184] CMS may not hold a dry run for every measure.

29.1 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 28, Measure Selection. Develop the coordination and rollout plan, which includes the following key parts:

- Timeline for quality measure implementation
- Plan for stakeholder meetings and communication
- The anticipated business processes model[185]
- The anticipated data management processes
- The audit and validation plan
- Plans for any necessary education.

The coordination and rollout plan is referenced in the MIDS Umbrella Statement of Work (USOW). The anticipated business processes model and anticipated data management processes together represent the implementation referenced in the MIDS USOW.

The COR will be 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 the following:

- State agencies
- Other CMS divisions
- Office of Information Services
- Software vendors
- Providers and provider organizations.

In addition to coordination with groups and individuals, coordinate 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. Some examples of communications strategies may include the following:

- Announcement to the Quality Improvement Organization community by a Healthcare Quality Information System (HQCIS) 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

[185] According to Williams (1967), business process models are graphical and textual representations 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. Williams, S. (1967). Business process modeling improves administrative control. Automation, December, 44, 50.

- Open door forums
- Other processes as determined by the COR.

A measure developer must consider the communication activities listed above when developing the initial timeline for quality measure implementation. The timeline is then reviewed for approval by the COR.

For some measures, CMS measure developers should develop an implementation algorithm (also referred to as the calculation algorithm). The calculation algorithm is an ordered sequence of data element retrieval and aggregation through which numerator and denominator events or continuous variable values are identified in the measure specifications. The algorithm is documented in the Measure Information Form under Measure Specifications as the Calculation Algorithm/Measure Logic. This documented process is expected to begin with the submission of data by the providers (for measures based on data abstraction) or the initiation of data collection (for measures based on administrative data) and end with the posting of the measures for public reporting. Measure developers should consult with their COR if uncertain about the need for an implementation algorithm.

Before implementing any process that involves the collection of new data, measure developers should consult with their COR regarding the PRA requirements. OMB approval is required before requesting most types of information from the public. Section 2, Chapter 1.3, Stakeholder Engagement includes a discussion of these requirements.

29.2 Implement the Rollout Plan

This step primarily applies to a new set of measures or a new use for an existing measure. The ultimate intended use of the measures will be a major factor in determining what is required for the rollout plan. Examples of activities that can be conducted during this step include the following:

- Develop the work processes and tools for data collection, rate calculation, and reporting.
- Develop the process for responding to questions about the measures.
- Identify which CMS divisions need to be involved to ensure that 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 a process for documenting questions and answers, so they can be monitored for trends and used to inform measure maintenance activities.

29.3 Implement the Data Management Processes

The data management processes that were created and tested during measure development must now be adapted for measures that are in use.

The major tasks in this step include the following:

- Translating the algorithm used with hypothetical or test data into one that can be used with actual quality data.
- Developing protocols and tools to receive data.
- Parallel processing of the data through the analysis program to ensure accuracy of the interpretation of the algorithm.
- Developing measure data collection quality control processes.

29.4 Develop the Auditing and Validation Plan

The measure developer will 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: the full measure or the individual data elements.

When auditing and validating data element results, consider the following:

- Have the data been collected correctly?
- Were the algorithm and all auxiliary instructions followed correctly? This is a particular 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 receiving a hysterectomy or a female diagnosed with prostate cancer.

When auditing and validating measure function, consider the following:

- If there are multiple databases used to calculate the rates, were they correctly linked?
- Was the sampling methodology correct?
- Were the 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 percent may indicate an error in the calculation algorithm or in the calculation programming. Similarly, unexpectedly low rates may indicate a problem as well.

29.5 Develop an Appeals Process

Before implementing a measure, CMS will determine if providers can appeal either the audit results or measure rates. The measure developer may be required to help develop and design these processes.

29.6 Implement Education Processes

Providers will likely need to be educated on exactly what is being measured and how to interpret the results. For example, QIN-QIO networks 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 the following:

- User guides and training manuals as indicated
- Conference calls and recordings of the calls
- Web-based presentations and recordings of the presentations
- Workshops at conferences or scientific society meetings
- Train-the-trainer events
- Other venues as determined by the COR.

29.7 Conduct the Dry Run

The dry run is the final stage of measure testing and the second to last stage in measure rollout.[186] 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 medical records data), and measurement calculation. It will verify that the measure design works as intended and begin to identify unintended consequences such as gaming or misrepresentation. The dry run also familiarizes relevant entities, such as CMS, the Quality Innovation Network-Quality Improvement Organizations (QIN-QIOs), and the providers with the reports of the measure results. This 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. It also identifies any issues with the report production process so the report production processes 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, though 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 should be addressed and resolved before the measure is fully implemented.

29.8 Submit Reports

CMS may request reports summarizing the rollout processes. These may include the following:

- 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[187]
    - Business processes model
    - Data management processes
    - Audit and validation processes
    - Educational processes for either data collectors or users of the measure results.

[186] The final stage in measure rollout is the first use of a measure in a CMS program or first results reporting.

[187] Recommendation for changes to the measure specifications should clearly document the proposed changes and also address (at a minimum) whether the change is material or not, whether the change requires public comment or publication in the Federal Register, and whether the change affects other harmonized measures.

30 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 the following:

- 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, or 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 41 diagrams the overall production and monitoring components of a measure that has 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 following subsections.

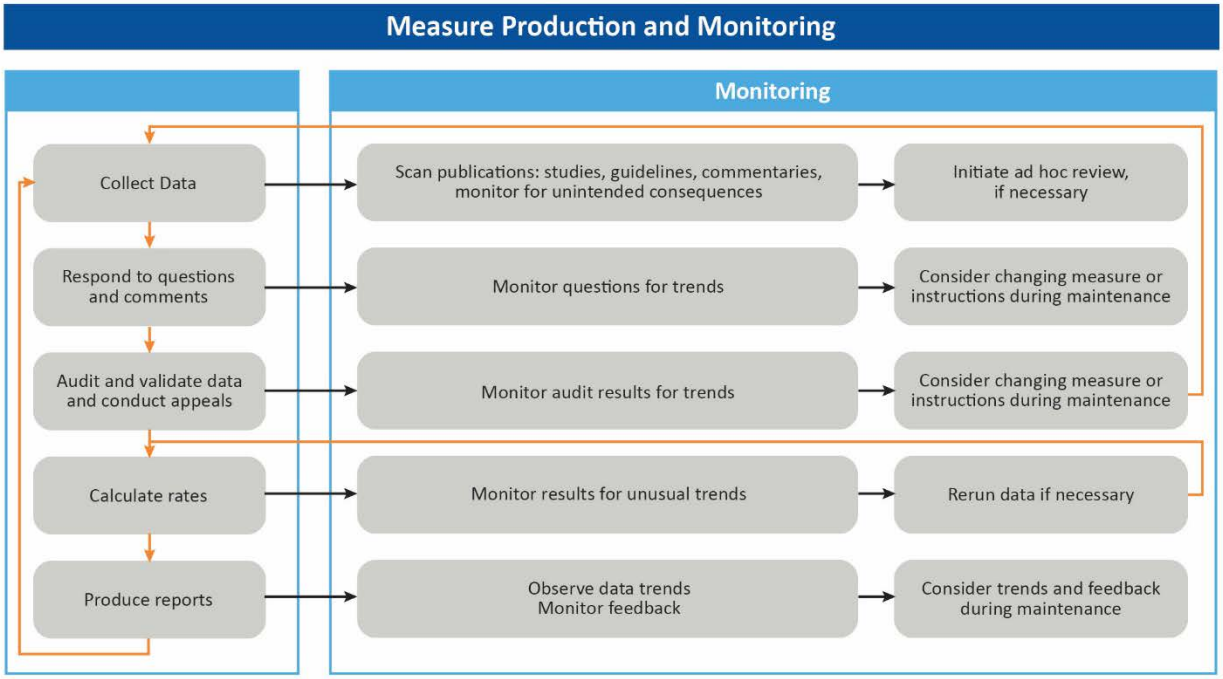

Figure 41. Overview of the Measure Monitoring Process

30.1 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. This means that the data are being collected, calculated, and publicly reported. As the measure is being used, the measure maintenance contractor should continue environmental scans of the literature about the measure. In addition to publications in medical and scientific publications, also watch the general media 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 an immediate action. Ongoing information surveillance is very similar to the information gathering stage of measure development as covered in Section 2, Chapter 1.1, 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. Pay particular attention to any organizations that issue clinical practice guidelines that are relevant, especially for process measures. If the measure is based on a particular 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, monitor for unintended consequences the measure might have on clinical practice or outcomes. Look for articles or studies describing unintended consequences in the literature and identify if any unusual trends in data suggest unintended consequences. If significant unintended consequences are identified, especially if patient safety is the concern, do not wait for scheduled annual or comprehensive review. An ad hoc review may be necessary and requested.

30.2 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, letting the submitter know that the 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.

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.

30.3 Produce Preliminary Reports

For public reporting programs, the results will be released to the providers before they are released to the public. The providers will be allowed a period of time (usually 30 days) to review and respond to the measure results.

The preliminary reports should be monitored for unusual trends both by CMS and by its measure developers. Investigate any trends that are discovered, rerunning the reports to check for errors in calculation. If the unexpected results are not due to error, the cause should be investigated and reported to CMS. If necessary, CMS has the option of suppressing some or all of the data from appearing on the website for a given reporting period (e.g., quarter or year). Data suppression might be necessary due to known problems with a given measure or measure set, or data collection issues with a particular provider or group of providers. The decision to suppress data may apply to the following:

- All measures in a given measure set(s)
- A particular measure (or measures)
- A group of providers (e.g., a state or a region)
- A particular provider (or providers).

30.4 Report Measure Results

Once the measure results are calculated and the providers have reviewed them (for public reporting or value-based purchasing programs), the results are released. Depending on the particular 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 if the other provider results are to be reported anonymously or not. The process by which the information is to be shared with providers and others, the format of the reports, and support for questions from providers should be established in the rollout plan before implementation.

If the 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, which requires federal agencies' electronic information to be accessible to people with disabilities.[188]

For value-based purchasing programs, the results will be shared with the appropriate areas within CMS responsible for calculating provider payments in addition to any requirements for public reporting of the data.

30.5 Monitor and Analyze the Measure Rates and Audit Findings

The measure performance rates and audit findings will be monitored and analyzed periodically and at least once a year for the following:

- Overall performance trends
- Variations in performance, gaps in care, and extent of improvement
- Disparities in the 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

[188] Department of Health and Human Services. *Section 508.* Available at: http://www.hhs.gov/web/508/index.html. Accessed on: March 14, 2016.

- Discretionary exclusion 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 the rates
- Impact of the measurement activities on providers
- Correlation of the performance 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 31.2, Comprehensive Reevaluation.

30.6 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 (denominator, numerator, and exclusion) are current, and to address other minor changes that may be needed. The standardized process for annual update is described in Section 3, Chapter 31, 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 31.2, Comprehensive Reevaluation.

As mentioned earlier, 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 31.3, CMS Ad Hoc Review, including the process for determining when an ad hoc review is necessary. For endorsed measures, request for ad hoc review may also come from NQF if there is evidence to justify such review.[189] 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 particular measure. Those decisions are described in Section 3, Chapter 31.5, Possible Outcomes of Maintenance Reviews and include whether to retain, revise, retire, remove, or suspend the measure in a particular program.

If NQF has endorsed the measure, the results of the maintenance review will be reported to NQF to reevaluate its endorsement at the time of the NQF maintenance review. The outcome of the NQF review may influence whether CMS continues using a particular measure in a program.

30.7 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, Priorities 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

[189] National Quality Forum. *Measuring Performance:* Maintenance of NQF-Endorsed® Performance Measures. NQF-Endorsed® is a registered trademark of National Quality Forum. Available at: http://www.qualityforum.org/Measuring_Performance/Endorsed_Performance_Measures_Maintenance.aspx. Accessed on: March 14, 2016.

use of the measures. The evaluation may also include options for alternative ways to interpret the measures and measure sets through the continuum of care. Performance trends of the measure can be used by the NQF MAP to evaluate the use of the same or similar measure in other settings or programs. This evaluation may be done as part of the pre-rulemaking process for the MUC list.

31 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 review.

31.1 Annual Update

The first type of measure reevaluation is the annual update. This 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 (Common Procedural Terminology [CPT], International Classification of Diseases, 10th Revision, Clinical Modification [ICD-10-CM], LOINC, etc.) used within the measure are updated when the code systems change. However, this is also the time to review and address feedback received about the measure's specifications, reliability, and validity. This review includes the reliability and validity of the measure's constituent data elements. Review the measure for opportunities for harmonization at this time.

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 in the 3-year endorsement period, the measure steward is responsible for informing 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, changed in what is being measured, inclusion of new data sources, expansion of the level of analysis or care settings).

Annual Update Deliverables

1. An updated Measure Information Form showing all recommended changes to the measure. If there are changes relevant to the Measure Justification, that form should be updated as well.
2. For measure developers maintaining electronic clinical quality measures (eCQMs), the revised specifications - Health Quality Measure Format (HQMF) file, Simple eXtensible Markup Language (XML) file, and human-readable Hyper Text Markup Language (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 Measure Information Form.
4. NQF Annual Update online submission, regardless if any change was made to the measure.
5. NQF submission documentation for any material changes to the measure.

Figure 42. Annual Update Deliverables

The annual update process ensures that the CMS 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 if revision is needed beyond updating the codes. Figure 42 contains a list of the annual update deliverables.

The annual update process involves three parts, below, divided into eight steps:

- Gathering information that has been generated since the last review (the comprehensive reevaluation, annual update, or measure development—whichever occurred most recently)
- Recommending action
- Approving and implementing the action(s).

31.1.1 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. 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 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 rationale for continuing with possibly duplicative measures.

31.1.2 Procedure

Follow these steps to conduct measure maintenance.

31.1.2.1 Review the Measure's Code Systems

Review the code systems used by the measure to determine the following:

- Whether new codes have been added to or deleted from the code systems that may affect the measure
- Whether 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. [190]

When maintaining eCQM value sets, it is important to align with the vocabulary recommendations made by HIT Standards Committee Clinical Quality Workgroup and Vocabulary Task Force. Section 3, Chapter 17, Codes, Code Systems, and Value Sets provides more information on the procedure.

31.1.2.2 Gather Information

The measure developer 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 NQF or CMS.

If the 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

[190] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *ICD-10: Statute and Regulations.* Available at: http://www.cms.gov/ICD10/02d_CMS_ICD-10_Industry_Email_Updates.asp. Accessed on: March 14, 2016.

Section 3, Chapter 31.3, 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 the following:

- National performance rates
- State and regional performance rates
- Variations in performance rates
- Validity of the measure and its constituent data elements
- Reliability of the measure and its constituent data elements.

31.1.2.3 Determine the Recommended Disposition of the Measure

Criteria which 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 31.5, Possible Outcomes of Maintenance Reviews.

The possible dispositions are:

- Retain
- Revise
- Remove
- Retire
- Suspend.

31.1.2.4 The COR Reviews the Recommendation for Approval

Forward the recommendations to the COR, along with an updated Measure Information Form (and eCQM files, if appropriate), and value sets, and Summary of Changes/Release Notes. If significant changes were made, an updated Measure Justification Form and Measure Evaluation Report may be necessary.[191]

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 as necessary. If the recommendation is approved, the COR notifies the measure developer of the approved course of action.

31.1.2.5 Implement the Approved Action

For measures that are proposed to be revised, suspended, 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 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. This may include announcements through usual communication modes for the project, arranging for reprogramming, notifying other CMS measure developers, notifying NQF if a measure is currently endorsed, or re-education of providers. Notify the Measures Manager to ensure that the CMS Measures Inventory is updated appropriately regardless of the disposition decision.

[191] The NQF submission may be acceptable for this deliverable. National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx Accessed on: March 14, 2016.

31.1.2.6 Assist the COR in Notifying NQF of the Updated Measure

After a measure is endorsed by NQF, CMS (as the measure steward) is required to submit a status report of the measure specifications to 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 is responsible to 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 may need to notify NQF of the changes each time they occur. If no changes are made, only one annual update is required.

NQF provides a standardized template for submission of an annual measure maintenance update that is prepopulated with measure information. CMS will direct NQF regarding the appropriate measure developer to contact for the annual update. The measure developer is responsible to prepare this report for NQF. The measure developer must also obtain COR review and approval before submitting the report to NQF. NQF may conduct its own ad hoc review if the changes materially affect the measure's original result. The measure developer responsible for measure maintenance should be aware of NQF's measure maintenance schedule and when the annual update is due to NQF. The due date for the measure developer's measures updates should be confirmed annually with NQF because schedules may change. The measure developer should also inform NQF of any contact information changes so that notifications can go to the correct recipients.

31.1.2.7 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 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 31.2, Comprehensive Reevaluation. The CMS measure developer is also responsible for updating any CMS documentation of the measure to reflect changes made by the measure owner 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 owner's consent. The measure developer will also be responsible for ongoing surveillance of the literature addressing the measure and alerting the COR to possible issues.

31.1.2.8 Submit the NQF Annual Status Update Report

The measure developer prepares the annual update report of the measure specifications, submits it to the COR for approval, then submits it online to NQF. Some measure maintenance contracts may require updates more than once a year. In those cases, measure developers should notify NQF of the changes as often as appropriate.

NQF staggers deadlines for annual maintenance submissions throughout the year. 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 may request a different quarter for their annual updates.

Confirm the deadline for each annual update with NQF. These update requirements also appear on measure developers' NQF dashboards. It is the responsibility of the measure developers 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 and COR may seek guidance from the Measures Manager during any stage of this process.

31.2 Comprehensive Reevaluation

Measure developers 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 here. 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 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 NQF.
Figure 43 lists comprehensive reevaluation deliverables.

Comprehensive Reevaluation Deliverables

1. An updated Measure Information Form detailing all recommended changes to the measure.
2. For measure developers maintaining eCQMs, the revised HQMF, SimpleXML, HTML, 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 with the updated MIF.
4. An updated Measure Justification Form documenting the environmental scan results, any new controversies about the measure, and any new data supporting the measure's justification.
5. An updated Measure Evaluation Report describing measure performance compared to the measure evaluation criteria and the performance of the measure.
6. An updated business case that reports on the measure performance trend and trajectory as compared to the projections made during measure development, including recommendations.
7. NQF endorsement maintenance online submission documentation (at the scheduled three-year endorsement maintenance).
8. If it is time for three-year maintenance review (comprehensive reevaluation) but the NQF project is not ready, an annual update report may be submitted online.

Figure 43. Comprehensive Reevaluation Deliverables

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 helps CMS maintain the best measures over time.

The comprehensive reevaluation process includes nine steps that fall into the following phases:

- Gathering information generated since the measure's development or since the last comprehensive reevaluation—whichever occurred most recently
- Measure evaluation and recommended action based on the evaluation
- Approving and implementing the action.

The comprehensive reevaluation process assumes that the measure developer has been monitoring the scientific literature and clinical environment related to the measure, including relevant clinical guidelines.

31.2.1 For Measures in CMS Programs Not Owned by CMS or Maintained by a CMS Measure Developer

Measure developers responsible for 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 (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.

31.2.2 Harmonization

During comprehensive reevaluation, make full consideration to determine if 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 if similar existing measures are related or competing is described in Section 3, Chapter 31.1.1, Potential for Harmonization. It is part of the National Quality Strategy to foster alignment of performance measures as much as possible, so these considerations are particularly important during comprehensive review.

31.2.3 Procedure

31.2.3.1 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:

- The rulemaking cycle for any regulatory process governing the measure set in question
- The NQF measure maintenance schedule.

31.2.3.2 Gather Information

During measure monitoring, ongoing surveillance is conducted. Summarize the findings of the environmental scan and update the Measure Justification Form. 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 the following:

- 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 subject matter or measurement experts
- Patients' perspective on the measures under review
- Reevaluation of the business case supporting the measure
- Feedback that has been received since the measure was last evaluated (either the initial evaluation or the last comprehensive reevaluation, whichever is most recent).

Obtain measure performance information including but not limited to the following:

- 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
- The 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
- The results of audits 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. The analysis should determine if any disparities identified earlier 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.

Compare the information gathered with the 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.

31.2.3.3 Convene a TEP

During comprehensive reevaluation, a TEP is usually convened to assess the measures. 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 on the membership. Section 3, Chapter 12, Technical Expert Panel (TEP) provides details of the standardized process of issuing a call for nominations and convening a TEP.

Present 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 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 24, Measure Evaluation describes the measure evaluation criteria. Measure selection criteria are discussed in Section 2, Chapter 4, Measure Implementation.

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.

31.2.3.4 Identify and Document Changes That Will Be Recommended

For each measure, compile the information gathered in the steps above using the measure evaluation criteria to update the Measure Justification Form. 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, consider converting any ICD-9 codes to ICD-10. 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 delays the compliance date for ICD-10 from October 1, 2014, to October 1, 2015, at the earliest.

When maintaining eCQMs value sets, it is important to align with the vocabulary recommendations made by HIT Standards Committee Clinical Quality Workgroup and Vocabulary Task Force. More information on these requirements is found in Chapter 17 – Codes, Code Systems, and Value Sets.

Update the MIF (HQMF including: XML, SimpleXML, HTML, and value sets for eCQM) 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 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.

31.2.3.5 Determine the Preliminary Recommended Disposition of the Measure

Criteria which 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 31.5, Possible Outcomes of Maintenance Reviews.

The possible dispositions include the following:

- Retain
- Revise
- Remove
- Retire
- Suspend.

31.2.3.6 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 18.

Table 18. Extent of Measure Evaluation as a Function of Prior Comprehensive Evaluation and Measure Use

	Measure in Use	Measure Not in Use
First comprehensive reevaluation	Measure developer should obtain data from the population where the measure was implemented and analyze it to augment previous evaluation findings obtained when the measure was initially developed and endorsed. If material changes are made at this time, the revised measure will need to be tested.	Measure developer should conduct expanded testing relative to the initial testing conducted during development (e.g., expand number of groups/patients included in testing compared to prior testing used to support the measure's initial development and submission for endorsement).
Subsequent comprehensive reevaluation	If measure has not materially changed, CMS may require minimal analysis and may use prior data for NQF maintenance if past results demonstrated a high rating for reliability and validity of the measure.	If measure has not materially changed, measure developer may submit prior testing data when past results demonstrated adequate reliability and validity of the measure.

If the measure requires testing, develop a plan. The components of a testing plan are described in Section 2, Chapter 3, Measure Testing.

31.2.3.7 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. Consult 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.

Analyze the comments received and refine the measure as indicated. Document any changes in the MIF (HQMF, SimpleXML, HTML, and value sets for eCQMs). If necessary, update the Measure Justification Form 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.

31.2.3.8 Implement the Approved Action

After review, the measure maintenance contractor may be responsible to help CMS implement the chosen measure disposition. For measures that are proposed to be revised, suspended, 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. This may include announcements through usual communication modes for the project, arranging for reprogramming, notifying other CMS contractors, or re-education of providers. Notify the Measures Manager to ensure that the CMS Measures Inventory is updated appropriately regardless of the disposition decision.

31.2.3.9 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 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 are responsible to 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.

Be aware that the 3-year maintenance reevaluation follows the NQF maintenance schedule and not necessarily the measures' dates of endorsement. Depending on the volume of measures being reevaluated during an endorsement cycle, committees and topic areas from one cycle may need to be extended and scheduled to convene in a subsequent cycle. As a result, measures may be subject to early or late 3-year maintenance reevaluation. Measures with initial endorsement dates falling within 18 months of the committee meeting are exempt from endorsement maintenance review until the topic's next endorsement cycle. These measures still require annual updates to be submitted to NQF.

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 the 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.

31.3 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 ad hoc review reflects those shifts; to preserve measure stability, it should be reserved for only those instances where new evidence indicates that 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. Review deliverables are listed in Figure 44.

31.3.1 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 come 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, 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 endorsement 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.

Ad Hoc Review Deliverables

1. Updated Measure Information Form, if the ad hoc review results in changes to the measure specifications.
2. For measure developers maintaining eCQMs, the updated HQMF, SimpleXML, HTML, and value sets, if the ad hoc review results in changes to the measure specifications.
3. Updated Measure Justification Form, 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, if the review resulted in a change to the measure's strengths and/or weaknesses.

Figure 44. Ad Hoc Review Deliverables

31.3.2 Deferring an Ad Hoc Review

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 a 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 particular programs which may have their own schedules (such as hospital measures which are governed by different rulemaking schedule requirements), a decision may take some time to be implemented in all the programs using a given measure.

31.3.3 Procedure

The CMS measure developer remains responsible to monitor the maintenance performed by the steward even if the measure developer is not the measure steward (that is, not the steward or 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-9-CM, and 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 with a measure for which CMS is not the steward, the measure developer responsible for monitoring the measure should bring the matter to the attention of the COR to determine what action, if any, is necessary. CMS may contact the steward to determine if the steward is aware of the concern and what action is being taken. If the measure is NQF-endorsed, CMS may consider requesting NQF to 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 if an ad hoc review should be conducted
- Conducting the review and recommending an outcome
- Approving and implementing the approved outcome.

31.3.3.1 Determine whether the Concern Is Significant

If the clinical practice underlying the measure is causing harm to the patients, the measure should be at least revised, if not suspended or retired. This includes harm caused by unintended consequences of the measure. Though there is no defined schedule for this process, CMS or NQF may require the measure developer 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 if 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. NQF may also be the source of the request for urgent ad hoc review depending on the nature and source of the concerns.

If the 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 (the remaining steps). If the measure maintenance contractor is different from the measure developer monitoring the measure, the measure maintenance contractor should be responsible for the review.

If the experts determine that the issue is not significant, the issue should be documented for consideration at the next scheduled review.

31.3.3.2 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.

Conduct 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 Chapter 6—Information Gathering.

31.3.3.3 Consult with the Experts, Especially the TEP

Consult 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 (TEP, guideline writers, or 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 both costs and time

31.3.3.4 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.

31.3.3.5 Recommend a Course of Action to the COR

Criteria which form the basis for the disposition decision for each measure and description of the possible outcomes are discussed in Section 3, Chapter 31.5, Possible Outcomes of Maintenance Reviews.

Depending on the findings from the previous steps, the recommendation may be one of the following:

- Retain
- Revise
- Remove
- Retire
- Suspend.

Submit 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.

31.3.3.6 The COR Reviews the Recommendation for Approval

Forward the recommendations to the COR, with the updated Measure Information Form and summary of changes or Release Notes as indicated. If significant changes were made, an updated Measure Justification Form and Measure Evaluation Report may be necessary.[192]

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.

[192] The NQF submission may be acceptable for this deliverable. National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: March 14, 2016.

31.3.3.7 Implement the Approved Action

For measures that are proposed to be revised, suspended, 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. This may include announcements through usual communication modes for the project, arranging for reprogramming, notifying other CMS measure developers, or re-education of providers. Notify the Measures Manager to ensure that the CMS Measures Inventory is updated appropriately regardless of the disposition decision.

31.4 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 or measure revision, notify 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 the 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 of the following 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 endorsement evaluations 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 justifying evidence to require such 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 maintenance review, NQF will notify the steward whether NQF has 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 responsible for helping 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.

31.5 Possible Outcomes of Maintenance Reviews

The following are potential measure dispositions that CMS can choose based on recommendations made as a result of any of the three maintenance review types discussed above.

- Retain—Keep the measure active with its current specifications and minor changes.
- Revise—Update the measure's current specifications to reflect new information.
- Retire—Cease to collect or report the measure indefinitely. This applies only to measures owned by CMS. CMS will not continue to maintain these measures. (When retiring a measure from a set, consider other measures that may complement the remaining set as a replacement.)
- Remove—A measure is no longer included in a particular 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 particular 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.
- Suspend—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.)

If the measure continues to meet the measure evaluation criteria (Section 3, Chapter 24, Measure Evaluation) and the measure selection criteria (Section 3, Chapter 28, 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, consider recommending other available qualifying measures as replacements.

Figures 45, 46 and 47 are adapted from the "Standard CMS Measure Implementation Determination Criteria" and lists the criteria CMS uses to make decisions regarding the various dispositions described above.[193]

[193] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Standard CMS Measure Implementation Determination Criteria*. [unpublished] Prepared by E. Garcia and K. Goodrich for the CMS Quality Measures Task Force, March 26, 2012.

Revise or Retain

Core Criteria		Optional Criteria
Measure is responsive to specific program goals and statutory requirements. 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. Measure addresses one or more of the six NQS priorities and the CMS Quality Strategy. Measure selection promotes alignment with CMS program attributes and across HHS programs. Measure reporting is feasible and measures have been fully developed and tested. Measure results and performance should identify opportunities for improvement. CMS will not select measures that are topped out. 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).	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.	Should be electronically specified whenever possible. Should be aligned with the EHR Incentive Programs where applicable.

Figure 45. CMS Criteria for Measure Disposition: Revise or Retain

Retire

Core Criteria	Optional Criteria
Measure is owned by CMS, and CMS will no longer maintain the measure.	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.

Figure 46. CMS Criteria for Measure Disposition: Retire

Remove	
Core Criteria	**Optional Criteria**
Measure is no longer used in a CMS program. If the measure is owned by CMS, CMS continues to maintain it even after removal. If another entity owns the measure, other payers/ purchasers/programs using the measure are responsible for determining if the owner is continuing to maintain the measure.	Collection or public reporting of a measure leads to negative unintended consequences other than patient harm. The measure, as currently specified, cannot be reported. A measure does not align with current clinical measure performance or 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.

Figure 47. CMS Criteria for Measure Disposition: Remove

Section 4. Forms and Templates

1 ENVIRONMENTAL SCAN OUTLINE

The following is an example of an environmental scan outline.

1. *Cover Page*, including the Task Order title and contract number, contractor 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 the development of the Measure Justification Form in the later phases of measure development or reevaluation.
   - Information gathered to build the business case for the measure:
     - Incidence/prevalence of 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 identifying 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 it responds to
   - 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:
     - If available, data source(s) used
     - 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)
      - Include 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
      - List of stakeholders and their relevance to the project
      - Manner of interaction (e.g., telephone call, face-to-face meeting, survey)
      - Summary of findings with recommendations

- Summary of Solicited and Structured Interviews, if applicable (might refer to any of the above 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

This form is a guide for measure developer 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 <measure developer name> to develop <measure (set) name or description>. The contract name is <insert contract name>. The contract 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.[194]

Measure uses (select all that apply):

Check all the current and planned uses for the measure.

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.[195]

Measure Impact on Care:

Estimate the expected performance of the measure on the quality of care. 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. During measure maintenance, compare the actual performance to the estimates and report the differences with analysis and recommendations.[196]

Measure Impact on Health Outcomes:

Estimate the expected performance of the measure on health outcome(s). Follow the approach detailed for Measure Impact on Care.

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.

Influencing Factors:

There may 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 as possible to document any observed influencing factors affecting measure implementation and/or performance.

[194] Concepts incorporated from Business Case sample prepared by Yale for THA/TKA resource use measure in 2014.
[195] Concepts incorporated from Business Case sample prepared by Yale for THA/TKA resource use measure in 2014.
[196] Concepts incorporated from Business Case sample prepared by QP Rhode Island for Antipsychotic Polypharmacy and Warfarin Monitoring in 2009.

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.

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.[197]

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.[198]

Description of model(s) and formulas used:

Describe the assumptions, variables, and formulas used to construct the business case.[199]

Limitations of analysis:

Describe any limitations in the data or the assumptions used in the business case.[200]

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.[201]

[197] Concepts incorporated from Business Case sample prepared by FMQAI for Glycemic Control measures in 2014.
[198] Ibid.
[199] Concepts incorporated from Business Case sample prepared by FMQAI for Glycemic Control measures in 2014.
[200] Ibid.
[201] 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 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. It is recognized that at the time of the initial submission some of the data and details may be limited, however, it is expected that the measure developer will have fully detailed information in the final submission.>*

CMS has intentionally aligned this form with NQF's 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 a NQF submission.

The business case is about more than the possible financial benefits of a measure. Please make sure to 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 <measure developer name> to develop <measure (set) name or description>. The contract name is <insert contract name>. The contract number is <project number.>

Date:

Information included is current on <Insert Date>.

Measure Description:

Use the Measure Title, as it is listed in the Measure Information Form. 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.6. Text from NQF submission form may be inserted here if available >

Denominator Statement:

Brief, narrative description of the target population being measured

<Field part of NQF Measure Submission Form, Field S.7. 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, 1b.1, 1b.4, and 1b.5c. 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.[202]

<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 >

Measure Impact on Care:

Describe the linkages and steps between the measure focus and anticipated improvements to care 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.

[202] Concepts incorporated from Business Case sample prepared by Yale for THA/TKA resource use measure in 2014.

<Information can be compiled from fields in NQF Measure Evidence Form, text from NQF submission may be inserted here if available >

Measure Impact on Health Outcomes:

Describe the linkages and steps between the measure focus and anticipated health outcomes. Include details on structure, process, intermediate outcome(s), and health outcome(s). 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, please state none. Please include a brief explanation of why there is still a strong business case for measure (i.e. lives saved, improved health outcomes, etc.) >

Influencing Factors:

<Please 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 4c.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, Measure Specification, including fields S.12-25. Text from NQF submission may be inserted here if available >

Limitations of analysis:

<Information provided as part of NQF Measure Submission. Text from NQF submission may be inserted here if available >

Net benefit:

Report benefits (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, including fields 4b.1 and 1b.2 and 1b.4.

Call for Measures Web Posting

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 <measure developer name> to develop <measure (set) name or description>. The contract name is <insert contract name>. The contract number is <project number.> As part of its measure development process, CMS 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, please include all required documentation following 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 Measure Information Form 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.

4 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. It tracks very closely to the NQF online measure submission, Version 7.0, and references corresponding fields from that submission form. 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 MIF for measures that are not submitted to NQF.

Measure Name

Use the Measure Title, as it will be required in the Specifications tab (item De.2.). It should be brief and include the measure focus and the target population.

Introduction Tab

This tab gives general information about the NQF measure submission. There are no fields to complete on this tab, but there are links to the measure evaluation criteria and the Measure Steward Agreement.

NQF Conditions Tab

This tab lists conditions that NQF requires before a proposed measure may even be considered for suitability as a voluntary consensus standard.

A. The measure is in the public domain or a Measure Steward Agreement is signed. (All non-government organizations must sign a Measure Steward Agreement even if measures are made publicly and freely available.)
B. The measure owner/steward verifies that there is an identified responsible entity and a process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years.
C. The intended use of the measure includes both accountability applications (including public reporting) and performance improvement to achieve high-quality, efficient healthcare.
D. The measure is fully specified and tested for reliability and validity.
E. The measure developer/steward attests that harmonization with related measures and issues with competing measures have been considered and addressed, as appropriate.
F. The requested measure submission information is complete and responsive to the questions so that all the information needed to evaluate all criteria is provided.

A box at the end of the page requires developers to attest that they have read and accepted the conditions.

Specifications Tab

Descriptive Information

Measure Type De.1.

Identify a measure type from among the following. PROs include health-related quality of life, functional status, symptom burden, experience with care, and health-related behavior.

- Process
- Process: Appropriate Use
- Outcome
- Cost / Resource Use
- Efficiency
- Outcome: PRO
- Structure
- Intermediate Clinical

Measure Title De.2.

Briefly convey as much information as possible about the measure focus and target population.

Brief Description of Measure De.3.

This should also be concise, but include type of score, measure focus, target population, and time frame.

If Paired or Grouped De.4.

Give the reason the measure must be reported with other measures to appropriately interpret results.

Measure Specifications

The following items follow the NQF requirements for measure submission and provide information required for measure evaluation.

Measure-specific Web Page S.1.

Provide a URL link to a Web page where current, detailed specifications can be obtained that include code lists, risk adjustment model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.

If This Is an eMeasure S.2a.

HQMF specifications must be attached. Attach the output from the MAT. If the MAT was not used, contact NQF for further directions. Use the specification fields from the online form for the plain-language description of the specifications. Use the MIF to collect information for use when entering data into the MAT.

Data Dictionary, Code Table, or Value Sets 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 NQF for further directions.

For Endorsement Maintenance S.3.

If this form is being used for endorsement maintenance, briefly describe any changes to the measure specifications since the last endorsement date, and explain the reasons for the changes.

Numerator Statement S.4.

Briefly describe the measure focus—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}

For outcome measures, state the outcome being measured. Describe calculation of the risk-adjusted outcome later in the calculation algorithm.

Numerator Details S.5.

Include all the information necessary to identify and calculate the cases from the target population with the target process, condition, event, or outcome. Give 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.

For outcome measures, describe how the observed outcome is identified and counted. The calculation algorithm should also describe how to calculate the risk adjustment.

Give the time period in which data will be aggregated for the measure, such as 12 months, 3 years, or another specified look-back period.

Denominator Statement S.6.

Give 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. For example:

> Patient's [age] with [condition] in [setting] during [time frame]

Denominator Details S.7.

Give all definitions and instructions needed to identify and calculate the target population. For measures based on a coded data set, identify the code set, the specific codes, descriptors, definitions, and specific data collection items 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.

Denominator Exclusion (NQF Includes "Exception" in the "Exclusion" Field) S.8.

Identify patients in the target population but who should not receive the process (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. For example:

> Patients in the [target population] who [have some additional characteristic, condition, procedure]

Denominator Exclusion Details (NQF Includes "Exception" in the "Exclusion" Field) S.9.

All information needed to identify and calculate exclusion from the denominator, such as definitions and/or specific data collection items and responses. For measures based on a coded data set, identify

the code set, the 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.

Stratification Details/Variables S.10.

Provide instructions for calculating the measure by category (for example, age) including the stratification variables, all codes, logic, and definitions. 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.

Risk Adjustment Type S.11.

Select risk adjustment type. Provide specifications for risk stratification in S.12 and for statistical model in S.14–15.

- No risk adjustment or risk stratification
- Stratification by risk category/subgroup
- Statistical risk model
- Other (S.13.a.)

Type of Score S.12.

- Count
- Rate/proportion
- Ratio
- Categorical (e.g., yes or no)
- CV (e.g., an average)
- Other (specify)

Interpretation of Score S.13.

Classifies whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score.

Calculation Algorithm/Measure Logic S.14.

Describe the sequence of steps necessary to calculate the measure score, including identifying the target population; exclusion; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; and any other calculations.

You may provide a diagram of the Calculation Algorithm/Measure Logic described in S.18. at a measure-specific Web page URL identified in S.1. or in an attached appendix.

Sampling S.15.

If the measure is based on a sample or survey, provide instructions for obtaining the sample, conducting the survey, and minimum response rate required. If the measure is a PRO-PM, identify whether (and how) proxy responses are allowed.

Survey/Patient-Reported Data S.16.

If the measure is based on a survey, provide instructions for conducting the survey 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 (imputation, delete case, etc.). This item is required for composite measures and PRO-PMs.

Data Source S.17.

Indicate all sources for which the measure is specified and tested.

- Claims (Only)
- Claims (Other)
- EHRs Hybrid
- Electronic Health Record (Only)
- Imaging-Diagnostic
- Laboratory
- Pharmacy
- Registry
- Provider Tool
- Management Data
- Paper Records
- Patient Reported Data
- Non-Medical Data
- Other

Data Source or Collection Instrument S.18.

Identify the specific data source/data collection instrument, such as the name of database, clinical registry, collection instrument, etc. If the measure is a PRO-PM, identify the specific tools being used to collect the measure information and standard methods, modes, and languages of administering the tools.

Data Source or Collection Instrument (Reference) S.19.

Give the reference for the data source or collection instrument. Either attach a copy or specify the URL where it can be found.

Level of Analysis 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

Care Setting 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
- Inpatient Rehabilitation Facility
- Long Term Acute Care
- Birthing Center
- No Applicable Care Setting
- Other

Composite Performance Measure 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.

5 BLANK MEASURE INFORMATION FORM TEMPLATE

Project Title:

<List the project title as it should appear.>

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) has contracted with <measure developer name> to develop <measure (set) name or description>. The contract name is <insert contract name>. The contract number is <project number.>

Date:

Information included is current on <Insert Date>.

Measure Name

Descriptive Information

Measure Name (Measure Title De.2.)

Measure Type De.1.

Brief Description of Measure De.3.

If Paired or Grouped De.4.

Measure Specifications

Measure-specific Web Page S.1.

If This Is an eMeasure S.2a.

Data Dictionary, Code Table, or Value Sets S.2b.

For Endorsement Maintenance S.3.1 and S.3.2

Numerator Statement S.4.

Numerator Details S.5.

Denominator Statement S.6.

Denominator Details S.7.

Denominator Exclusion (NQF Includes “Exception” in the “Exclusion” Field) S.8.

Denominator Exclusion Details (NQF Includes “Exception” in the “Exclusion” Field) S.9.

Stratification Details/Variables S.10.

Risk Adjustment Type S.11.

Type of Score S.12.

Interpretation of Score S.13.

Calculation Algorithm/Measure Logic S.14.

Sampling S.15.

Survey/Patient-Reported Data S.16.

Data Source S.17.

Data Source or Collection Instrument S.18.

Data Source or Collection Instrument (Reference) S.19.

Level of Analysis S.20.

Care Setting S.21.

Composite Performance Measure S.22.

6 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 and references corresponding fields from that submission. 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.

Measure Name

Give the measure name that was used for the MIF. Use the Measure Title, as it will be required in the Specifications tab (item De.2.). It should be brief and include the measure focus and the target population.

Type of Measure

Identify a measure type from the following listed items. Patient-reported outcomes (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.

- Composite
- Cost-resource use
- Efficiency
- Outcome
- Patient-reported outcome (PRO)
- Process
- Structure

Importance Tab

NQF considers importance to measure and report a threshold criterion. Measures must meet all three subcriteria or they will not pass the criterion to be recommended for endorsement.

Evidence (subcriterion 1a)

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.

- Complete 1a.1 and 1a.12 for all measures.
- Complete **EITHER 1a.2, 1a.3 or 1a.4**, as applicable for the type of measure and evidence.
- For composite performance measures:
- A separate evidence form is required for each component measure unless several components were studied together.
- If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.
- All information needed to demonstrate meeting the evidence subcriterion (1a.) must be in this form. An appendix of supplemental materials may be submitted, but there is no guarantee that NQF will review it.
- For NQF submissions, the evidence information should not exceed a maximum of 20 pages.

- Information on evidence should be sufficient for recipients (the NQF Steering Committee or the Centers for Medicare & Medicaid Services [CMS]) to understand to what degree the evidence for the measure meets evaluation criteria.

Summary of Measure Evaluation Subcriterion 1a. Evidence to Support the Measure Focus (for reference only)

The measure focus is a health outcome or is evidence-based, demonstrated 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 clinical outcome—a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence that the measured intermediate clinical outcome leads to a desired health outcome.
- The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or GRADE guidelines.
- Experience with care—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 (NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care;[203] AQA Principles of Efficiency Measures[204]).

1a.1.—This is a Measure of: *(should be consistent with type of measure entered in De.1)*

Outcome.

Health outcome: Name the health outcome. Health outcome includes PROs such as health-related quality of life, functional status, symptom or burden, experience with care, and health-related behaviors.

Intermediate clinical outcome: Name the intermediate outcome.

Process: Name the process.

Structure: Name the structure.

Other: Name what is being measured.

1a.12—Logic Model

Briefly state or diagram the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s).

[203] National Quality Forum. *Measurement Framework: Evaluating Efficiency Across Patient-Focused Episodes of Care*. Washington, DC: Available at: http://www.qualityforum.org/Publications/2010/01/Measurement_Framework__Evaluating_Efficiency_Across_Patient-Focused_Episodes_of_Care.aspx. Accessed on: May 31, 2015.
[204] American Quality Alliance. *AQA: AQA Principles of "Efficiency" Measures*. American Quality Alliance: Revised Jun 2009. Available at: http://www.aqaalliance.org/files/PrinciplesofEfficiencyMeasurement.pdf. Accessed on: May 31, 2015.

Respond to only one section below—Either 1a.2, 1a.3, or 1a.4*

1a.2.—Outcome Measures including Patient Reported Outcomes

Briefly state or diagram the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

1a.3.—Systematic Review(s) of the Evidence for Intermediate Outcome, Process, or Structure Performance Measures (If the evidence is not based on systematic review, go to section 1a.4)

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. (IOM)

- Clinical Practice Guideline recommendation (with evidence review)
- U.S. Preventive Services Task Force 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.

Source of Systematic Review (SR): • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
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?	

1a.4.—Other Source of Evidence

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which performance measure is based.

1a.4.1.—Briefly SYNTHESIZE the evidence that supports the measure

A list of references without a summary is not acceptable.

1a.4.2.—What process was used to identify the evidence?

Identify guideline recommendation number and/or page number and quote verbatim the specific guideline recommendation.

1a.4.3.—Provide the citation(s) for the evidence

Grade assigned to the quoted recommendation with definition of the grade.

For Maintenance of Endorsement (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.

Performance Gap – Opportunity for Improvement (1b.)

1b.1.—Rationale

Briefly explain the rationale for this measure (i.e., the improvements in quality envisioned by use of this measure).

If a PRO-PM (eg, HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. (Describe how and from whom their input was obtained.)

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 1c.3.

1b.2.—Performance Scores

Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. (This is required for endorsement maintenance. 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 (4b.1.) under Usability and Use.

1b.3.—Summary of Data Indicating Opportunity

If no or limited performance data on the measure as specified is reported in 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.

1b.4.—Disparities

Provide data on how the measure, as specified, addresses disparities (current and over time) by population group: race or ethnicity, gender, age, insurance status, social risk factors, and disability. This is also required for endorsement maintenance. 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. This information also will be used to address the subcriterion on improvement (4b) under Usability and Use.

> 1b.5.—If no or limited data on disparities from the measure as specified is reported in 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 1b.4.

Scientific Acceptability Tab

This section is now submitted as an MS Word document attachment about measure testing. The template is available on the NQF Submitting Standards website. Items from that document are reproduced here for reference and other submission purposes.

- Measures must be tested for all the data sources and levels of analyses that are specified.
- For all measures, sections 1, 2a.2., 2b.2., 2b.3., and 2b.5. must be completed.
- For outcome and resource use measures, section 2b.4. also must be completed.
- If specified for multiple data sources/sets of specifications (such as claims and EHRs), section 2b.6. also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a.2.) and validity (2b.2.–2b.6.) must be in this form. An appendix for supplemental materials may be submitted, but there is no guarantee that NQF will review it.
- For NQF submissions, the Scientific Acceptability information should not exceed a maximum of 20 pages.
- Information on scientific acceptability should be sufficient for recipients (the NQF Steering Committee or CMS) to understand to what degree the testing results for the measure meet evaluation criteria for testing.

Reliability and Validity Evaluation Criteria Summary (for reference only)

The numbers correspond to the reliability and validity items on NQF's measure testing attachment. The numbers also correspond to the NQF Measure Testing subcriteria available on NQF's website.

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.

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).

2b.2.—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 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.

2b.3.—Exclusion is supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion. 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.

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). Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

2b.4.—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. 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 socioeconomic status (SES) factors or sociodemographic status (SDS) factors).

OR

Rationale/data support no risk adjustment/ stratification.

2b.5.—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 1 percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent versus 75 percent) is clinically meaningful, or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

OR

There is evidence of overall less-than-optimal performance.

2b.6.—If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b.7.—For eMeasures, composites, and PRO-PMs (or other measures susceptible to missing data), 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 nonresponders) and how the specified handling of missing data minimizes bias.

Data Sample Description (1.)

This description should be the same as what is submitted with the MIF. The first five questions apply to all measure teasing. For NQF submissions, all Scientific Acceptability items are entered on NQF's Measure Testing template as an attachment.

1.1.—What Type of Data Were Used for Testing?

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 S.23)
- Abstracted from paper record
- Administrative claims
- Clinical database/registry
- Abstracted from EHR
- eMeasure (HQMF) implemented in EHRs
- Other:
- Measure Tested With Data From:
- Abstracted from paper record
- Administrative claims
- Clinical database/registry
- Abstracted from EHRs
- eMeasure (HQMF) implemented in EHRs
- Other:

1.2.—Identify the Specific Dataset

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).

1.3.—What Are the Dates of the Data Used in Testing?

Enter the date range.

1.4.—What Levels of Analysis Were Tested?

Testing must be provided for all the levels specified and intended for measure implementation (e.g., individual clinician, hospital, and health plan).

- Measure Specified to Measure Performance of: (must be consistent with data sources entered in S.26)
- Individual clinician
- Group/practice
- Hospital/facility/agency
- Health plan
- Other:
- Measure Tested at Level of:
- Individual clinician
- Group/practice
- Hospital/facility/agency
- Health plan
- Other:

1.5.—How Many and Which Measured Entities Were Included in the Testing and Analysis?

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).

1.6.—How Many and Which Patients Were Included in the Testing and Analysis?

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.

1.7.—Sample Differences, if Applicable

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.

Reliability Testing (2a.2)

If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required—in 2a.2.1., check critical data elements; in 2a.2.2., enter “see section 2b2 for validity testing of data elements”; and skip 2a.2.3. and 2a.2.4.

2a.2.1.—Level of Reliability Testing

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)

2a2.2.—Method of Reliability Testing

Describe the method of reliability testing for each level used. Describe the steps—do not just name a method. What type of error is it testing? What statistical analysis was used?

2a2.3.—Statistical Results from Reliability Testing

What were the statistical results from reliability testing? 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.

2a2.4.—Interpretation

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?

Validity Testing (2b.2)

2b2.1.—Level of Validity Testing

What level of validity testing was conducted? (It may be more than one level.)

- Critical data elements (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).

2b2.2.—Method of Validity Testing

For each level tested, describe the method of validity testing and what it tests (describe the steps—do not just name a method; what was tested [e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected]; and what statistical analysis was used).

2b2.3.—Statistical Results from Validity Testing

Provide statistical results and assessment of adequate validity, such as correlation, *t* test.

2b2.4.—Interpretation

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?

Exclusions Analysis (2b3)

If there are no exclusions, indicate that this section is not applicable and skip this section.

2b3.1.—Method of Testing Exclusion

Describe the method of testing exclusion and what it tests. Describe the steps—do not just name a method; what was tested (e.g., whether exclusion affects overall performance scores); what statistical analysis was used.

2b3.2.—Statistical Results From Testing Exclusion

What were the statistical results from testing exclusion? Include overall number and percentage of individuals excluded, frequency distribution of exclusion across measured entities, and impact on performance measure scores.

2b3.3.—Interpretation

What is your interpretation of the results in terms of demonstrating that exclusion 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 exclusion.

Risk Adjustment or Stratification for Outcome or Resource Use Measures (2b4)

If the measure is not an intermediate, or health outcome, or PRO performance measure, or resource use measure, skip to section 2b5.

2b4.1.—Method of controlling for differences

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:

2b4.2.—Rationale Why Risk Adjustment Is Not Needed

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 (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3.—Conceptual, Clinical, and Statistical Methods

Describe the conceptual, clinical, and statistical methods and criteria used to select patient factors **(clinical factors or sociodemographic 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 x or higher; patient factors should be present at the start of care and not related to disparities).

2b4.4a—Statistical Results

What were the statistical results of the analyses used to select risk factors?

2b4.4b

Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5.—Method Used to Develop the Statistical Model or Stratification Approach

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; what statistical analysis was used).

2b4.6., 2b4.7., 2b4.8., 2b4.9.—Results

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix). If stratified, skip to 2b4.9.

2b4.6. Statistical Risk Model Discrimination Statistics (such as, c-statistic, R^2)

2b4.7. Statistical Risk Model Calibration Statistics (such as the Hosmer-Lemeshow statistic)

2b4.8. Statistical Risk Model Calibration—Risk decile plots or calibration curves

2b4.9. Results of Risk Stratification Analysis

2b4.10.—Interpretation

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)?

2b4.11.—Optional Additional Testing for Risk Adjustment

This is not required, but it would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set, sensitivity analysis for missing data, other methods.

2b5. Identification of Meaningful Differences in Performance

The intent of this section is to go beyond demonstrating a performance gap and address statistical significance, if possible.

2b5.1.—Method

Describe the method for determining if 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 1b. Performance Gap.

2b5.2.—Statistical Results

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?

2b5.3.—Interpretation

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?

2b6. Comparability of Multiple Data Sources/Methods

This item is directed to measures that are risk-adjusted (with or without SDS 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 eMeasures). 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 SDS 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.

2b6.1.—Method

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?

2b6.2.—Statistical Results

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)?

2b6.3.—Interpretation

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?

2b7. Missing Data Analysis and Minimizing Bias

2b7.1—Method

Describe the method of testing 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 nonresponders) and how the specified handling of missing data minimizes bias. Describe the steps—do not just name a method. What statistical analysis was used?

2b7.1—Missing Data Analysis

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.

2b7.3—Interpretation

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 nonresponders) 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?

Feasibility Tab

The numbering for this section of the submission corresponds to the Feasibility section of the NQF Measure Evaluation criteria found on NQF's website.

Data Elements Generated as Byproduct of Care Processes (3a.)

3a.1. How are the data elements needed to compute measure scores generated? List all that apply from the following.

Data used in the measure are:

- Generated by and used by healthcare personnel during the provision of care (for example, blood pressure, lab value, medical condition).

- Coded by someone other than the person obtaining original information (for example, DRG, ICD-10 codes on claims).
- Abstracted from a record by someone other than the person obtaining original information (for example, chart abstraction for quality measure or registry).

3a.2. If "Other," please describe.

Electronic Sources (3b.)

3b.1. Are the data elements needed for the measure as specified available electronically (elements that are needed to compute 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.

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.

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make it available at a measure-specific URL.

Data Collection Strategy (3c.)

3c.1. Describe what you have learned or modified 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 a PRO-PM, consider the implications of burden for both individuals providing the data (patients, service recipients, respondents) and those whose performance is being measured.

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.

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.

Current and Planned Use (4.1)

Select all the uses that apply from the following list. Identify if 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
- Public health/disease surveillance
- Regulatory and accreditation programs
- Professional certification or recognition program
- Quality improvement with external benchmarking to multiple organizations
- Quality improvement internal to a specific organization

4a.1. For each current use listed previously, provide the name of the program and sponsor; the purpose, geographic area, and number and percentage of accountable entities; and patients included.

4a.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?

4a.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—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.)

Progress on Improvement

4b. Progress on improvement. (Not required for initial endorsement unless available.)

Performance results on this measure (current and over time) should be provided in 1b.2. and 1b.4. Discuss or document the following:

- Progress (trends in performance results, number, and percentage of people receiving high-quality healthcare)
- 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.

Unexpected Findings

4c.1. Explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

4c.2. Explain any unexpected benefits from implementation of this measure.

Vetting of the measure by those being measured by others

This is a new sub‑criterion for use and usability in 2016. It is not a must‑pass criterion. It will be used to consider whether the measure is eligible for the "Endorsement+" designation.

4d1.1. 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.

4d1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

4d2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1. Describe how feedback was obtained.

4d2.2. Summarize the feedback obtained from those being measured.

4d2.3. Summarize the feedback obtained from other users.

4d.3. Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why

Related and Competing Measures Tab

Relation to Other NQF-Endorsed Measures (5)

If there are related measures (either same measure focus or target population) or competing measures (both the same measure focus and same target population), list the NQF number and title of all related and/or competing measures.

Harmonization (5a.)

5a.1. If this measure conceptually addresses either the same measure focus or the same target population as NQF-endorsed measure(s): Are the measure specifications completely harmonized?

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

Competing Measures (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] Tab

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

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

Other Additional Information

Ad.1. Workgroup/Expert Panel Involved in Measure Development

List the workgroup/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

7 BLANK MEASURE JUSTIFICATION FORM TEMPLATE

Project Title:

<List the project title as it should appear.>

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) has contracted with <measure developer name> to develop <measure (set) name or description>. The contract name is <insert contract name>. The contract number is <project number.>

Date:

Information included is current on <Insert Date>.

Measure Name

Type of Measure

Importance

1a—Opportunity for Improvement

1a.1. This is a Measure of

1a.2.—Linkage

1a.2.1 Rationale

1a.3.—Linkage

1a.3.1. Source of Systematic Review

1a.4.—Clinical Practice Guideline Recommendation

1a.4.1. Guideline Citation

1a.4.2. Specific Guideline

1a.4.3. Grade

1a.4.4. Grades and Associated Definitions

1a.4.5. Methodology Citation

1a.4.6. Quantity, Quality, and Consistency

1a.5.—United States Preventative Services Task Force Recommendation

1a.5.1. Recommendation Citation

1a.5.2. Specific Recommendation

1a.5.3. Grade

1a.5.4. Grades and Associated Definitions

1a.5.5. Methodology Citation

1a.6.—Other Systematic Review of the Body of Evidence

1a.6.1. Review Citation

1a.6.2. Methodology Citation

1a.7.—Findings from Systematic Review of Body of the Evidence Supporting the Measure

1a.7.1. Specifics Addressed in Evidence Review

1a.7.2. Grade

1a.7.3. Grades and Associated Definitions

1a.7.4. Time Period

1a.7.5. Number and Type of Study Designs

1a.7.6. Overall Quality of Evidence

1a.7.7. Estimates of Benefit

1a.7.8. Benefits Over Harms

1a.7.9. Provide for Each New Study

1a.8.—Other Source of Evidence

1a.8.1. Process Used

1a.8.2. Citation

1b.—Evidence to Support Measure Focus

1b.1. Rationale

1b.2. Performance Scores

1b.3. Summary of Data Indicating Opportunity

1b.4. and 1b.5. Disparities

1c.—High Priority

1c.1. Demonstrated High-Priority Aspect of Health Care

1c.3. Epidemiologic or Resource Use Data

1c.4. Citations

1c.5. PRO-PM

Scientific Acceptability

1.—Data Sample Description

1.1. What Type of Data was Used for Testing?

1.2. Identify the Specific Dataset

1.3. What are the Dates of the Data Used in Testing?

1.4. What Levels of Analysis Were Tested?

1.5. How Many and Which Measured Entities Were Included in the Testing and Analysis?

1.6. How Many and Which Patients Were Included in the Testing and Analysis?

1.7. Sample Differences, if Applicable

2a.2—Reliability Testing

2a2.1. Level of Reliability Testing

2a2.2. Method of Reliability Testing

2a2.3. Statistical Results from Reliability Testing

2a2.4. Interpretation

2b2—Validity Testing

2b2.1. Level of Validity Testing

2b2.2. Method of Validity Testing

2b2.3. Statistical Results from Validity Testing

2b2.4. Interpretation

2b3—Exclusion Analysis

2b3.1. Method of Testing Exclusion

2b3.2. Statistical Results From Testing Exclusion

2b3.3. Interpretation

2b4—Risk Adjustment or Stratification

2b4.1. Method of controlling for differences

2b4.2. Rationale why Risk Adjustment is not Needed

2b4.3. Conceptual, Clinical, and Statistical Methods

2b4.4. Statistical Results

2b4.5. Method Used to Develop the Statistical Model or Stratification Approach

2b4.6. Statistical Risk Model Discrimination Statistics (e.g., c-statistic, R^2)

2b4.7. Statistical Risk Model Calibration Statistics (e.g., Hosmer-Lemeshow statistic)

2b4.8. Statistical Risk Model Calibration—Risk decile plots or calibration curves

2b4.9. Results of Risk stratification Analysis

2b4.10. Interpretation

2b4.11. Optional Additional Testing for Risk Adjustment

2b5—Identification of statistically significant and clinically meaningful differences

2b5.1. Method for determining

2b5.2. Statistical Results

2b5.3. Interpretation

2b6—Comparability of performance scores

2b6.1. Method of testing conducted to demonstrate comparability

2b6.2. Statistical Results

2b6.3. Interpretation

Feasibility

3a.1. How are the data elements needed to compute measure scores generated

3b.1. Are the data elements needed for the measure as specified available electronically

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment

3c.1. Describe what you have learned or modified as a result of testing

3c.2. Describe any fees, licensing, or other requirements

Usability and Use

4.1—Current and Planned Use

4a.1. Program, sponsor, purpose, geographic area, accountable entities, patients

4a.2. If not publicly reported or used for accountability, reasons

4a.3. If not, provide a credible plan for implementation

4b.1. Progress on improvement

4b.2. If no improvement was demonstrated, what are the reasons

Related and Competing Measures

5—Relation to Other NQF-Endorsed Measures

5.1a. The measure titles and NQF numbers are listed here

5.1b. If the measures are not NQF-endorsed, indicate the measure title

5a—Harmonization

5a.1. Are the measure specifications completely harmonized

5a.2. If not completely harmonized, identify the differences rationale, and impact

5b—Competing measures

5b.1 Describe why this measure is superior to competing measures

Additional 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

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

Ad.1. Workgroup/Expert Panel Involved in Measure Development

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 MEASURE EVALUATION CRITERIA AND INSTRUCTIONS

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. Chapter 20—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.[205]

Additional instruction for evaluating the specific measure types—Composite Performance Measure Evaluation Guidance,[206] Cost-Resource Use Measures, eMeasures, and Patient-Reported Outcomes in Performance Measurement[207]—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 each of the following types of measures are included where they vary from or are added to the general criteria. The following type-specific criteria are labeled wherever they appear.

- Composite performance measures
- Cost and resource use measures
- eMeasures
- Patient-reported outcome-based performance measures

Evidence and Performance Gap—Importance to Measure and Report:

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.

1a. Evidence to support the measure focus

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. Because multiple processes influence a health outcome, health outcomes generally do not require empirical evidence linking them to a known

[205] https://login.qualityforum.org/login.aspx.
[206] National Quality Forum. *Composite Performance Measure Evaluation Guidance*. Washington, DC: National Quality Forum; Apr 2013; Contract No. HHSM-500-2009-00010C. Available at: http://www.qualityforum.org/Publications/2013/04/Composite_Performance_Measure_Evaluation_Guidance.aspx. Accessed on: May 31, 2015.
[207] National Quality Forum. *Patient-Reported Outcomes In Performance Management.* National Quality Forum; Dec 2012. Available at: http://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx. Accessed on: May 31, 2015.

process or structure of care.[208] For other (non-outcome) types of measures, there must be a high-to-moderate degree of certainty, as demonstrated by the evidence, that the measure focus is linked to positive outcomes.

The measure focus is evidence-based, demonstrated as follows:

- Health outcome—a rationale supports the relationship of the health outcome to processes or structures of care.
- Intermediate clinical outcome—a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence that the measured intermediate clinical outcome leads to a desired health outcome.

Note: The preferred systems for grading the evidence are the USPSTF grade definitions[209] and methods,[210] or GRADE guidelines.[211]

- 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.

Note: Clinical care processes typically include multiple steps: assess —> identify problem or potential problem —> choose, then plan intervention (with patient input) —> provide intervention —> evaluate its impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement.

- 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.
- Experience with care—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 not required for the resource use component.

Patient-reported outcome-based performance measures (PRO-PMs)

In addition to evidence required for any outcome measure, evidence should demonstrate that the target population values the measured PRO and finds it meaningful (see Table 13 under Guidance on Evaluating Patient-Reported Outcome Performance Measures).

Process Measures incorporating Appropriate Use Criteria

See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well. (see Guidance on Evaluating Evidence for Appropriate Use Measures).

---

[208] National Quality Forum. *Guidance for Evaluating Evidence Related to the Focus of Quality Measurement and Importance to Measure and Report.* Jan 2011. Available at: https://www.qualityforum.org/Measuring_Performance/Improving_NQF_Process/Evidence_Task_Force_Final_Report.aspx. Accessed on: May 31, 2015.

[209] United States Preventive Services Taskforce Office. *U.S. Preventive Services Taskforce: Grade Definitions.* Available at: http://www.uspreventiveservicestaskforce.org/uspstf/grades.htm. Accessed on: May 31, 2015.

[210] United States Preventive Services Taskforce Office. *U.S. Preventive Services Taskforce: Methods and Processes.* Available at: http://www.uspreventiveservicestaskforce.org/Page/Name/methods-and-processes. Accessed on: May 31, 2015.

[211] Grading of Recommendations Assessment, Development and Evaluation Working Group. *GRADE Working Group: Lists of GRADE Working Group Publications and Grants.* Available at: http://www.gradeworkinggroup.org/publications/index.htm. Accessed on: May 31, 2015.

Note that Measures of efficiency combine the concepts of resource use and quality (NQF's Measurement Framework: Evaluating Efficiency Across Patient-Focused Episodes of Care;[212] American Quality Alliance's Principles of Efficiency Measures).[213]

Composite performance measures

For composite measures, the evidence subcriterion (1a) must be met for each component of the composite (unless NQF-endorsed under the current evidence requirements). The evidence could be for a group of interventions included in a composite performance measure (e.g., studies in which multiple interventions are delivered to all subjects and the effect on the outcomes is attributed to the group of interventions).

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.

1b. Performance Gap

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.

The performance gap criterion (1b) 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 little opportunity for improvement, justification for why it should be included in the composite is required (e.g., increase reliability of the composite, clinical evidence).

Demonstration of quality problems and opportunity for improvement, that is, data that demonstrate:

- Considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

Note: 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 (for example, expert panel rating) and judged to be a quality problem.

Cost and resource use measures

Cost/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.

[212] National Quality Forum. *Measurement Framework: Evaluating Efficiency Across Patient-Focused Episodes of Care*. Washington, DC: Available at: http://www.qualityforum.org/Publications/2010/01/Measurement_Framework__Evaluating_Efficiency_Across_Patient-Focused_Episodes_of_Care.aspx. Accessed on: May 31, 2015.
[213] American Quality Alliance. *AQA: AQA Principles of "Efficiency" Measures*. American Quality Alliance: Revised Jun 2009. Available at: http://www.aqaalliance.org/files/PrinciplesofEfficiencyMeasurement.pdf. Accessed on: May 31, 2015.

1c. Explicit logic for composite measures

Composite performance measures

Subcriterion 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, the following must be explicitly articulated and logical:

- 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.
- The rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.
- How the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.

Reliability and Validity—Scientific Acceptability of Measure Properties:

Extent to which the measure, as specified, produces consistent (reliable) and credible (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.

Reliability

The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allow for comparability. eMeasure specifications are based on the QDM and HQMF specifications.

> Note: Measure specifications include the target population (denominator) to whom the measure applies, identification of those from the target population who achieved the specific measure focus (numerator, target condition, event, outcome), measurement time window, exclusion, risk adjustment/stratification, definitions, data source, code lists with descriptors, sampling, and scoring/computation.

Composite performance measures

Composite measure specifications include component measure specifications (unless individually endorsed); scoring rules (that is, 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 (that is, 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 the following items when evaluating the measure's reliability:

- Construction logic (Detail logic steps used to cluster, group, or assign claims beyond those associated with the measure's clinical logic.)
- Clinical logic (Detail any clustering and the assignment of codes, including the grouping methodology, the 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 End Stage Renal Disease {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 (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.)

Reliability Testing

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.

Note that 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 (for example, signal-to-noise).

eMeasures

eMeasures should be specified in the HQMF and must use the QDM and value sets vetted through the NLM's VSAC. eMeasure specifications include data type from the QDM, value sets and attributes, measure logic, original source of the data (person or system) and recorder (person).

Patient-reported outcome-based performance measures

For PROs, data collection instruments (tools) should be identified (e.g., specific PROM instrument, scale, or single item). If multiple data sources (i.e., PROMs, 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, interunit 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.

Validity

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 (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 exclusion are supported by the evidence.

> Note: Measure specifications include the target population (denominator) to whom the measure applies, identification of those from the target population who achieved the specific measure focus (numerator, target condition, event, outcome), measurement time window, exclusion, risk adjustment/stratification, definitions, data sources, code lists with descriptors, sampling, and scoring/computation.

2b2. Data elements correct

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.

> Note: 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 measure's scores indicate quality of care—for example, 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 (for example, 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.

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.

Patient-reported outcome-based performance measures

For PROs, response rates can affect validity and should be addressed in testing. Differences in individuals' PROM values related to PROM instruments or methods, modes, and languages of administration need to be analyzed and potentially included in risk adjustment.

2b3. Exclusion supported by clinical evidence

Exclusion is supported by clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion.

> Note: 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.

If patient preference (for example, 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 (for example, numerator category computed separately, denominator, exclusion category computed separately).

> Note: 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.

2b4. Risk adjustment strategy

For outcome measures and other measures when indicated (for example, resource use):

- An evidence-based, risk adjustment strategy (for example, risk models, or risk stratification)
  - is specified;
  - Is based on factors (including clinical and sociodemographic risk factors) that influence the measured outcome and are present at start of care; and
  - 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 exclusion.

In late 2014, the NQF Board of Directors approved, for a trial period, a change in the policy that prohibited the use of sociodemographic factors in statistical risk models. During the trial period, risk-adjusted measures submitted to NQF for evaluation may include both clinical and sociodemographic factors in the risk adjustment models. See section on SDS Trial Period.With adequate 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 percent versus 75 percent) is clinically meaningful; or whether a statistically significant difference of $25

in cost for an episode of care (e.g., $5,000 versus $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Composite performance measures

Applies to outcome component measures (unless NQF-endorsed).

2b5. Meaningful differences

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.

Composite performance measures

Applies to composite performance measures.

2b6. Comparable results for multiple data sources

If multiple data sources or methods are specified, there is demonstration that they produce comparable results.

Composite performance measures

Applies to component measures of the composite.

Cost and resource use

Cost and resource use measures are assessed on the following 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.

Composite performance measures

Applies to composite performance measures.

2d. Empirical support for composite measures

Composite performance measures

For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

- The component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and
- The aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible; and
- 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).

Subcriterion 2d 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 15, 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.

eMeasure-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 eMeasures:

- 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 based on the HQMF specifications.
- eMeasure specifications include data type from the QDM (value sets and attributes), code lists, EHR field, measure logic, original source of the data (person or system), and recorder (person entering the data).
- Data element: 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 retooled EHR measure specifications with scores produced by the original measure specifications demonstrated similarity within tolerable error limits.
- A crosswalk of the EHR measure specifications (QDM quality data elements, code lists, and measure logic) is needed if there is a case where a measure needs to be retooled. Please note that comparability is only an issue if maintaining two sets of specifications.

Note: 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.

Feasibility

This criterion evaluates the extent to which the 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.

eMeasures

For eMeasures, 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."

Patient-reported outcome-based performance measures

For patient-reported measures, the burdens of data collection, including those related to use of proprietary PROMs, are minimized and do not outweigh the benefit of performance measurement.

3a. Byproduct of care (clinical measures only)

For clinical measures, the required data elements are routinely generated and used during care delivery (for example, blood pressure, lab test, diagnosis, medication order).

Patient-reported outcome-based performance measures

For patient-reported measures, the burden to respondents (people providing the PROM data) should be minimized (e.g., availability and accessibility enhanced by multiple languages, methods, modes).

3b. Data elements are available in EHRs or other electronic sources

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.

3c. Data collection strategy can be implemented

Demonstration that the data collection strategy (source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (i.e., already in operational use, or testing demonstrates that the strategy is ready to put into operational use).

Note: 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.

Patient-reported outcome-based performance measures

There should be infrastructure to collect PROM data and integrate that collection into workflow and EHRs, as appropriate.

Composite performance measures

Criteria 3a, 3b, and 3c apply to composite performance measures as a whole, taking into account all component measures.

3d. eMeasure Feasibility Assessment Summary

In the Report from the NQF: eMeasure Feasibility Assessment published in April 2013, the description of the feasibility criteria is slightly adjusted for eMeasures: "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."

The report outlines four areas for evaluating the feasibility of an eMeasure in a standard scorecard:

3d.1 Data availability

The extent to which the data are readily available in a structured format across EHR systems.

3d.2 Data accuracy

The extent to which the information contained in the data is correct. This would include whether the most accurate data source is used and/or captured by the most appropriate healthcare professional or recorder.

3d.3 Data standards

The extent to which the data element is coded using a nationally accepted terminology standard (vocabulary) as recommended by the Vocabulary Task Force of the HITSC's Clinical Operations Workgroup. Standard data elements, associated definitions and code sets, and mapping to the QDM are expected.

3d.4 Workflow

The extent to which capturing the data element impacts the typical workflow for that user is evaluated here. For example, it is of little benefit to have the capability of capturing certain patient symptoms if it requires five clicks and three screens during a busy clinical encounter, for the end result will likely be missing data.

The assessment should be made based on current implementation capabilities and future (3–5) years' implementation for data elements that have a current low score for feasibility:

- Current—Rate the characteristics of data elements using the 1–3 scale for current feasibility based on the assumptions and reference indicated.
- Future—Rate the characteristics of the data element using the 1–3 scale for feasibility in 3–5 years and indicate what is required to reach the future state if necessary.

The assessment should use quantitative methods whenever possible, such as data profiling, structured interview surveys, and questionnaires from providers in a variety of settings. Measure developers should use aggregate feedback of multiple scorecards to guide further development of the eMeasure.

Usability and Use

Evaluation of a measure's usability and use involves an assessment of the extent to which intended audiences (for example, consumers, purchasers, providers, and policy makers) could use or are using performance results. The results should be usable 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

Note that 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).

Patient-reported outcome-based performance measures

Adequate demonstration of the criteria supports usability and ultimately the use of a PRO-PM 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.

4a. Accountability and Transparency

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.

Note: 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 at large (generally on a public website). At a minimum, the data on performance results about identifiable, accountable entities are available to the public (for example, in an unformatted database). The capability to verify the performance results adds substantially to transparency.

Note: This guidance is not intended to be construed as favoring measures developed by organizations that are able to implement their own measures (such as government agencies or accrediting organizations) over equally strong measures developed by organizations that may not be able to do so (such as researchers, consultants, or academics). So, measure developers may request a longer time frame with appropriate explanation and justification.

Note: 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.

Composite performance measures

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.

4b. Improvement

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.

Note: 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.

Composite performance measures

Applies to composite performance measures.

4c. Benefits

The benefits of the performance measure in facilitating progress toward achieving high-quality efficient healthcare outweigh the evidence of unintended consequences to individuals or populations (if such evidence exists).

Composite performance measures

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.

4d. Vetting of the measure by those being measured and others is demonstrated when:

- those being measured have been given performance results and 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 and implementation
- this feedback has been considered when changes are incorporated into the measure

Subcriterion 4d will be one of the criteria used to determine whether a measure is granted the "Endorsement +"

4e. Measure Deconstruction

Cost and resource use measures

For cost and resource use measures, data and result detail are maintained such that the resource use measure, including the clinical and construction logic for a defined unit of measurement, can be deconstructed to facilitate transparency and understanding.

Comparison to Related or Competing Measures

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.

5a. Related measures

The specifications for this measure are harmonized with related measures.

Note: Measure harmonization refers to the standardization of specifications for:

- Related measures with the same measure focus (for example, influenza immunization of patients in hospitals or nursing homes);
- Related measures with the same target population (for example, eye exam and HbA1c for patients with diabetes);
- Definitions applicable to many measures (for example, age designation for children) so that they are uniform or compatible, unless differences are justified (for example, 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;

OR

The differences in specifications are justified.

5b. Competing measure

The measure is superior to competing measures (for example, a more valid or efficient way to measure quality);

OR

Multiple measures are justified.

Composite performance measures

Criteria 5a and 5b apply to composite performance measures as a whole as well as the component measures.

9 BLANK MEASURE EVALUATION REPORT TEMPLATE

Project Title:

<List the project title as it should appear.>

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) has contracted with <measure developer name> to develop <measure (set) name or description>. The contract name is <insert contract name>. The contract 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 Measure Information Form and Measure Justification Form, 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.

1. Evidence, Performance Gap, and Priority (Impact)—Importance to Measure and Report

Subcriteria	Anticipated NQF Rating	Rating Improvement Plan (if Low/Moderate)
1a. Evidence to Support the Measure Focus/Measure Intent		
1b. Performance Gap		
1c. High Priority (previously referred to as High Impact)		
1d. Explicit Logic (Composite Measures only)		

Summary Rating for Importance:

Fail: At least one of the subcriteria above is not rated as high.

Pass: Measure is important; all of 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 in order 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

Subcriteria	Anticipated NQF Rating	Rating Improvement Plan (if Low/Moderate)
2a. Reliability		
2a1. Precisely Specified		
2a2. Reliability Testing		
Cost and Resource Use Measure-Specific **Subcriteria**		
2a3. Construction Logic		
2a4. Clinical Logic		
2a5. Adjustments for Comparability—Inclusion/Exclusion Criteria		
2a6. Adjustments for Comparability—Risk Adjustment		
2a7. Adjustments for Comparability—Costing Method		
2a8. Adjustment for Comparability—Scoring		
2b. Validity		
2b1. Specifications		
2b2. Validity Testing		
2b3. Exclusion		
2b4. Risk Adjustment		
2b5. Meaningful		
2b6. Comparable Results		
2b7. Missing Data (eMeasures, Composite Measures, and PRO-PMs)		
Cost and **Resource Use Measure-Specific Subcriteria**		
2b8. Construction Logic		
2b9. Clinical Logic		
2b10. Adjustments for Comparability—Inclusion/Exclusion Criteria		
2b11. Adjustments for Comparability—Risk Adjustment		
2b12. Adjustments for Comparability—Costing Method		
2b13. Adjustment for Comparability—Scoring		
2b14. Significant Differences in Performance		
2b15. Comparability of Multiple Data Sources		
2c. Disparities		
2c1. Disparities Identified		
2d. Empirical Analysis **(Composite Measures Only)**		
2d1. Quality Construct		
2d2. Aggregation and Weighting Rules		
2e. eMeasure-**specific Subcriteria**		
2e1. HQMF Specifications		
2e2. Data Types		
2e3. Data Element Validity		
2e4. Comparability Analysis		

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 in order 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

Subcriteria	Anticipated NQF Rating	Rating Improvement Plan (if Low/Moderate)
3a. Data are a Byproduct of Care		
3b. Electronic Sources		
3c. Data Collection Strategy		
3d. eMeasure Feasibility Assessment Summary		For data elements that score low on current feasibility, indicate the anticipated feasibility score in 3–5 years based on a projection of the maturation of the EHR, or maturation of its use.
3d1. Data Availability		
3d2. Data Accuracy		
3d3. Data Standards		
3d4. Workflow		

Summary Rating for Feasibility/eMeasure 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

Subcriteria	Anticipated NQF Rating	Rating Improvement Plan (if Low/Moderate)
4a. Accountability and Transparency		
4b. Improvement		
4c. Benefits		
4d. Measure Deconstruction (Cost and Resource Use Measure Only)		

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

Subcriteria	Anticipated NQF Rating	Rating Improvement Plan (if Low/Moderate)
5a. Related Measure		
5b. Competing Measure		

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.

	High	Medium	Low	Insufficient
1. Importance to Measure and Report				
2a. Overall Reliability				
2b. Overall Validity				
2c. Disparities of Care				
3. Feasibility				
4. Usability and Use				
5. Comparison to Related or Competing Measures				

Recommendation:

Explanation:

10 CMS MEASURES MANAGEMENT SYSTEM WEBPAGE POSTING INSTRUCTIONS

Posting Submissions

- Send submissions to MMS Support inbox (MMSSupport@battelle.org) for posts on:
  - Public Comment WebPage
  - Technical Expert Panel (TEP) WebPage

- Posting Timeframe:
  - Please allow **at least 5 business days** for processing your post. Posts *may* be posted prior to this timeframe, if your post needs to be published on a specific date, please note this in your email and we will work with CMS to accomplish by this date/time.

- Posting Format:
  - **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 **508-compliant** and submitted in **PDF format.** Please 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 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)

11 PUBLIC COMMENT CALL WEB POSTING

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 *<measure developer name>* to develop *<measure (set) name or description>*. The contract name is <insert contract name>. The contract number is <project number.> As part of its measure development process, CMS 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:

The following documents are provided for your review and comment. The files are found below in the Download section.

<Name the measures and documents for comment.>

Expiration Notice:

This notice expires on < list the date>.

Project Specific Instructions:

<Project specific instructions>

Send your comments to <insert email address>.

12 PUBLIC COMMENT SUMMARY WEB POSTING

Project Title: *<List the project title as it should appear on the Web posting.>*

Dates:

The Call for Public Comment period closed on < l**ist the date**>.

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) has contracted with *<measure developer name>* to develop *<measure (set) name or description>*. The contract name is <insert contract name>. The contract number is <project number.> As part of its measure development process, CMS 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:

The following documents including a summary of public comments and the original measures are found below in the Download section.

<List the document names.>

Expiration Notice:

This notice expires on < list the date>.

13 PUBLIC COMMENT SUMMARY REPORT TEMPLATE

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 *<measure developer name>* to develop *<measure (set) name or description>*. The contract name is <insert contract name>. The contract number is <project number.> As part of its measure development process, CMS 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 general public of comment period>.

<Volume> responses were received on this topic.

Stakeholder Comments—General and Measure-Specific

<Summary of general comments>

<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.>

Date Posted	Measure Set or Measure	Text of Comments	Name, Credentials, and Organization of Commenter	Email Address*	Type of Organization*	Response*

*Optional

14 TECHNICAL EXPERT PANEL (TEP) CALL FOR TEP WEB POSTING

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**.> Please submit all nomination materials by the closing date.

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) has contracted with *<measure developer name>* to develop *<measure (set) name or description>*. The contract name is <insert contract name>. The contract number is <project number>. As part of its measure development process, CMS asks measure developers to convene 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 the following perspectives and areas of expertise:

Subject matter expertise: *<insert specific topic>*

Consumer/patient/family (caregiver) perspective

Healthcare disparities

Performance measurement

Quality improvement

Purchaser perspective

Potential TEP members must be aware that participation on the TEP is voluntary. As such, individuals wishing to participate on the TEP should understand that their input will be recorded in the meeting minutes. Proceedings of the TEP will be summarized in a report that is disclosed to the general public. If a participant has disclosed private, personal data by his or her own choice, then that material and those communications are not deemed to be covered by patient-provider confidentiality. If patient participants (only) wish to keep their names confidential, that request can be accommodated. Any questions about confidentiality will be answered by the TEP organizers.

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 (conference call, webinar, in person, etc.)>

Expiration Notice:

This notice expires on <list the date>.

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 two pages) highlighting experience/knowledge relevant to the expertise described above and involvement in measure development.

Curriculum vitae or a summary of relevant experience for a maximum of 10 pages.

Patient participants are not required to submit a curriculum vitae and may elect to keep their names confidential in public documents.

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, please complete the form and email it to: <measure developer contact information>

15 TECHNICAL EXPERT PANEL (TEP) NOMINATION FORM TEMPLATE

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 *<measure developer name>* to develop *<measure (set) name or description>*. The contract name is <insert contract name>. The contract number is <project number.> As part of its measure development process, CMS asks measure developers to convene 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 the following areas of expertise and perspectives:

Subject matter expertise: <insert specific topic>

Consumer/patient/family (caregiver) perspective

Performance measurement

Quality improvement

Purchaser perspective

Healthcare disparities

Instructions:

Applicants/nominees must submit the following documents ***with this completed and signed form***:

A letter of interest (not to exceed two pages) highlighting experience/knowledge relevant to the expertise described above and involvement in measure development.

Curriculum vitae or a summary of relevant experience (including publications) for a maximum of 10 pages. (Patient participants may elect to keep their names confidential in public documents.)

Please 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 must be aware that participation on the TEP is voluntary. As such, individuals wishing to participate on the TEP should understand that their input will be recorded in the meeting minutes. Proceedings of the TEP will be summarized in a report that is disclosed to the general public. If

a participant has disclosed private, personal data by his or her own choice, then that material and those communications are not deemed to be covered by patient-provider confidentiality. If potential patient participants wish to keep their names confidential, that request can be accommodated. Any questions about confidentiality will be answered by *<measure developer name>*.

*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:

Professional Role:

Organizational Affiliation:

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 will request the required information from the nominee.

Name:

Credentials:

Professional Role:

Organizational Affiliation:

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: _____

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's 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, please describe (grant/research support, consultant, speaker's bureau, and major stock shareholder, other financial or material support). Please 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, please 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.

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 that my participation on the TEP is voluntary. As such, I understand that my input will be recorded in the meeting minutes. Proceedings of the TEP will be summarized in a report that is disclosed to the general public. If I have disclosed private, personal data by my own choice, then that material and those communications are not deemed to be subject to any confidentiality laws.

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

16 TECHNICAL EXPERT PANEL (TEP) COMPOSITION (MEMBERSHIP) LIST TEMPLATE

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 *<measure developer name>* to develop *<measure (set) name or description>*. The contract name is <insert contract name>. The contract number is <project number.> As part of its measure development process, CMS asks measure developers to convene groups of stakeholders and experts who contribute direction and thoughtful input to the measure developer during measure development and maintenance. The following individuals were selected and have agreed to serve as the Technical Expert Panel for this project.

Name, Credentials, and Professional Role	Organizational Affiliation, City, State	Consumer Perspective	Clinical Content	Performance Measurement	Coding and Informatics	Conflict of Interest Disclosure

*Note: the table above cannot be changed. All columns must be completed.

17 TECHNICAL EXPERT PANEL (TEP) COMPOSITION (MEMBERSHIP) LIST WEB POSTING

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 *<measure developer name>* to develop *<measure (set) name or description>*. The contract name is <insert contract name>. The contract number is <project number.> As part of its measure development process, CMS asks measure developers to convene 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 the following perspectives and areas of expertise:

Subject matter expertise: *<insert specific topic>*

Consumer/patient/family (caregiver) perspective

Healthcare disparities

Performance measurement

Quality improvement

Purchaser perspective

TEP Expected Time Commitment:

<anticipated meeting dates and time frame for measure development activities >

<types of meetings (conference call, webinar, in person, etc.)>

Expiration Notice:

This notice expires on < list the date>.

18 TECHNICAL EXPERT PANEL (TEP) CHARTER TEMPLATE

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 *<measure developer name>* to develop *<measure (set) name or description>*. The contract name is <insert contract name>. The contract number is <project number>. As part of its measure development process, CMS asks measure developers to convene 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 or 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 Technical Expert Panel Membership List.>

19 TECHNICAL EXPERT PANEL (TEP) SUMMARY WEB POSTING

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 *<measure developer name>* to develop *<measure (set) name or description>*. The contract name is <insert contract name>. The contract number is <project number>. As part of its measure development process, CMS asks measure developers to convene 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>

Expiration Notice:

This notice expires on < list the date>.

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.

Adapted measures

An existing measure is changed to fit the current purpose or use. This may mean changes to the numerator or denominator, or changing a measure to meet the needs of a different care setting, data source, or population. Or, it may mean adding specifications to fit the current use.

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 as "encouraging the use of similar standardized performance measures across and within public and private sector efforts."[214] 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.

Attribution

Assignment of the results of a measure to an individual, group, or organization responsible for the decisions, costs, and outcomes.[215]

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 (for example, 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.

[214] National Quality Forum. *Changes to NQF's Harmonization and Competing Measures Process: Information for Measure Developers*. National Quality Forum; Washington, DC, 2013. Available at: https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=72716. Accessed on: March 14, 2016.
[215] Krumholz, H., et al. Standard for measures: efficiency in healthcare. *Journal of the American College of Cardiology*. 2008;52 (number 18):1522.

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.[216] 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 Quality Language (CQL)

CQL is an HL7 draft standard for trial use (DSTU). It is part of the effort to harmonize standards between eCQMs and clinical decision support (CDS). CQL provides the ability to express logic that is human readable yet structured enough for processing a query electronically. In the future, CQL is to be used in all of the CQM HQMF electronic specifications. It will replace the logic expressions currently defined in the QDM and QDM (v5.0) will include only the method for defining the data elements (the data model). For the most current information see the Clinical Quality Language page on the eCQI Resource Center.

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.

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.

Common use case

A use case describes the flow of quality information through an EHR system for the purpose of quality measurement, feedback and reporting.[217]

Comparable data

The accuracy, reproducibility, risk-adjustability, and validity of the measure should not be affected by differences in information technology 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.

[216] Leatherman, S., Berwick, D., et al. The business case for quality: case studies and an analysis. *Health Affairs.* 2003; 22, (number 2):18. Available at: http://content.healthaffairs.org/content/22/2/17.full.pdf. Accessed on: March 14, 2016.
[217] HHS (2007, May 6).Quality Draft Detailed Use Case. Office of the National Coordinator for Health Information Technology. Available from http://www.health.state.mn.us/e-health/standards/qualityusecase.pdf.

Competing measures

Competing measures address the same topic and the same population. This term is used when considering harmonization. See also "related measures."

Composite performance measure

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.[218] Also called composite measures.

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

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 (for example, 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.

[218] National Quality Forum. *Composite Performance Measure Evaluation Guidance*. Washington, DC: National Quality Forum; Apr 2013; Contract No. HHSM-500-2009-00010C. Available at: http://www.qualityforum.org/Publications/2013/04/Composite_Performance_Measure_Evaluation_Guidance.aspx Accessed on: March 14, 2016.

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 (for example, diagnosis, active).

Data sources

The primary source document(s) used for data collection (for example, billing or administrative data, encounter form, enrollment forms, medical record).

Denominator

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. Continuous variable 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. Denominator exception allows for adjustment of the calculated score for those providers with higher risk populations. 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. Denominator exception is used only in proportion measures.

These cases are removed from the denominator; however the number of patients with valid exception 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 if numerator criteria are met. Denominator exclusion 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.)

Denominator statement

A statement that describes the population evaluated by the performance measure.

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 Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement – Requirements for Endorsing eCQMs.

Direct referenced code

A specific code that is referenced directly in the eCQM logic to describe a data element or one of its attributes. Direct referenced code 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.[219]

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. The AHRQ has developed a typology or analytical framework of efficiency measures: perspective, outputs, and inputs.[220]

Electronic clinical quality measure (eCQM)

An eCQM is a clinical quality measure that is expressed and formatted to use data from EHRs and/or HIT systems to measure healthcare quality, ideally data captured in structured form during the process of patient care. So they can be reported from an EHR, the HQMF is used to format the eCQM content using the QDM to express the logic and the data elements needed to evaluate a provider or organization's performance. These electronic specifications include:

[219] Truman B., Smith CK, Roy K, et al. *CDC Health Disparities and Inequalities Report – United States, 2011*. CDC, Morbidity and Mortality Weekly Report; United States Government Printing Office; Jan 14, 2011; Report No. 2011-723-011/21019 . Available at: http://www.cdc.gov/mmwr/pdf/other/su6001.pdf. Accessed on: March 14, 2016.

[220] Department of Health and Human Services, Agency for Healthcare Research and Quality. *Health care efficiency measures: identification, categorization, and evaluation. Final report*. Rockville, MD: Agency for Healthcare Research and Quality; Apr 2008; Report No. 08-0030. Available at: http://archive.ahrq.gov/research/findings/final-reports/efficiency/ . Accessed on: March 14, 2016.

- **HTML** - A web-facing, human readable rendition 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** – Specific code systems to capture clinical concepts and patient data in the EHR system. Value sets provide definitions of the codes necessary to calculate the eCQM. The value sets for each measure are stored by The NLM VSAC. Through the VSAC, providers, implementers, and developers can access the value sets for each eCQM.

Electronic Health Record (EHR)

This is also known as the electronic patient record, electronic medical record, or computerized patient record. As defined by Healthcare Information Management and Systems Society, "the electronic health record (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."[221]

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 American Society for Testing and Materials 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. (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 different 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

See denominator exception.

Exclusion

See denominator exclusion and numerator exclusion.

[221] Healthcare Information and Management and Systems Society. *Resource Library: Electronic Health Records.* Available at: http://www.himss.org/library/ehr/?navItemNumber=13261 . Accessed on: March 14, 2016.

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 (for example, whether the proportion of patients with BP < 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.

Guidelines

Clinical practice guidelines are systematically developed statements to support practitioner and patient decisions about appropriate healthcare for specific clinical circumstances.

Harmonization

The standardization of specifications for related measures with the same measure focus (for example, influenza immunization of patients in hospitals or nursing homes); related measures for the same target population (for example, eye exam and HbA1c for patients with diabetes); or definitions applicable to many measures (for example, age designation for children) so that they are uniform or compatible, unless differences are justified (in other words, 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 (HIT)

Provides the umbrella framework to describe the comprehensive management of health information across computerized systems and its secure exchange between consumers, providers, government and quality entities, and insurers.[222]

Health Information Technology for Economic and Clinical Health (HITECH) Act

A provision within the American Recovery and Reinvestment Act which authorizes incentive payments through Medicare and Medicaid to hospitals and clinicians toward meaningful use of EHRs. See Meaningful Use.

Health Level 7 (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.

HQMF

A standards-based representation of quality measures as electronic documents. A quality measure expressed in this way is also referred to as an eCQM.

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 patient-reported outcomes, 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 (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 (for example, 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 longer-term health outcome. There should be a body of evidence that the measured intermediate clinical outcome leads to a desired health outcome.

[222] Wikipedia. *Health information technology*. Available at: http://en.wikipedia.org/wiki/Health_information_technology. Accessed on: May 31, 2015.

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 also pertains 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 Healthcare-associated Infections (HAIs) as a standardized infection ratio (SIR). The SIR compares the actual number of HAIs (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 a 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 JIRA to track issues with quality measures. This system is commonly referred to as JIRA.[223]

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.[224]

Logic

The criteria used to define a quality measure and its key components.[225]

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.

[223] http://en.wikipedia.org/wiki/JIRA. Accessed on: May 31, 2015.
[224] http://en.wikipedia.org/wiki/Kaizen . Accessed on: May 31, 2015.
[225] http://public.qualityforum.org/hitknowledgebase/Pages/eMeasures.aspx . Accessed on: May 6, 2015.

Meaningful Use

A provision within the American Recovery and Reinvestment Act which authorizes the CMS to provide a reimbursement incentive for physician and hospital providers who are successful in becoming "meaningful users" of an EHR. These incentive payments began in 2011 and have gradually phased out. Starting in 2015, providers are expected to have adopted and be actively using an EHR in compliance with the "meaningful use" definition or they will be subject to financial penalties under Medicare.

Measure Applications Partnership (MAP)

A 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 four workgroups 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.[226]

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 (for example, proportion of patients who died, average lab value attained) for the entity being measured (hospital, health plan, home health agency, clinician, etc.). The measure specifications designate the entity that is being measured and to whom the measure applies.

Measure steward

Also called measure owner, this is 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.[227]

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.

[226] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Measure Authoring Tool.* Available at: https://www.emeasuretool.cms.gov. Accessed on: May 31, 2015.
[227] National Quality Forum. *Phrasebook: a Plain Language Guide to NQF Jargon.* Washington, DC: National Quality Forum; 2013. Available at: http://public.qualityforum.org/NQFDocuments/Phrasebook.pdf. Accessed on: May 31, 2015.

Measure (performance measure or quality measure)

A mechanism to assign a quantity to an attribute by comparison to a criterion.[228] A measure may stand alone or belong to a composite, subset, set, and/or collection of measures. 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.[229]

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 [Affordable Care] Act.[230]

Medical record (data source)

Data obtained from the records or documentation maintained on a patient in any healthcare setting (for example, 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 diseased 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)

Patients 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.[231]

---

[228] Department of Health and Human Services, Agency for Healthcare Research and Quality. *National Quality Measures Clearinghouse: Glossary*. Available at: http://www.qualitymeasures.ahrq.gov/about/glossary.aspx. Accessed on: May 31, 2015.
[229] National Quality Forum. *Phrasebook: a Plain Language Guide to NQF Jargon.* Washington, DC: National Quality Forum; 2013. Available at: http://public.qualityforum.org/NQFDocuments/Phrasebook.pdf . Accessed on: May 31, 2015.
[230] Section 3014 of the Affordable Care Act (ACA) (P.L. 111-148) created the new Section 1890A(a)(2) of the Social Security Act.
[231] National Quality Forum. *Multiple chronic conditions measurement framework. Final report*. Washington, DC: 2012 Available at: http://www.qualityforum.org/Publications/2012/05/MCC_Measurement_Framework_Final_Report.aspx. Accessed on: May 31, 2015.

National Priorities Partnership (NPP)

A multi-stakeholder group organization convened by the NQF that offers consultative support to the HHS on setting national priorities and goals for HHS's NQS.

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 exclusion

Defines instances that should not be included in the numerator data. Numerator exclusions are used only in ratio and proportion measures.

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.

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.[232]

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 general public. Measure developers should be familiar with the PRA before implementing any process that involves the collection of new data. [233]

Patient-reported Outcome (PRO)

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. This definition reflects the key domains[234] of:

- Health-related quality of life (including functional status).
- Symptoms and symptom burden (e.g. pain, fatigue).

[232] National Quality Form. *Measure Evaluation Criteria and Guidance Summary Tables*. Available at http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx. Accessed on: May 31, 2015.
[233] 104th Congress of the United States. *Paperwork Reduction Act of 1995.* United States Government Printing Office. Available at: http://www.gpo.gov/fdsys/pkg/PLAW-104publ13/pdf/PLAW-104publ13.pdf. Accessed on: July 7, 2015.
[234] Note that CMS and other HHS agencies define and use the term "domain" differently from one another. Therefore, within the Blueprint you will see the term "domain" defined differently in different contexts, depending on the relevant agency within the discussion.

- Experience with care.
- Health behaviors (e.g., smoking, diet, exercise).[235]

Patient-reported Outcome Measure (PROM)

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.[236]

Patient-reported outcome-based performance measure (PRO-PM)

A performance measure that is based on PROM data aggregated for an accountable healthcare entity.[237]

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 (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 (for example, 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.[238]

---

[235] National Quality Forum. *Patient-reported outcomes (PROs) in performance measurement*. National Quality Forum; Jan 2013; Contract No. HHSM-500-2009-00010C. Available at: https://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx. Accessed: May 31, 2015.
[236] Ibid.
[237] Ibid.
[238] *Merriam-Webster's Online Dictionary, Public Domain.* Available at: http://www.merriam-webster.com/dictionary/public+domain. Accessed on: May 31, 2015.

Quality Data Model (QDM)

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.[239]

Quality measure (or performance measure)

Numeric quantification of healthcare quality for a designated accountable healthcare entity, such as hospital, health plan, nursing home, clinician, etc. 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 HIT systems.
- Can be used to exchange eCQM data between systems.
- Are 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.[240]

R^2 statistic

Values for R^2 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 (for example, the number of patients with central lines who develop infection divided by the number of central line days).

Receiver-operating characteristic (ROC) curve

The graph that provides the c-statistic value is the ROC curve. The ROC curve graphs the predictive accuracy of a logistic regression model.

Related measures

Related measures address either the same topic or the same population. This term is used when considering harmonization. See also competing measures.

[239] Quality Data Model. Available at: https://ecqi.healthit.gov/qdm. Accessed on May 24, 2016.
[240] QRDA. Available at: https://ecqi.healthit.gov/qrda. Accessed on May 24, 2016

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 (signal versus noise).

Respecified electronic clinical quality measure (eCQM)

Respecified measures are eCQMs developed from current claims/registry/paper measures that are placed into HQMF format. Hence, a respecified measure will often include different data elements than the original CQM created based on the same evidence.

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—for example, 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 (denominator) and how they are selected.

Scientific acceptability of the measure properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented.

Score (see 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 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, etc.). The measure specifications designate the entity that is being measured and to whom the measure score applies.[241]

Scoring

Method(s) applied to data to generate results/score. Most quality measures produce rates; however, other scoring methods include categorical value, continuous variable, count, frequency distribution, non-weighted score/composite/scale, ratio, and weighted score/composite/scales.

Semantic validation

A 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, sensitivity refers to the proportion of actual positives that are correctly identified as such (for example, the percentage of people with diabetes who are correctly identified as having diabetes). See Specificity.

Specifications

Measure instructions that address the following: 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, specificity refers to the proportion of negatives that are correctly identified (for example, the percentage of healthy people who are correctly identified as not having the condition). Perfect specificity would mean that the measure recognizes all actual negatives (for example, all healthy people will be recognized as healthy). See 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 measure

A structural measure is one that assesses features of a healthcare organization or clinician relevant to its capacity to provide healthcare.

[241] National Quality Forum. *Glossary of Terms.* Available at: *https://www.qualityforum.org/Measuring_Performance/Submitting_Standards/NQF_Glossary.aspx*. Accessed on: June 25, 2015.

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. (NQF Glossary)

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 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 (cases) and denominator (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

A time frame used to determine cases for inclusion in the denominator, numerator, or exclusion. The time frame 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 criteria

Extent to which intended audiences (for example, consumers, purchasers, providers, policy makers) can understand the measure's results and find them useful for quality improvement and decision making. Usability criteria ask if the measure is strong enough to be used for various types of measurement programs, including public reporting, whether it leads to actual improvement for patients, and whether the benefits of the measure outweigh any potential harms.[242]

Validation

Testing to determine if the measure accurately represents the concept being evaluated and achieves the purpose for which it is intended (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 (to measure quality). For example, the measure:

[242] National Quality Forum. *Phrasebook: a plain language guide to NQF jargon.* Washington, DC: National Quality Forum; 2013. Available at: http://public.qualityforum.org/NQFDocuments/Phrasebook.pdf. Accessed on: May 31, 2015.

- Clearly identifies the concept being evaluated (face validity).
- Includes all necessary data elements, codes, and tables to detect a positive occurrence when one exists (construct validity).
- Includes all necessary data sources to detect a positive occurrence when one exists (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

A value set is a 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)

The VSAC is a 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 Data Element Catalog.[243] 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.

[243] Value Set Authority Center (VSAC). Available at: https://ecqi.healthit.gov/ecqm-tools/tool-library/value-set-authority-center-vsac. Accessed on May 24, 2016.

ACRONYMS

AARP	American Association of Retired Persons
ACA	The Patient Protection and Affordable Care Act of 2010
ACE	Angiotensin Converting Enzyme
AHRQ	Agency for Healthcare Research and Quality
AMA	American Medical Association
AMI	Acute Myocardial Infarction
ANOVA	Analysis of Variance
APM	Alternative Payment Model
ARRA	American Recovery and Reinvestment Act of 2009
ASPE	HHS Office of the Assistant Secretary for Planning and Evaluation
AUC	Area Under the ROC Curve
CABG	Coronary Artery Bypass Graft
CAD	Coronary Artery Disease
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CAUTI	Catheter-associated Urinary Tract Infections
C-CDA	Consolidated Clinical Document Architecture
CDA	Clinical Document Architecture
CDC	Centers for Disease Control and Prevention
CDS	Clinical Decision Support
CEHRT	Certified EHR Technology
CHIP	Children’s Health Insurance Program
CLABSI	Central-line Associated Blood Stream Infection
CLD	Content Logical Definition
CMS	Centers for Medicare & Medicaid Services
COPD	Chronic Obstructive Pulmonary Disease
COR	Contracting Officer’s Representative
CPOE	Computerized Physician Order Entry
CPT	Current Procedural Terminology
CPT4	Current Procedural Terminology, Fourth Edition

CPT I	CPT Category I
CPT II	CPT Category II
CQL	Clinical Quality Language
CQM	Clinical Quality Measure
CQS	Composite Quality Score
CSAC	NQF Consensus Standards Approval Committee
CT	Computerized Tomography
CV	Continuous Variable
CVC	Central Venous Catheter
DENEX	Denominator Exclusion
DENEXEP	Denominator Exception
DENOM	Denominator
DOD	Department of Defense
DSTU	Draft Standard for Trial Use
ECG	Electrocardiogram
eCQI	Electronic Clinical Quality Improvement
eCQM	Electronic Clinical Quality Measures
ED	Emergency Department
EH	Eligible Hospital
EHR	Electronic Health Record
EP	Eligible Professional
ESRD	End Stage Renal Disease
FACA	Federal Advisory Committee Act
FFS	Fee for Service
FOTO	Focus On Therapeutic Outcomes, Inc.
GPCK	Generic Pack
GRADE	Grading of Recommendation, Assessment, Development, and Evaluation
HAC	Hospital-acquired Condition
HAI	Healthcare-associated Infection
HCPCS	Healthcare Common Procedure Coding System

HHS	U.S. Department of Health and Human Services
HIPAA	Health Insurance Portability and Accountability Act
HIT	Health Information Technology
HITECH	Health Information Technology for Economic and Clinical Health
HITEP	Health Information Technology Expert Panel
HITSC	Health Information Technology Standards Committee
HL7	Health Level Seven
HOS	Health Outcomes Survey
HQCIS	Healthcare Quality Information System
HQI	Hospital Quality Incentive
HQMF	Health Quality Measure Format
HTML	Hyper Text Markup Language
HTN	Hypertension
ICC	Intraclass Correlation Coefficient
ICD	International Classification of Diseases
ICD-9-CM	International Classification of Diseases, 9th Revision, Clinical Modification
ICD-10-CM	International Classification of Diseases, 10th Revision, Clinical Modification
ICD-10-PCS	International Classification of Diseases, 10th Revision, Procedure Coding System
ICF	International Classification of Functioning, Disability, and Health
ICU	Intensive Care Unit
IHI	Institute for Healthcare Improvement
IOM	Institute of Medicine
IPOP	Initial Population
IRB	Institutional Review Board
IRF	Inpatient Rehabilitation Facility
IT	Information Technology
LA	LOINC Answers
LBBB	Left Bundle Branch Block
LDL	Low-density Lipoprotein
LOINC	Logical Observation Identifiers Names and Codes

LTC	Long-term Care
MacPAC	Medicaid and CHIP Payment and Access Commission
MACRA	Medicare Access and CHIP Reauthorization Act
MAP	Measure Applications Partnership
MAT	Measure Authoring Tool
MCC	Multiple Chronic Conditions
MCCB	MAT Change Control Board
MDP	Measurement Development Plan
MDS	Minimum Data Set
MedPAC	Medicare Payment Advisory Committee
MER	Measure Evaluation Report
MIDS	Measure and Instrument Development and Support
MIF	Measure Information Form
MIPS	Merit-based Incentive Payment System
MJF	Measure Justification Form
MMS	Measures Management System
MSRPOPL	Measure Population
MSRPOPLEX	Measure Population Exclusion
MUC	Measures Under Consideration
NCHS	CDC National Center for Health Statistics
NIH	National Institutes of Health
NLM	National Library of Medicine
NPP	National Priorities Partnership
NQF	National Quality Forum
NQMC	National Quality Measures Clearinghouse
NQS	National Quality Strategy
NUMER	Numerator
NUMEX	Numerator Exclusion
OASIS	Outcome and Assessment Information Set
OID	Object Identifier

OMB	Office of Management and Budget
ONC	Office of the National Coordinator for Health Information Technology
PACU	Post-anesthesia Care Unit
PCORI	Patient-Centered Outcomes Research Institute
PFAC	Patient Family Advisory Council
PHDSC	Public Health Data Standards Consortium
PHIN VADS	CDC Public Health Information Network Vocabulary Access and Distribution System
PPACA	Patient Protection and Affordable Care Act
PRA	Paperwork Reduction Act
PRO	Patient-reported Outcome
PROM	Patient-reported Outcome Measure
PROMIS	Patient-Reported Outcomes Measurement Information System
PRO-PM	Patient-Reported Outcome-based Performance Measurement
PSI	Patient Safety Indicator
QA	Quality Assurance
QDM	Quality Data Model
QIN-QIO	Quality Innovation Network-Quality Improvement Organizations
QRDA	Quality Reporting Document Architecture
RAI	Resident Assessment Instrument
RCT	Randomized Controlled Trial
ROC	Receiver Operating Characteristic
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
SNOMED CT	Systematized Nomenclature of Medicine—Clinical Terms
SR	Systematic Review

SSI	Surgical Site Infection
STS	Society of Thoracic Surgeons
STU	Standard for Trial Use
TEP	Technical Expert Panel
UCUM	Unified Code for Units of Measure
UMLS	Unified Medical Language System
URL	Uniform Resource Locator
US	United States
USCRS	US SNOMED CT Content Request System
USOW	Umbrella Statement of Work
USPSTF	United States Preventive Services Task Force
VHA	Veterans Health Administration
VSAC	Value Set Authority Center
VSD	Value Set Definition
XML	eXtensible Markup Language

Section 6. Appendices

APPENDIX A: ECQM LOGIC QUALITY ASSURANCE CHECKLIST

This checklist can be used to review the logic used in eCQMs.

Table 19. Logic Review Checklist

#	Mandatory?	Reviewed?	Passed?	Item	Comment
L-1	Y			Is the intent of the measure described in the measure description articulated/captured in the measure logic?	
L-2	Y			Do the logic elements map to definitions in the measure narrative, data dictionary, or supporting reference documentation?	
L-3	Y			Do the populations in the narrative align with the populations defined in the logic?	
L-4	Y			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 applying recurring logic patterns and Boolean logic rules.	
L-5	Y			Are nested Boolean structures expressed correctly? (For example, no extraneous OR and AND statements.)	
L-6	Y			Are all QDM elements time-bound (either directly or indirectly)?	
L-7	Y			Do the 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.)	
L-8	Y			Are operator precedence rules followed as specified in the eCQM Logic and Implementation Guidance?	
L-9	Y			Have subset operators been applied correctly as defined in the eCQM Logic and Implementation Guidance?	
L-10	Y			Are time intervals represented in similar units (e.g., hours)?	
L-11	Y			Are annotations included in the logic for sections of the measure that have been updated/changed?	
L-12	Y			Does the measure demonstrate at least 80% coverage of test patients with at least one positive and one negative patient for each population?	
L-13	Y			Has the measure author provided the Bonnie account (email address) where the measure was tested?	
L-14	Y			Are there value sets present in the measure that cause the logic to evaluate incorrectly, or in an unexpected way?	
L-15	N			Additional comments / issues / suggestions?	

APPENDIX B: ECQM METADATA

Header Data Elements	Definition	Measure Developer Guidance	Preferred Term (Required, None, Not Applicable)
eMeasure Title[244]	The title of the quality eCQM.		Required
eMeasure Identifier (MAT)	Specifies the eCQM identifier generated by the MAT.	Field is auto-populated by the MAT.	Required
eMeasure Version Number	A value used to indicate the version of the eCQM.	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)	Required
NQF Number	Specifies the NQF number.	"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.	Not Applicable
GUID	Represents the globally unique measure identifier for a particular eCQM.	Field is auto-populated by the MAT.	Required
Measurement Period	The time period for which the eCQM applies.	MM/DD/20xx—MM/DD/20xx	Required
Measure Steward	The organization responsible for the continued maintenance of the eCQM.	CMS is the measure steward for measures developed under CMS contracts.	Required
Measure Developer	The organization that developed the eCQM.		Required
Endorsed By	The organization that has endorsed the eCQM through a consensus-based process.	All endorsing organizations should be included (not specific to just NQF).	None
Description	A general description of the eCQM intent.	A brief narrative description of the eCQM, such as "Ischemic stroke patients with atrial fibrillation/flutter who are prescribed anticoagulation therapy at hospital discharge."	Required

[244] At the time of publication, the MAT uses the term "eMeasure". The MAT will use eCQM in a future version.

Header Data Elements	Definition	Measure Developer Guidance	Preferred Term (Required, None, Not Applicable)
Copyright	Identifies the organization(s) who own the intellectual property represented by the eCQM.	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 American Medical Association. All Rights Reserved").	None
Disclaimer	Disclaimer information for the eCQM.	This should be brief.	None
Measure Scoring	Indicates how the calculation is performed for the eCQM. (e.g., proportion, CV, ratio).		Required
Measure Type	Indicates whether the eCQM is used to examine a process or an outcome over time.		Required
Measure Item Count	An item count is a means to establish what the measure is counting. For example, if a measure is interested in an episode of care, using item count allows the measure developer to establish what episode(s) of care the measure is interested in.	An optional field 'Items Counted' is added to the Measure Details page. All QDM elements applied to the selected measure display in the alphabetical Items Counted list. Scroll through the list of QDM elements and select the desired QDM element to be used for 'Item Count' by selecting the checkbox positioned to the left of the QDM element name. The number of QDM elements selected for the item count is populated to the right of the 'Items Counted' list box.	

Header Data Elements	Definition	Measure Developer Guidance	Preferred Term (Required, None, Not Applicable)
Stratification	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 2 sub-strata: those with a primary mental health diagnosis, and those with a primary diagnosis of sexually transmitted infection); and (3) Evaluate the eCQM based on different locations within a facility. (e.g., evaluate the overall rate for all intensive care units. Some strata may include additional findings such as specific birth weights for neonatal intensive care units).		None
Risk Adjustment	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.	Provide a brief description with instructions where the complete risk adjustment methodology may be obtained.	None

Header Data Elements	Definition	Measure Developer Guidance	Preferred Term (Required, None, Not Applicable)
Rate Aggregation	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).	For eCQMs that do not have rate aggregation, enter 'None'.	None
Rationale	Succinct statement of the need for the measure. Usually includes statements pertaining to Importance criterion: impact, gap in care, and evidence.		Required
Clinical Recommendation Statement	Summary of relevant clinical guidelines or other clinical recommendations supporting this eCQM.		Required
Improvement Notation	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).		None
Reference(s)	Identifies bibliographic citations or references to clinical practice guidelines, sources of evidence, or other relevant materials supporting the intent and rationale of the eCQM.		None
Definition	Description of individual terms, provided as needed.	This field may be removed in the future.	None
Guidance	Used to allow measure developers to provide additional guidance for implementers to understand greater specificity than could be provided in the logic for data criteria.		None

Header Data Elements	Definition	Measure Developer Guidance	Preferred Term (Required, None, Not Applicable)
Transmission Format	Uniform Resource Locator (URL) or hyperlinks for the transmission formats specified for a particular reporting program.	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.	None
Initial Population	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.	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 QA 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.	Required
Denominator	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 eCQM. Different measures within an eCQM set may have different Denominators. CV eCQMs do not have a Denominator, but instead define a Measure Population.	For proportion/ratio measures, include the text "Equals Initial Population" where applicable.	Not Applicable (for CV eCQMs)

Header Data Elements	Definition	Measure Developer Guidance	Preferred Term (Required, None, Not Applicable)
Denominator Exclusion	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.		None *(for proportion or ratio eCQMs)* Not Applicable (*for CV eCQMs*)
Numerator	Numerators are used in proportion and ratio 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.		Not Applicable (for CV eCQMs)
Numerator Exclusion	Numerator Exclusion is 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 1000 catheter days were to exclude infections with a specific bacterium, that bacterium would be listed as a numerator exclusion.	Numerator Exclusion is 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.	None (*for ratio and proportion eCQMs*) Not Applicable (*for CV and cohort eCQMs*)

Header Data Elements	Definition	Measure Developer Guidance	Preferred Term (Required, None, Not Applicable)
Denominator Exception	Denominator Exception is 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 Exception allow for adjustment of the calculated score for those providers with higher risk populations on individual patient characteristics or preferences. Denominator Exception are used only in proportion eCQMs. They are not appropriate for ratio or CV eCQMs. Denominator Exception 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 System reasons	Be specific for all categories of denominator exception reasons.	None (*for proportion eCQMs*) Not Applicable (*for ratio or CV eCQMs*)
Measure Population	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.	For CV eCQMs, include the text "Equals All in Initial Population." Then add any specific additional criteria if needed.	Not Applicable (*for ratio or proportion eCQMs*)
Measure Population Exclusion	Measure Population Exclusion is 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.	Measure population exclusions are used only in CV eMeasures. It is a narrative description of the eMeasure population to exclude.	None (*for CV eCQMs*) Not Applicable (*for ratio or proportion eCQMs*)

Header Data Elements	Definition	Measure Developer Guidance	Preferred Term (Required, None, Not Applicable)
Measure Observation	Measure Observation 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.	Measure observations are generally described using a statistical methodology such as: count, etc.	Not Applicable (for proportion eCQMs)
Supplemental Data Elements	CMS defines four required Supplemental Data Elements (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.	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 the following 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.	Required

APPENDIX C: 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.

Section/Chapter	Changes
ALL	Overhaul of eCQM throughout entire document • Removed eCQM specific section (formerly Section 4) and several eCQM specific appendices • Integrated eCQM throughout document • Added three new chapters to Section 3: Introduction to eCQM; eCQM Standards and Tools; and Values, Value Sets, and the VSAC • Added significant new content to Chapters 13 and 21 • Changed all instances of “emeasure” to “eCQM”
ALL	Overhaul of all figures and tables to ensure aesthetic consistency
ALL	Incorporated almost 1,000 diverse comments and suggestions from measure developer reviewers on various topics throughout the Blueprint
ALL	Changed language with respect to socioeconomic status according to recommendations based on the recent ASPE Report to Congress
Section 3, Chapter 8	Added Master Deliverables List
Section 3, Chapter 11	Refined Business Case Development Section and Template
Section 3, Chapter 12	Refined language and added best practices on TEPs
Section 4, Measure Submission Forms	Incorporated NQF review of measure submission forms, to be fully completed in version 13.1