A Blueprint

for the

CMS Measures Management System

Version 11.2 January 2016

Table of Contents

Table of Contents **3**
List of Figures **10**
List of Tables **11**
Preface: How to Use the Interactive Blueprint **12**
Structure of the Blueprint 12
How Sections 1, 2, and 3 Work Together for eMeasure Processes 12
How to Use the Glossary, Appendices, and Tools 14
How External Resources Are Referenced 16
How To Formally Submit Changes For Upcoming Versions 16
Section 1—Measure Development Concepts **18**
National Quality Strategy 18
CMS Quality Strategy 19
Crosswalk Between the National Quality Strategy and the CMS Quality Strategy 20
Chapter 1—The CMS Measures Management System Blueprint *22*
1.1—Role of the Measure Developer 24
1.2—Role of the Measure Developer's COR 25
1.3—Role of the Measures Manager 26
1.4—Measure Governance 27
Chapter 2—Priorities Planning *28*
2.1—Alignment, Harmonization, and Prioritization 29
2.2—CMS Measure Planning Activities 30
2.3—Role of the Measure Developer in Priorities Planning 33
2.4—Role of the Measures Manager in Priorities Planning 34
Chapter 3—Stakeholder Input *35*
3.1—Role of Person and Family Engagement and Participation 36
3.2—Enhancing Person and Family Member Engagement 37
3.3—Person/Family Member Engagement Best Practices 41
3.4—Role of the Technical Expert Panel in Stakeholder Input 46
3.5—Role of Public Comment in Stakeholder Input 50
Chapter 4—Harmonization *53*
4.1—Adapted Measures 55
4.2—Adopted Measures 56
4.3—New Measures 57
Chapter 5—Concepts of Measure Evaluation *59*
5.1—Measure Evaluation Criteria and Subcriteria 60
5.2—Applying Measure Evaluation Criteria 61
5.3—Report of the Measure Developer's Evaluation of the Measures 62
5.4—Developing and Evaluating the Business Case 63
5.5—Evaluation Considerations for Special Types of Measures 69
5.6—Overview of Measure Evaluation Tools 72
Chapter 6— Measure Classification and Selected Specific Measure Types *73*
6.1 – Measure Classification 73
6.2—Cost and Resource Use Measures 78
6.3—Composite Performance Measures 79
6.4—Patient-Reported Outcome (PRO) Measures 81
6.5—Multiple Chronic Conditions Measures 85

Section 2—Measure Lifecycle 91

Chapter 1—Measure Conceptualization 94

Chapter 1.A.—Information Gathering 95

Chapter 1.B.—Technical Expert Panel 117

Chapter 1.C.—Public Comment 127

Chapter 2—Measure Specification 133
2.1—Deliverables 133
2.2—Technical Specification 134
2.3—Harmonizing Measure Specifications 157
2.4—Risk-Adjusted Outcome Measures 159
2.5—Cost Resource Use Measure Technical Specifications 174
2.6—Composite Measure Technical Specifications 177

Chapter 3—Measure Testing 181
3.1—Deliverables 182
3.2—Alpha and Beta Testing 182
3.3—Procedure 185
3.4—Testing and Measure Evaluation Criteria 191
3.5—Testing Considerations for Selected Measure Types 199

Chapter 4—Measure Implementation 203
4.1—National Quality Forum Endorsement 204
4.2—Measure Selection 212
4.3—Measure Rollout 216

Chapter 5—Measure Use, Continuing Evaluation, and Maintenance 223
5.1—Measure Production and Monitoring 224
5.2—Measure Update 233
5.3—Comprehensive Reevaluation 238
5.4—Ad Hoc Review 245
5.5—Possible Outcomes of Maintenance Reviews 250

Section 3—eMeasure Lifecycle 257

eMeasure Components 258
Overview of the eMeasure Lifecycle 259

Chapter 1—eMeasure Conceptualization 263
1.1—Deliverables 264
1.2—eMeasure Feasibility Assessment 264
1.3—Information Gathering for eMeasures 265
1.4—Additional or Adapted Evaluation Subcriteria for Scientific Acceptability of the eMeasure Properties 266
1.5—Technical Expert Panel/Subject Matter Expert 267

Chapter 2—eMeasure Specification 269
2.1—Deliverables 269
2.2—JIRA Basics for eCQM Developers 270
2.3—Develop eMeasure Specifications 271
2.4—Procedure 273
2.5—eCQM Specifications Review Process 297
2.6—Innovations in eSpecifications 304

Chapter 3—eMeasure Testing 305
3.1—Deliverables 306
3.2—Types of eMeasure Testing 306
3.3—Phases of eMeasure Testing 314

3.4—Tools for Testing eMeasures 315
3.5—Procedure 315
3.6—Testing Challenges 317

Chapter 4—eMeasure Implementation 319
4.1—Deliverables 320
4.2—Preparing eMeasure Specifications for Public Release and Public Comment 321
4.3—NQF eMeasure Endorsement 322
4.4—eMeasure Rollout, Implementation, and Publication 323

Chapter 5—eMeasure Reporting 329
5.1—Deliverables 329
5.2—Quality Reporting Document Architecture (QRDA) Category I 329
5.3—Quality Reporting Document Architecture (QRDA) Category III 330
5.4—Procedure 331
5.5—Continued Improvement and Maintenance 332

Chapter 6—eMeasure Use, Continuing Evaluation, and Maintenance 335
6.1—Deliverables 336
6.2—eMeasure Evaluation and Testing After Implementation 337
6.3—Annual Update Process for eMeasures 339
6.4—Comprehensive Reevaluation for eMeasures 341
6.5—Ad Hoc Review for eMeasures 342

Section 4—Tools, Appendices, and Forms 343

Acronyms and Abbreviations 343

Glossary 345

Appendix A: XML View of a Sample eMeasure 367
Population criteria 368
Data criteria 369
Reporting stratification 370
Supplemental data elements 370

Appendix B: Electronic Specification (eMeasure) 372

Appendix C: eMeasure QA Checklists 374

Appendix D: eMeasure Review Process Quick Reference 376

Appendix E: Time Unit and Time Interval Definitions 378

Appendix F: Time Interval Calculation Conventions 380
Timing relationships with no calculation unit defined 383
Other date/time-related calculations using QDM functions 385

Appendix G: Proportion Measure Calculations (exclusion, exception, and aggregation) 386

Appendix H: Continuous Variable and Ratio Measure Calculations 392
Measure Populations 392
Individual Observations 394
Measure Aggregate Calculations 394
Ratio Measure Example 395
Measure aggregate calculations 396
Continuous Variable Measure Example 397

Appendix I: Harmonization of Inpatient Encounter Value Sets 399

Appendix J: Summary of Changes to Blueprint, version 11.2 402
..... 403

Business Case Form Instructions **405**
Project Title: 405
Project Overview: 405
Date: 405
Measure Name: 405
Summary: 405
Incidence and prevalence: 405
Current performance, including any disparities: 406
Major clinical, functional status and other benefits: 406
Successful intervention is feasible: 406
Costs of clinical care: 406
Complications or untoward effects: 406
Estimated size and trajectory of anticipated improvement: 407
Model and formulas used: 407
Limitations: 407
Net benefit: 407
Blank Business Case Form Template **408**
Project Title: 408
Project Overview: 408
Date: 408
Measure Name: 408
Summary: 408
Incidence and prevalence: 408
Current performance, including any disparities: 408
Major clinical, functional status and other benefits: 408
Successful intervention is feasible: 408
Costs of clinical care: 408
Complications or untoward effects: 409
Estimated size and trajectory of anticipated improvement: 409
Model and formulas used: 409
Limitations: 409
Net benefit: 409
Call for Measures Web Posting **410**
Project Title: 410
Dates: 410
Project Overview: 410
Project Objectives: 410
Instructions: 410
Measure Information Form Instructions **411**
Measure Name 411
Introduction Tab 411
NQF Conditions Tab 411
Specifications Tab 412
Blank Measure Information Form Template **422**
Project Title: 422
Project Overview: 422
Date: 422
Measure Name 423
Descriptive Information 423
Measure Specifications 423
Measure Justification Form Instructions **424**

Measure Name 424
Type of Measure 424
Importance Tab 424
Scientific Acceptability Tab 432
Feasibility Tab 440
Usability and Use Tab 441
Related and Competing Measures Tab 443
Additional [Information] Tab 443

Blank Measure Justification Form Template 445

Project Title: 445
Project Overview: 445
Date: 445
Measure Name 445
Type of Measure 445
Importance 445
Scientific Acceptability 447
Feasibility 449
Usability and Use 449
Related and Competing Measures 449
Additional Information 450

Measure Evaluation Criteria and Instructions 451

1. Evidence, Performance Gap, and Priority (Impact)—Importance to Measure and Report: 451
2. Reliability and Validity—Scientific Acceptability of Measure Properties: 455
3. Feasibility: 463
4. Usability and Use: 465
5. Comparison to Related or Competing Measures: 467

Blank Measure Evaluation Report Template 469

Project Title: 469
Project Overview: 469
Date: 469
Measure Name: 469
Measure Set (or Setting): 469
Measure Developer: 469
1. Evidence, Performance Gap, and Priority (Impact)—Importance to Measure and Report 469
2. Reliability and Validity—Scientific Acceptability of Measure Properties 470
3. Feasibility 471
4. Usability and Use 472
5. Comparison to Related or Competing Measures 472
Preliminary Recommendation for Endorsement 472

Public Comment Call Web Posting 474

Project Title: 474
Dates: 474
Project Overview: 474
Project Objectives: 474
Documents and Measures for Comment: 474
Expiration Notice: 474
Project Specific Instructions: 474

Public Comment Summary Web Posting 475

Project Title: 475
Dates: 475
Project Overview: 475

Project Objectives: 475
Comment Summary: 475
Expiration Notice: 475

Public Comment Summary Report Template 476

Project Title: 476
Dates: 476
Project Overview: 476
Project Objectives: 476
Information About the Comments Received: 476
Stakeholder Comments—General and Measure-Specific 476
Preliminary Recommendations 476
Overall Analysis of the Comments and Recommendations 476

Technical Expert Panel Call for TEP Web Page Posting 478

Project Title: 478
Dates: 478
Project Overview: 478
Project Objectives: 478
TEP Requirements: 478
TEP Expected Time Commitment: 479
Expiration Notice: 479
Required Information: 479

Technical Expert Panel (TEP) Nomination Form Template 480

Project Title: 480
Project Overview: 480
Project Objectives: 480
TEP Expected Time Commitment: 480
TEP Requirements: 480
Instructions: 480
Applicant/Nominee Information (Self-Nominations Are Acceptable): 482
Person Recommending the Nominee: 482
Applicant/Nominee's Disclosure: 483
Applicant/Nominee's Agreement: 483

Technical Expert Panel Composition (Membership) List Template 484

Project Title: 484
Dates: 484
Project Overview: 484

Technical Expert Panel Composition (Membership) List Web Page Posting 485

Project Title: 485
Dates: 485
Documents: 485
Project Overview: 485
Project Objectives: 485
TEP Requirements: 485
TEP Expected Time Commitment: 485
Expiration Notice: 486

Technical Expert Panel Charter Template 487

Project Title: 487
Dates: 487
Project Overview: 487
Project Objectives: 487
TEP Objectives: 487

Scope of Responsibilities: 487
Guiding Principles: 487
Estimated Number and Frequency of Meetings: 487
Date Approved by TEP: 487
TEP Membership: 488
Technical Expert Panel Summary Web Page Posting 489
Project Title: 489
Dates: 489
Documents: 489
Project Overview: 489
Project Objectives: 489
Expiration Notice: 489

List of Figures

Figure 1: eMeasure Icon 13
Figure 2: Glossary Definition Pop-up for Terms 14
Figure 3: Navigating PDF Bookmarks 15
Figure 4: Navigation Pane Activation 15
Figure 5: External Hyperlinks 16
Figure 6: CMS Measures Management System Model 24
Figure 7: CMS Priorities Planning and Measure Selection 30
Figure 8: Applying Measure Evaluation Criteria 61
Figure 9: Inputs and Uses for the Business Case 65
Figure 10: MMS Business Case Best Practices 67
Figure 11: Measurement Settings 77
Figure 12: Relationship Between Resource Use, Value, and Efficiency 79
Figure 13 : Flow of the Measure Lifecycle 92
Figure 14: Timeline of Measure Lifecycle 93
Figure 15: Flow of the Measure Lifecycle - Measure Conceptualization 94
Figure 16: The Posting Process 96
Figure 17: Environmental Scan Data Sources 99
Figure 18: Flow of the Measure Lifecycle - Measure Specification 133
Figure 19: Structure Process Outcome Relationship 159
Figure 20: Example of ROC Curves 172
Figure 21: Flow of the Measure Lifecycle - Measure Testing 182
Figure 22: Overview of Measure Testing 186
Figure 23: Flow of the Measure Lifecycle - Measure Implementation 203
Figure 24: Measure Submission 205
Figure 25: Overview of the Measure Rollout Process 217
Figure 26: Flow of the Measure Lifecycle - Measure Use, Continuing Evaluation, and Maintenance 224
Figure 27: Overview of the Measure Monitoring Process 225
Figure 28: Measure Review Cycle Timeline 233
Figure 29: eMeasure Lifecycle 261
Figure 30: eMeasure Conceptualization Tools and Stakeholders 263
Figure 31: eMeasure Specification Tools and Stakeholders 269
Figure 32: eMeasure Specification Development Process 274
Figure 33: Vision for Robust Value Set Management 292
Figure 34: eMeasure Testing Tools and Stakeholders 306
Figure 35: eMeasure Implementation Tools and Stakeholders 320
Figure 36: QRDA Category I Sample Report 330
Figure 37: QRDA Category III Sample Report 331
Figure 38: eMeasure Use, Continuing Evaluation, and Maintenance Key Stakeholders 336
Figure 39: Annual Update Process for eMeasures 341
Figure 40: Proportion Measure Populations 386
Figure 41: Ratio Measure Populations 392
Figure 42: Continuous Variable Measure Populations 393

List of Tables

Table 1: Crosswalk between Sections 2 and 3 13
Table 2: National Quality Strategy and CMS Quality Strategy 20
Table 3: Options for Person and Family Member Engagement by Measure Development Stage 40
Table 4: Harmonization Decisions During Measure Development 55
Table 5: NQMC Clinical Quality Measure Domains 73
Table 6: Examples of Measures Addressing Each of the National Quality Strategy Priorities 75
Table 7: Differences Between Related and Competing Measures 95
Table 8: Attributes of Risk Adjustment Models 161
Table 9: Types of Composite Measure Scoring 177
Table 10: Features of Alpha and Beta Testing 183
Table 11: Harmonization Decisions during Measure Maintenance 231
Table 12: Extent of Measure Evaluation as a Function of Prior Comprehensive Evaluation and Measure Use 243
Table 13: CMS Criteria for Measure Disposition 252
Table 14: eCQM Metadata 276
Table 15: Diagnoses Available for Use in eCQMs 282
Table 16: Conventions for Constructing Age Logic 283
Table 17: ONC HIT Standards Committee Recommended Vocabulary Summary 286
Table 18: ONC HIT Standards Committee Transition Vocabulary Standards Summary and Plan 286
Table 19: Allergy and Intolerance Value Set Naming Convention 287
Table 20: Quality Data Model Categories With ONC HIT Standards Committee Recommended Vocabularies 288
Table 21: Measure Populations Based on Type of Measure Scoring 292
Table 22: Value Set Review Areas and Remedial Actions 300
Table 23: Sample eMeasure Identifier 323
Table 24: Sample File Naming 325
Table 25: Application of Evaluation Methods 337
Table 26: Example of XML View 367
Table 27: Sample Electronic Specification 372
Table 28: Logic Review Checklist 374
Table 29: Clinical Review Checklist 375
Table 30: Naming File Attachments 376
Table 31: Comment Boxes 376
Table 32: Time Unit and Time Interval Definitions 378
Table 33: Time Interval Calculation Conventions 380
Table 34: Harmonization of Inpatient Encounter Value Sets 399

Preface: How to Use the Interactive Blueprint

Structure of the Blueprint

The Blueprint is divided into four sections. The first section contains measure development concepts, the second describes procedural steps for measure development and maintenance, the third covers procedural information specific to eMeasures, and the fourth is a collection of tools and forms to be used during the measure lifecycle. Below is a more detailed explanation of each section of the Blueprint:

- **Section 1—Measure Development Concepts**—covers concepts that apply throughout the measure lifecycle and provides background and a more in-depth description of those concepts. This material is not repeated in the procedure sections but is referenced wherever applicable.
- **Section 2—Measure Lifecycle**—describes the procedural steps necessary to develop and maintain a measure.
- **Section 3—eMeasure Lifecycle**—is constructed parallel to Section 2 because it also describes measure development and maintenance procedures, but material in this section is specific to eMeasures. The section references conceptual material from Section 1 and procedural material from Section 2 where they apply to eMeasures.
- **Section 4—Tools, Appendices, and Forms**—contains all of the tools, appendices, and forms referenced in the Blueprint. It also contains a glossary and a list of acronyms used in the Blueprint.

The Blueprint is now designed primarily as an electronic document with navigation bookmarks and internal and external hyperlinks. Resources and forms are referenced throughout the Blueprint where they are required, and key terms are linked to their appropriate glossary entries. Any text formatted in blue represents an internal or external hyperlink.

How Sections 1, 2, and 3 Work Together for eMeasure Processes

Due to the increased use of eMeasures in CMS quality programs, the eMeasure Specification section from previous editions of the Blueprint has been expanded to cover the complete eMeasure lifecycle from development through maintenance. The concepts of measure development and maintenance covered in Section 1 are linked to the procedures for general measure development and maintenance in Section 2, and are referenced with the specifics of eMeasure development and maintenance covered in Section 3.

Section 1 describes concepts that apply throughout the measure lifecycle. Where the concepts are used procedurally, they are referenced to the appropriate information in sections 2 and 3. Section 2—Measure Lifecycle and Section 3—eMeasure Lifecycle are structured in a parallel manner because many aspects of the process are used regardless of whether the measure is based on paper medical records or an electronic health record. These common processes are not repeated for both types of measures, but the topics are referenced wherever needed and linked to the procedure description. The material in Section 3 covers procedures for all types of eMeasures

including new (*de novo*), retooled (respecified), and reengineered (adapted) eMeasures, through all phases of the eMeasure Lifecycle. General procedures for Measure Conceptualization, Measure Specification, Measure Testing, Measure Implementation, and Measure Use and Continuing Evaluation are not repeated in Section 3, but links are provided back to the more in-depth discussion of the general processes in Section 2. Table 1 below provides a crosswalk between Sections 2 and 3. Correspondingly, eMeasure development concepts or processes that deviate from the way other types of measures are developed are denoted with the icon shown in Figure 1 (scaled to fit the text).

Figure 1: eMeasure Icon

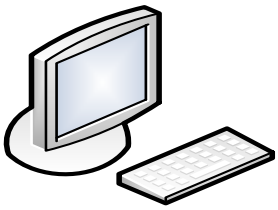


Table 1: Crosswalk between Sections 2 and 3

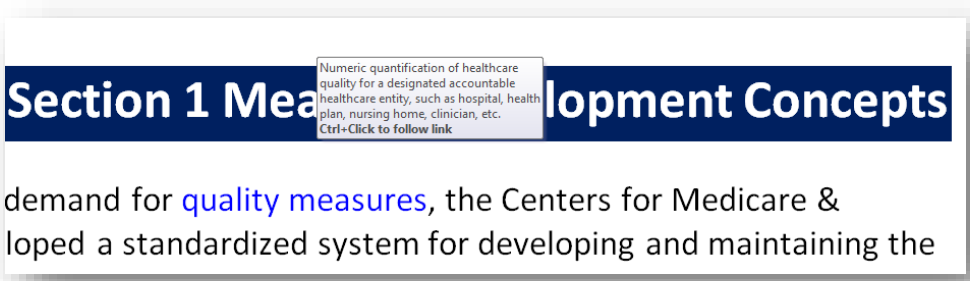
Section 2—Measure Lifecycle	Section 3—eMeasure Lifecycle
Chapter 1—Measure Conceptualization **Chapter 1.A.—Information Gathering** **Chapter 1.B.—Technical Expert Panel** **Chapter 1.C.—Public Comment**	**Chapter 1—eMeasure Conceptualization** • 1.1—Deliverables • 1.2—eMeasure Feasibility Assessment • 1.3—Information Gathering for eMeasures • 1.4— Additional or Adapted Evaluation Subcriteria for Scientific Acceptability of the eMeasure Properties • 1.5—Technical Expert Panel/Subject Matter Expert
Chapter 2—Measure Specification • 2.1—Deliverables • 2.2—Technical Specification • 2.3—Harmonizing Measure Specifications • 2.4—Risk-Adjusted Outcome Measures • 2.5—Cost Resource Use Measure Technical Specifications • 2.6—Composite Measure Technical Specifications	**Chapter 2—eMeasure Specification** • 2.1—Deliverables • 2.2—JIRA Basics for eCQM Developers • 2.3—Develop eMeasure Specifications • 2.4—Procedure • 2.5—eCQM Specifications Review Process • 2.6—Innovations in eSpecifications
Chapter 3—Measure Testing • 3.1—Deliverables • 3.2—Alpha and Beta Testing • 3.3—Procedure • 3.4—Testing and Measure Evaluation Criteria • 3.5—Testing Considerations for Selected Measure Types	**Chapter 3—eMeasure Testing** • 3.1—Deliverables • 3.2—Types of eMeasure Testing (data element, logic, etc.) • 3.3—Phases of eMeasure Testing • 3.4—Tools for Testing eMeasures (Bonnie, Cypress) • 3.5—Procedure • 3.6—Testing Challenges

Section 2—Measure Lifecycle	Section 3—eMeasure Lifecycle
Chapter 4—Measure Implementation • 4.1—National Quality Forum Endorsement • 4.2—Measure Selection • 4.3—Measure Rollout	**Chapter 4—eMeasure Implementation** • 4.1—Deliverables • 4.2—Preparing eMeasure Specifications for Public Release and Public Comment • 4.3—NQF eMeasure Endorsement • 4.4—eMeasure Rollout, Implementation, and Publication
	Chapter 5—eMeasure Reporting • 5.1—Deliverables • 5.2—Quality Reporting Document Architecture (QRDA) Category I • 5.3—Quality Reporting Document Architecture (QRDA) Category III • 5.4—Procedure • 5.5—Continued Improvement and Maintenance
Chapter 5—Measure Use, Continuing Evaluation, and Maintenance • 5.1—Measure Production and Monitoring • 5.2—Measure Update • 5.3—Comprehensive Reevaluation • 5.4—Ad Hoc Review • 5.5—Possible outcomes of maintenance reviews	**Chapter 6—eMeasure Use, Continuing Evaluation, and Maintenance** • 6.1—Deliverables • 6.2—eMeasure Evaluation and Testing After Implementation • 6.3—Annual Update Process for eMeasures • 6.4—Comprehensive Reevaluation for eMeasures • 6.5—Ad Hoc Review for eMeasures

How to Use the Glossary, Appendices, and Tools

Definitions are provided for terms used throughout the Blueprint. The terms are included in the Glossary, but they are also linked so that when the cursor hovers over the word, a short definition pops up as illustrated in Figure 2: Glossary Definition Pop-up for Terms.

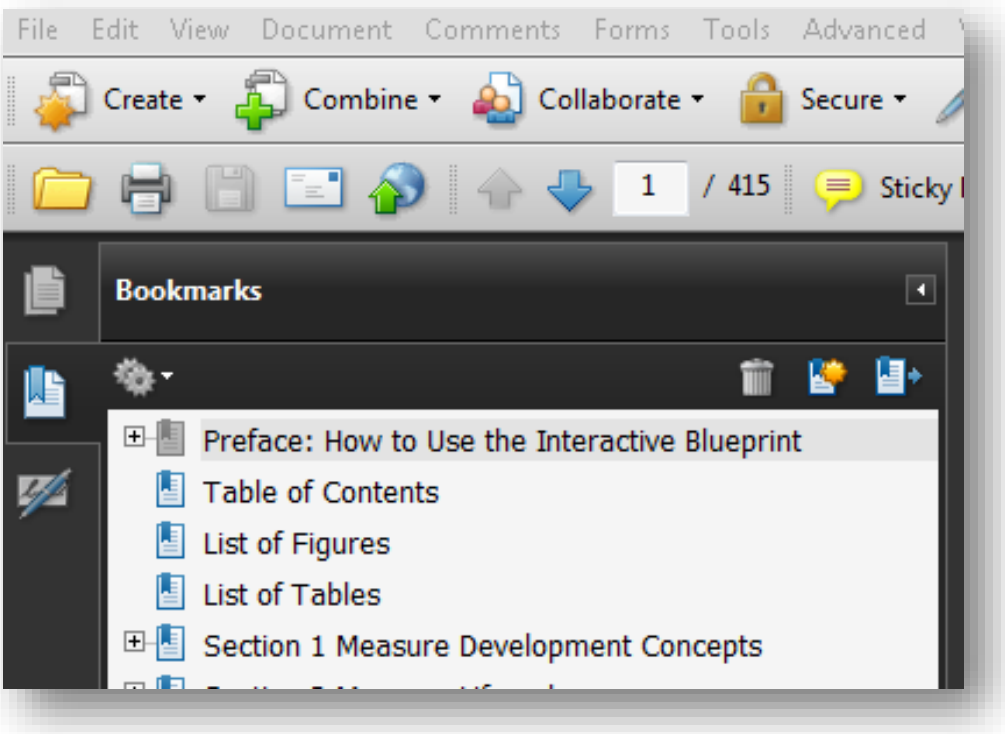
Figure 2: Glossary Definition Pop-up for Terms



Pressing the keyboard "Ctrl" button and left clicking on the text will bring the reader to the appropriate section, appendix, form, or other resource within the Blueprint. To return to the original place in the document, press the "Alt" key and the left arrow together either in Word or in the PDF version.

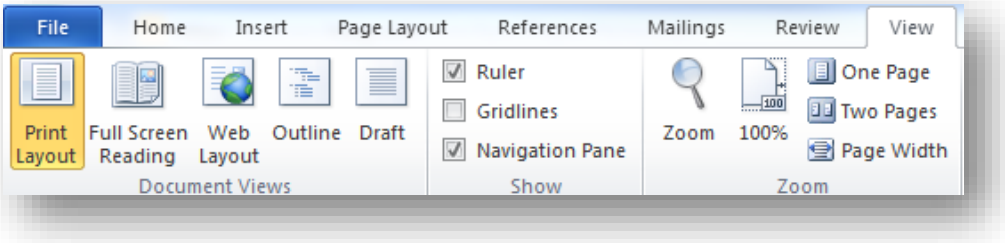
When viewed as a PDF in Adobe Reader, the bookmarks at the left of the screen correspond to the sections, chapters, and subheadings listed in the Table of Contents. Clicking on the icon with the ribbon reveals the bookmarks as depicted in Figure 3: Navigating PDF Bookmarks.

Figure 3: Navigating PDF Bookmarks



These bookmarks can also be viewed in the Word document if the Navigation Pane is activated from the View menu tab as illustrated in Figure 4: Navigation Pane Activation.

Figure 4: Navigation Pane Activation

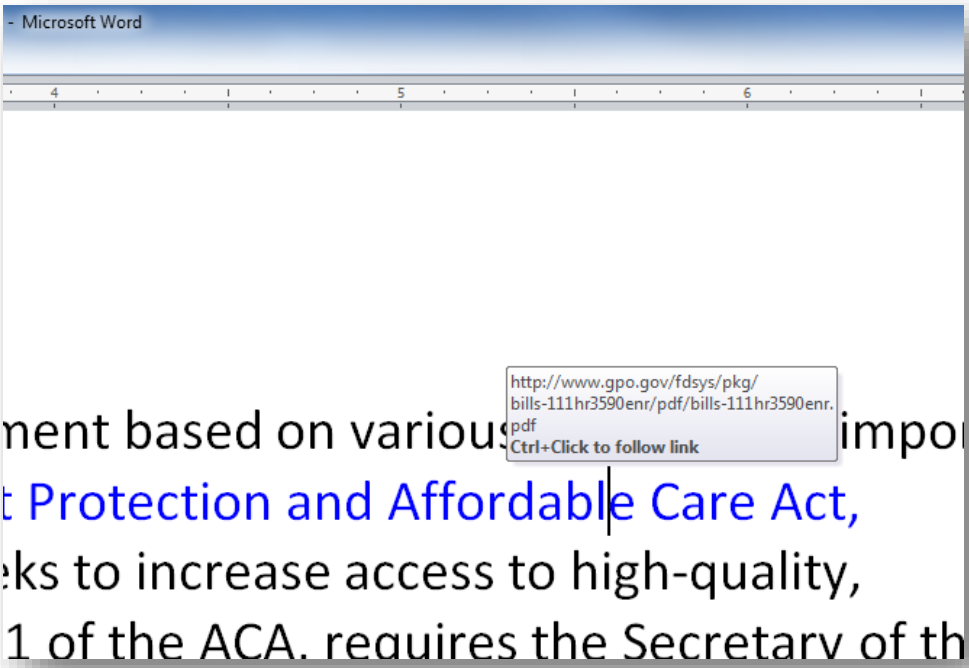


Using either the Navigation Pane or the PDF bookmarks, the reader can move easily between sections of interest within the Blueprint.

How External Resources Are Referenced

References to documents or Web pages available online are linked in the text. Pressing the keyboard "Ctrl" button and left-clicking on the hyperlinked blue text will connect the user to the selected location. To reveal the destination URL, hover the cursor over the hyperlinked blue text. The full URL is also spelled out in a corresponding footnote. Figure 5 depicts an external hyperlink.

Figure 5: External Hyperlinks



How To Formally Submit Changes For Upcoming Versions

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

- 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 accepted year-round, and incorporated into the next proximate cyclical review of the document.

Section 1—Measure Development Concepts

In response to an ever-increasing demand for quality measures, the Centers for Medicare & Medicaid Services (CMS) has developed a standardized system for developing and maintaining the quality measures used in its various quality initiatives and programs. Known as the Measures Management System (MMS),[1] CMS-funded measure developers (or contractors) should follow this core set of business processes and decision criteria when developing, implementing, and maintaining quality measures. The primary goal of the Measures Management System is to provide information to measure developers to help them produce high-caliber quality measures. This manual, the CMS Measures Management System Blueprint, Version 11.2 (the Blueprint), documents the full Measures Management System core set of business processes.[2]

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 (National Quality Strategy). The Patient Protection and Affordable Care Act of 2010, commonly called the Affordable Care Act (ACA),[3] seeks to increase access to high-quality, affordable healthcare for all Americans. Section 3011 of the ACA requires the Secretary of the Department of Health and Human Services (HHS) to establish a National Quality Strategy that sets priorities to guide this effort and includes a strategic plan for how to achieve it. The initial National Quality Strategy 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.

---

[1] 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.

[2] Centers for Medicare & Medicaid Services. *Measures Management System: Blueprint.* Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/MeasuresManagementSystemBlueprint.html. Accessed on: May 6, 2015.

[3] 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: May 6, 2015.

- Making quality care more affordable for individuals, families, employers, and governments by developing and spreading new healthcare delivery models.[4]

The National Quality Strategy's first annual progress report to Congress,[5] published in April 2012, elaborated on these 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.[6]

At the HHS level, the Secretary has articulated Delivery System Reform goals. These goals focus on how Medicare and other payers pay providers, deliver care and distribute information, to achieve a future state that is patient-centered, provides incentives for patient outcomes, is sustainable, emphasizes coordinated care and shared decision-making with patients and relies on transparency of quality and cost information. This future state promotes value-based payment systems including value-based purchasing, Accountable Care Organizations, episode-based payments and Patient-Centered Medical Homes. During January 2015, HHS set internal goals for value-based payments within the Medicare Fee for Service (FFS) system and invited private sector payers to match or exceed HHS goals:

- Goal 1: **30%** of Medicare payments are tied to quality or value through **alternative payment models** by the end of 2016, and **50%** by the end of 2018.
- Goal 2: **85%** of all Medicare fee-for-service payments are **tied to quality or value** by the end of 2016, and **90%** by the end of 2018.

CMS Quality Strategy

The CMS Quality Strategy[7] pursues and aligns with the three broad aims of the National Quality Strategy as well as Delivery System Reform payment goals. Like the National Quality Strategy, 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. The CMS Quality Strategy pursues and aligns with the three broad aims of the National Quality Strategy and its six priorities. Each of these 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.

---

[4] 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.

[5] Department of Health and Human Services. Agency for Health Research and Quality, National Quality Strategy. *2011 Report to Congress: National Strategy for Quality Improvement in Health Care. Mar 2011*. Available at: http://www.ahrq.gov/workingforquality/nqs/nqs2011annlrpt.htm. Accessed on: May 6, 2015.

[6] Department of Health and Human Services, Agency for Health Research and Quality, National Quality Strategy. *2012 Annual Progress Report to Congress National Strategy for Quality Improvement in Health Care*. Available at: http://www.ahrq.gov/workingforquality/nqs/nqs2012annlrpt.htm. Accessed on: May 6, 2015.

[7] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *CMS Quality Strategy 2013 — Beyond*. Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/CMS-Quality-Strategy.pdf. Accessed on: May 15, 2015.

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, and drug and health plans. Beginning in 2012, CMS efforts expanded the quality reporting programs to include 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 in which care is delivered."[8]

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),[9] university researchers, health services research organizations, and consulting groups. The Measures Manager supports the CMS Contracting Officer Representatives (CORs) and their various measure developers in their work implementing the Measures Management System.

Crosswalk Between the National Quality Strategy and the CMS Quality Strategy

The following table compares the National Quality Strategy with the goals and objectives of the CMS Quality Strategy: [10]

Table 2: National Quality Strategy and CMS Quality Strategy

National Quality Strategy 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.

[8] Conway PH, Mostashari F, Clancy C. The future of quality measurement for improvement and accountability. *JAMA.* 2013;309(21):2215-2216.
[9] Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityImprovementOrgs/index.html?redirect=/QualityImprovementOrgs. Accessed on: May 16, 2015.
[10] 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: May 25, 2015.

National Quality Strategy Priorities	CMS Quality Strategy Goals and Objectives
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. • 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.

Chapter 1—The CMS Measures Management System Blueprint

CMS recognizes the need for quality measures of the highest caliber, maintained throughout their lifecycle to ensure they retain the highest level of scientific soundness, importance, feasibility, and usability. The Blueprint documents the set of processes and decision criteria used by CMS to oversee the development, implementation, and maintenance of healthcare quality measures. The primary goal of the Blueprint is to help measure developers carry out their contracts and produce quality measures that are appropriate for use by CMS. Through application of the Blueprint processes, CMS ensures that it has a coherent, transparent system for measuring the quality of care delivered to its beneficiaries.

CMS launched the Blueprint in October, 2003. The Blueprint was originally designed based on the quality measurement work at CMS, augmented by best practices of other major measure developers. These best practices were collected and implemented through the first Blueprint and are updated annually. Though there are specific processes and supporting documentation tools provided for each phase of measure development, the sequence of steps can be flexible and the tools can be customized. The measure developers can add processes at the direction and approval of their COR.

From Version 1 through the current Version 11.2, the Blueprint has been updated to incorporate changes in the regulatory environment, changes 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 in Version 11.2, see Appendix J.

Inputs to Blueprint updates include:

- CMS suggestions.
- Issues identified from environmental scans of journal articles.
- Suggestions from measure developers.
- Innovative and best practices observed from measure developers and other developers.
- Suggestions received from other stakeholders.
- Alignment with National Quality Forum (NQF) policies, processes, and position papers.
- Industry trends identified in the quality measurement community.
- Issues and redesigned processes identified through various Lean Kaizen events.

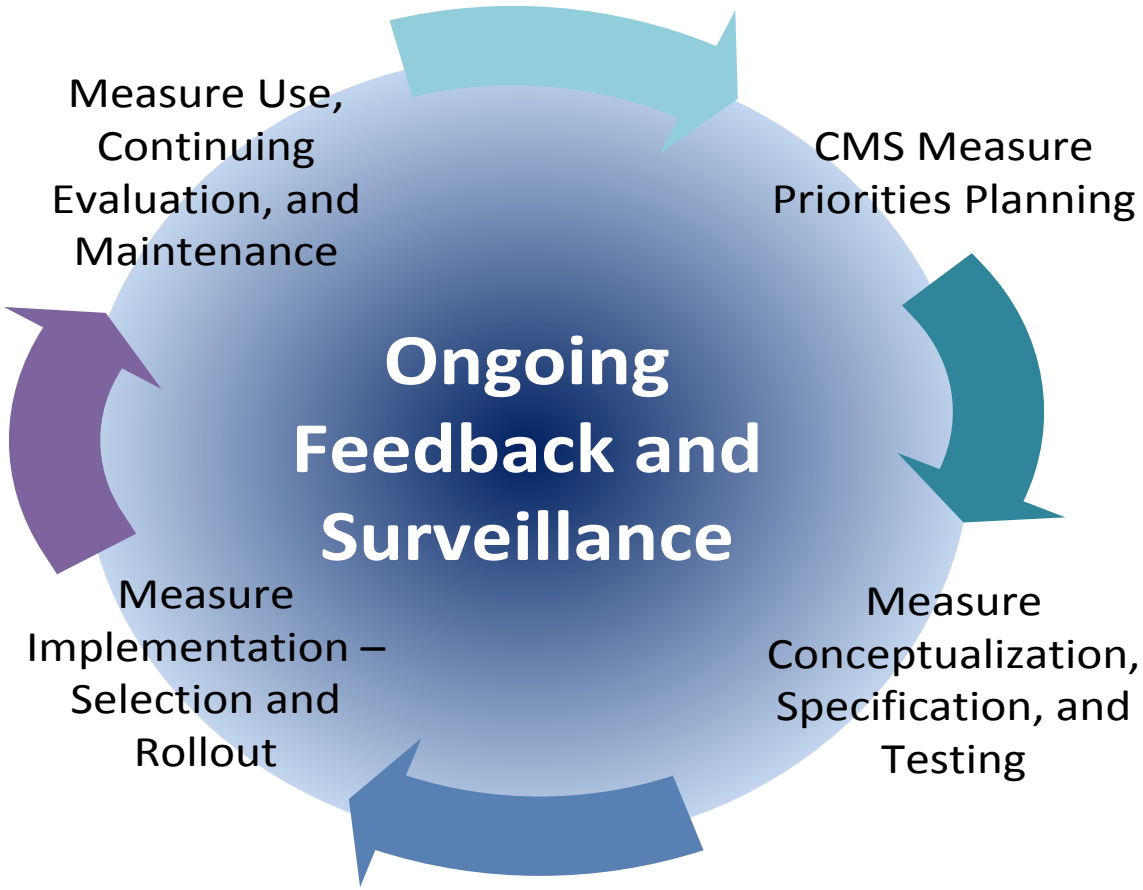
Several principles guide the content of the Blueprint. These principles include:

- Standardize core processes for measure development and maintenance to ensure development of high caliber quality measures.
- Make the decision-making processes of measure development and maintenance transparent to consumers and key stakeholders.

- Ensure the most efficient processes by applying Lean principles without compromising the quality of the measures produced.
- Permit flexibility and innovation in sequencing of steps and customization of tools and forms.
- Establish clear roles and responsibilities relative to the measure development process for CMS, measure developers, and the Measures Manager.
- Ensure stakeholder involvement and input throughout the measure development lifecycle.
- Increase communication, coordination, and collaboration between CMS, measure developers, and across HHS agencies.

Figure 6: CMS Measures Management System Model depicts how the measure lifecycle includes the following major categories of activities: measure conceptualization, specification, and testing; implementation; and use and continuing evaluation. These activities are augmented by ongoing surveillance of the measures environment, including the scientific basis and performance trends of the measures and the ongoing management of feedback on the measures or measure sets. Each of these measure development, implementation, and maintenance activities involve specific tasks as detailed in this Blueprint.

Figure 6: CMS Measures Management System Model



1.1—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 the Blueprint is designed as a guide for entities holding a measure development and maintenance contract with CMS, the Blueprint 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. For example, eMeasure developers may use the term measure developer to refer to the entity that "develops the electronic form of the measure, assembles the electronic Clinical Quality Measure (eCQM) review package, creates a record in the Office of the National Coordinator for Health Information Technology (ONC) JIRA

tracking system for the eCQM, and remediates eCQM issues discovered during the review process."[11]

Another term used for entities involved with measures is measure steward. NQF defines measure steward as follows: "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."[12] CMS will be the steward for most measures developed under contract for CMS. However, for NQF-endorsed measures, the measure developer will be responsible for carrying out the tasks required by the Measure Steward Agreement.[13]

Tools to guide and help measure developers are provided in the Blueprint. These tools are intended to be integrated into the measure developers' work and to help produce materials that serve as deliverables to inform CMS and document measure developers' progress.

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

- Using the processes and forms detailed in the Blueprint.
- Giving attention to Blueprint updates as provided by the Measures Manager.
- Reviewing the Blueprint requirements in context of their measure contract and good business practice. If the context requires flexible interpretation of the activities specified in the Blueprint, discuss options with the 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 Measures Management System.
- 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

1.2—Role of the Measure Developer's COR

Though the measure developers are responsible as authors 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 the Blueprint, it is the ultimate responsibility of the

[11] Technical Authority for the Unified Clinical Quality Improvement Framework (Tacoma). *Electronic Clinical Quality Measures (eCQM) Review Process Overview.* Centers for Medicare & Medicaid Services, Office of the National Coordinator for Health Information Technology. p. 2.
[12] 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.
[13] http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: May 25, 2015.

measure developer's COR to ensure that tasks in the measure development, implementation, and maintenance contracts are completed successfully. The COR achieves this mission by:

- Notifying the Measures Manager 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.

The COR Handbook for the CMS Measures Management Blueprint (COR Handbook), which is an essential companion document used by the COR, describes the specifics of the COR's role when managing measure development and maintenance contracts.

1.3—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 crosscutting perspectives to CMS and measure developers regarding measures and measurement methods and strategies.
- Scanning the measurement environment to ensure CMS is informed of issues related to quality measures. A journal scan report is submitted to CMS monthly.
- 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 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 Developer communications.

1.4—Measure Governance

Measure developers, as authors, and their CORs, as stewards, have distinct roles and responsibilities of governance throughout the lifecycle of a measure. Authors perform editing functions, whereas Stewards approve the work of the Authors and submit measures for publication. An elaboration on those roles are listed below, as adapted from the National Library of Medicine (NLM) Value Set Authority Center (VSAC)[14] doctrine on value sets governance, which closely mirrors governance of all measures, not just eMeasure value sets:

Author and Steward Roles

Authors create, edit, and submit measures to a designated Steward. Stewards approve, reject, and publish submitted measures from Authors. Authors may choose to act as their own Stewards. Authors have permissions to create, edit, and delete their own draft measures, as well as to submit measures to their assigned Stewards for approval, and withdraw measures from approval. Authors may create and edit measures, and they may submit or withdraw a measure (for review) to a Steward. It is also the responsibility of the author to socialize (or circulate, in order to gain feedback) their measure content and to collaborate on potential measure changes suggested by other authors or other entities.

Stewards have permissions to approve, reject, and publish measures that their assigned Author groups create and submit. Stewards provide overall coordination and management of the measures created by Authors under a specific program or for specific purpose. Stewards are responsible for approving measure content.

[14] https://vsac.nlm.nih.gov/. Accessed on: July 14, 2015.

Chapter 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 of clinical practice guidelines 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 three National Quality Strategy aims and six priorities[15], and emphasizes the five quality domains. 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."[16] 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, and there is room for improvement, 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 Measure Applications Partnership (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 National Quality Strategy and CMS Quality Strategy in measure development has contributed to significant improvement in this area – closing gaps and generating parsimonious measure sets. The recent Institute of Medicine (IOM) Vital Signs report[17] and the 2015 Impact Report[18] "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.

---

[15] 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: July 16, 2015.

[16] 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: July 16, 2016.

[17] 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 July 13, 2015.

[18] 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 July 13, 2015.

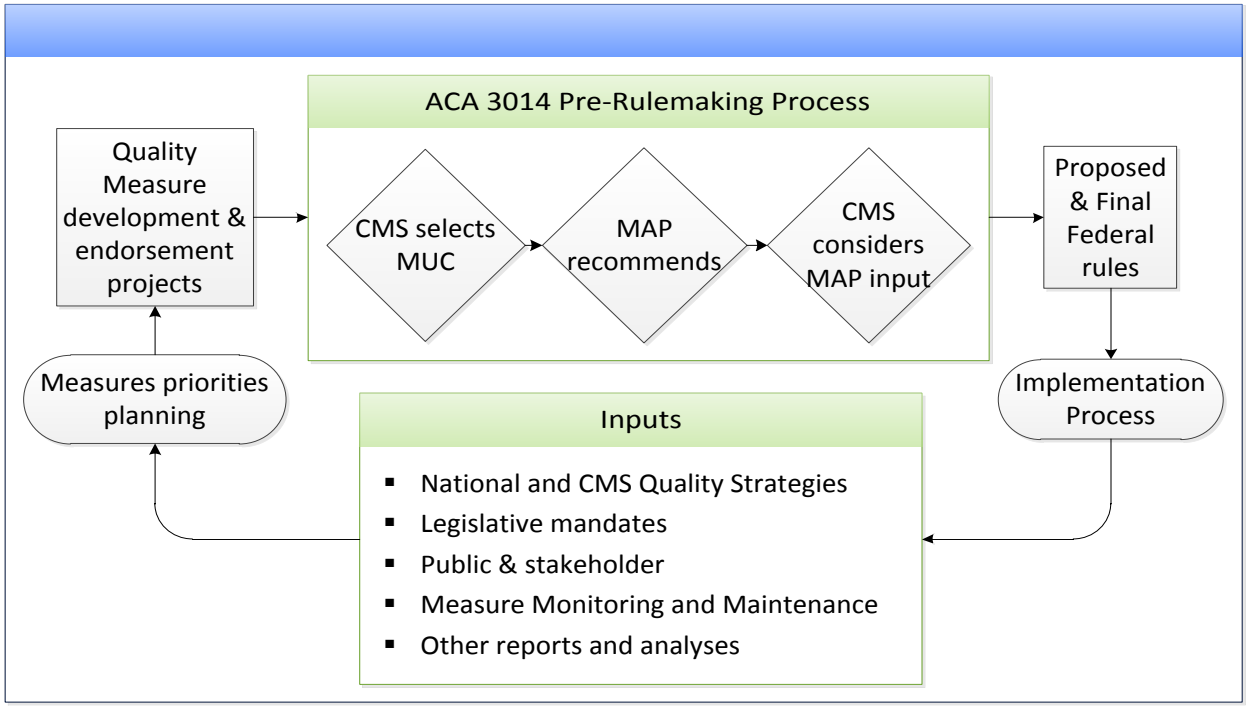
2.1—Alignment, Harmonization, and Prioritization

Figure 7 (below) 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 Medicare program(s). The Measures under Consideration (MUC) list is made available to the public no later than December 1st.

Around the second quarter of each year, through a call for quality and efficiency measures, CMS begins the annual pre-rulemaking cycle of collecting and compiling MUC. 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 provide input to HHS on the selection of quality and efficiency measures. The NQF, the entity with a 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 four workgroups including: 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. General requirements for informal rulemaking are summarized in a step by step fashion at produced by ICF Consulting in this Map of Informal Rulemaking.

Figure 7: CMS Priorities Planning and Measure Selection



2.2—CMS Measure Planning Activities

Input: National Quality Strategy and CMS Quality Strategy

The National Quality Strategy 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.

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[19] is built on the foundation of the National Quality Strategy, as described in the beginning of this section. Based on this foundation, CMS makes strategic choices in measure development and maintenance contracts as well as targeted measure selection for programs.

Input: Legislative mandates

CMS uses priorities mandated under several laws to drive its measure domains. Recently, Congress passed and the President signed the Medicare Access and Children's Health Insurance Program (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 both 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' quality measurement priorities, which have led to the broad payment reform and a quality-based payment model[20]. The transition is successfully underway and in 2015 Congress mandated that the quality reporting incentives will phase out in 2018, while the Merit-based Incentive Program (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 professionals. This supports 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 the inpatient rehabilitation facilities, hospices, long-term care hospitals, psychiatric units and hospitals, 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 measure activities. MACRA also supports the gains made under the ARRA. The ARRA launched a period of significant funding for the development of standards for electronic health records (EHRs) and the widespread adoption of meaningful use of EHR systems across providers. MACRA continues this with a mandate for widespread interoperability among these systems with requirements for CMS to develop metrics for successful interoperability and develop incentives and payment penalties to encourage rapid achievement of that goal.

Input: Patients, public, and other stakeholders

CMS conducts its measurement activities in a transparent manner. The information gathered through various methods described in the Stakeholder Input chapter informs HHS and CMS about future measurement needs. Additionally, MACRA of 2015 requires the CMS to solicit, accept, and respond to input from stakeholders including physician specialty societies, applicable practitioner

---

[19] 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: May 31,2015.

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

organizations, and other stakeholders for episode groups (i.e., care episode groups and patient condition groups). CMS and its contracted measure developers will take the lessons learned during measure development and maintenance gathered from public and stakeholder input and apply it to the planning of measurement priorities. 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 the Stakeholder Input chapter.

Input: 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 contract.

Measure maintenance reports yield information that CMS leadership may find valuable for setting priorities, such as 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 program evaluations to determine the effectiveness of its various programs. Many of these programs use quality measures and evaluate the usefulness of the measures as they are used in the programs.

The triennial national CMS Measures Impact Assessment 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 that is a 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:

- MedPAC and MacPAC—quality reports.
- Agency for Healthcare Research and Quality (AHRQ)—National Healthcare Quality and Disparities Reports.
- CMS Center for Strategic Planning—Chronic Conditions among Medicare Beneficiaries.

- Universities, researchers, and healthcare facilities—journal articles, 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 measurement 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. Refer to the Stakeholder Input chapter for 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 related or similar to those they are contracted 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 who 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 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 to determine 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 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 five categories:[21]
    - **Current Measures**—refers to measures that are currently implemented in one or more CMS programs and measures that have been finalized for implementation in one or more CMS programs.
    - **Previously Used Measures**—refers to fully developed measures that CMS no longer uses for a variety of reasons; for example, those that were retired or replaced.
    - **Proposed Not Finalized**—refers to measures that were listed in a proposed rule posted on the Federal Register but were not finalized at the time of the inventory posting.
    - **Pipeline Measures**—refers to measures which are currently in development, whether or not they have been included in a MUC list or in a proposed rule or Final Rule.
    - **Fully Developed Future**—refers to fully developed measures that CMS has yet to use in a CMS program.

[21] 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.

Chapter 3—Stakeholder Input

CMS measures are selected for implementation in quality initiatives that include public reporting, value-based purchasing, and other payment incentive and accountability programs. Stakeholder input is a key process to ensure that measures developed and maintained are effective for accountability, for quality improvement, and are useful to consumers. Gathering input from stakeholders is an effective way to develop measures that matter.[22] Transparency that includes input from those being held accountable is another key component, ensuring that the value added by the measure outweighs the burdens of collecting and reporting the data.

In the past, stakeholder input in terms of TEPs and public comment opportunities were specified as discrete events in a measure's lifecycle. However, in an increasingly connected environment and in keeping with Lean principles of eliminating rework, measure developers are encouraged to solicit input early in measure development and often throughout the lifecycle. In this way, measure developers can avoid wasting time and resources on measures that do not carry meaning for consumers and that are unduly burdensome for providers.

Many processes have been established to solicit input from the public and stakeholders throughout the measure development, selection, implementation, and maintenance cycle. These processes include:

- Posting calls for nominations for TEPs convened by the measure developers on the CMS MMS website.[23] Instructions for all postings to the CMS MMS website are included in Section 2, Chapter 1.A.—Information Gathering.
- Posting the proposed or candidate measures for public comment and posting the comments that were received on the CMS MMS website.[24]
- Posting candidate measures by NQF on their website during the endorsement process.
- Posting Measures under Consideration for public comment as part of the MAP input[25] to the pre-rulemaking process.
- Publicly soliciting comments on implemented measures that have undergone extensive changes during measure maintenance. Instructions are included in Section 2, Chapter 5—Measure Use, Continuing Evaluation, and Maintenance.
- Soliciting ongoing feedback on implemented measures. Instructions are included in Section 2, Chapter 4—Measure Implementation.

---

[22] Statement by Patrick Conway before Committee on Finance U.S. Senate. Wednesday, Jul 17, 2013. Available at: http://www.hhs.gov/asl/testify/2013/07/t20130717c.html. Accessed on: May 31, 2015.

[23] Centers for Medicare & Medicaid Services. *Measures Management System: Technical Expert Panels.* Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/TechnicalExpertPanels.html. Accessed on: May 31, 2015.

[24] 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: May 31, 2015.

[25] National Quality Forum. *Measuring Performance: Measure Applications Partnership.* Available at: https://qualityforum.org/map/. Accessed on: May 31, 2015.

In addition to the input obtained by measure developers, CMS provides opportunities for stakeholder input in the following ways:

- Publicly posting the MUC list by December 1 of each year as part of the pre-rulemaking process. Comments on the MUC list in preparation for the MAP deliberations are collected via an NQF-hosted website.[26]
- Engaging the NQF-convened MAP (a multi-stakeholder panel) to make recommendations on the MUC list (described in greater detail in Section 2, Chapter 4—Measure Implementation).
- Soliciting comments on proposed rules and publishing the final rules in the Federal Register.
- Holding CMS Open Door Forums and other provider information sessions about quality measures that are implemented in programs.
- Posting the Blueprint to inform the public and stakeholders about the processes that CMS measure developers use to develop and maintain quality measures. The Blueprint is accessible on the CMS Measures Management website.[27]

As public reporting expands, there is a need for enhanced involvement of persons and their family members in measure development, selection, and implementation. As used here, the term *person* refers to a non-healthcare professional representing those who receive healthcare. This version of the Blueprint provides new guidelines for collecting input from persons and their family members to ensure that measures reported publicly are of value and are useful to these individuals.

3.1—Role of Person and Family Engagement and Participation

In keeping with CMS's overall goal to assess the quality of healthcare in ways that are meaningful and person-centered, person and family member input must be integrated proactively during the measure development process. CMS has firmly asserted that listening to persons' voices, and those of their family members is a core mission of its quality strategy.[28]

Meaningful involvement of persons and their family members may occur throughout the measurement development process—from conceptualization to continuing evaluation, as appropriate. CMS measure development contracts specify that the person and family member perspective must be explicitly incorporated in measure development plans through participation in activities such as TEPs. To the greatest extent possible, this participation should involve actual persons and their family members, not just the organizations who represent them. As described below, additional person and family member engagement options are now recommended for use during the measure development process.

Persons and their family members are uniquely qualified to provide input regarding measure-related topics or concepts based on their experience. These individuals can help identify issues

[26] National Quality Forum. *Measure Applications Partnership.* Available at: http://www.qualityforum.org/map/. Accessed on: May 31, 2015.
[27] Centers for Medicare & Medicaid Services. *Measures Management System: Measures Management System Blueprint.* Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/MeasuresManagementSystemBlueprint.html. Accessed on: May 31, 2015.
[28] Centers for Medicare & Medicaid Services. *CMS Quality Strategy 2013 – Beyond. Nov 18, 2013.* Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy.html. Accessed on: May 31, 2015.

that are important to them and that they find useful to better understand the quality of their care. Measure developers should begin considering this perspective early in the information gathering process to maximize person and family member input into the measure topic selection.

Effective engagement of persons and their family members in the measure development process requires a well-formulated plan. Measure developers are encouraged to develop a person and family member engagement plan recognizing that input on technical measurement issues may not be a good use of persons' or their family members' time. This plan should describe how input from person and family member stakeholders will be incorporated during each of the five stages of measure development (*Measure Conceptualization*; *Measure Specification*; *Measure Testing*; *Measure Implementation*; and *Measure Use, Continuing Evaluation, and Maintenance*).

The Patient-Centered Outcomes Research Institute (PCORI) has developed a person Engagement Rubric. The rubric encourages researchers to consider how persons will be engaged throughout the measure development process.

Some of the concepts highlighted by PCORI that are applicable to person/family member engagement in the measure development process include:

- **Reciprocal Relationships**: Roles and decision-making authority of all involved are defined collaboratively and clearly stated.
- **Co-learning:** It is important to ensure that person stakeholder partners understand the measure development process and that all participants understand person 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.

3.2—Enhancing Person and Family Member Engagement

CMS strives to adopt new and more inclusive methods for explicitly incorporating the person perspective in the measure development process. CMS would like measure developers to consider additional engagement options beyond inclusion in a TEP to ensure that persons and their family members have readily accessible opportunities for involvement throughout the process.

With person- or family-member-specific forms of engagement, developers can first learn what persons or their family members would like to know and what information they need to make informed decisions before beginning the development process.

Options for Person and Family Member Engagement

CMS supports increased person/family member engagement by instituting new requirements for their inclusion in measure development. This section includes a menu of options for enhanced person and family member inclusion.[29]

New measure developers should work to make persons/family members more central to the process by incorporating at least two of the options below. Current measure developers are encouraged to incorporate what they can of the following options into their current plans pending existing deadlines and contract obligations.

Person or Family Member Advisory Panels

In the context of measure development, a *person or family advisory panel* is a group of individuals who voluntarily share their thoughts on measure-related topics for the purpose of informing measure-development efforts. Advisory panels provide developers with input reflecting the person/family member perspective that developers can subsequently use to shape the measure(s) they are constructing. Developers should first provide participants with materials that describe in lay terms what each proposed measure is intended to communicate. Panelists can then engage in informed discussions on how they would potentially use the measure.

Advisory panels can be conducted either in person or through a web-based format such as WebEx. Virtual web meetings allow measurer developers to reach participants in various parts of the country representing diverse viewpoints and population segments without the need for travel. Persons or family members can participate from the comfort of their home, making them more likely to take part and eliminating significant burden on them.

Person or family member advisory panels can be conducted via three approaches:

1. *Person (or family member)-only TEPs* in which participants discuss issues covered by subject matter expert panels but in a forum that is less formal and in which participants are more similar to them.
2. *Traditional focus groups* in which a facilitator guides a group of 3-9 persons or family members through a discussion.
3. *Condition-based groups* involving guided discussions among persons who have the health condition relevant to the measure under development.

It is recommended that panels be composed solely of <u>either</u> persons <u>or</u> family members because each of these groups is likely to have a different perspective on the discussion topics.[30]

Virtual meeting technologies and new recruitment options make it feasible to have person/family member-specific panels for measure development without extended timelines or cost. Even small

[29] Many of the person/family engagement options presented were recommended by the Physician Compare group led by A. Hovatter and D. St. Clair based on their experiences working with persons and family members, 2015.

[30] K. Campbell, Person and Family Engagement in the Measure Development Process. Panel presentation at the CMS Quality Conference. Baltimore, MD, December 2015.

groups of up to 9 persons/family members can provide measure developers with extensive information that they can use to draw out common themes to inform measurement development, selection, and implementation.

Person/Family Member of the Technical Expert Panel

The Institute for Patient and Family Care recommends including multiple patients on a panel, noting that including a single patient is not usually successful.[31] Including multiple persons or family members on a TEP reduces the likelihood that an individual will feel isolated or intimidated during their TEP participation experience. Therefore, expert panel composition should include at least two individuals that represent the person/family member perspective. [32]

Concept Testing

Concept testing is the process of evaluating consumer response to measurement-related topics before they are translated into fully formed measures. The purpose of concept testing is to evaluate person/family member interest in and understanding of possible measure ideas prior to deciding which measures to develop. Concept testing can also be used during the development process to determine whether consumers would use a particular measure for making healthcare-related decisions.

Cognitive Testing

The purpose of cognitive testing within the measure development context is to evaluate persons'/family members' understanding of measure concepts and definitions. Cognitive testing involves presenting individuals with measure-related definitions and concepts and asking them to interpret the terms in their own words. This technique is particularly useful during the formative testing process and for appraising measures that are designed to be patient-reported measures.

Plain Language Testing

In plain language testing, measure developers investigate whether the language used in measures and measure descriptions leads to accurate understanding and interpretation by the average person/family member and is not too technical to be valuable. This is particularly useful for evaluating measures planned for public reporting.

Engagement in a Text-based Virtual Community

A website community can promote discussion and commentary among persons/family members about measure development through use of focused questions and topic threads. This technique may provide valuable insight into persons' or family members' viewpoints. At all points in the measure development lifecycle, persons/family members can be engaged in the online panel to review and comment on information related to the measure and its development. An advantage

[31] 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.

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

of the virtual community is that it presents low burden to participants and to measure developers. However, this technique should only be used when individuals can reasonably be expected to participate given their potential literacy, socioeconomic, or technology-related constraints. Another caveat is that text-based virtual community discussions may not yield responses that are representative of the person or family member populations at large.[33]

Table 3 displays possible person/family engagement options indexed by measure development stage. Persons/family members should be engaged throughout the measurement development cycle; however, some engagement options may be more appropriate for particular stages of development.

Table 3: Options for Person and Family Member Engagement by Measure Development Stage

Measure Development Stage	Options for Person and Family Member Engagement					
	Person Advisory Panel	Person member(s) of TEP	Cognitive testing	Plain language testing	Concept testing	Virtual community/online panel
Measure conceptualization	✓	✓		✓	✓	✓
Measure specification	✓	✓	✓			✓
Measure testing	✓	✓				✓
Measure implementation	✓	✓	✓			✓
Measure use, continuing evaluation, and maintenance	✓	✓	✓			✓

[33] K. Campbell, personal communication. December 2015.

3.3—Person/Family Member Engagement Best Practices

In a recent survey of persons and family members who participated in expert panels for measure development, most stated that they would do it again and would recommend participation to other persons and consumers. They felt that they were listened to by the other experts and that their input was valued. Though the practice of including persons on these kinds of panels is relatively new, the people who have been included were glad they were consulted.[34]

Reaching Persons and Family Members

There are increasingly diverse options for reaching persons and family members given the capabilities of social media. However, even with these new opportunities, it can still be a challenge to find individuals who are willing and able to participate in measure development. Recruitment strategies used to find other experts (such as the TEP recruitment strategies listed in Section 2) may be used, but other sources and other methods may also be required. The following list includes some possible recruitment approaches.

Examples:

- 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 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[35] or the Michael J. Fox Foundation for Parkinson's Research[36]) 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)[37] have person engagement representatives who are experienced mentors and know of persons who are able to participate.

The following websites include consumer advocacy organizations and support groups which 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

[34] Unpublished report from NQF Kaizen Stream 1 Patient and Consumer Participation survey. Feb 2014
[35] American Diabetes Association. *Advocacy.* Available at: http://www.diabetes.org/advocacy/. Accessed on: May 31, 2015.
[36] Michael J. Fox Foundation for Parkinson's Research. Available at: https://www.michaeljfox.org/. Accessed on: May 31, 2015.
[37] Patient-Centered Outcomes Research Institute. *Advisory Panel On Patient Engagement.* Available at: http://www.pcori.org/get-involved/advisory-panels/advisory-panel-on-patient-engagement/. Accessed on: May 31, 2015.

- Patient Voice Institute
- AgingCare.com
- Caring.com
- Connecticut Center for Patient Safety
- Daily Strength
- Informed Medical Decisions Foundation
- MD Junction
- Med Help
- Patients Like Me
- CMS Quality Measures Public Comment
- People For Quality Care
- National Quality Forum

Various social media options are also available to recruit persons/family members who are willing and able to inform measure development and maintenance. The websites listed above and similar sites often include contact information including social media sources that can be included. Social networking pages including Twitter, Facebook, and numerous other social media hosts can also be used to recruit persons/family members. These forms of recruitment are low cost and can be very effective. Since the use of social media for recruitment is still somewhat new, measure developers should review the plan with their COR to ensure the message is appropriate to the requirements of the contract.

Maximizing Input and Minimizing Burden

It is important to engage stakeholders in a manner that enables them to provide meaningful input in the measurement development process, and not simply to involve persons or family members as a "check the box" requirement. To engage persons/family members effectively it is important that the measure developer carefully consider the types of concerns these individuals may have related to healthcare quality and ensure they are engaged in conversations that are meaningful to them.

Example: Organize specially designed workgroup call(s) or individual phone call(s) for persons/family members to obtain their input separately.

Timing It Right

Measure developers also should be mindful of when it is most meaningful to include person or family member stakeholders in the measure development process, and establish meeting agendas and attendance requirements accordingly. Past experience indicates that often it is most meaningful to engage person or family member stakeholders early in the measure development process, during the more formative stages of measure development, such as measure conceptualization. Person/family engagement can be helpful during this phase to understand what health and quality of care issues are of greatest concern and understand what these individuals feel is worth measuring and evaluating.

Example: 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 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 in 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.

Making Person/Family Member Participation Accessible

Persons/family members may have time and health limitations that make attending in-person meetings too burdensome. Stakeholders' schedules and transportation constraints should be taken into consideration when planning the times and locations of the meetings. Some stakeholders may require reimbursement for transportation costs in order to participate in on-site meetings. Measure developers should ensure that person/family members' engagement opportunities (for example, TEP or advisory panel meetings) are conducted in a manner such that participants who are not able to attend in person can still participate in the discussion. This can be achieved by offering options such as a conference call dial-in number and/or web-based connections to meetings and by providing web-based options for interaction beyond these meetings, such as through virtual communities.

Example: A measure developer will be holding 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 online via web-based meeting software.
- Guaranteed parking and mileage reimbursement for those that choose to drive to the meeting.

Incorporating Input

In the event that person or family member feedback is solicited outside of the formal TEP meeting, it is important that this information is disseminated at the formal TEP meeting. Presenting information gathered from persons or their family members outside of the formal TEP should be included on the formal agenda, and the measure developer is responsible for sharing key findings at the TEP meeting.

Example: Persons' input was incorporated into the Business Case that was reviewed by the TEP prior to prioritizing concepts for development.

Communicating Transparently

It is the responsibility of the measure developer to ensure that all discussions that include person or family member stakeholders are conducted so that laypersons can understand and engage in the conversation. Technical jargon should be minimized, and in instances where it is unavoidable the measure developer should make sure that the term is explained. Meeting materials should use plain language with a readability grade level that matches the literacy of the participants.

Example: A TEP meeting that includes person stakeholders is scheduled to take place next month. The measure developer required that all PowerPoint slide presentations be submitted in advance of the meeting for review. During the review process, plain language testing was conducted with a handful of person representatives to ensure that the materials could be understood by a lay audience.

Meeting Planning

Not much literature is available specifically on the inclusion of persons or their family members in measure development. However, a recent survey of a limited number of persons who participated in TEPs revealed that participants wanted more pre-meeting preparation.[38] Resources are available on preparing persons to participate in other healthcare quality activities such as advisory boards or hospital governance boards. A best practice identified by organizations for these other types of involvement has been to assign a mentor to lay participants.

Examples:

- Inform potential person/family member participants during recruitment about the time commitment requirements and the nature of the input being sought from them.
- Assess their ability to participate in on-site or in-person meetings as well as teleconferences, virtual web meetings, or virtual communities.
- Assure that any materials given to lay participants are clearly understandable and they are given ample opportunity to review the materials and ask questions.

Confidentiality

When persons or family members are involved in TEP or advisory panel deliberations, extra confidentiality considerations may need to be taken. The measure developer should 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. The guidance on confidentiality developed and contained in the Blueprint is provided with the understanding that the authors and publishers are not rendering legal advice. If special

[38] Unpublished report from NQF Kaizen Stream 1 survey. Feb 2014.

confidentiality arrangements or other releases are required, the measure developer should seek appropriate legal counsel.

Example: In developing the TEP Nomination Form the measure developer included language informing participants that meeting notes would be made publicly available, and if a person shared input during a TEP that information may be included in the notes, including health history-related details. The form also indicated that persons had the right to request that their name be kept confidential.

Meeting Support

As noted above, assigning a mentor or facilitator to support person or family members assisting with measure development activities is recommended. The mentor or facilitator can welcome participants and serve as a primary contact for questions, concerns, or points of clarification that may arise during the meeting or testing session. Similar support would be beneficial in all person or family member advisory groups, focus groups, and one-on-one testing.

Example: Each family member participating in a TEP convened to gather input on several functional status quality measures was assigned a mentor who called each participant 2 days before the TEP to introduce themselves and answer questions about the meeting materials that had been mailed to participants. On the day of the TEP, the mentor welcomed each participant to the group and answered their questions during the meeting breaks.

The following best practices are adapted from a toolkit created by the Institute for Patient and Family Care.[39] Before the meetings, facilitators can:

- Assess and resolve any physical accessibility needs.
- Help with meeting logistics, such as reminding participants of the date and time of the meeting 1-2 days prior to the meeting, providing directions to on-site locations, and welcoming and escorting the participants to the meeting room when they arrive.
- Help with technology needs, such as email/ internet access, web conferencing, and teleconference meetings.
- Provide printed materials in advance if required by participants without email or Internet access.
- Review meeting materials, in advance, with person or family participants to clarify terms or other concepts as needed.

During the meetings, facilitators can:

- Ensure that introductions clarify the purpose of the meeting and the role that each participant will play.
- Ensure that jargon or technical terms are explained in lay language.
- Continue assisting with technology needs for virtual or teleconference meetings, as needed.

[39] 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.

- Ensure participants feel comfortable participating in the discussion and stress that everyone's input is important.
- Convey that the group should hear and respect each participant's perspective.
- 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.

After the meetings, facilitators can:

- Debrief the person or family member participants, ensuring that they know their input was valued.
- Listen to the participants' suggestions to improve their experience.

Innovations

During an NQF-sponsored Kaizen event in fall 2013, patient engagement in the measure development process was one of four key initiatives identified. In keeping with Lean principles, recommendations from that event included creating a training program, and developing a pool of stakeholders and persons who could serve on measure developers' expert panels. NQF is also considering adding patient input to support the importance criterion when evaluating measures. Further resources provided by this and other groups adding to resources for person or family involvement will be included in future updates to the Blueprint.

3.4—Role of the Technical Expert Panel in Stakeholder Input

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 subject matter experts (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.

Balanced Membership

The TEP's input helps the measure developer ensure transparency during measure development. The TEP process is designed to solicit balanced, multi-stakeholder input. The following people (in addition to those discussed previously) should be considered when recruiting a balanced TEP:

People with clinical content expertise:

- Practicing clinicians
- Facility administrators
- Provider organizations—specialty societies, trade associations
- Clinical researchers

People with performance measurement expertise:

- Experts in quality improvement
- Methodologists, experts in quality measurement
- Statisticians
- Persons with expertise in healthcare disparities

People with medical coding and informatics expertise, especially for EHR-based measures:

- Informaticists
- EHR experts
- Programmers
- EHR/Health IT Vendors

Conflict of Interest

As part of transparency, potential conflicts of interest should be considered as the TEP membership is assembled. Though the balance brought by forming a group with diversity of perspectives helps, it is also important that the perspectives are made "explicit" so the group can consider each individual's contributions appropriately. Potential for conflict of interest is not solely a reason to exclude an individual from participation on a TEP, because the membership should also be balanced with applicable points of view and backgrounds. The measure developer should, however, give preference to individuals who will not be inappropriately influenced by any particular special interest.

The measure developer should inform CMS of a TEP membership nominee with potential conflicts of interest to determine if the person may be selected to serve on the TEP. If the expertise of the nominee is needed and the decision is made to include the candidate on the TEP, the potential conflict of interest is then disclosed at the beginning of the TEP meetings. This allows all parties to be aware of potential biases that may arise during discussions. The measure developer, the COR, the TEP chair, and other TEP members will then need to decide how to mitigate the effect of that individual's interest or relationships during discussions or recommendations. For example, the individual TEP member with a potential conflict may be recused from voting on a measure that would encourage the use of a product for which a member might have a financial interest while allowing the member to vote on other measures for which no conflict exits.

Because TEPs may meet multiple times over an extended period during measure development and maintenance, it is necessary to repeat the opportunity for disclosure of conflict of interest before each TEP meeting. TEP members should also be informed that if a potential conflict of interest arises during the period, the member must bring it to the TEP chair's and measure developer's attention immediately.

Some of the methods the measure developer may use to publicize the Call for TEP include, but are not limited to:

- Posting the notice on the MMS TEP page.[40]
- Posting the notice on the program-specific CMS site.
- Announcing the notification during appropriate CMS workgroup or Open Door Forum calls and sending the notice to the related distribution list.
- Sending notice via email to the stakeholders' email lists or having the stakeholder organizations post a notice on their websites.
- Recruiting on patient support group websites and other consumer organizations as discussed earlier in this chapter.
- Using social media (e.g., Twitter, Facebook, YouTube, and LinkedIn). Contact the COR for the appropriate process.

Timing of TEP input

If the TEP is to be held early during the contract period, then the measure developer should post a call for the panel as soon as the contract is awarded. The Call for TEP is usually done concurrently with the environmental scan, literature review, and other tasks so that the findings are available to present at the TEP meetings. Occasionally, it may be necessary to convene a smaller, more focused group of subject matter experts SMEs than the entire TEP to provide specific expertise (e.g., on technical aspects of coding measure specifications). These SMEs can provide information on feasibility to the larger TEP for consideration.

TEP input should be considered at the following points during the measure lifecycle:

Measure conceptualization

- Information gathering—to give input on topics and importance
- Refining the candidate measure list
- Applying the measure evaluation criteria to the candidate measures
- Feasibility assessment, especially for eMeasures

Measure specification

- Constructing technical specifications
- Risk adjusting outcome measures

Measure testing

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

[40] 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: May 31, 2015.

Measure implementation

- Responding to questions or suggestions made by NQF Steering Committee

Measure use, continuing evaluation, and maintenance

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

For most measure development contracts, measure developers will convene several TEP meetings, either by teleconference or face-to-face. During early TEP meetings, the members will be asked to review the results of the environmental scan and clarify measure concepts. They will be asked to evaluate the list of potential measures and narrow them down to candidate measures. During subsequent meetings, the TEP will be asked to 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 that can be reviewed by a TEP. During maintenance, TEPs should also compare measure performance to the business plan predictions of impact on quality. Details of the procedures for TEP involvement in comprehensive reevaluation, annual updates, and ad hoc reviews are included in Section 2, Chapter 5—Measure Use, Continuing Evaluation, and Maintenance.

In addition to developing measures that address measurement gaps, the measure developer should keep an overall vision for discerning the breadth of quality concerns and related goals for improvement. The TEP should be directed and encouraged 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 measure developer should be able to show how the recommended measures relate to overall HHS goals including the National Quality Strategy priorities, CMS Quality Strategy, measurement priorities, and relevant program goals.

Including a patient or caregiver representative on the TEP roster is an effective way to ensure input on the quality issues that are important to patients. TEP participation by consumer and patient advocacy organizations may be desirable but should not be used in lieu of actual patients. In addition to the best practices listed in this section, procedures for recruiting, choosing, and convening TEPs are detailed in Section 2, Chapter 1.B.—Technical Expert Panel**.**

3.5—Role of Public Comment in Stakeholder Input

The public comment process is an essential way that CMS ensures that 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 by the TEP. Public comments obtained during measure development (and maintenance) are separate from, and complement the public comment obtained during the NQF endorsement process, as described in Section 2, Chapter 4.1—National Quality Forum Endorsement.

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. Early and frequent opportunities for public comment during measure development are a best practice because the measure developer can address the issues raised by public comments 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—commenting 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 report[41] is also posted for public comment.
- Public comment opportunities are part of the NQF Consensus Development Process.[42]
- Proposed federal rules are posted for public comment.
- Federal Register Notices are posted for public comment.

[41] 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.

[42] National Quality Forum. *Consensus Development Process*. Available at: http://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process.aspx. Accessed on: May 31, 2015.

- Feedback can be obtained during CMS listening sessions, Open Door Forums,[43] Special Open Door Forums,[44] 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.[45]
- Summaries of TEP meetings held during measure maintenance are posted for public comment.

Consideration when soliciting public comments

The Paperwork Reduction Act (PRA) mandates that all federal government agencies must obtain approval from the Office of Management and Budget (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. Measure developers should consult with their COR regarding the PRA to confirm if OMB approval is required before requesting most types of information from the public. HHS also maintains an additional website with frequently asked questions and answers about the PRA.[46]

Federal Rulemaking

The federal rulemaking process also includes a public comment period. The public comment period during rulemaking is a time when CMS receives feedback on its measures, because most CMS quality programs are included in rulemaking. However, the federal rulemaking process should not be the only time when public comments are sought and addressed. The federal rulemaking process usually occurs after the measure is developed and is being proposed for implementation; therefore, the measure developer misses 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.

[43] https://www.cms.gov/Outreach-and-Education/Outreach/OpenDoorForums/index.html?redirect=/OpenDoorForums/01_Overview.asp.
[44] https://www.cms.gov/Outreach-and-Education/Outreach/OpenDoorForums/ODFSpecialODF.html.
[45] National Quality Forum. *Quality Positioning System.* Available at: http://www.qualityforum.org/QPS/QPSTool.aspx. Accessed on: May 31, 2015.
[46] Department of Health and Human Services, Office of the Chief Information Officer. *Frequently Asked Questions About PRA/Information Collection.* Available at: http://www.hhs.gov/ocio/policy/collection/infocollectfaq.html. Accessed on: May 31, 2015.

Chapter 4—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. Procedures for completing harmonization in each phase of the measure lifecycle are discussed throughout Section 2 whenever they apply.

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.[47]

Measure alignment is defined as, "Encouraging the use of similar standardized performance measures across and within public and private sector efforts."[48] 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.[49] A parsimonious core set of measures increases "signal" for public and private recognition and payment programs.[50] 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

[47] 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.

[48] Ibid.

[49] 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.

[50] 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.

the achievement for population health across the multiple settings. If functional status measurement is harmonized and the measure use aligned across CMS programs, it would be possible to compare gains across the continuum of care. Consumers and payers are enabled to choose based on measures calculated in similar ways. In these (and other) ways, harmonization promotes:

- Coordination across settings in the continuum of care.
- Comparisons of population health outcomes.
- Clearer choices for consumers and payers.

The CMS Quality Strategy[51] (as it implements the National Quality Strategy) provides a framework for alignment by focusing development on the six priorities and goals. More detail on the CMS and National Quality Strategies is located in Section 1. All quality measures chosen for use by CMS should address one or more of the goals of the CMS Quality Strategy. To accomplish these goals, measure developers should consider both harmonization and alignment throughout the measure lifecycle; and whether to adapt 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.

The following table summarizes ways to identify whether measures are related, competing, or new, and indicates the appropriate action based on the type of harmonization issue.

[51] 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: May 31, 2015.

Table 4: 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.
Numerator: Same measure focus Denominator: Different target population	Related measures	▪ Harmonize on measure focus (adapted), or ▪ Justify differences, or ▪ Adapt existing measure by expanding the target population.
Numerator: Different measure focus Denominator: Same target population	Related measures	▪ Harmonize on target population, or ▪ Justify differences.
Numerator: Different measure focus Denominator: Different target population	New measures	▪ Develop measure.

Procedures for harmonization at each stage of the measure lifecycle are included in the respective chapters of Section 2.

4.1—Adapted Measures

If the measure developer changes an existing measure to fit the current purpose or use, 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. Or, 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 re-specified and tested. Before deciding to adapt 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 a measure is to assess the applicability of the measure focus to the population or setting of interest. Is the focus of the existing measure applicable to the quality goal of the new measure population or setting? 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. Chapter 2—Measure Specification, in Section 2, describes the standardized process for developing sound technical specifications.

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.

4.2—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.

4.3—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[52] (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 Chapter 1.A—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. Refer to the Information Gathering chapter (Section 2, Chapter 1.A.) for details of the procedure.

[52] 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.

Chapter 5—Concepts of 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.[53] 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. The self-evaluation report documents for CMS the extent to which the measure meets the criteria. This report is usually required by the CMS Measure & Instrument Development and Support (MIDS) contract and is called the Measure Evaluation Report in that contract. 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 October 2013.[54] A Measure Evaluation Report template is available in Section 4—Tools, Appendices, and Forms.

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.

[53] The criteria and evaluation guidance given in the Blueprint align with the NQF criteria and guidance.
[54] 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.

- Conducting comprehensive reevaluation.

5.1—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 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,[55] 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.

The Blueprint provides general guidance for applying these evaluation criteria to individual measures. It also covers specific types of measures such as composite performance measures; eMeasures; patient-reported outcome-based performance measures (PRO-PM); and cost and resource use measures. There may be minor differences in evaluation criteria application for these

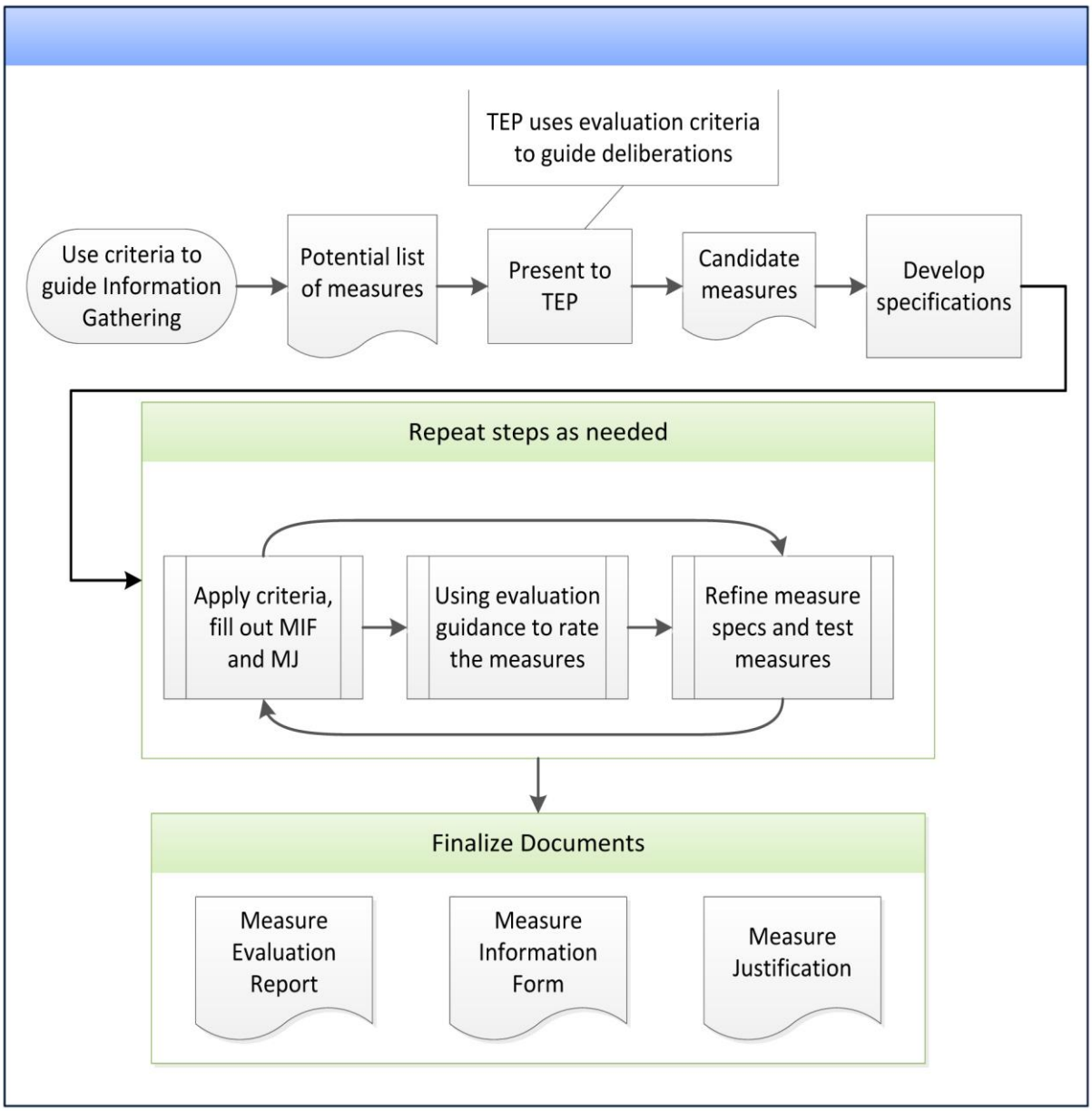
[55] National Quality Forum. *Measure Evaluation Criteria and Guidance Summary Tables.* National Quality Forum; Revised Oct 2013. Available at: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=73365. Accessed on: June 1, 2015.

special measure types, but these are noted wherever appropriate in the Measure Evaluation Criteria and Instructions.

5.2—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. Figure 8: Applying Measure Evaluation Criteria diagrams the process of applying the measure evaluation criteria.

Figure 8: Applying Measure Evaluation Criteria



Measure evaluation criteria are applied:

- 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, Measure Information Form, and Measure Justification Form.[56]

5.3—Report of the Measure Developer's Evaluation of the Measures

A sample Measure Evaluation Report blank template, a full description of the criteria in the Measure Evaluation Criteria and Instructions, and links to the NQF measure evaluation guidance document[57] are provided in Section 4. The report is designed to help the measure developer apply the measure evaluation criteria on an ongoing basis and report the measure evaluation ratings in a formal, standardized way at specific stages of measure development and maintenance. The Measure Evaluation Report is developed by the measure developer based on the measure developer's application of the measure evaluation criteria to measure information documented in the Measure Information and Measure Justification Forms, during measure testing activities, and TEP input; and from any public comments received.

Timing of Measure Evaluation Report (measure developer's self-evaluation)

In addition to the three formal reports (indicated below), consider evaluating measures and completing a Measure Evaluation Report during the following 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

- 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

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

[56] 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: June 1, 2015.
[57] Ibid.

Measure Testing

- Apply the evaluation criteria when analyzing test results.
- Review updated measure specifications and justification updated according to the evaluation criteria (formal report required).

Measure Implementation

- 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

- 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 eMeasures, composite measures, and cost and resource use measures.

5.4—Developing and Evaluating the Business Case

There is significant potential for well-designed and strategically selected measures to enhance the quality of healthcare delivered to consumers. Reporting the results of these measures also makes the healthcare system more transparent to consumers and payers. Prudent choices, the responsibility of the measure developer in teamwork with CMS, during development and selection of quality measures will support national quality initiatives. The National Quality Strategy refers to measurement and public reporting as two of the nine levers that help implement the strategy.[58]

[58] 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: June 1, 2015.

The CMS Quality Strategy also lists "measuring and publicly reporting providers' quality performance" among the drivers and policy levers of quality.[59]

Though the contract deliverable is called a business case, it is not limited to a description of economic benefits. The business case documents the anticipated impact of a quality measure by predicting the return on the investment of resources into developing and implementing measures. Returns on investment apply to lives saved, costs reduced, complications prevented, clinical practice improved, and patient experience enhanced. The levers named in the CMS Quality Strategy drive choices that address this broad set of priorities. The business case supporting these choices should be made as clearly and strongly as possible by the measure developer developing the measures. 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. For example, to reduce mortality through early detection and treatment, there may be increased costs and potential complications of screening tests.

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 contractor 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 return on investment in the particular measure.

The business case for a measure applies information gathered, as well as supports the measure importance evaluation criterion by providing supplementary information to create a model that predicts performance of the measure and the impact it will have on health and financial outcomes. The formal business case for a measure supports measure evaluation during its initial development and facilitates reevaluation during measure maintenance. 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. The importance criteria in the Measure Evaluation Report and in the NQF Measure Evaluation Criteria[60] contain requirements for information that will be used to begin a business case. The

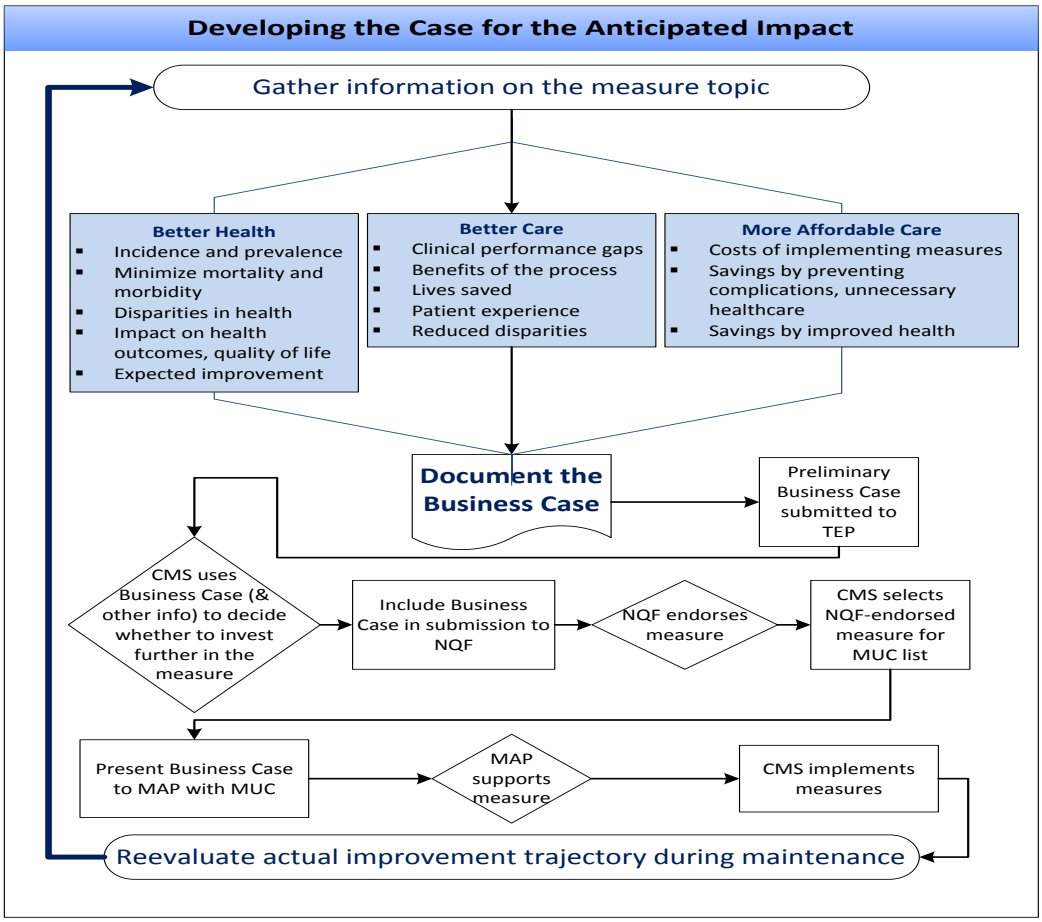
[59] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *CMS Quality Strategy 2013 — Beyond*. Nov 11, 2013. Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy.html. Accessed on: June 1, 2015.

[60] National Quality Forum. *Measure Evaluation Criteria.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards/Measure_Evaluation_Criteria.aspx. Accessed on: June 1, 2015.

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.

Figure 9: Inputs and Uses for the Business Case diagrams business case inputs and the impact of the business case throughout the measure lifecycle.

Figure 9: Inputs and Uses for the Business Case



The business case should be evaluated during measure development and maintenance. Evaluation of the strength of the business case is ongoing during measure development and 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.

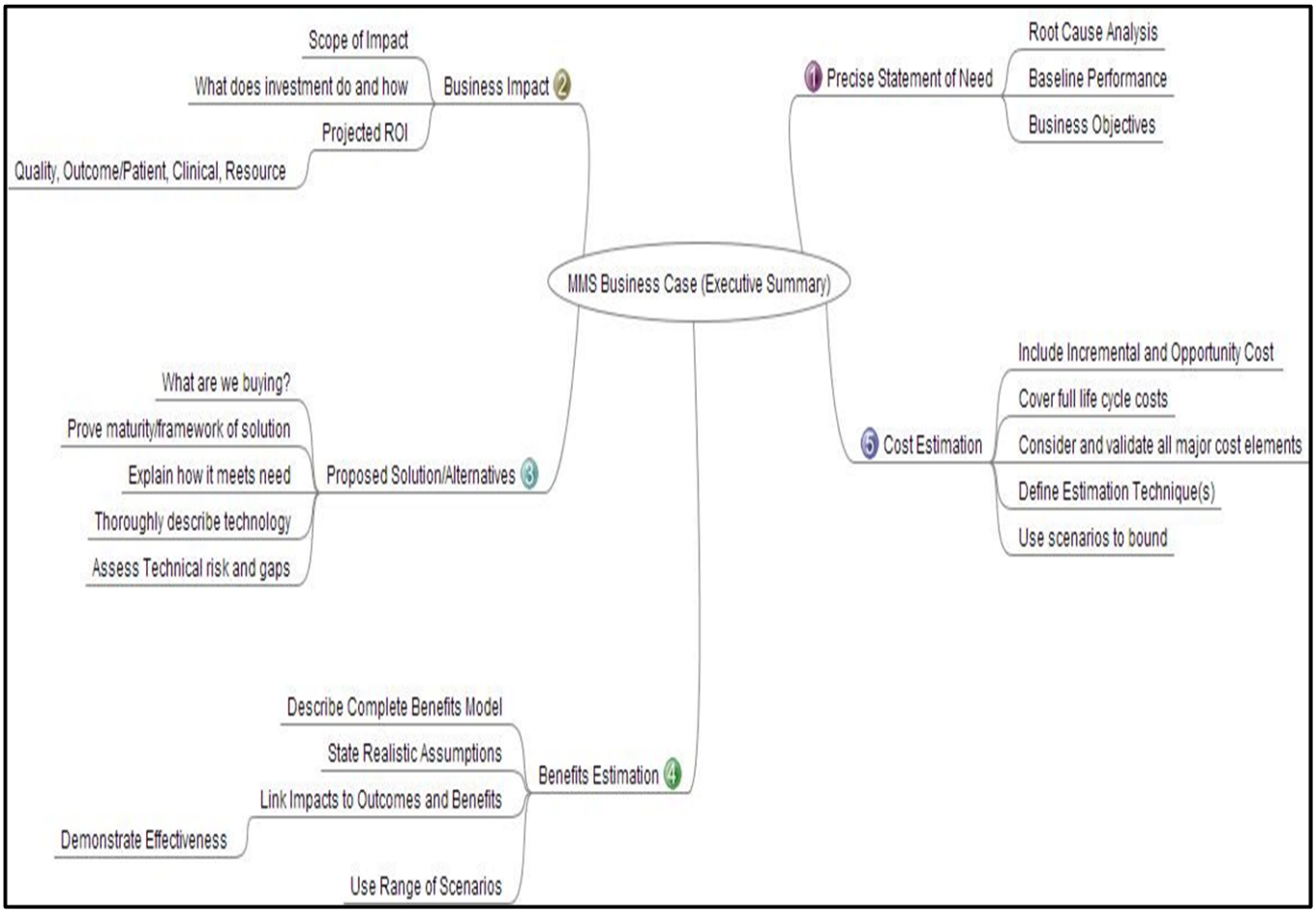
The business case and the predictions about measure performance used to inform decision making during measure development and selection for use should be compared against actual performance after the measure is implemented. If anticipated improvements in health, provider care performance, and increased 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.

MMS Business Case Best Practices

There are five key elements in a well-constructed business case, as shown in Figure 10: MMS Business Case Best Practices).

Figure 10: MMS Business Case Best Practices



The executive summary should focus on what is available and concisely (maximum of 300 words) provide a high-level overview of the following:

- **Precise Statement of Need**
  - Concise statement of the problem being solved
  - Leverage information gathering report
  - Outcomes of gap analysis between current processes and identified best practices
  - Baseline performance assessment with proposal of how to close gaps
  - Failure modes, effects, and criticality analysis/root cause analysis recommendations in case of failures
  - Impact on the “measured” and patients subject to care

- **Business Impact**
  - What is the actual business case, what is the purpose?

  - What is benefit vs the toll? [61]
  - Scope of Impact
  - State added value in allocating additional resources
  - Projected return on investment (ROI) expressed as improved quality, outcome/patient, clinical, resource
  - What are the risk mitigation approaches?

- **Proposed Solution/Alternatives**
  - Comprehensive statement of the problem
  - What is the current solution?
  - How does the proposed solution meet the need?
  - Prove maturity/framework of measure solution, show linkage(s) to outcomes (if not an outcome measure)
  - Explain how solution meets need/fills gaps
  - Assess technical risk and gaps

- **Benefits Estimation**
  - Describe complete benefits model
  - State realistic assumptions including "human toll" (including data collection burden)
  - Link impacts to outcomes and benefits
    - Benefit/Cost analysis if monetized
    - Effectiveness/Cost analysis if counted
    - ROI analysis if can be monetized in terms of human toll (i.e., is the improvement worth the burden)
  - Use a range of scenarios

- **Cost Estimation**
  - Define estimation technique(s)
  - What will it cost the government in both cost and schedule
  - Include incremental and workflow interruption or smoothing cost
  - Cover full life cycle costs, such as development cost, implementation costs, data collection, maintenance and upgrade costs
  - Consider and validate all major cost elements (e.g., human resources, technology)
  - Use scenarios based on measure implementation to bound cost estimating

5.5—Evaluation Considerations 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.

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.[62] Composite performance measures fall into two main types:

- Measures with two or more individual performance scores combined into one score for an accountable entity are considered a composite.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity are also considered composites.

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.

General considerations when 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:
- 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.

[62] 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: June 1, 2015.

- 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.[63]

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.

The following are general considerations related to evaluation of 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.

[63] Ibid.

- 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 a chronic obstructive pulmonary disease (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.

Evaluating eMeasures

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 without the burden of data collection. Because clinical data are entered directly into 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, however, still must meet all of the evaluation criteria as any other measure. Detailed procedures specific to eMeasure evaluation are discussed in the eMeasure Lifecycle section.

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*,[64] 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 patient-reported outcomes 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 have to be responsive to specific healthcare services or intervention.

[64] 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: June 1, 2015.

- 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.

5.6—Overview of Measure Evaluation Tools

Measure Evaluation Criteria

The full set of evaluation criteria and sub-criteria is included in Section 4—Tools, Appendices, and Forms. These criteria are appropriate for all types of measures. Evaluation criteria descriptions for composite performance measures, cost and resource use measures, eMeasures, and PRO-PMs are included where they vary from the general guidance. The specifics for each type of measure are labeled with headings.

Guidance for applying the criteria

NQF developed a document with guidance for evaluating the criteria as applied to specific measure attributes that is available on the NQF Measure Evaluation website.[65]

Measure Evaluation Report Template

This template provides a structure for documenting the formal review of each measure against the appropriate set of evaluation criteria. The template includes space for evaluation of all types of measures. Form spaces not relevant to the specific measure may be deleted. The instructions at the beginning of the document should also be deleted when the document is submitted.

Business Case Form Instructions and Template

The Business Case Form Instructions and Template provide 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. It also 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.

[65] National Quality Forum. *Measure Evaluation Criteria.* Available at: http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx. Accessed on: June 1, 2015.

Chapter 6— Measure Classification and Selected Specific Measure Types

Section 1 of the Blueprint has covered concepts that apply throughout the standardized Measures Management System for developing and maintaining the quality measures. The concepts around priorities planning, stakeholder input, harmonization, and measure evaluation apply to all aspects of the measure lifecycle. It is intended that measure developers use standardized processes (as described in Section 2—Measure Lifecycle) when developing quality measures. Measure developers should follow these processes as closely as possible because they are designed to systematically guide the development of high-caliber measures suitable for NQF submission and CMS implementation.

However, some categories of measures and topics are conceptually different, including cost and resource use measures, composite performance measures, patient-reported outcome measures, and multiple chronic conditions measures.

This chapter defines the measure classification process as well as general approaches to selected measure types when they vary from other measures. Specific procedures for development and maintenance of these measures are covered in Section 2 at the procedural point where they apply.

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 electronic health record 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.

6.1 – Measure Classification

Measures may be classified according to a variety of schemes, including measurement domain, by the National Quality Strategy priority(ies) addressed, and by measurement setting. Elements of these classification schemes and examples are provided in Tables 5 and 6. A list of measurement settings is shown in Figure 11.

Table 5: NQMC Clinical Quality Measure Domains[66]

Measurement Domain	Definition	Example
Process	A process of care is a health care-related activity performed for, on behalf of, or by a patient. Process measures are supported by evidence that the clinical process—that	The percentage of patients with chronic stable coronary artery disease (CAD) who were prescribed lipid-lowering therapy

[66] National Quality Measures Clearinghouse. http://www.qualitymeasures.ahrq.gov/about/domain-framework.aspx. Website accessed 10/21/2015.

Measurement Domain	Definition	Example
	is the focus of the measure—has led to improved outcomes. These measures are generally calculated using patients eligible for a particular service in the denominator, and the patients who either do or do not receive the service in the numerator.	
Access	Access to care is the attainment of timely and appropriate health care by patients or enrollees of a health care 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 health care. 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 health care 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 health care organization or clinician related to the capacity to provide high quality health care. Structure measures are supported by evidence that an association exists between the measure and one of the other clinical quality measure domains. These measures can focus on either health care organizations or individual clinicians.	Does the health care 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)?

Measurement Domain	Definition	Example
Patient Experience	Experience of care is a patient's or enrollee's report of observations of and participation in health care, 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.

Table 6: Examples of Measures Addressing Each of the National Quality Strategy Priorities[67]

National Quality Strategy (NQS) Priority	Examples of Measures Addressing Each NQS Priority
Effective Communication and Care Coordination	Adult depression in primary care: percentage of patients who have a follow-up contact within three 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. Care coordination: percentage of children who needed care coordination help but did not receive all that they needed.
Health and well-being of communities	Annual dental visit: percentage of members 2 to 21 years of age who had at least one dental visit during the measurement year. Acute care prevention of falls: percentage of patients who receive appropriate falls prevention interventions based upon the results of their falls risk assessment.
Making Care Safer	Acute care prevention of falls: rate of inpatient falls per 1,000 patient days.

[67] http://www.qualitymeasures.ahrq.gov/browse/by-national-quality-strategy.aspx

National Quality Strategy (NQS) Priority	Examples of Measures Addressing Each NQS Priority
	Long-stay nursing home care: percent of residents who received the seasonal influenza vaccine.
Making Quality Care More Affordable	Total knee replacement: percentage of patients undergoing a total knee replacement with radiographic evidence of arthritis within one year prior to the procedure. Imaging efficiency: percentage of brain CT studies with a simultaneous sinus CT.
Person- and Family-centered Care	Asthma: average number of lost workdays and/or school days in the past 30 days. Behavioral health care patients' experiences: percentage of adult patients who reported whether they were provided information about treatment options.
Prevention and Treatment of Leading Causes	Acute myocardial infarction (AMI)/chest pain: percentage of 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 DVT prophylaxis prescribed within 24 hours of hospital admission.

Figure 11: Measurement Settings[68]

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 - Other
- 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 Use Treatment Programs/Centers
- Transition

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

6.2—Cost and Resource Use Measures

The CMS strategic goals for 2013–2017 include providing better care and lower costs of care for all Americans.[69] 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 with the objective being used as tools to direct efforts to make healthcare more affordable. Some terms related to measures addressing affordable care include:

- 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.
- 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, or standardized prices) to each unit of resource use—that is, monetize the health service or resource use units.
- Quality of care—Quality measures assess performance on the six healthcare aims specified by the IOM: 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.[70]
- Value of care—This type of measure includes a specified stakeholder's preference-weighted assessment of a particular 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,[71] 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,

---

[69] 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: June 1, 2015.

[70] 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: June 1, 2015.

[71] Murray CJL, Frenk J. Ranking 37th –– Measuring the Performance of the U.S. Health Care System. *New England Journal of Medicine.* 2010;362(2):98-99.

effectiveness, equity, and patient-centeredness."[72] NQF used the following figure (adapted) to illustrate the relationship between resource use, efficiency, and value.

Figure 12: 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.

6.3—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."[73] These measures are useful for a variety of purposes. Composite performance measures can simplify and summarize a large number of measures or indicators into a more succinct piece of information. Stakeholders can then track broader ranges of information without being overwhelmed by data elements.

[72] National Quality Forum. *Endorsing Cost and Resource Use Measures*. Washington, DC: NQF; 2014. Available at: http://www.qualityforum.org/Publications/2014/01/Endorsing_Cost_and_Resource_Use_Measures_-_Phase_1.aspx. Accessed on: June 1, 2015.

[73] 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: June 1, 2015.

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.

Purpose of composite measures

For measures to be grouped as a composite, they must have a purpose for which the composite will be used (for example, comprehensive assessment of adult cardiac surgery quality of care). There also needs to be a delineated quality construct to be measured, for example, the four domains of cardiac surgery quality which include perioperative medical care, operative care, operative mortality, and postoperative morbidity.[74]

Composite performance measure development should follow these principles:[75]

- 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 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.[76]

[74] Ibid.

[75] 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.

[76] 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.

Component performance measures

The following are some considerations for selecting measures to be included in a composite:[77]

- Components should be justified based on clinical evidence.
- NQF endorsement is not required; however, measures will need to be justified in terms of 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.

Assess components of the composite for internal consistency. However, 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.

6.4—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 their 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 patient-reported outcomes and use them in quality and accountability programs is still under construction. Tools to collect these data (such as the PROMIS tools) have mostly been used in academic settings and are being tested for clinical application.[78]

Patient-reported outcomes

Patient self-reported outcomes 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,

[77] 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: June 1, 2015.

[78] National Institutes of Health. *Patient Reported Outcomes Measurement Information System (PROMIS).* http://www.nihpromis.org/measures/availableinstruments. Accessed on: June 1, 2015.

without interpretation of the patient's response by a clinician or anyone else."[79] Self-reported patient data provide a rich data source for outcomes. This definition reflects the key domains listed in the NQF report on PRO-PM:

- Health-related quality of life (including functional status).
- Symptoms and symptom burden (e.g., pain, fatigue).
- Experience with care.
- Health behaviors (e.g., smoking, diet, exercise).[80]

Patient-reported outcome measurement (tools)

Tools that are used to collect patient-reported outcomes have been called Patient-Reported Outcome Measurements (PROMs). 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.[81]
- 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 reporting, and health improvement. All managed care plans with Medicare Advantage contracts must participate.[82]
- 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.[83]

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 has to be constructed that applies the outcome data collected by the tools to measure the quality of care.

---

[79] 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: June 1, 2015.

[80] 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: June 1, 2015.

[81] National Institutes of Health. *Patient Reported Outcomes Measurement Information System (PROMIS)*. Available at: http://www.nihpromis.org/measures/availableinstruments. Accessed on: June 1, 2015.

[82] Centers for Medicare & Medicaid Services. *Medicare Health Outcomes Survey*. Available at: http://www.hosonline.org/. Accessed on July 8, 2014.

[83] Focus on Therapeutic Outcomes Inc. Available at: http://www.fotoinc.com/. Accessed on: June 1, 2015.

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.[84]

Approaches to developing patient-reported outcome performance measures

Though patient-reported outcomes 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. Also, refer to the subsection on Risk-Adjusted Outcome Measures in Chapter 2—Measure Specification of Section 2 for the procedure for risk adjusting outcome measures. NQF outlined a pathway for patient-reported outcomes to move from simple patient-reported data to measurement, to performance measurement, and finally to endorsed measures in use for reporting and accountability.

Choose and define a PRO

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, mortality would not be an appropriate measure for cataract surgery as it is an extremely rare and nearly irrelevant outcome for this type of surgery. A better outcome measure in this instance might be a clinically meaningful measure of improvement in vision.

Outcome performance measures also have to 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. If a TEP is to be used, refer to Section 2, Chapter 1.B.—Technical Expert Panel for the standardized processes.

Determine the appropriate way to collect the PRO using a Patient-Reported Outcome Measure (tool) (PROM)

Measure development always begins with an environmental scan and literature review. Identify if there are existing tools to collect the outcome in the target population. Many tools in this area have been developed for research and have existing psychometric data establishing reliability and validity. With further testing in clinical settings, they can be used for PROMs. Feasibility must also be tested for the relevant clinical applications.

---

[84] 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: June 1, 2015.

Determine the appropriate performance measure: the Patient-Reported Outcome-Based Performance Measure (PRO-PM)

Patient-reported outcomes can be aggregated into benchmarks. The outcomes for target populations can be reported as average change or percentage improvement determined by the topic of interest. All have to be tested for reliability, 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 has to capture the results of the care given and not the influence of co-morbidities or other extraneous variables. However, as in any other outcome measurement, risk adjustment should not be allowed to mask disparities. Refer to the subsection on Risk-Adjusted Outcome Measures in Chapter 2—Measure Specification of Section 2 for discussion on determining the need for risk adjustment and development, and evaluation of risk adjustment models.

Evaluate the outcome measure

Outcome measures, including those based on patient-reported outcomes, must be evaluated against standard criteria in the same way that all measures under development have to be evaluated. Detailed specifications must be submitted, using the Measure Information Form.[85]

Some of the unique considerations (in addition to the others in each category) that apply to evaluating patient-reported outcome performance measures 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 have to be established not only for the data measurement (PROM) but also for the derived performance measurement.
- Feasibility—Burden to respondents must be minimized. Illness may complicate accessibility issues. Language 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 benefits of the information should outweigh the burden to both patients and providers.

Measure developers developing eMeasures for patient-reported outcomes should submit detailed specifications using an eMeasure XML file; SimpleXML file; eMeasure human-readable rendition (HTML file, and value sets); Measure Justification Form, and the summarized evaluation using the Measure Evaluation Report Template. Refer to the eMeasure Lifecycle section for more

[85] 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: June 1, 2015.

information. Special considerations for outcome measures are included in the criteria descriptions and in the report forms.

The NQF endorsement criteria for PRO-PMs are enumerated in NQF's final report, *Patient-Reported Outcomes (PROs) in Performance Measurement*.[86] Documentation of all of these items should be submitted to the COR at appropriate times as specified in the contract. Ensure strong justification when submitting outcome measures for endorsement.

6.5—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.[87] Though the numbers vary by counting methods, the CDC also found the prevalence of multiple chronic conditions (MCC) in all adults had increased from 21.8 percent in 2001 to 26.0 percent in 2010.[88] 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.[89] In 2011, CMS found that Medicare beneficiaries with multiple chronic conditions were the heaviest users of healthcare services.[90] For those with six or more chronic conditions, two-thirds were hospitalized and they accounted for about half of Medicare spending on hospitalizations.[91] However, very few measures exist that are specifically designed to evaluate the quality of care provided to these people.[92]

Multiple chronic conditions measures address the challenge of evaluating care and outcomes for persons who have "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."[93] Specific considerations for measure development in these circumstances will also be described.

---

[86] 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: June 1, 2015.

[87] 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: June 1, 2015.

[88] 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: June 1, 2015.

[89] 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

[90] 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: June 1, 2015.

[91] Ibid.

[92] 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: June 1, 2015.

[93] National Quality Forum. *Multiple Chronic Conditions Measurement Framework.* 2012. Available at: http://www.qualityforum.org/Projects/Multiple_Chronic_Conditions_Measurement_Framework.aspx. Accessed on: June 1, 2015.

Multiple chronic conditions 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.
- 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 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.[94]

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.[95] Current quality measures are largely based on performance standards derived from clinical practice guidelines for management of specific diseases.[96] Patients with multiple chronic conditions 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.[97] Few measures exist to evaluate inappropriate care in these situations.

Considerations for measure development targeting persons with MCC

What to consider when choosing appropriate measure concepts

Without evidence-based guidelines specifically directed to care of persons with multiple chronic conditions, 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 multiple chronic

[94] Ibid. p. 7.
[95] Ibid.
[96] 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.
[97] Ibid.

conditions. The following measurement concepts were identified as having potential for high-leverage in quality improvement for patients with MCC:[98]

- Optimizing function, maintaining function, or preventing further decline in function
- Seamless transitions between multiple providers and sites of care
- Patient-important outcomes (includes patient-reported outcomes 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 crosscutting 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.

When determining how to address key issues

Guiding principles

The NQF Framework identified that quality measures for persons with multiple chronic conditions should be guided by several principles. Quality measures should:[99]

- Promote collaborative care among providers.
- Incorporate several types of measures that address appropriateness of care.
- Prioritize optimum outcomes that are jointly established by considering patient preferences.
- Address shared decision making.
- Assess care longitudinally, changes in care over time, such as delta measures of improvement or maintenance rather than attainment.
- Be as inclusive as possible, not excluding persons with multiple conditions.
- 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.

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 multiple conditions, it is very difficult to know where to attribute

[98] Ibid., p. 9.
[99] Ibid.

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.

Attribution issues to consider

Issues of attribution are compounded when adding the factor of multiple chronic conditions. 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 multiple chronic conditions 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.

Methodological issues to consider

Quality measures for this population should be designed to be as inclusive as possible. Instead of excluding these individuals from the denominator of measures, 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 multiple chronic conditions. Risk adjustment should be used with caution in the situation of multiple chronic conditions. Stratification may allow quality comparison across populations without masking important distinctions of access, care coordination, and other issues. Refer to the subsection on Risk Adjusted Outcome Measures in Chapter 2—Measure Specification of Section 2 for 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.

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 the work of breathing and the work of communicating could make responding to a survey more difficult. Interpretation of different types of data is needed as the data may come from multiple providers, multiple sources, in multiple types, and over extended periods. It is important for measure developers to standardize data collection methods.

When testing and evaluating measures for persons with MCC

Evaluation methods described elsewhere in the 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.

Section 2—Measure Lifecycle

The first section of the Blueprint introduced key concepts of measure development, implementation, and maintenance that are applied throughout the measure lifecycle. The second section will explain the procedural steps involved throughout the measure lifecycle. The procedures in Section 2 implement the concepts described in Section 1, which are linked where they are used.

The end product of measure development is a precisely specified, valid and reliable measure that is directly linked to the CMS quality goals. This precisely specified measure is then documented in a Measure Information Form, and Measure Justification Form.[100] These documents provide standardized details of the measure, to allow others to understand the rationale of the measure, and produce consistent interpretation and implementation. The measure documentation is updated iteratively throughout the measure lifecycle as the specifications are refined and improved.

Figure 13 : Flow of the Measure Lifecycle and Timeline of the Measure Lifecycle, respectively, show 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. CMS may accomplish this through one or more contracts. Key points where CMS makes major decisions during the measure lifecycle are indicated in red text. 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 some steps concurrently provided the changes are approved by the COR.

The chapters titles in Section 2 correspond to the headings in the blue boxes listed in Figure 13 3. Each chapter in Section 2 also begins with a figure highlighting the corresponding portion of

Figure 13 3, indicating the phase of the Measure Lifecycle that is being discussed. In Section 3, the eMeasure Lifecycle is depicted with a similar figure. The corresponding chapters in Section 3 also begin with a figure that references the phase that applies to each particular chapter. The lifecycle processes for all measures are parallel with variations in details as required by measure specifics.

[100] 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: June 1, 2015.

Figure 13 : Flow of the Measure Lifecycle

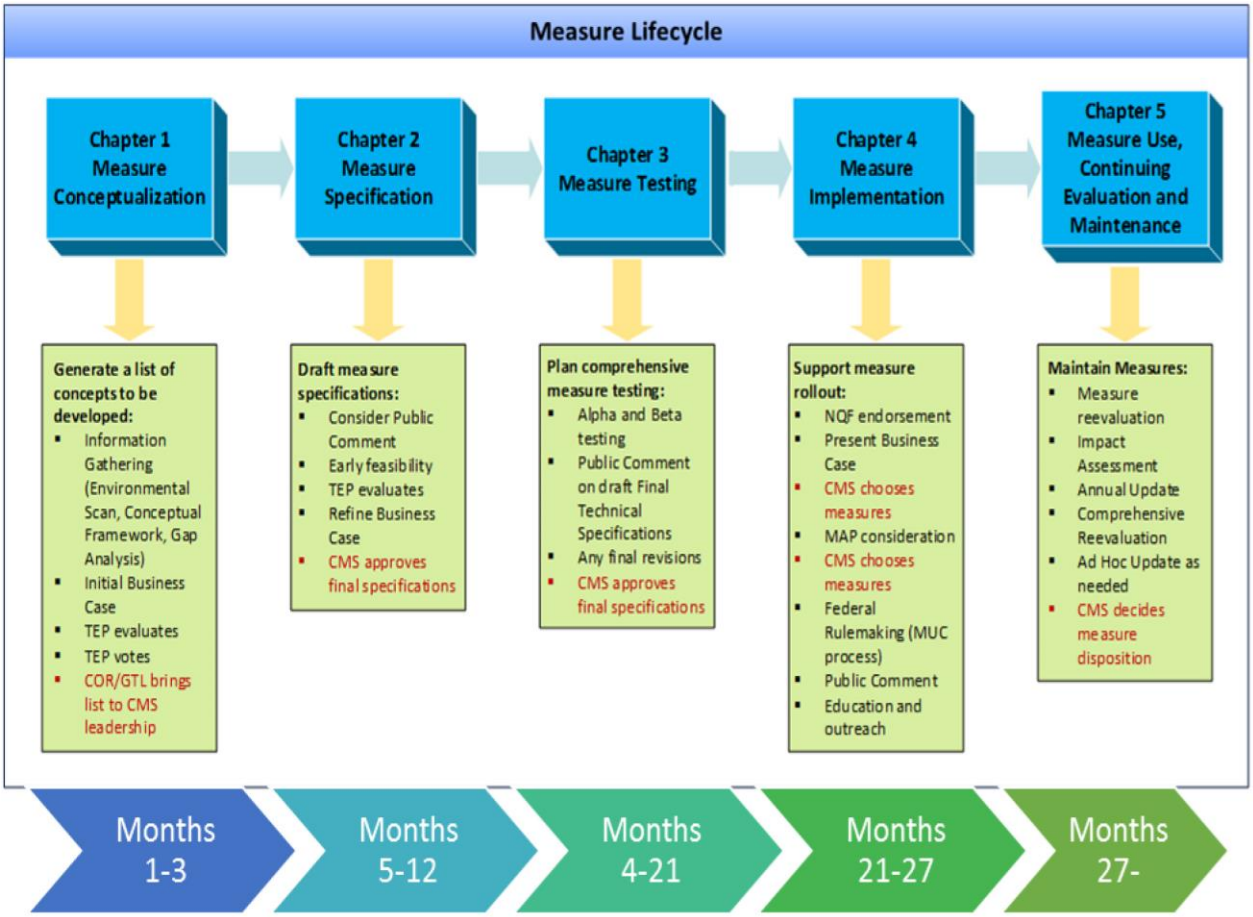
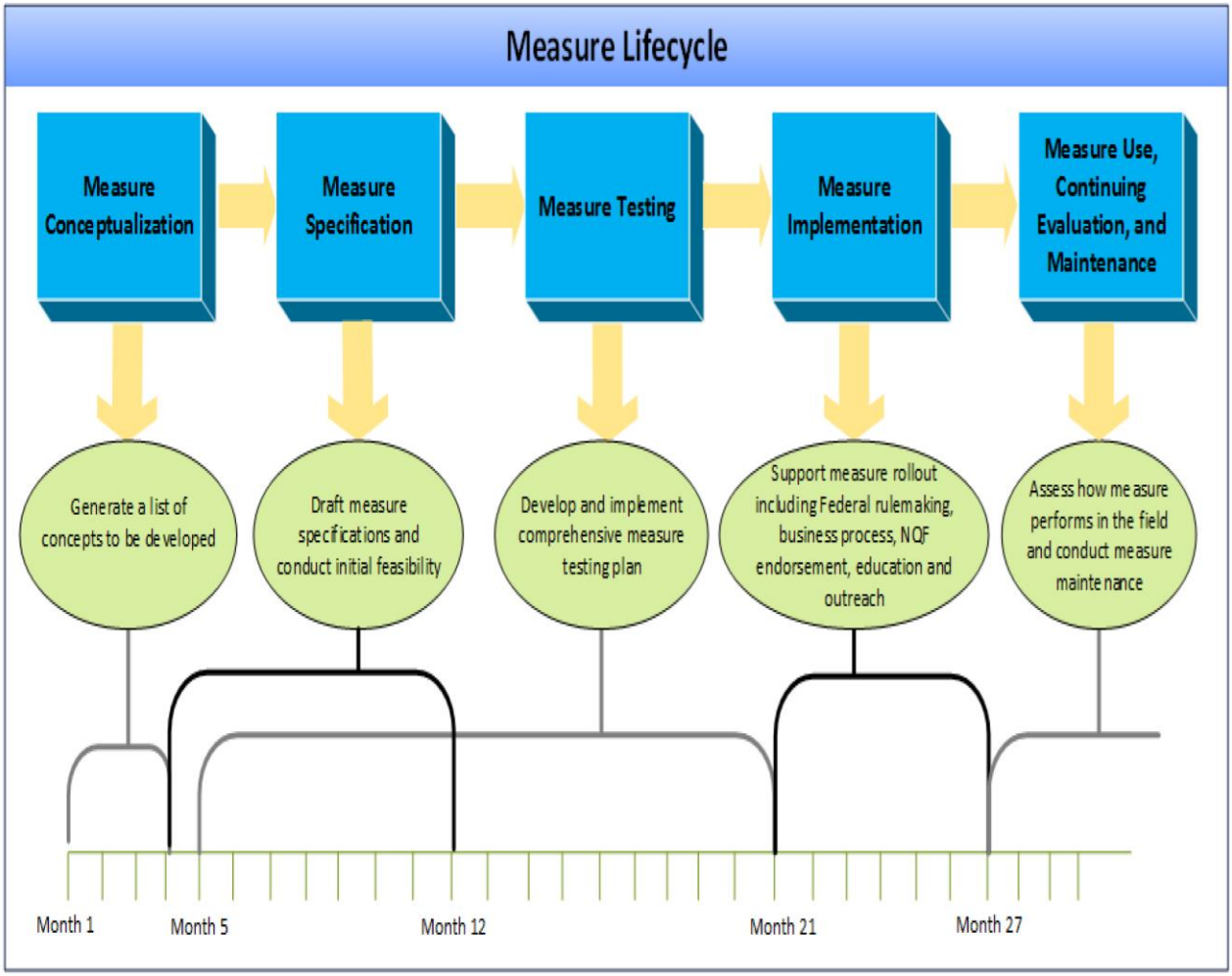


Figure 14: Timeline of Measure Lifecycle



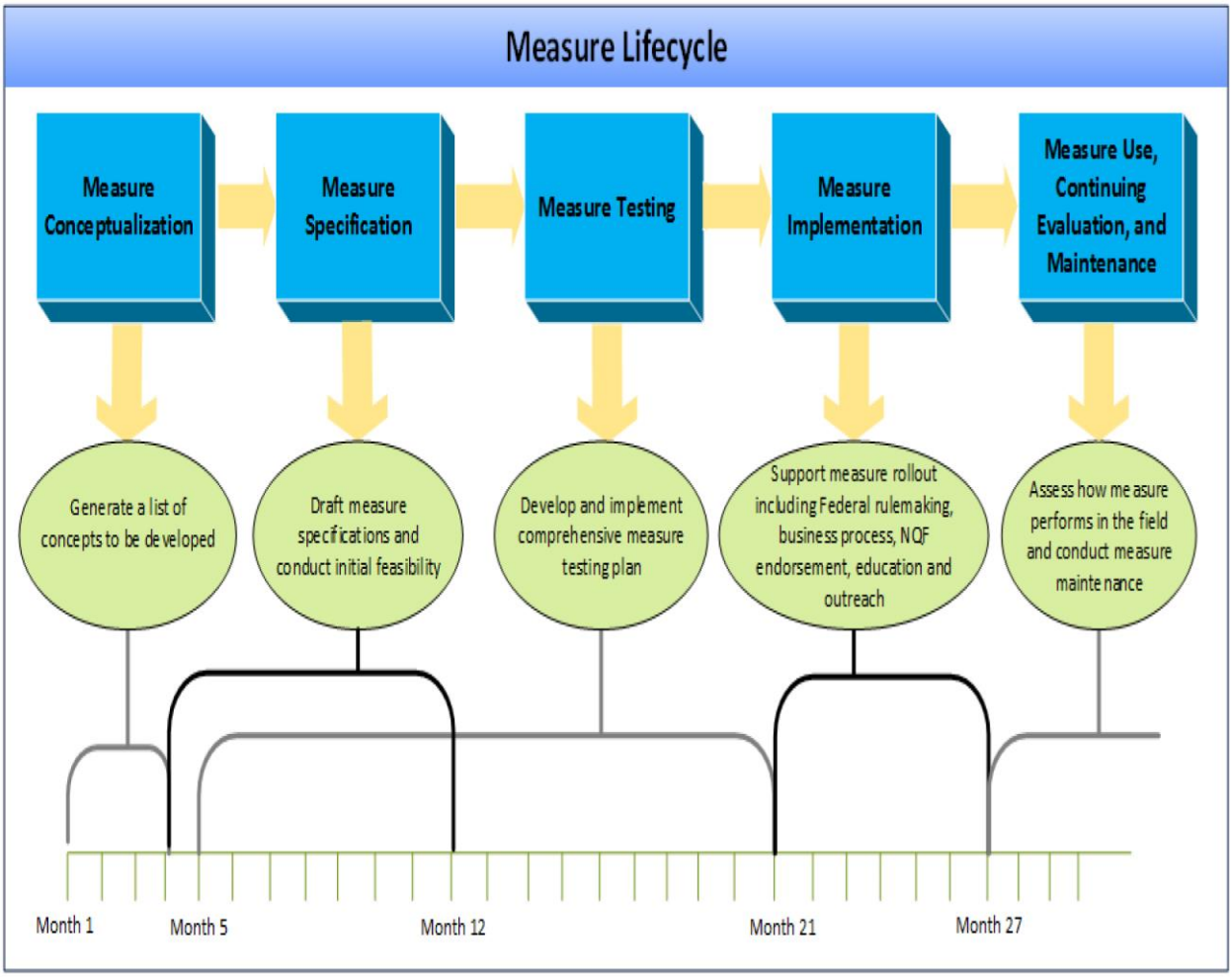
Chapter 1—Measure Conceptualization

In the first phase of the measure lifecycle, the evidence base for the concept and basic elements of the measures are generated.

- The initial measures list created during information gathering is reviewed and the measures are rated to become the List of Candidate Measures evaluated by the TEP and CMS Leadership, using the measure evaluation criteria.
- The TEP then forwards the measures it recommends as candidate measures to the measure developer, who then proposes the candidate measures to CMS.

Figure 155 lists the major tasks of Measure Conceptualization. The process flow may be cyclical and iterative as the candidate measure list is narrowed.

Figure 15: Flow of the Measure Lifecycle - Measure Conceptualization



Chapter 1.A.—Information Gathering

Information gathering is the first of three steps in the measure conceptualization process (the other two steps being TEPs, Chapter 1.B., and Public Comment, Chapter 1.C). 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.

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, as well as 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. Guidance on when to consider related and or competing measures is covered in Table 7.

Table 7: Differences between Related and Competing Measures

	Same focus	**Different focus**
Same population	Competing measures	Related measures
Different population	Related measures	Non-competing, non-related measures

Deliverables

- Information Gathering Report—Summary Report of Environmental Scan and Empirical Analysis
- List of potential candidate measures.
- Measure documentation[101], including:

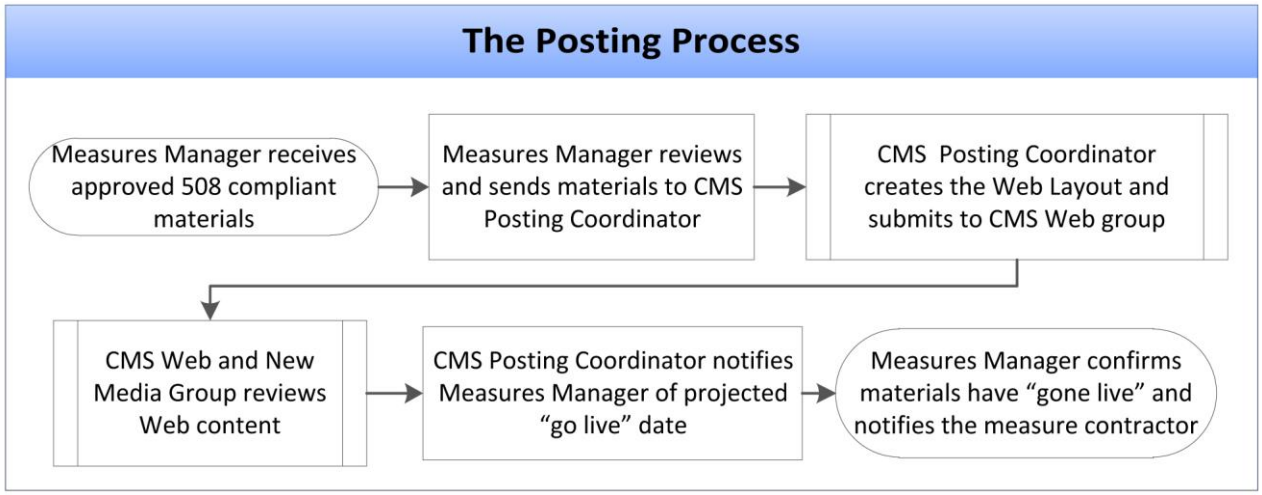
[101] Note: The document that is exported from the online NQF measure submission may be accepted in lieu the MIF and MJF, when available and at the discretion of the COR, following NQF submission and favorable review of the measure.

- Measure Information Form
- Measure Justification Form
- Business case for measure concepts, topics, or candidate measures.
- Expert Input Report (if expert input obtained during information gathering).[102]

CMS Measures Management System Posting Process

The procedure depicted in Figure 16 and described below, are used for all postings to the pages (Call for Measures, Technical Expert Panels and Call for Public Comment) at the MMS Overview site.[103]

Figure 16: The Posting Process



1. The measure developer assembles 508 compliant materials to be posted and obtains final approval from the COR. Information about CMS 508 compliance is available on HHS's website.[104] If files are to be added to the download section, they should be zipped together and the folder named with the project title and document set name.
   a. If JIRA is to be used for comments, ensure that the link and email box are activated before sending materials for posting.
2. Upon receipt by the Measure Manager, the material to be posted is reviewed again for 508 compliance, completeness of information, confirmed COR approval, and compliance with formatting requirements.
3. The materials are then sent to the CMS Website Posting Coordinator.
4. CMS Website Posting Coordinator creates the updated web page layout and submits it to the CMS Web group for posting.

[102] Expert input may include Technical Expert Panels, patient input, subject matter experts, or other relevant types of experts. This input may be obtained several times throughout the measure lifecycle.
[103] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Measures Management System.* Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/index.html. Accessed on: June 1, 2015.
[104] Department of Health and Human Services. *Section 508.* Available at: http://www.hhs.gov/web/508/index.html. Accessed on: June 1, 2015.

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

Posting Timeframe
Please allow **at least 5 business days** for processing your submission prior to the day you would like it to be posted online.

Submissions *may* be posted prior to this timeframe. If you **do not wish** your submission to be posted prior to a specific date, please note this in your email.

Posting Format
Every submission must include a **web posting document**. The web posting document should be submitted in **Word format**.

All **other documents/ attachments** should be submitted in **PDF format.** Please note: tables must have repeated headers on every page.

Posting Template
All submissions must follow the relevant Blueprint template to be compliant. If they do not, we will ask you to revise them before submitting it as a final post. Templates for the web posting documents can be found after Appendix J of the Blueprint.

Procedure

Follow the 6 steps below and document progress in the information gathering report, the Measure Information Form, and the Measure Justification Form.[105] The steps are not meant to be sequential but may be undertaken simultaneously.

Step 1 of 6: Conduct an environmental scan

The environmental scan is an essential part of building the case for quality measures. It builds on the needs assessment and 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

[105] 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: June 1, 2015.

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.[106]

The scan should consider CMS Quality Measurement Task Force 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, 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 shall explore the various dimensions of quality to develop informative quality measures. The resulting report of the environmental scan will include several findings:

1. Identification of related, similar, or competing measures, including opportunities for harmonization and alignment

2. Listing of clinical guidelines pertinent to the clinical domain or topic specified in the Task Order.

3. Review of studies that document the success of particular measures in the same or similar health care setting or domain covered by the Task Order.

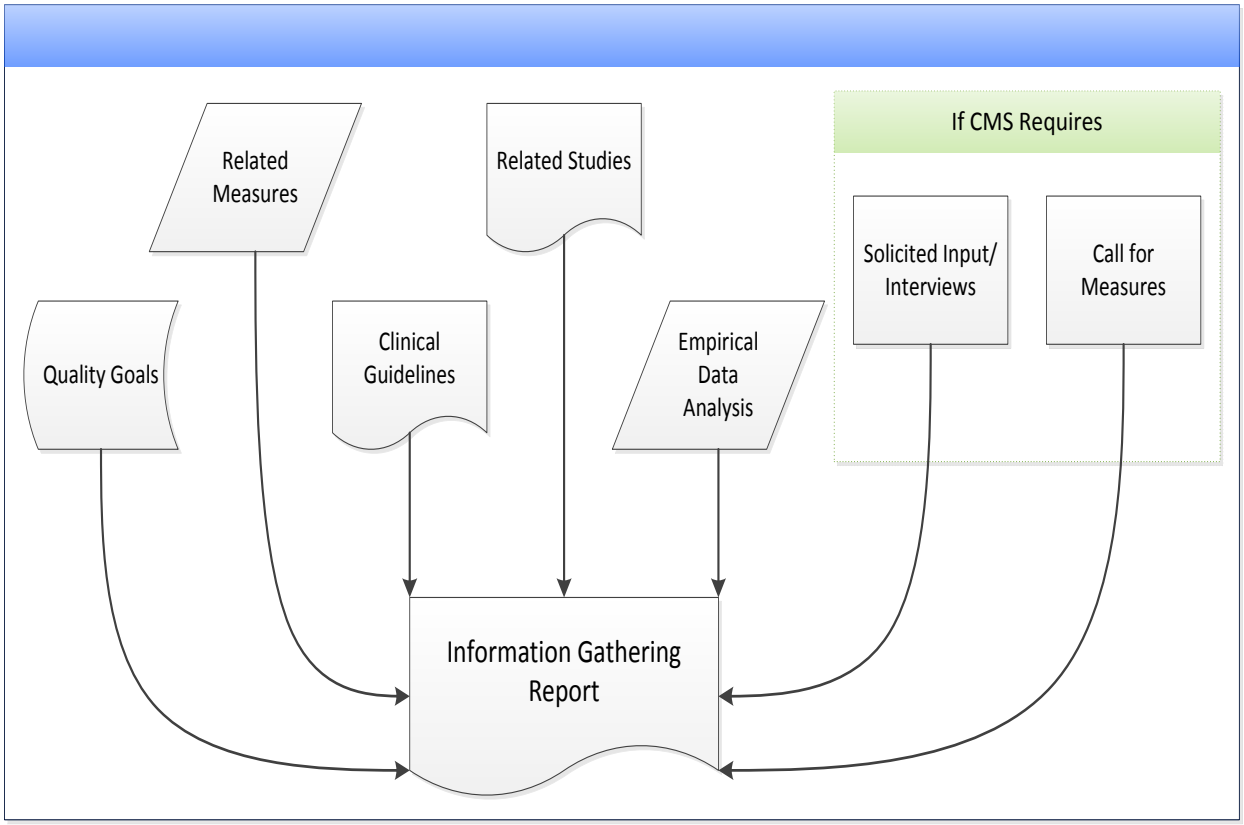
4. Discussion of scientific evidence supporting clinical leverage points that might serve as a basis for the measure (e.g., importance).

[106] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Measures Management System: Call for Measures.* Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/CallForMeasures.html. Accessed on: June 1, 2015.

Environmental Scan Outline

The environmental scan includes a literature review (white and grey), clinical practice guidelines review, review of regulations and their implications on measurement (e.g., MACRA), existing measures evaluation, expert input (including TEP and other experts), and stakeholder input–inclusive of all relevant stakeholders, including patients (Figure 17).

Figure 17: Environmental Scan Data Sources



The following is an example of an outline for the scan:

1. *Cover Page*, including the Task Order title and contract number, contractor contact information, and contracting officer's representative'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*

    - Technical expert panel

        - 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*

In developing the environmental scan, we recommend following the steps:

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 (i.e., 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 using the techniques of Constant Comparison Analysis, Domain Analysis, Theme Analysis, and Taxonomic Analysis. 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, we propose the following recommendations:

- Be strategic in planning and managing your scan
- Formalize your scanning process
- Design your scan in collaboration with domain experts

- Manage the information obtained as a core process[107]

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.[108] Whenever possible, include the electronic versions of articles or publications when submitting the report.

Use the measure evaluation criteria described in the Concepts of Measure Evaluation chapter 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 Affordable Care[109] 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 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.

Review the 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

---

[107] 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.
[108] Suggested outline provided in Step 5.
[109] http://www.ahrq.gov/workingforquality/about.htm

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.[110] 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 (such as race, ethnicity, age, socioeconomic status, income, region, sex, 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:

- Studies:
    - Published in peer-reviewed journals.
    - Published in journals from respected organizations.
    - Written recently (within the last five years).
    - Based on data collected within the last 10 years.
    - Unpublished studies or reports such as those described as grey literature. Governmental agencies such as the AHRQ, CMS, and the CDC produce unpublished studies and reports.
    - If available, systematic literature reviews[111] to assess the overall strength of the body of evidence for the measure contract topic. Evaluate each study to report the grade of the body of evidence for the topic.

Other resources for grading the body of evidence:

[110] 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: June 1, 2015.

[111] A systematic literature review is 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.

- IOM report: Finding What Works in Health Care Standards for Systematic Reviews.[112]
- NQF report: Guidance for Evaluating the Evidence Related to the Focus of Quality Measurement and Importance to Measure and Report.[113]

Clinical practice guidelines review

Search for the most recent clinical practice guidelines applicable to the measure topic (i.e., written within the past five 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 one guideline over another.

Sources for clinical practice guidelines review:

- National Guideline Clearinghouse[114]
- IOM report: Clinical Practice Guidelines We Can Trust[115]

---

[112] Institute of Medicine. *Finding What Works in Health Care Standards for Systematic Reviews.* Washington, DC: The National Academies Press, 2011. Available at: http://www.iom.edu/Reports/2011/Finding-What-Works-in-Health-Care-Standards-for-Systematic-Reviews.aspx. Accessed on: June 1, 2015.

[113] 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: June 1, 2015.

[114] Department of Health and Human Services, Agency for Healthcare Research and Quality. *National Guideline Clearinghouse.* Available at: http://www.guideline.gov/. Accessed on: June 1, 2015.

[115] Institute of Medicine. *Clinical Practice Guidelines We Can Trust.* Washington, DC: The National Academies Press, 2011. Available at http://www.iom.edu/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx. Accessed on: June 1, 2015.

Search for 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 particular topic/condition or setting. Determine what measurement gaps exist for the topic area, as well as existing measures that may be adopted or adapted for the project. For example, if a contract objective is the development of immunization measures for use in the home health setting, it will be necessary to identify and review existing home health measures. In addition, it might also be helpful to analyze 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. Below are links to a few readily available sources:

- National Quality Measures Clearinghouse
- HHS Measures Inventory
- CMS Measures Inventory and Pipeline
- NQF's Quality Positioning System
- AHRQ
- 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.

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 submit the plan for COR review. The Role of Patient Engagement and Participation is also discussed in Section 1 of the Blueprint, including 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 when the focus can be more technical. Patient input may be obtained during the earlier, less technical, phases of TEP discussions. These actions should be discussed with the COR in the 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, 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 early on before actual measure development begins. Details of how to conduct a TEP and other stakeholder meetings are covered later in this chapter.

Call for Measures (or measure concepts)

While conducting the environmental scan, if insufficient numbers or types of measures have been identified, discuss the situation with COR to determine if a Call for Measures is needed. If CMS approves, the measure developer may issue a Call for Measures to the general 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 if 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. Calls for measures during the

implementation phase of development seek fully developed measures that will be considered for implementation in CMS programs. The Measure Implementation chapter 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.[116] Use the Call for Measures Web Posting form to document the following:

- Overview of the measure development project.
- Specific project objectives.
- Types of measures needed (outcomes, patient experience of care, care coordination, etc.).
- Instructions on what type of measure information is needed, how long the Call for Measures will last, etc.
- Measure developer's email address (or other contact information) where measures and any questions should be sent.

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

Step 2 of 6: Conduct 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 particular 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 Measure Justification form,[117] and in the business case.

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

The eMeasure Conceptualization and eMeasure Testing chapters of the eMeasure Lifecycle section provide more information about empirical data analysis, especially early feasibility testing for eMeasures.

[116] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Measures Management System: Call for Measures.* Available at: https://www.cms.gov/MMS/13_CallForMeasures.asp. Accessed on: June 1, 2015.

[117] 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: June 1, 2015.

Step 3 of 6: Evaluate information collected during environmental scan and empirical data analysis

If there are related measures, evaluate the measures to assess if they meet the needs of the measure development contract. If the measure cannot be adapted or adopted to meet the needs of the contract, consider harmonization issues related to similar data elements (e.g., age ranges, time of performance, allowable values for medical conditions or procedures, allowable conditions for inclusion in the denominator, reasons for exclusion from the denominator, risk adjustment methods). A detailed description of harmonization concepts is covered in the Harmonization chapter in Section 1.

Regarding adapted measures, if the measure developer changes an existing measure to fit the current purpose or use, the measure is considered adapted. 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.

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 soundness and feasibility before a determination regarding use in a different setting can be made. The Measure Specification chapter 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 re-specified for the new setting in order 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 may also be required for the measure to receive NQF endorsement. The Measure Testing chapter 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 for example when the exclusionary conditions have dramatically increased.
- 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 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:

- A measure for screening adult patients for depression is found. The current contract requires mental health screening measures for adolescents. The owner of the adult depression screening measure may be willing to expand the population in the measure to the adolescent population. If the owner is not willing to expand the population, it may be necessary to develop a new measure specific to the adolescent population which will be harmonized with the existing measure.

Regarding adopted measures, if the proposed 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, such as data submission instructions, the measure is considered adopted.

Examples:

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

If a 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 measures, include the detailed specifications for the measure.

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.[118] Through this analysis, the measure developer may identify existing measures that can be adopted or adapted, or identify new measures that need to be developed.

Determine the appropriate basis for creation of new measures

If no existing measures are suitable for adoption or adaption, then new measures must be developed, and the measure developer will determine the appropriate basis for the new measures

[118] National Quality Forum. *Measuring Performance: Measure Evaluation Criteria.* Available at: http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx. Accessed on: June 1, 2015.

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.

- 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 the Measure Specification chapter.
- 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.

Step 4 of 6: Apply measure evaluation criteria and propose a list of potential measures

If a large number of measures or concepts were identified, narrow down the list of potential measures by applying the measure evaluation—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 the Measure Information Form and Measure Justification Form should begin as early as possible during the development

process.[119] Before presenting measures to the TEP, the measure developer may choose to use a modified Measure Information Form and Measure Justification Form to display partial information as it becomes available. At the end of the project, document each potential measure on the Measure Information and Measure Justification Forms. The Measure Information and Measure Justification Forms are aligned with the NQF measure submission. By the end of measure development, these forms should be completed in their entirety for new measures or measures that are significantly changed from the original.[120]

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 Measure Justification Form. Measures that are adopted and NQF-endorsed do not require further documentation in the Measure Justification Form.

The Measure Justification Form should be completed for adapted 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 and validity, feasibility, and usability as well.

Step 5 of 6: 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:

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, including all 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 (population, study size, data sources, study type, and method).
- Which measure evaluation criteria (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.
- Identify the major benefits of the process or intermediate outcome under consideration for the measure.

[119] 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: June 1, 2015.
[120] National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: June 1, 2015.

- 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.

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

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

Provide a summary of findings and measurement gaps:

- Existing related measures
- Gap analysis

Empirical Data Analysis Summary

For new measures:

- If available, data source(s) used
- 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.

Summary of Solicited and Structured Interviews, if applicable

Include, at a minimum:

- 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 used.

Step 6 of 6: 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, 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, measure steward, etc. An example of such a grid can be provided upon request. 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 (candidate lists, etc.) to the Measures Manager, who will help the measure developer identify potential harmonization opportunities.

Design a business case for the candidate measures

The CMS MIDS contract requires a business case to be developed for each candidate measure. Further detail about the business case is found in Chapter 5—Concepts of Measure Evaluation in Section 1. Most of the information needed to develop the business case will have been obtained through the initial information gathering process. Document the business case for each candidate measure with the other measure documentation.

The following types of information should be systematically evaluated to build the business case.[121]

Better Health

- Incidence and prevalence of the condition in the population.
- Disparities in population health.
- Indirect costs from wages or salaries lost when ill.

[121] 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: June 1, 2015.

Better Care

- The major benefits of the process or intermediate outcome under consideration for the measure (e.g., heart attacks not occurring, hospital length of stay decreased) and the expected magnitude of the benefits.
- Untoward effects of a process or intermediate outcome and the likelihood of their occurrence (e.g., bleeding from anticoagulation, death from low blood glucose levels).
- Current process performance, intermediate outcomes, and performance gaps.

More Affordable Care

- Cost of implementing the measures.
- Cost of implementing the clinical processes.
- Savings from preventing complications.
- Savings from preventing unplanned readmissions.
- Savings from improved health.

The business case should estimate the amount of improvement, including:

- Magnitude of the expected improvement.
- Time frame for the expected benefits of improvement.
- Projected measure performance trajectory, including estimation of when performance may top out.

Multiple models can be used to design the business case. Based on the measure focus, determine the appropriate model. To evaluate potential quality measures' aggregate effect of cash inflows and outflows accruing to an organization as a result of implementing a specific process or treatment that is the measure focus can be used. This model presents a single, easily interpretable result, 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 design 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 can also be measured with quantifiable anticipated benefits assigned to a model that can then be tested. Using the example mentioned above, improved care coordination may not only reduce expenses associated with unnecessary readmissions but it may also reduce mortality in selected populations, and may improve patient satisfaction. Regardless of whether the model is financial, epidemiological, or another kind, it should be stated in explicit terms as a hypothesis that can be used for later comparison. At a minimum, predict how the measure will perform (the trajectory) so comparisons can be made during measure use, continuing evaluation, and maintenance.

The Business Case Form Instructions and the Blank Business Case Form Template are provided in Section 4. Document the business case on the template and in the Importance section of the Measure Justification form. Evaluate the business case against the criteria provided in the Measure Evaluation Criteria and Instructions.

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.

Chapter 1.B.—Technical Expert Panel

The TEP is the second part of the measure conceptualization process, but continues throughout the measure lifecycle. A TEP is a group of stakeholders and experts who contribute direction and thoughtful input to the measure developer during measure development and maintenance. Since 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.

One note of caution when obtaining TEP input is that it cannot be framed in terms of advising CMS directly. The Federal Advisory Committee Act (FACA) has specific rules about advising the government directly, so it is important to be familiar with them.[122] Be clear in all materials and references that the TEP is advising the measure developer and not CMS directly.

Deliverables

- Call for TEP
- TEP Nomination forms
- List of stakeholders
- TEP Charter
- TEP Composition Documentation (TEP Membership List)
- TEP meeting schedule
- Meeting minutes
- Potential measures presented to the TEP
- Measure Evaluation Report
- Updated Measure Information Form and updated Measure Justification Form[123]
- Expert Input Report (TEP Summary Report) including list of candidate measures

Procedure

The following 12 steps should be performed when convening the TEP and conducting the TEP meetings. The exact order and level of detail required for the steps may vary depending on the phase of the measure lifecycle (for example, initial conceptualization versus reevaluation during measure maintenance), but the general process should be followed. Updates to the Measure Information and Measure Justification Forms may be made after the TEP deliberations. For

[122] General Services Administration. *Federal Advisory Committee Act (FACA) Management Overview.* Available at: http://www.gsa.gov/faca. Accessed on: June 1, 2015.

[123] 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: June 1, 2015.

example, information regarding the importance of the measure can be documented in the Measure Justification Form.

The TEP process involves three postings to the dedicated Measures Management System page on the CMS website.[124] These three postings include:

- Technical Expert Panel (Call for TEP) nominations.
- The TEP Composition Documentation with meeting dates.
- The TEP Summary Report.

The measure developers will communicate and collaborate with the Measures Manager to achieve these postings. The website posting process is detailed in Chapter 1.A.—Information Gathering, and will take up to five working days.

Step 1 of 12: Draft the TEP Charter and Consider Potential TEP Members for Recruitment

Draft the charter using the TEP Charter Template. This draft will later 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:

- The TEP's goals and objectives.
- The TEP's scope of responsibilities and how its input will be used by the measure developer.
- How the TEP will use the Measure Evaluation criteria.
- The estimated number and frequency of meetings.

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 in the charter, how issues of confidentiality, particularly for patient 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. Details on the types of balance needed for an effective TEP are described in the Stakeholder Input chapter in Section 1.

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. The Stakeholder Input chapter in Section 1 provides more details about patient and caregiver input into TEP deliberations.

[124] 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: June 1, 2015.

Step 2 of 12: 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 the following 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 (TEP Nomination form, letter of intent)
- Information on confidentiality of TEP proceedings and the way the TEP summary will be used
- Measure developer's email address where TEP nominations and any questions are to be sent

Step 3 of 12: 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.
- Scientific organizations related to the measure topic.
- Provider groups that may be affected by the measures.
- NQF measure developer groups.
- Other measure developers.

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

Step 4 of 12: 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.[125]

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.

If an insufficient pool of candidates has been 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.

Step 5 of 12: 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 specifically includes nationally recognized experts in the relevant fields, including clinicians (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 include explicit incorporation of the patient perspective in measure development through patient and/or caregiver input into quality issues that are important to patients.[126] The measure developer then proposes the list of measures 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.

[125] 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: June 1, 2015.

[126] Requirement quoted from MIDS contract 2013 language.

- Fair balance—Reasonable effort should be made 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. TEP members need to be accessible throughout the performance period of the measure developer's contract.
- Potential conflict of interest—TEP members are asked to disclose any potential conflict of interest during the nomination process, or if one arises after selection, to notify the TEP chair and the measure developer immediately. The TEP Nomination form template includes conflict of interest disclosure language.

Important considerations for patient/caregiver participation on the TEP

Plan for the way patient input will be solicited, gathered, and meaningfully incorporated into the measure development processes and submit the plan for COR review. One way of soliciting patient and caregiver input is to include them on the TEP. Whenever patient input is solicited or included, give consideration to confidentiality issues. Guidance and forms included in the Blueprint are not intended as legal advice, and any questions about how to handle confidentiality should be discussed with appropriate legal counsel as necessary. The following criteria can be used to select appropriate patients and/or caregivers for the TEP:

- Availability and accessibility of TEP meeting location, or willingness to participate via a conference call.
- Interest in or personal experience with the measure topics and concepts proposed by the project team.
- Within the age range (patient) or caring for someone within the age range of the population that will be measured.
- Comfortable speaking and participating in group discussions.
- Agree/consent that their input to the TEP will be shared in project reports that will be made public. For patient participants only, they may request that their names not be disclosed on TEP rosters or other project reports.

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 general 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 particular TEP member was selected. Additional information, such as TEP member biographies, may also be sent to the COR. Notify the COR about the TEP membership list within one week after the close of the posting. Confirm each member's participation on the TEP.

Step 6 of 12: 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.

Step 7 of 12: 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, to post the list following the process as described in Chapter 1.A. 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.

Step 8 of 12: Arrange TEP meetings

Organize and arrange all TEP meetings and conference calls. TEP meetings may occur face-to-face, via telephone conferencing, or via a combination of the two. If an in-person meeting is required, the measure developer plans the meeting dates, times, and venue, and helps participants with travel and hotel arrangements, etc., as needed.

The measure developer may decide that additional subject matter experts and staff may be needed to support the TEP. The areas of expertise may include, but are not limited to, data management and coding representatives, EHR experts, health informatics personnel, and statisticians/health services researchers. These subject matter experts can contribute summarized technical information to the TEP for consideration.

Step 9 of 12: 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. 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 Measure Information and the Measure Justification forms.[127]
    - Measure developers may modify the Measure Information Form to suit their particular contract needs. For example, the contract may not require the developer to develop detailed specifications, so a much shorter summary of the measure information could be used. Alternatively, measure developers who have identified a large number of 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.
    - For lay (patients and caregivers) members of the TEP, 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.
- 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.

Step 10 of 12: Conduct the TEP meetings

It is recommended that TEP discussions be held in 2 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.

[127] 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: June 1, 2015.

Phase 1 TEP meetings:

This phase 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 particular 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.

Patient/Caregiver Participation:

As described in the chapter on Stakeholder Input, it is important to ensure that patients and any other lay participants in TEP meetings are adequately prepared and supported during the meetings. It is best to assign a dedicated staff member to ensure participants are comfortable with the materials and that their needs are met during the meetings.

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.

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, adequacy of measure specifications, etc. 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 of the measure evaluation criteria as part of its deliberations. The Measure Evaluation chapter 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.

Step 11 of 12: 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 also delivered to CMS 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.

Step 12 of 12: Post the TEP Summary Report

Communicate and coordinate with the Measures Manager to post the approved TEP Summary Report 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 shall review the comments received and recommend appropriate action, particularly in regard to 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.

Chapter 1.C.—Public Comment

The public comment process is the third step under measure conceptualization, and like the TEP, it continues throughout the measure lifecycle. 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. Public comment also provides opportunities to refine measure specifications, etc. Times when public comment should be considered are discussed in Section 1, Chapter 3—Stakeholder Input. The following procedures will apply whenever public comment is obtained.

Deliverables

- Public Comment Call Web Posting template form[128]
- List of stakeholders for notification
- Measure Information and Measure Justification Forms (for candidate measures)[129]
- eMeasure specifications (as appropriate)
- Public Comment Summary report, including verbatim comments
- Responses to Public Comment

Procedure

The following 8 steps are essential to successfully soliciting public comment. Deviation from the following procedure requires COR approval.

The Call for Public Comment involves several postings to the dedicated CMS MMS website.[130] The measure developers will develop materials to send to the Measures Manager to post the call. Website postings involve two CMS divisions, and the process to post the materials will take approximately five working days.

Measure developers must plan accordingly for the deadlines for submitting information to be posted for public comment, and 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.

[128] All forms are included in Section 4 of the Blueprint.

[129] 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: June 1, 2015.

[130] 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: June 1, 2015.

Step 1 of 8: 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 at this time after contents have been approved by the COR.

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.[131] The chapter on Stakeholder Input in Section 1 includes a discussion of these requirements.

Step 2 of 8: 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 who 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 (American Quality Alliance and others).
- 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 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, and LinkedIn). Contact the appropriate COR for the process.

[131] 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: June 1, 2015.

Step 3 of 8: Post the measures following COR approval

After obtaining COR approval, work with the Measures Manager to post the Measure Information and Measure Justification Forms on the dedicated CMS MMS website.[132] Using the Call for Public Comment form and the posting process described in Chapter 1.A.—Information Gathering. When submitting the forms, prominently mark the Measure Information and Measure Justification Forms as "draft."[133]

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

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

- The specific objectives of the measure development contract.
- The 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).
- Measure testing.
- Posting for public comment.
- Refining measures as needed.
- Objectives of the solicitation, such as to help determine measure importance, to refine specifications and to comment on usability and feasibility.
- The Measure Information and Measure Justification Form. Include information on the development stage of the measures.
- A 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.

The posting process is the same as described earlier in this chapter.

Step 4 of 8: Collect information

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

[132] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Measures Management System: Overview.* Available at: https://www.cms.gov/MMS/17_CallforPublicComment.asp. Accessed on: June 1, 2015.

[133] 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: June 1, 2015.

Step 5 of 8: Summarize comments and produce report

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

The report should include:

- 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 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.

Step 6 of 8: 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 as a result of the public comments. This may be done via email, teleconference, or in-person meeting.

Step 7 of 8: 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, such as 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.

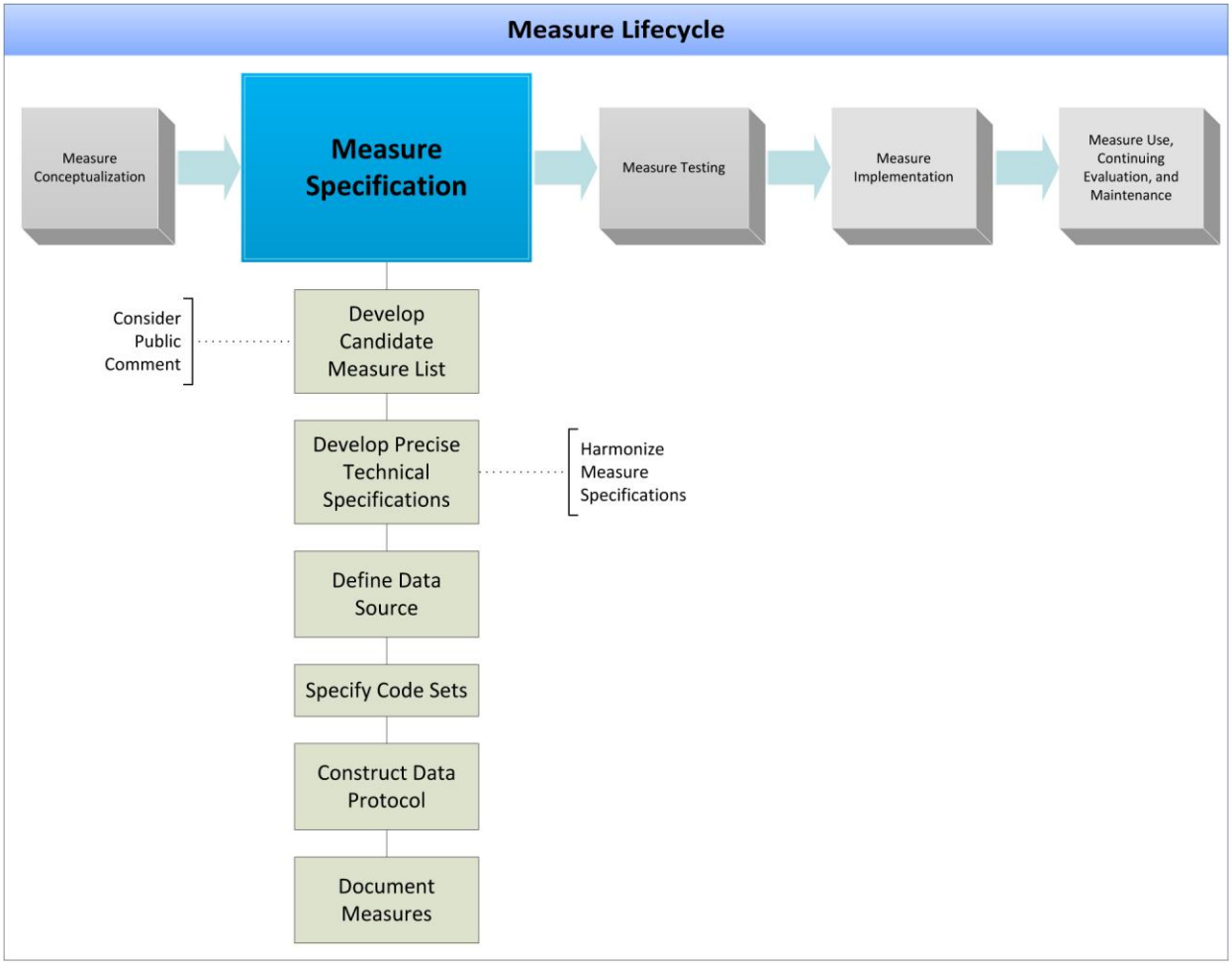
Step 8 of 8: Arrange for the final Public Comment Summary Report to be posted on the website

After obtaining COR approval, work with the Measures Manager to post the final Public Comment Summary Report including verbatim comments within three 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 listed in Chapter 1.A. This posting will remain on the website for a minimum of 21 days. After the 21-day period, the posting may be removed from the website.

Chapter 2—Measure Specification

This chapter[134] provides guidance to the measure developer to ensure that measures developed for CMS have complete technical specifications that are detailed and precise. Final technical specifications provide the comprehensive details that allow the measure to be collected and implemented consistently, reliably, and effectively. Note that measure developers intending to develop eMeasures must follow the processes and guidance provided in the eMeasure Lifecycle section. Figure 18 depicts the measure specification portion of the measure lifecycle that is discussed in this chapter.

Figure 18: Flow of the Measure Lifecycle - Measure Specification



2.1—Deliverables

- List of potential measures to be developed and timelines

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

- Measure Information Form or equivalent document[135] as approved by the COR.
- Measure Justification Form (or equivalent document[136] as approved by the COR) 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 existing NQF-endorsed measures, the Measure Justification Form is not required for measures that are being proposed for use by CMS because the justification is part of the endorsement process.
- For measure developers developing eMeasures, the eMeasure Specification chapter of the eMeasure Lifecycle section provides standardized processes for developing electronic specifications and value sets.

2.2—Technical Specification

The process of developing measure specifications occurs throughout the measure development process. During information gathering, the measure developer identifies whether existing measures may be adopted or adapted to fit the desired purpose. If no measures are identified that match the desired purpose for the measure, the measure developer must communicate and collaborate with its TEP to develop new measures. Depending on the information gathering findings, including initial feasibility assessments, the TEP will consider potential measures that are newly proposed or are derived from existing measures. The measure developer submits the list of candidate measures, selected with TEP input, to the COR for approval. Upon approval from the COR, the measure developer proceeds with the development of detailed technical specifications for the measures.

The Measure Information Form is used to document the technical specifications of the measures. At this stage, the technical specifications chapter is likely to include high-level numerator and denominator statements and initial information on potential exclusion, if applicable, and will continue to be completed throughout the development process as more information is obtained. The fields in the Measure Information Form have been aligned with the NQF measure submission, and collecting information on them throughout the development process will make the measure submission process easier.[137] Depending on the stage of measure development, the Measure Information Form can be modified to display only the available information or the developer may use a different format to display multiple measures (i.e., to present to the TEP). The ultimate goal of this process is to collect the information necessary to make the NQF measure submission process easier.

Developing technical specifications is an iterative process:

[135] 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: June 1, 2015.
[136] Ibid.
[137] 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: June 1, 2015.

- 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 advise the measure developer of the recommended changes.
- If directed by the COR, public comments may be obtained regarding the draft measures. Comments received during the public comment period will be reviewed and taken into consideration by the measure developer, CMS, and the TEP and will often result in revisions to the measure specifications.
- 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. The Measure Testing chapter provides details of the procedures for beta testing. As a result, technical specifications will continue to evolve, becoming more detailed and precise.

Though it may be advisable to obtain public comments at several points during measure development, after measure testing is complete is a key time to get additional public comments. Use the formal process outlined in the Chapter 1.A.—Information Gathering.

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 of these factors will make the technical specifications more precise and make the measure more valid and reliable. 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, exclusion, and measure logic. Then, the measure concept is more precisely specified with increasing amounts of detail, including the appropriate codes sets and/or detailed and precisely defined data elements.

Below are the key components of technical specifications:

- Measure name/title

- Measure description
- Initial population
- Denominator
- Numerator
- Exclusion and exception
- Data sources
- Key terms, data elements, and codes
- Unit of measurement or analysis
- Sampling
- Risk adjustment
- Time windows
- Measure results
- Calculation algorithm

Procedure

The following 6 steps are performed to develop the full measure technical specifications. The fields within the Measure Information Form are completed and updated as the measure progresses from information gathering to measure testing, and finally when submitted to NQF for endorsement consideration.

Step 1 of 6: 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 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. Candidate measures may be 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 (often satisfied with a checkbox, date, or code).

Examples of such measures include:

- A completed assessment.
- A completed care plan.
- An instruction such as teaching, 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 patients' perspective.[138] Although documenting having counseled or educated a patient on a specific issue is relatively easy, it is important, though 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 which may not have the intended effect in measuring a patients understanding or competence. For this reason, it is important to base performance measures on the patient's perspective as much as possible.

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. Criteria for measure evaluation are discussed in the Testing and Measure Evaluation Criteria subsection of Chapter 3—Measure Testing.

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. Also, at this stage, 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 and validity of the measure or portions of the measure. EHR data experts and informaticists would also be valuable resources for conducting early feasibility testing. The Measure Testing chapter includes more details of this process.
- After determining any areas for potential harmonization (as described in the Harmonization chapter), the measure developer develops the detailed specifications.

Adapted measures

If the measure developer changes an existing measure to fit the current purpose or use, the measure is considered adapted. This process includes changes to the numerator or denominator specifications, or revising 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.

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

[138] National Quality Forum. *CSAC Guidance on Quality Performance Measure Construction*. May 2011. Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: June 1, 2015.

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 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 and possibly respecified and tested. The Harmonization chapter 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 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?

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 the numerator and denominator statement to fit the particular use, the detailed specifications will be developed.

Adopted measures

For a measure to be considered adopted, it must have the same numerator, denominator, and data source as its 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.

Examples:

- 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 hospital measure from The Joint Commission not developed by CMS is now added to the CMS measure set.

If a measure is copyright protected, there may be issues relating to its ownership or issues related to proper referencing of the measure. The measure owner may have the right to review and approve any portrayal of the measure before it is disseminated. The measure owner may need to be contacted for details. Measure developers should beware of the copyright status of any measure proposed for adoption. Before contacting the measure owner, measure developers should discuss the situation with their COR.

Composite performance measures

Select the component measures to be combined in the composite performance measure. Chapter 6—Concepts of Selected Specific Measure Types in Section 1, describes the way component measures are used to construct composite performance measures.

Step 2 of 6: 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, or adopted. At a minimum, measures should be specified with the broadest applicability (target population, setting, level of measurement/analysis) as supported by the evidence.[139]

Develop measure name and description

The **measure name** should be a very brief description of the measure's focus and target population.

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

Examples:

- Patients with diabetes who received an eye exam
- Long-stay residents with a urinary tract infection
- Adults who received a Body Mass Index assessment

For measures based on Appropriate Use criteria addressing over use 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}

[139] National Quality Forum. *CSAC Guidance on Quality Performance Measure Construction*. May 2011. Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: June 1, 2015.

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 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 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. 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.

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.

Example:

- The population of the Acute Myocardial Infarction (AMI) measure set is identified using four data elements:
- ICD-9-CM principal diagnosis code
- admission date
- birth date
- discharge date
- Patients admitted to the hospital for inpatient acute care with an ICD-9-CM principal diagnosis code for AMI, identified by patient age (admission date minus birth date) greater than or equal to 18 years, and a length of stay (discharge date minus admission date) less

than or equal to 120 days, are included in the AMI initial population and are eligible to be sampled.

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 window.
- 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 last six 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., a hospital, skilled nursing facility, or rehabilitation facility) and seen within 60 days following discharge in the office by the physician providing ongoing care.

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, and 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 window 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 who received a foot exam including visual inspection, sensory exam with monofilament, or pulse exam.
- The number of denominator-eligible patients who had documentation of receiving aspirin within 24 hours before emergency department arrival or during their emergency department 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).

Determine if denominator exception or denominator exclusion is allowable

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 multiple chronic conditions, exceptions and exclusions will determine if the care for this potentially vulnerable group is examined.

Exception permits the exercise of clinical judgment and imply 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.[140] For this reason, most measures apply exception only to cases where the numerator is not met.

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

[140] 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.

- COPD is an allowable exclusion for the performance measure of the use of beta blockers for patients with heart failure; however, 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 COPD. This decision may also be made with the patient's preferences and functional goals in mind.[141] Therefore if a beta blocker were administered to a patient with heart failure and COPD, this would be an exception and the patient would remain in both the numerator and denominator of the performance measure. For clarity sake and to distinguish types of exceptions this would be known as a "positive exception" in that care was given, and for which, the patient remains in the numerator and denominator of the performance measure.
- A second type of exception would be where the numerator criteria are not met, that is, the expected care is not given but the lack of care being given is not the responsibility of the Eligible Professional. In these "defensible exceptions" there is some other defensible reason why the care is not given and therefore the individual or patient is removed from both the numerator and denominator and the performance measure altogether.

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
- Patients' reasons for not receiving the service specified may be an exception to allow for patient preferences. Caution needs to be exercised when allowing this type of exception.
- 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 and 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 patients' refusal in the record
- System reason: A vaccine shortage prevented administration of the vaccine).

[141] Tinetti ME, Studenski SA, Comparative Effectiveness Research and Patients with Multiple Chronic Conditions. *New England Journal of Medicine*. 2011; 364;26. 2478-2481.

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, 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.

Exception may sometimes be reported as numerator positives rather than being removed from the denominator (as described above as a "positive exception"). This is sometimes done to preserve denominator size when there is an issue of small numbers, and also out of respect for the clinical judgment and autonomy of the eligible professional. 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 measure population and denominator before determining if numerator criteria are met. Exclusion is absolute, meaning that the treatment is not applicable and would not be considered. Missing data should not be specified as an exclusion. Missing data in itself may indicate a quality problem, so excluding those missing cases may present an inaccurate picture of quality. Systematic missing data (when poor performance is selectively not reported) also reduces the validity of conclusions that can be made about quality.[142]

Example of an exclusion:

- 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.

[142] National Quality Forum. *CSAC Guidance on Quality Performance Measure Construction*. May 2011. Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: June 1, 2015.

- Evidence that the exclusion significantly improves the measure validity.

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

There is a significant amount of discussion on the use of exclusion and exception, particularly the ability to capture exception in EHRs. There is no agreed-upon approach; however, additional discussion on the use of exception in eMeasures is provided in more detail in the eMeasure Specification chapter of the eMeasure Lifecycle section.

Step 3 of 6: Define the data source

The data source used to calculate a measure can determine reliability, 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.

Administrative data

Electronic data often include transactional data that have been created for the purpose of 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.

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 electronic health record but is not coded in such a way that a computer program can extract the information, and thus requires manual abstraction, is not considered electronic data. 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 purpose of the measure. The eMeasure Lifecycle section provides details of how to use electronic clinical data in quality measurement.

Medical records (paper-based or electronic)

Medical records are a traditional source of clinical data for measures, and the data may be documented on paper or electronically. This data source, however, requires labor-intensive manual extraction. Information manually abstracted from an EHR, which may include 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 similar to a paper patient record.

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 of 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.

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.

Surveys

Survey data may be collected directly from a patient, such as the Consumer Assessment of Healthcare Providers and Systems (CAHPS) surveys that assess the satisfaction and experiences of patients in various settings of care.[143] Surveys can provide the following advantages:

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

Step 4 of 6: Specify the code sets (including ICD-10 discussion)

Most CMS measures rely at least in part on the use of various code sets for classifying healthcare provided in the United States. Any codes 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 certain positions, or on claims from specific provider types. Some code sets may require copyright statements to accompany their use.[144] The Current Procedural Terminology (CPT) code sets are owned and maintained by the American Medical Association (AMA) and require current copyright statements to accompany their use. Logical Observation Identifiers Names and Codes (LOINC), copyrighted by the Regenstrief Institute, and Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) are other proprietary code sets that should be properly presented.

NOTE: Often existing code sets are not adequate to capture a clinical concept required by a measure. Addition to, or modification of, an existing code set should be noted in the technical specifications when necessary.

The VSAC is a public use tool sponsored by the National Institutes of Health for maintaining code set lists used in eCQM using object identifiers or OIDs. However, OIDs may also be used in the

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

maintenance of non-eCQM measure code lists, and measure developers should explore the use of OIDs for all measures, not just eCQM measures, as a best practice for code list maintenance.

Below are some commonly used code sets along with some considerations for their use.

International Classification of Diseases (ICD)

ICD is used for identifying data on claims records, data collection for use in performance measurement, and reimbursement for Medicare/Medicaid medical claims. ICD is an epidemiological classification used to identify diagnoses (diseases, injuries, and impairments). The U.S. version also includes procedures (surgical, diagnostic, and therapeutic).

Although the ICD-9-CM Coordination and Maintenance Committee is a federal committee, suggestions for modifications come from both the public and private sectors. Interested parties are asked to submit recommendations for modification prior to a scheduled meeting. Proposals submitted for a new code should include a description of the code being requested and rationale for why the new code is needed, along with supporting references and literature. Proposals should be consistent with the structure and conventions of the classification.

ICD-9-CM Coordination and Maintenance Committee meetings are open to the public; comments are encouraged both at the meetings and in writing. Recommendations and comments are carefully reviewed and evaluated before any final decisions are made. No decisions are made at the meetings. The committee's role is advisory. All final decisions are made by the National Center for Health Statistics (NCHS) director and the CMS administrator. Final decisions are made after the December meeting and become effective October 1 of the following year.[145]

On January 16, 2009, HHS released a 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.[146]

The current ICD-9-CM system does not provide the necessary detail for patients' medical conditions or the procedures and services performed on hospitalized patients. The new classification system provides significant improvements through greater detailed information and the ability to capture additional advancements in clinical medicine.

---

[145] Centers for Disease Control and Prevention. *Classification of Diseases, Functioning and Disability: ICD-9-CM Coordination and Maintenance Committee*. Available at: http://www.cdc.gov/nchs/icd/icd9cm_maintenance.htm. Accessed on: June 1, 2015.

[146] 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: June 1, 2015.

ICD-10-CM/PCS consists of two parts:

- ICD-10-CM—Diagnosis classification system developed by the CDC for use in all U.S. healthcare treatment settings. Diagnosis coding under this system uses 3–7 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 only in the U.S. for inpatient hospital settings. The new procedure coding system uses 7 alpha or numeric digits, whereas the ICD-9-CM coding system uses 3 or 4 numeric digits.[147]

When a developer submits ICD-10 codes, then the following requirements should also be met:

- Provide a statement of intent for the selection of ICD-10 codes, chosen from the following:
- Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.
- Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.
- The intent of the measure has changed.
- Provide a spreadsheet, including:
- A full listing of ICD-9 and ICD-10 codes, with code definitions.
- The conversion table (if there is one).
- Provide a description of the process used to identify ICD-10 codes, including:
- Names and credentials of any experts who participated.
- Name of the tool used to identify/map to ICD-10 codes.
- Summary of stakeholder comments received.[148]

Current Procedural Terminology, Fourth Edition (CPT4®)

CPT is a registered trademark of the AMA for the Current Procedural Terminology, Fourth Edition (CPT4). The CPT Category I (or CPT codes) is 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 will thereby provide an effective means for reliable nationwide communication among physicians, patients, and third parties.[149] This code set is updated annually.

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.

---

[147] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *ICD-10 CM/PCS – An Introduction*. Medicare Learning Network. ICN 901044;Apr 2013. Available at: http://www.cms.gov/ICD10/Downloads/ICD-10Overview.pdf. Accessed on: June 1, 2015.

[148] National Quality Forum. *ICD-10-CM/PCS Coding Maintenance Operational Guidance: A Consensus Report.* National Quality Forum. Oct 2010. Available at: http://www.qualityforum.org/Publications/2010/10/ICD-10-CM/PCS_Coding_Maintenance_Operational_Guidance.aspx. Accessed on: June 1, 2015.

[149] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Data Submission Specifications Utilizing HL7 QRDA Implementation Guide Based on HL7 CDA Release 2.0 Version: 2.0* Last Modified: July 01, 2010 Available at: https://www.cms.gov/PQRI/20_AlternativeReportingMechanisms.asp#TopOfPage. Accessed on: June 1, 2015.

CPT Category II (or 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 II codes are only updated annually and are not modified or updated during the year.

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.[150]

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.

SNOMED CT

Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) is a registered trademark of SNOMED International.[151] 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 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, and ICD-9-CM mapping; and the Technical Reference Guide.

Each SNOMED record corresponds to a single observation. The SNOMED codes are not intended to transmit all possible information about an observation or procedure. They are only intended to identify the observation or procedure. The SNOMED code for a name is unique and permanent.

SNOMED CT combines the content and structure of the SNOMED Reference Terminology (SNOMED RT) with the United Kingdom's Clinical Terms Version 3 (formerly known as the Read Codes).

---

[150] American Medical Association/Centers for Medicare & Medicaid Services. Current Procedural Terminology (CPT) Copyright Notices and Disclaimers and Point and Click License. 2011.

[151] Health and Human Services, Centers for Medicare & Medicaid Services. *Data Submission Specifications Utilizing HL7 QRDA Implementation Guide Based on HL7 CDA Release 2.0 Version: 2.0.* Last Modified: July 01, 2010. Available at: https://www.cms.gov/PQRI/20_AlternativeReportingMechanisms.asp#TopOfPage. Accessed on: June 1, 2015.

For information on obtaining the standard, contact:

SNOMED International
College of American Pathologists
325 Waukegan Rd
Northfield, IL 60093-2750
http://www.ihtsdo.org

Logical Observation Identifier Names and Codes (LOINC®)

LOINC[152] codes are available for commercial use without charge, subject to the terms of a license that assures the integrity and ownership of the codes. The LOINC database provides sets of universal names and ID codes for identifying laboratory and clinical observations and other units of information meaningful in cancer registry records.

Each LOINC record corresponds to a single observation. The LOINC codes are not intended to transmit all possible information about a test or observation. They are only intended to identify the observations. The LOINC code for a name is unique and permanent. LOINC codes must always be transmitted with a hyphen before the check digit (e.g., "10154-3"). The numeric code is transmitted as a variable length number, without leading zeros.[153] More information can be found at the LOINC website.

To obtain the LOINC database, contact:

Regenstrief Institute
1001 West 10th Street RG-5
Indianapolis, IN 46202

RxNorm

RxNorm[154] is the recommended national standard for medication vocabulary for clinical drugs and drug delivery devices produced by the NLM. RxNorm is intended to cover all prescription medications approved for human use in the United States.

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.

---

[152] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Data Submission Specifications Utilizing HL7 QRDA Implementation Guide Based on HL7 CDA Release 2.0 Version: 2.0.* Last Modified: July 01, 2010. Available at: https://www.cms.gov/PQRI/20_AlternativeReportingMechanisms.asp#TopOfPage. Accessed on: June 1, 2015.

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

[154] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Data Submission Specifications Utilizing HL7 QRDA Implementation Guide Based on HL7 CDA Release 2.0 Version: 2.0.* Last Modified: Jul 01, 2010. Available at: https://www.cms.gov/PQRI/20_AlternativeReportingMechanisms.asp#TopOfPage. Accessed on: June 1, 2015.

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. More information can be found at the RxNorm website.[155]

Step 5 of 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 lists

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 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."

Medical record data from EHRs (for eMeasures, 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. The eMeasure Specifications chapter of the eMeasure Lifecycle section describes the process in more detail. Information captured by an EHR, but is not coded in such a way that a computer program can extract the information, and thus requires manual abstraction, is not considered electronic data.

[155] National Institutes of Health, National Library of Medicine. *Unified Medical Language System® (UMLS®)*. ®UMLS is a registered trademark of the National Library of Medicine. Available at: http://www.nlm.nih.gov/research/umls/rxnorm/docs/index.html. Accessed on: June 1, 2015.

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: the problem list, the history and physical, and the 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-9-CM code 410.xx, excluding those with 410.x2 (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.

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 (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.

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.

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.[156]

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 Measure Information Form[157] 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. The Risk-Adjusted Outcome Measures section of this chapter provides details of the procedure.

Clearly define any time windows

Time windows must be stated whenever they are used to determine cases for inclusion in the denominator, numerator, or exclusion. Any index event used to determine the time window is to be defined. 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.
- Exactly state the 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. Appendix E: Time Unit and Time Interval Definitions 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 clinical quality measures.[158]

---

[156] National Quality Forum. *CSAC Guidance on Quality Performance Measure Construction*. May 2011. Available at: https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=68999. Accessed on: June 1, 2015.

[157] 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: June 1, 2015.

[158] International Organization for Standardization *Data elements and interchange formats — Information interchange — Representation of dates and times.* ISO 860:3rd ED:2004. Available at: http://dotat.at/tmp/ISO_8601-2004_E.pdf. Accessed on: June 1, 2015.

Example:

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

Describe how the measure results are scored and reported

Most quality measures produce rates; however, there are other scoring methods such as categorical value, continuous variable, 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 pre-defined, discrete, non-continuous classes (male, female), (board certified, not board certified). Categories may reflect a natural order, in which case they are called ordinal (cancer stage: I, II, III, or IV), (hospitals rankings: good, better, best).
- Continuous variable: A measure score in which each individual value for the measure can fall anywhere along a continuous scale (for example, 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 (for example, 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 (in other words, 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).[159]

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: race, ethnicity, age, socioeconomic status, income, region, sex, primary language, disability, or other classifications.

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 continuous variable 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 Measure Information Form 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

[159]National Quality Forum. *CSAC Guidance on Quality Performance Measure Construction*. May 2011. Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: June 1, 2015.

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.

Appendices G and H provide explicit instructions for calculating proportion, ratio, and continuous variable eMeasures and the process to be used to determine individual and aggregate scores. This ensures that all implementers will obtain the same scores, given the same data and same eMeasures.

Step 6 of 6: Document the measures and obtain COR approval

Complete the detailed technical specifications, including any additional documents required to produce the measure as it is intended. The complete specifications, including all attachments, are documented in the Measure Information and Measure Justification forms.[160]

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 Measure Information and Measure Justification Forms to the COR for approval.

The Measure Information and Measure Justification Forms 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.

2.3—Harmonizing Measure Specifications

When specifying measures, measure developers should consider if a similar measure exists for the same condition, process of care, outcome, or care setting. This consideration ensures that the measure under development or maintenance is harmonized with existing measures, and promotes harmonization throughout the measure lifecycle. Measures should be harmonized unless there is a compelling reason for not doing so. 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

[160] 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: June 1, 2015.

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.
- Performance time period.
- Allowable values for medical conditions or procedures; code sets, descriptions.
- Allowable conditions for inclusion in the denominator; code sets, descriptions.
- Exclusion categories, whether exclusion is from the denominator or numerator, whether optional or required.
- Calculation algorithm.
- Risk adjustment methods.

Examples:

- Ambulatory diabetes measures exist, but the new diabetes measure is for a process of care different from existing measures.
- 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 attributes 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 may be that:

- 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.

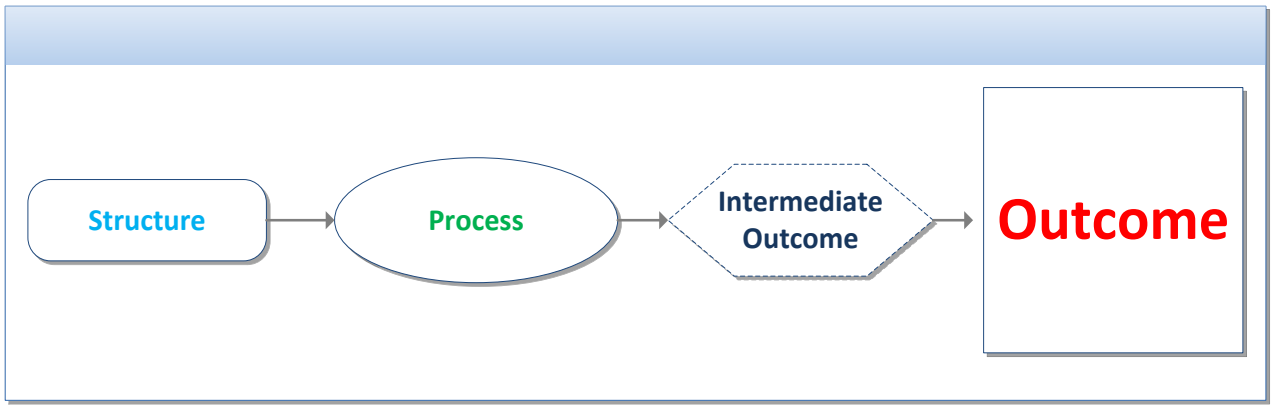
Examples:

- 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.

2.4—Risk-Adjusted Outcome Measures

Performance measures include three basic types of measures used to assess the quality of healthcare: structure, process, and outcome[161]. 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 19 depicts the relationship between structures, processes, and outcomes that measures should evaluate.[162]

Figure 19: Structure Process Outcome Relationship



Though multiple provisions in the ACA support the development of quality measures, Section 10303 specifically calls for outcome measures to be developed.[163] Outcomes capture what patients and society care most about—the results of care. However, when constructing performance measures using outcomes, the outcomes have to 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. The Technical Specifications chapter of the Blueprint discusses issues around development of risk-adjusted outcome measures including selecting appropriate outcomes, time intervals, clinically relevant variables, and risk adjustment models. There is much agreement on the need for outcome measures, but the

[161] 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.

[162] 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: June 1, 2015.

[163] 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: June 1, 2015.

infrastructure to capture patient-reported outcomes routinely and use them to measure performance is still being built.

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 combinations to "level the playing field" when comparing healthcare outcomes achieved by healthcare services and providers.

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.

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.[164] 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. Adjusting for these risk factors can help to avoid misleading comparisons. However, risk adjustment models for publicly reported quality measures should not obscure disparities in care associated with race, socioeconomic status, or gender. 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)[165] 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

[164] 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.

[165] NQF policy generally precludes using risk factors that obscure disparities in care associated with factors such as race, socioeconomic status, or gender.

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.[166] 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.

Measure developers developing risk-adjusted eMeasures should note that risk variables in risk-adjusted outcome eMeasures are currently represented as supplemental data elements and are represented as measure observation in the July 2014 Measure Authoring Tool (MAT) release. The MAT's capabilities for representing risk-adjusted eMeasures were expanded in the Health Quality Measures Format (HQMF) Release 2.

Deliverables

- 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.
- Measure Information Form with completed risk adjustment sections for each measure.[167]
- Measure developers developing eMeasures must submit electronic specifications (eMeasure XML file, SimpleXML file, eMeasure human-readable rendition [HTML] file) that include instructions where the complete risk adjustment methodology may be obtained.

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 the following table, which was partially derived from a description of preferred attributes of models used for publicly reported outcomes.[168] Each of the attributes listed in the table are described in detail in the sections below.

Table 8: 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

[166] Boston, MA. 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.
[167] 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: June 1, 2015.
[168] 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.

Attribute	Description
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

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.

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).

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 are as follows:

- The data were 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 was 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.

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 subject matter experts. 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 rarely is 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.

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.[169] 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.

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. Recent HHS policies emphasize transparency.[170] NQF policy on the endorsement of proprietary measures promotes the full disclosure of all aspects of a risk adjustment model used in measure development.[171]

The risk adjustment method used, performance of the risk adjustment model and its components, algorithms, as well as 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.

Procedure

The following 7 steps are recommended in the development of a risk adjustment model. 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.

Step 1 of 7: 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., re-admission rates, mortality, and

[169] 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.

[170] 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: June 1, 2015.

[171] National Quality Forum. *Measuring Performance: Measure Evaluation Criteria. Available at:* http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx. Accessed on: June 1, 2015.

length of stay).[172] 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 the Measure Testing chapter. 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 the chapter on Stakeholder Input, 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 variables in risk-adjusted outcome eMeasures are currently represented as supplemental data elements and are represented as measure observation in the July 2014 MAT release. Refer to the eMeasure Lifecycle section of the Blueprint for more details.

Step 2 of 7: 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 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:

- 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. Chapter 1.B.—Technical Expert Panel covers the standardized process used for convening a TEP.

[172] 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: June 1, 2015.

- When appropriate data are available, automated computer routines can be used to identify potential factors for consideration by subject matter experts.

Step 3 of 7: 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, socioeconomic status, 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.

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.

Medicaid status

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

Sex

Males and females often show differences relative to treatment, utilization, risk levels for mortality and disease, and needed level of care. Though restricting a measure to a single sex using

exclusion criteria may be reasonable, inclusion of sex in a risk model should be avoided to prevent obscuring disparities in care. It is preferable to stratify by sex, rather than use it in a risk adjustment model.

Step 4 of 7: Acquire data (sample, if necessary)

Health care 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 Chapter 3—Measure Testing provides more details of the processes.

Step 5 of 7: 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.

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 in order to consistently return the same model statistics across samples. If the outcome is uncommon, then the number of cases required could be much larger.[173] 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.

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

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

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).

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.[174] 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.

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.[175] 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.

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

[174] 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).

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

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.

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, non-profit 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.[176]

Step 6 of 7: 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.

[176]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: June 1, 2015.

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.

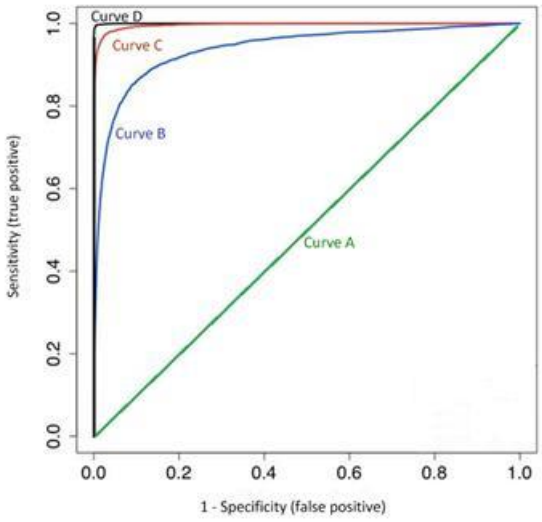
ROC curve, AUC, and C-statistic

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 20. Curves approaching the 45-degree diagonal of the graph represent less desirable models (see curve A) when compared to curves falling to the left of this diagonal that indicate higher overall accuracy of the model (see curves B and C). A test with nearly perfect discrimination will show a ROC curve that passes through the upper-left corner of the graph, where sensitivity equals 1, and 1 minus specificity equals zero (see curve D).

The power of a model to correctly classify outcomes into two categories (i.e., discriminate) is often quantified by the area under the ROC 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 20: Example of ROC Curves



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.

Step 7 of 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:

- 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.[177]
- 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 Measure Information Form. 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

[177] 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.

open source requirements of NQF's Conditions for Consideration,[178] 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.

2.5—Cost Resource Use Measure Technical Specifications

It is important to submit instructions and analytic steps for aggregating data when designing cost and resource use measures.[179] 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

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.

Measure construction logic

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 window for measure reporting is the calendar year.

---

[178] National Quality Forum. *Measuring Performance: Measure Evaluation Criteria. Available at:* http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx. Accessed on: June 1, 2015.

[179] 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: June 1, 2015.

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.

Adjusting for comparability

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.

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.

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

Measure reporting

Attributing resource use measures

Resource 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.

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.

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 multi-level regression are also used.

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.

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.

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.

2.6—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 and Measure Justification forms 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 Chapter 5—Measure Evaluation in Section 1.

Methodology and considerations for scoring

- Ensure that the weighting and scoring of the components support the goal that is articulated for the measure.
- 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 6. 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.[180] The five types discussed are:

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

Table 9: 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.	Promotes a high standard of excellence. Patient centric. Fosters a system perspective.	May waste valuable information. May weight common but less important processes more heavily than infrequent but important processes.	Minnesota Community Measurement Optimal Diabetes Care measure. QIO 8th SOW Appropriate Care Measure (ACM) used hospital AMI, HF, and PN process measures.[181]

[180] Much of the information in the table was obtained from Peterson et al. (2010).

[181] ACM Calculation: Add the total number of patient ACM numerators and divide by the total number of patient ACM denominators across the three topics (AMI, HF, PN) to calculate the ACM percentage. Example: a hospital has four cases submitted to the QIO clinical warehouse. Case #1 is an AMI patient that is qualified (eligible) for 3 of the 5 AMI measures and passes 2 of them. Case #2 is a HF patient that is qualified (eligible) for

Type of Scoring	Advantages	Disadvantages	Examples/Evidence
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.	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).	The provider who achieved 4 of 5 measures appears the same as the provider who achieved none of 5 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.	IHI Bundles: ventilator, central line. STS Perioperative Medical Care, a process bundle of 4 medications (preoperative beta blockade and discharge anti-platelet, beta blockade, and lipid-lowering agents). Study using Premier 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. [182]
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.
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	Premier/CMS Hospital Quality Incentive Demonstration uses a composite of process and outcome to measure quality for CABG. The composite quality score (CQS) was based on an equally weighted combination of 7 measures (4 process measures and 3 outcome measures). The actual

both of the 2 HF measures and passes both of them. Case #3 is a PN patient that is qualified (eligible) for 2 of the 3 PN measures and passes both of them. Case #4 is a HF patient that is not qualified for either of the 2 HF measures. For the ACM calculation, cases #1, #2, and #3 are in the denominator. Cases #2 and #3 are in the ACM numerator. The hospital’s ACM rate is 2/3 or 66.7%.

[182] 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
		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.	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 3 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 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.
Opportunity Scoring Opportunity scoring counts the number of times a given care process was actually	Provides an alternative to simple averaging often used for aggregating individual process measures.	Rate is influenced by the most common care processes, regardless of	The opportunity model was developed for the Hospital Core Performance Measurement Project for

Type of Scoring	Advantages	Disadvantages	Examples/Evidence
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.	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.	whether they are the most important methods.	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 eMeasures using weighted methodologies (such as regression-based) cannot be represented in current standards for HQMF. There is no way to:

- Express composite score calculation
- Reference the component measures

Approaches are being explored for HQMF to help resolve these issues. However, HQMF Release 2.1 supports composite measure metadata.

Chapter 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.[183] 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 scales (and the items in the scales if applicable), and the performance score.

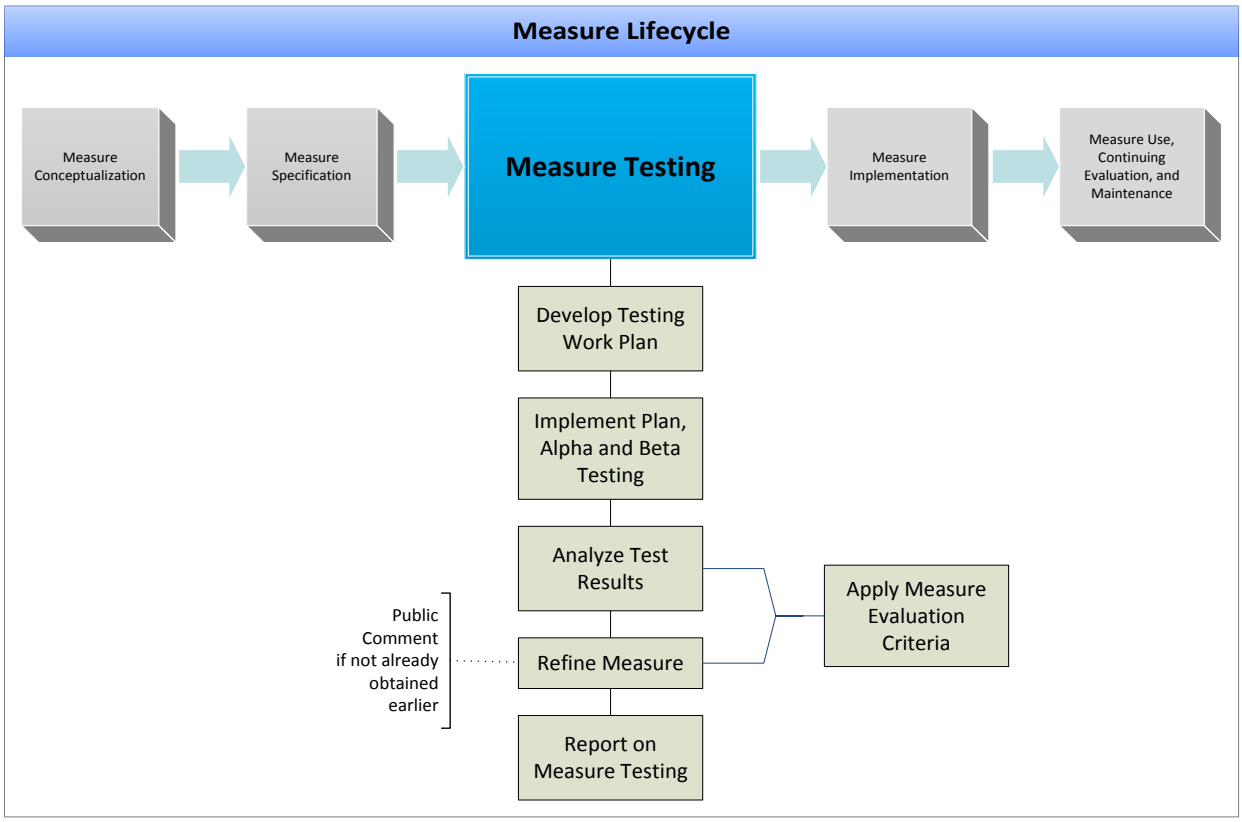
Properly conducting measure testing and analysis is critical to approval of a measure by CMS and endorsement by the NQF. This chapter describes the types of testing that may be conducted during measure development (alpha and beta testing), the procedure for planning and testing under the direction of the CMS COR, and key considerations when analyzing and documenting results of testing and analysis.

The information in this chapter is not meant to be prescriptive or exhaustive, as 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.

Figure 21: Flow of the Measure Lifecycle - Measure Testing describes how testing fits into the flow of the measure lifecycle measure. Though both alpha and beta testing are discussed in this chapter, in the diagram, the measure testing box shows where beta testing occurs in the process. Alpha testing may occur as early as information gathering and is repeated iteratively during the development of measure specifications.

[183] NQF Measure Evaluation Criteria. Available at: https://www.qualityforum.org/docs/measure_evaluation_criteria.aspx, updated October 11, 2013). Accessed on: June 1, 2015.

Figure 21: Flow of the Measure Lifecycle - Measure Testing



3.1—Deliverables

- Measure Testing Plan
- Measure Testing Summary Report
- Updated Measure Information Form[184]
- Updated Measure Justification Form
- Updated Measure Evaluation Report

3.2—Alpha and Beta Testing

A measure should be tested one or more times during the development process. 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. Initial testing during development (sometimes referred to as pilot testing) is generally conducted within the framework of alpha and beta tests. Attributes of each type of test are shown below in Table 10: Features of Alpha and Beta Testing, and these may be used as considerations when developing a work plan for alpha or beta tests.

[184] 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: June 1, 2015.

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 data source to be used for the concept, 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.

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.

Table 10: 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.

	Alpha Testing	Beta Testing
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.
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.

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:

- 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.

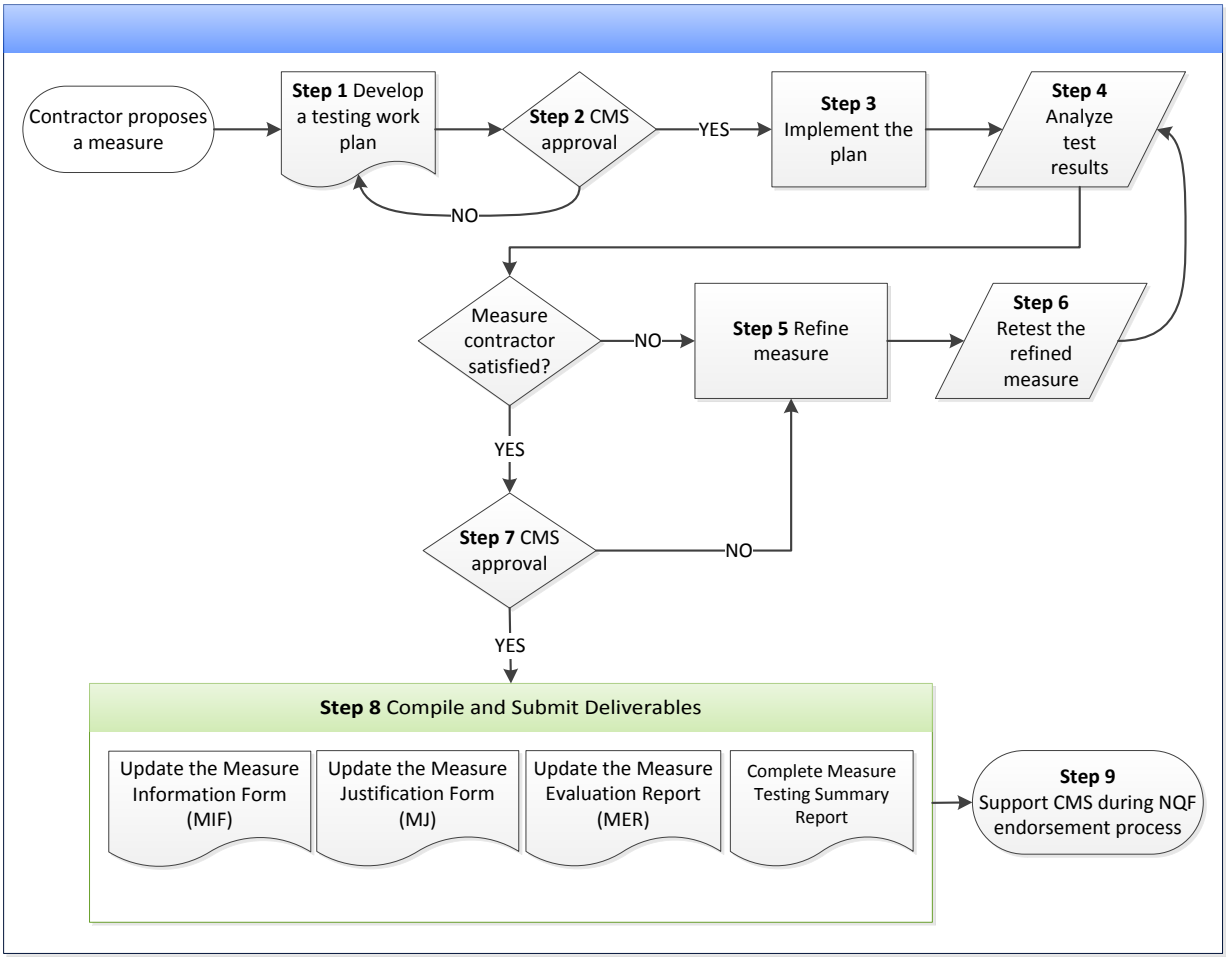
However, when determining the appropriate sample size during testing, it is necessary to evaluate the burden placed on providers and/or beneficiaries to collect the information. The PRA mandates that all federal government agencies obtain approval from the OMB before collection of information that will impose a burden on the general public. However, with the passage of the Medicare Access and CHIP Reauthorization Act of 2015, data collection for quality measure development is now exempt from PRA requirements. [185] Measure developers should consult with their COR about the ramifications of the PRA and MACRA exemption before requesting information from the public.

3.3—Procedure

When developing a measure (or set of measures) for CMS, a measure developer is required to submit specific reports and is encouraged to follow the steps listed below. The first few steps address planning and implementation of the testing, and are identical for alpha and beta testing. The last steps address reporting and follow-up after the conclusion of testing. 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 22: Overview of Measure Testing shows the relationships between the 8 steps of measure testing.

[185] 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.

Figure 22: Overview of Measure Testing



Step 1 of 8: 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)
- 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 the following 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 Paperwork Reduction Act, if relevant.[186]
- 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

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

[186] 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: June 1, 2015.

Step 3 of 8: Implement the plan

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

Step 4 of 8: 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 and face validity of the measures. The findings of all testing analyses will be presented in a final summary report described in Step 10 and discussed with the COR.

Step 5 of 8: 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. For example:

- Following alpha testing, measure re-specification 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.

Step 6 of 8: 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.

Step 7 of 8: 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.[187]

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.

[187] 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: June 1, 2015.

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:

- 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 information:

- 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 the section on Risk-Adjusted Outcome Measures.
  - 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
- For limitations specific to eMeasures, refer to Chapter 3—eMeasure Testing of the eMeasure Lifecycle section.

Step 8 of 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.

3.4—Testing and Measure Evaluation Criteria

Step 4 of the measure testing procedure describes the analysis of testing results. The results 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.

Application of the testing results to each of the four measurement areas (importance, scientific acceptability, usability, and feasibility) is discussed below.

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 measured structure 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 intra-class correlation

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.[188] 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.
- **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 subject matter experts 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.

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

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.

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 of 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.

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 intra-class correlations for ratings on a scale, where variation between raters is quantified for raters randomly selected to rate each occurrence.

- eMeasures that are implemented as direct queries to EHR databases may not use abstraction. Therefore, inter-rater reliability may not be needed for eMeasures.

Form equivalence reliability (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 (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.[189] 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 intra-class 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.

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 we are trying to assess) and process (i.e., how we actually assess it).

[189] 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.

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.[190]

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

[190] 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.

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 modified Delphi approach where participants systematically rate their agreement, and formal aggregating and consensus failure processes are followed.[191] 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. [192]This type of formal process can also be used when addressing whether specifications of the measure are consistent with medical evidence.

Testing 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.[193]

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., ICD, CPT) can be compared to manual abstraction from a sample of medical charts.
- 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 clinical record information—EHR information extracted using automated processes based on measure technical specifications can be compared to manual abstraction of the entire EHR (not just the fields specified by the measure). 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.

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

[191] 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.
[192] National Quality Forum. Review and Update of Guidance for Evaluating Evidence and Measure Testing, Technical Report, Approved by CSAC on October 8 2013: 12.
[193] Ibid.

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 may be used to calculate the measure score or computed measure score since the two concepts are both mathematically and conceptually related.[194] 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.

NQF provides the following guidance:[195]

- 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.

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 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.
- Evidence that measure elements (e.g., codes) used to identify exclusion/exception are valid.

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 reported both with and without the exception, and the proportion of exception should be included for any group-level tabulations.

---

[194] 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.

[195] 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.

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 the Risk-Adjusted Outcome Measures subsection of the Measure Specification chapter.

Usability

Formal usability 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 respect to intended audiences (e.g., consumers, purchasers, providers, and policy makers), a variety of methods are available. These include:

- 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.

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

3.5—Testing Considerations for Selected Measure Types

For more information about special considerations pertaining to eMeasure testing, see the eMeasure Testing chapter of the eMeasure Lifecycle section.

Adapted measures

When adapting a measure for use in a new domain (e.g., new setting or population), construct the measure testing to detect important changes in the functionality or properties of the measure. As applicable, review changes in the:

- 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 eMeasures are adapted for use in different settings or with different populations, the adapted measures must also be tested and evaluated. Procedures for testing eMeasures are discussed in Chapter 3—eMeasure Testing in Section 3—eMeasure Lifecycle.

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 composites create 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.

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.

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.[196]

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 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 include:

- 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.

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

- 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.

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.

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, 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.

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 with regard to the decisions on component retention and weighting.

If the majority of the composite's variation is the result of only a subset of the components used for the composite, also provide information (e.g., a table) on the contribution of each of the components to the composite (e.g., regression coefficients or factor loadings to address which subset of components is contributing to the majority of the aggregate's variation. The variation (i.e., information content) of a composite might be conveyed in a variety of ways, such as through

reporting of regression results, factor loadings, and percentages of shared variation explained from a principal components analysis.

The results of the composite evaluation process might not be well aligned with the separate results for each of the components in the composite measure, as the composite may primarily reflect a minority of the components of the composite. For example, group differences on an emergency room composite measure may be largely determined by emergency department 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.

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.

Chapter 4—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:

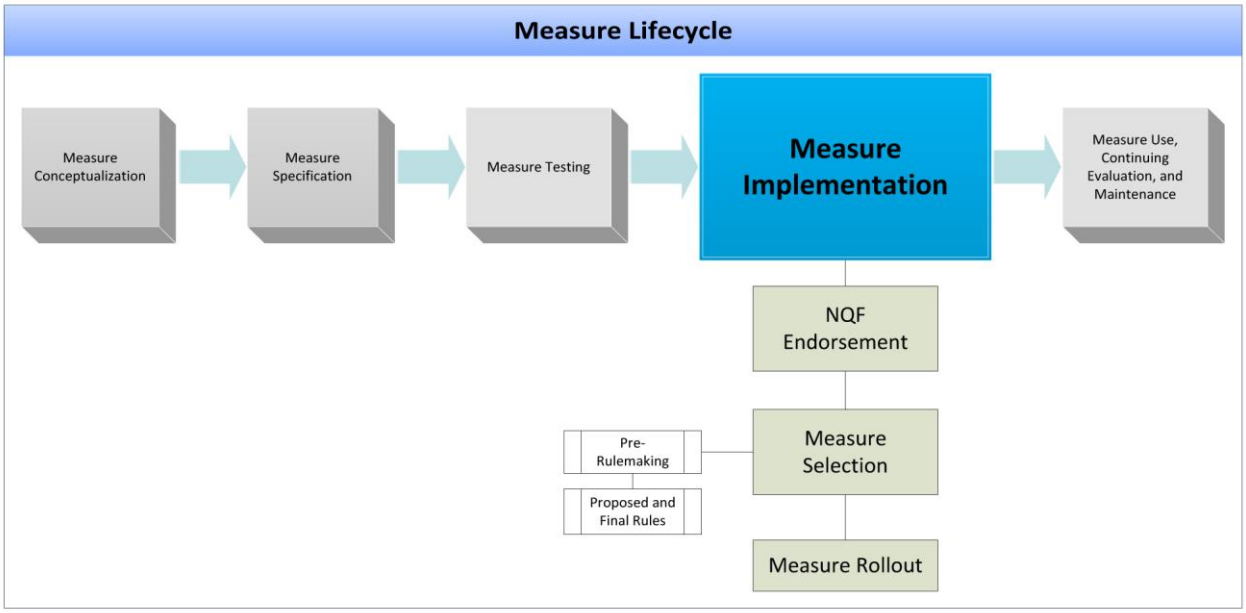
- Scope of measure implementation
- Health care provider being measured
- Data collection processes
- Ultimate use of the measure (e.g., quality improvement, public reporting, pay-for-reporting, or value-based purchasing)
- Program into which the measure is being added

The scope of measure implementation could be one of the following:

- A measure or measure set is implemented in a new program
- A measure or measure set is added to an existing program

This chapter describes the initial implementation of measures. Ongoing use of the measures is described in the next chapter on measure use, continuing evaluation, and maintenance. Depending on the scope of implementation factors listed above, Figure 23 depicts the process of measure implementation described in this chapter.

Figure 23: Flow of the Measure Lifecycle - Measure Implementation



4.1—National Quality Forum Endorsement

The National Quality Forum (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.

The information about the measure developer's role is based on NQF's Consensus Development Process, version 1.9. However, measure developers are responsible for reviewing the latest version of the NQF measure endorsement processes and resources found on the NQF website.[197]

NQF works to continuously improve its processes. In 2012, NQF created a task force charged with:

- Reviewing different approaches to reaching and establishing consensus
- Identifying the strengths and weaknesses of the current process
- Recommending enhancements to the current process

NQF is currently redesigning its consensus development process to incorporate Lean principles and is coordinating Kaizen improvement efforts with CMS and the ONC. This process is evolving, and measure developers are strongly encouraged to review updates on the NQF website and attend measure developer webinars and meetings hosted by NQF.

NQF is currently piloting consensus development process improvements that they plan to implement more broadly in the future. All projects are piloting standing steering committees, as compared with the previous process of selecting topic-specific steering committees for each project. More frequent measure submission is being piloted in the endocrine endorsement project and continuous commenting on measures is being piloted in the readmissions endorsement project. Based on the results from the pilots, NQF may adopt these as part of its regular consensus development process.

Procedure for measure submission

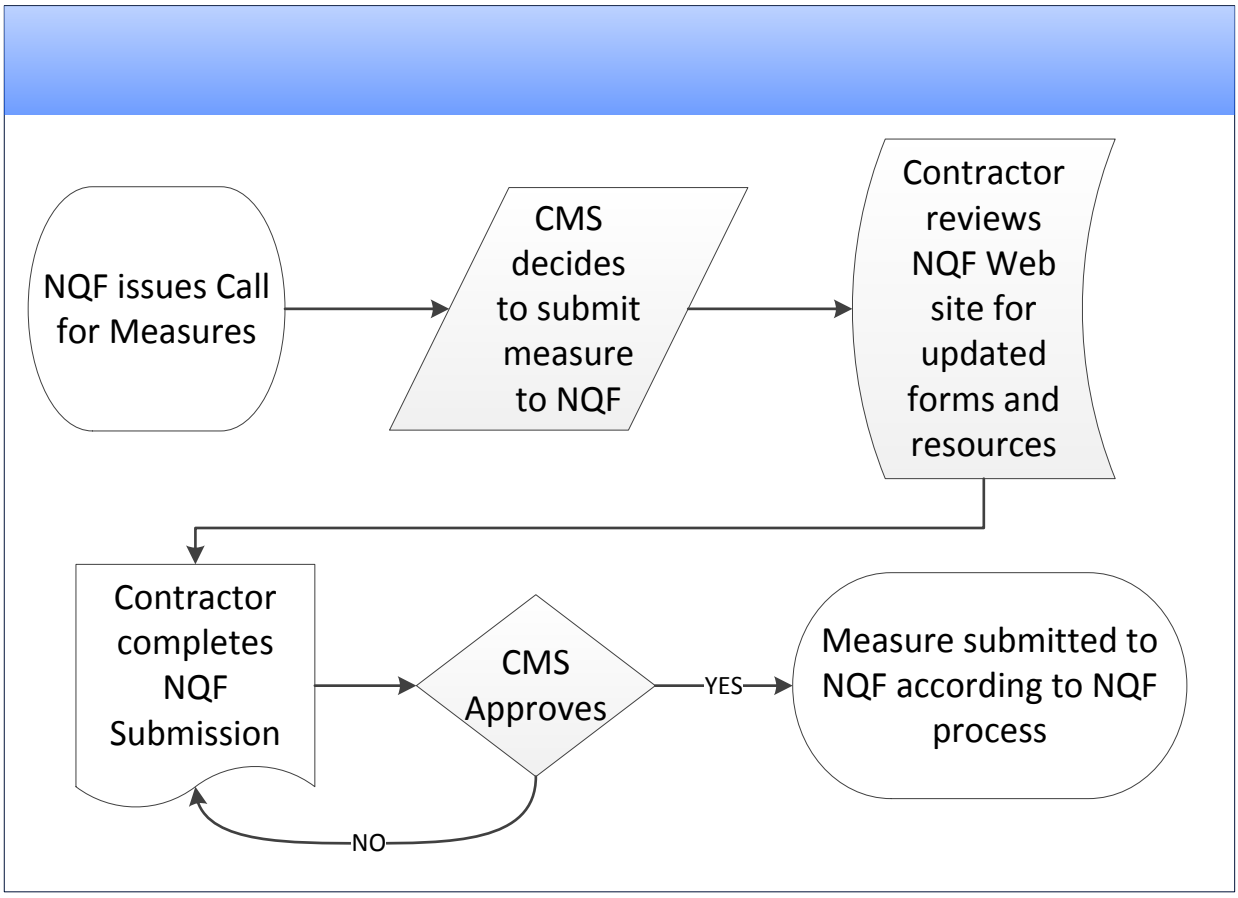
NQF's measure endorsement process is standardized in a regular cycle of topic-based measure consideration. NQF endorses measures as part of a larger consensus development project to seek consensus standards (measures) for a given healthcare condition or practice. NQF follows a three-year schedule that outlines the review and endorsement of measures in major topic areas such as Patient Safety, Person and Family (caregiver)-Centered Care, and Disparities. NQF revises these topic and subtopic areas as indicated to address current measurement needs. A public call for nominations is posted on the NQF website to select appropriate members for the Steering

[197] http://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process.aspx

Committee. NQF projects follow a seven-month timeline from measure submission to the appeals period.

The measure submission process is summarized in Figure 24: Measure Submission and in steps 1 through 5 that follow. Steps 6 through 9 cover the NQF review and endorsement processes.

Figure 24: Measure Submission



Step 1 of 9: NQF issues a Call for Measures

At least two 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 a large number of 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.

Step 2 of 9: 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.

With the introduction of the EHR incentive programs, there is a movement toward the development and/or retooling of measures specified for use in the electronic health record (or eMeasures). The COR will provide guidance as to which eMeasures are candidates for NQF submission. These eMeasures, which are encoded in the HQMF, must meet specific NQF submission criteria. Measure developers are responsible to monitor NQF's eMeasure requirement policies prior to submission. The eMeasure Implementation chapter of the eMeasure Lifecycle section provides more detailed information.

Step 3 of 9: The measure developer completes the NQF measure submission

Measure developers must submit their measures via an online measure submission.[198] The online form is available on the NQF website and allows users to:

- Gain secure access to the submission form from any location with an Internet connection.
- Save a draft version of the form and return to complete it at their convenience.
- Print a copy of the submission form for reference or other uses, if desired.

When initiating an online measure submission, the measure developer may contact NQF and request access for additional users to enter data in the online form. 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 over-written.

The CMS Measure Information Form and Measure Justification Form have been aligned with the most recent NQF measure submission 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 Measure Information Form and Measure Justification Form are CMS forms designed to present measures in a standardized way. Every effort has been made to ensure that the Measure Information Form and Measure Justification Form are aligned with the NQF measure submission to facilitate online information entry.[199]

[198] http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx

[199] 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: June 1, 2015.

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 Steering 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. Listed below are a few tips for successful submissions:

- 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 exclusion.
- 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, or for other technical support, NQF maintains a website with Frequently Asked Questions and other help. Questions about the content or information required by the online submission form should be directed to the NQF project director whose name and contact information appear on the project's Information page on the NQF website. Measure developers are expected to have worked closely with the Measures Manager throughout development—and provided all measure development deliverables—to ensure that no duplication of measure development occurs and to identify potential harmonization opportunities prior to NQF submission.

The search for related and competing measures should be conducted early during the Information Gathering phase of development and again just prior to submission to NQF. Before NQF will even consider a measure that is submitted, the measure developer must attest that harmonization with related measures and issues with competing measures have been considered and addressed, as appropriate. Measure Evaluation criterion 5 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 Steering Committee reviewing the measures would then request that both responsible measure developers create a harmonization plan addressing the possibility and challenges of harmonizing their respective measures. NQF will consider the response and decide whether to recommend the measure for endorsement.

Step 4 of 9: 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.

Step 5 of 9: 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 in steps 6 through 8.

Step 6 of 9: 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 three years.
- The intended use of the measure includes both accountability applications (including public reporting) and performance improvement to achieve high-quality, efficient healthcare.
- The measure is fully specified and tested for reliability and validity.
- 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.[200]

Once this is done, the measures are accepted and are considered ready for review by the project Steering Committee.

Step 7 of 9: Steering Committee and Technical Advisory Panel Review

After the measures are accepted, the measures are reviewed by a Steering Committee that has been selected from nominations submitted during a public call for nominations. This Steering Committee oversees the work of NQF consensus development projects, offers expert advice, ensures that input is obtained from relevant stakeholders, and makes recommendations to the NQF membership about measures that are proposed for endorsement.

A technical advisory panel may also be used depending on the technical expertise of the Steering Committee. The technical advisory panel members are experts in their field and provide guidance to the Steering Committee around specific technical issues related to some or all of the measures under review.

The Steering Committee evaluates each submitted measure using the NQF Measure Evaluation Criteria. The Steering Committee evaluates each measure against the importance criterion first. Only if the measure meets the Importance criterion is it then evaluated against the other criteria. After review, the Steering Committee recommends that a measure either continues through the consensus development process or is returned to the measure steward and/or developer for further development and/or refinement. The measure developer should be available to provide an overview and respond to Steering 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 Steering 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 approved measures are then sent to the Consensus Standards Approval Committee (CSAC), a standing committee of the NQF Board of Directors. 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.

Step 8 of 9: NQF determines type of endorsement appropriate for the measure

The CSAC reviews the recommendations of the Steering Committee and the results of the NQF member voting period. As during the Steering Committee review, the measure developer and COR (and possibly a member from the TEP) involved in measure development should attend the CSAC

[200] National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: June 1, 2015.

meeting in person or by teleconference to answer any questions that the CSAC members may have about the measures.

After review of the measure, the CSAC may recommend one of the following:

- Grant full endorsement—requiring no further documentation. According to NQF, full endorsement is given to measures meeting NQF endorsement requirements.
- Defer endorsement—this designation is generally used when the CSAC requires additional information from the measure steward until such information can be submitted and reviewed.
- Decline to endorse the measure altogether—according to NQF, measures are not endorsed when they do not fully meet NQF criteria for endorsement.
- For eMeasures, CSAC may make a fourth determination of a measure's status. CSAC may grant a trial use approval status, which approves (not endorses) a candidate eMeasure for a limited time period if the measure meets all NQF requirements except adequacy of testing. Prior to granting trial use approval status to an eMeasure, the CSAC reviews and approves the measure developer's plan and timeline for testing and provision of results to the CSAC.

All of the CSAC's approval decisions are forwarded to the NQF Board of Directors for ratification. After a consensus standard (measure) has been formally endorsed by NQF, any interested party may file an appeal of the endorsement decision with the NQF Board of Directors. CSAC will make a recommendation to the NQF Board of Directors regarding the appeal. The Board of Directors will take action on an appeal within seven calendar days of its consultation with the CSAC.

Step 9 of 9: Review endorsement results

The measure developer meets with the measure developer's COR and the Measures Manager to review the NQF endorsement results and identify lessons learned.

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 Measures Management System processes, and discuss potential next steps.

Measure developer's role during NQF review

During its review, NQF may have questions about the submitted measure. These questions may come from NQF staff during initial measure submission 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 contactor's team who is prepared to explain and defend the measure should attend the Steering 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 eMeasures, the measure developer, with support from a HQMF or eMeasure standards subject matter expert, will communicate and collaborate with NQF during the review.

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. Refer to the eMeasure Implementation chapter of the eMeasure Lifecycle section for more information.

During the NQF review, 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.

Trial use approved measures

On a very limited basis, NQF may grant an eMeasure trial use approval. Measures with trial use approval will lose that status and be de-endorsed if the measure developer fails to provide testing results or seek an extension before the trial endorsement period expires. They will also be de-endorsed if the testing results fail to meet the NQF's measure evaluation criteria. The NQF Trial Use Approval Policy can be found on the NQF website.[201]

Expedited NQF review

In order to meet emerging national needs, NQF may expedite its consensus development process for certain measures. NQF will obtain CSAC approval before starting the expedited review. The entire endorsement process for expedited review may be completed within four to five months. The COR and measure developer will need to respond to requests quickly to meet this schedule.

NQF requires all of the following criteria to be met prior to CSAC consideration for an expedited review:

- The measures under consideration must have been sufficiently tested and/or in widespread use. eMeasures will require only semantic validation testing using EHR simulated data sets, sometimes known as test beds.

[201] http://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process/CSAC_Decision.aspx

- The scope of the project/measure set is relatively narrow.
- There is a time-sensitive legislative or regulatory mandate for the measures.

Measure Maintenance for NQF

Once NQF has endorsed a measure, the measure developer supports ongoing maintenance of the measure (if that support is part of the measure developer's scope of work). Chapter 5, Measure Use, Continuing Evaluation, and Maintenance, describes the standardized measure maintenance processes and the processes involved to maintain NQF endorsement.

4.2—Measure Selection

CMS identifies and selects measures it is considering for use in its Medicare programs through a transparent process. CMS adopted a set of criteria, identified below, to ensure a consistent approach. Measures selected for use in Medicare programs must meet all of the criteria listed below. Harmonization is addressed in the CMS Measure Selection Criteria: to have measures that are aligned across CMS and HHS. When considering a measure for a topic already measured in another program, CMS prefers to use the same measure or a harmonized measure. The MAP also considers program and measure alignment when deciding which measures to recommend. After considering the MAP recommendation,[202] CMS chooses which measures to implement.

Pre-Rulemaking Process

Section 3014 of the ACA[203] 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 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.

Measures under Consideration

Over the past few years, CMS has articulated a number of measure selection criteria in its Federal Rules for various programs. The term "measure selection" typically applies to determining 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.

---

[202] National Quality Forum. *Measuring Performance: Measure Applications Partnership.* Available at: https://qualityforum.org/map/. Accessed on: June 1, 2015.

[203] 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: June 1, 2015.

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.
- eMeasures must be fully developed and tested.
- eMeasures must be entered into the MAT, created in the HQMF.
- eMeasures must undergo reliability and validity testing, including review of the logic and value sets by the CMS partners, including, but not limited to, MITRE and the NLM.
- eMeasures must have successfully passed feasibility testing.
- eMeasures must meet CMS EHR system infrastructure requirements, as defined by the EHR Incentive Program.
- The data collection mechanisms must be able to transmit and receive requirements as identified in the EHR Incentive Program rule. For example, eMeasures must meet QRDA Category I standards for transmission format.

Applying the measure selection criteria listed above, CMS develops the MUC list.[204] 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.

MAP recommendations

The MAP input to HHS on the list of quality and efficiency measures under consideration by the Medicare program is due by February 1 of each year as a recommendation report. The *MAP 2014 Recommendations on Measures for More Than 20 Federal Programs* gave input on 231 unique measures that were submitted for potential use in the Medicare program.[205] To note, 234

[204] 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.

[205] 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: June 1, 2015.

measures were initially submitted, but 3 were withdrawn prior to MAP review. In 2014, the MAP provided HHS with three general categories of feedback for the measures on the MUC list.

- Support—the measure should be added to program measure sets during the current rulemaking cycle.
- Do Not Support—the measure or measure concept is not recommended for inclusion in program measure sets.
- Conditionally Support—the measure or measure concept should be added to program measure sets after specific conditions are met.

Additionally, the MAP reviews measures which are currently used or finalized for use in the Medicare program. The MAP provided HHS with two general categories of feedback for these measures.

- Retain—the measure should remain in the program measure set.
- Remove—the measure should be removed from a program measure set, following a proper timeline.

The full report can be found on the MAP pages on the NQF website.[206]

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.

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 rule—CMS writes the proposed rule and publishes it in the Federal Register. The proposed rule is generally available for public comment for 60 days.
2. Final rule—CMS considers the comments that were received and publishes the final rule in the Federal Register.

Rollout, Production, and Monitoring of Measures

When the measures are finalized in the rule, CMS prepares a plan for implementation including the initial rollout, data management and production, audit and validation, provider education, dry

[206] http://www.qualityforum.org/Setting_Priorities/Partnership/Measure_Applications_Partnership.aspx

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.

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 eMeasures, ONC 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 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.

Impact assessment of Medicare quality measures

Also mandated by Section 3014 of the ACA, once every three years, the Secretary 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 National Quality Strategy'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.

4.3—Measure Rollout

This section describes the major tasks involved in the implementation of measures at the national level. The process of measure implementation may vary from one measure contract to another; however, this section provides a basic framework of the key steps that are necessary when implementing measure(s).

The scope of measure implementation could be one of the following:

- A measure or measure set is implemented in a new program.
- A measure or measure set is added to an existing program.

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". [207] The measure development and maintenance procedures detailed in the Blueprint also comply with HHS's Guidelines for Ensuring the Quality of Information Disseminated to the Public. That law adds that, "CMS has developed administrative mechanisms to allow affected persons to seek and obtain correction of disseminated information that does not comply with OMB, HHS, and CMS guidelines."[208]

Deliverables

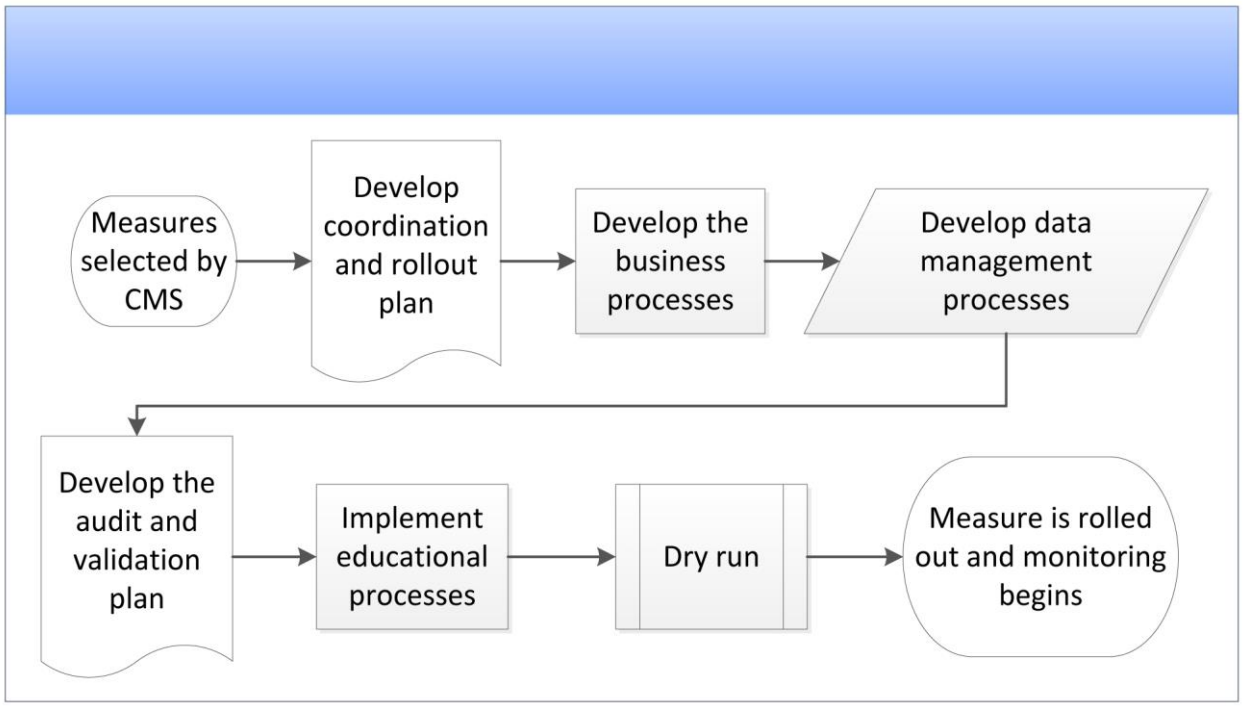
- Public Description of Quality Measures
- Timeline for Data Item and/or Quality Measure Implementation
- Implementation Stakeholder Meetings
- Questions and Answers Support
- Implementation Process Roadmap
- Measure Calculations/Results
- Pre-Posting Preview Results
- Compare Sites Files and Measures (as applicable)
- Implementation Algorithm (also called the Calculation Algorithm/Measure Logic)

[207] http://www.fws.gov/informationquality/section515.htm
[208] http://www.aspe.hhs.gov/infoquality/Guidelines/CMS-9-20.shtm

- Data Use Agreement

Figure 25: 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 25: Overview of the Measure Rollout Process



Once the COR has approved a measure for use in a particular program, several tasks have to be completed for rollout. Perform the following 9 steps simultaneously whenever possible to achieve an efficient timeline.

Procedure

Step 1 of 9: 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 Federal Rulemaking process. This process is conducted in three phases:

- Pre-rulemaking
- Publishing the proposed rule for public comment
- Issuing the final rule

Measure developers will support CMS with the rulemaking processes. For example, during the pre-rulemaking process, measure developers may need to provide information about their measures for presentation to the NQF-convened MAP. During the next phase, publishing the proposed rule

for comment, the measure developers will monitor the comments submitted on their measures and begin drafting responses for CMS. When the final rule is being issued, measure developers may also be asked to provide additional documentation on their measures.

Step 2 of 9: Develop the coordination and rollout plan

The coordination and rollout plan includes the following key parts:

- Timeline for quality measure implementation
- Plan for stakeholder meetings and communication
- The anticipated business processes model
- 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:

- 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:

- 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.
- Open door forums.
- Other processes as determined by the COR.

A measure developer must consider all of 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.[209] OMB approval is required before requesting most types of information from the public. The chapter on Stakeholder Input in Section 1 includes a discussion of these requirements.

Step 3 of 9: 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:

- 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.

[209] 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: June 1, 2015.

Step 4 of 9: 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:

- 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.

Step 5 of 9: 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:

- 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 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:

- 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.

Step 6 of 9: 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.

Step 7 of 9: 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:

- 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.

Step 8 of 9: Conduct the dry run

The dry run is the final stage of both measure testing and measure rollout. 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 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.

Step 9 of 9: Submit reports

CMS may request reports summarizing the rollout processes. These may include:

- Reports describing the business processes.
- Results of any education that was conducted.
- Results of the dry run, including, but not limited to:
- Analysis of the measure's success in meeting CMS's intentions for it.
- Recommendations regarding:
- Measure specifications.[210]
- Business processes model.
- Data management processes.
- Audit and validation processes.
- Educational processes for either data collectors or users of the measure results.

[210] 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.

Chapter 5—Measure Use, Continuing Evaluation, and Maintenance

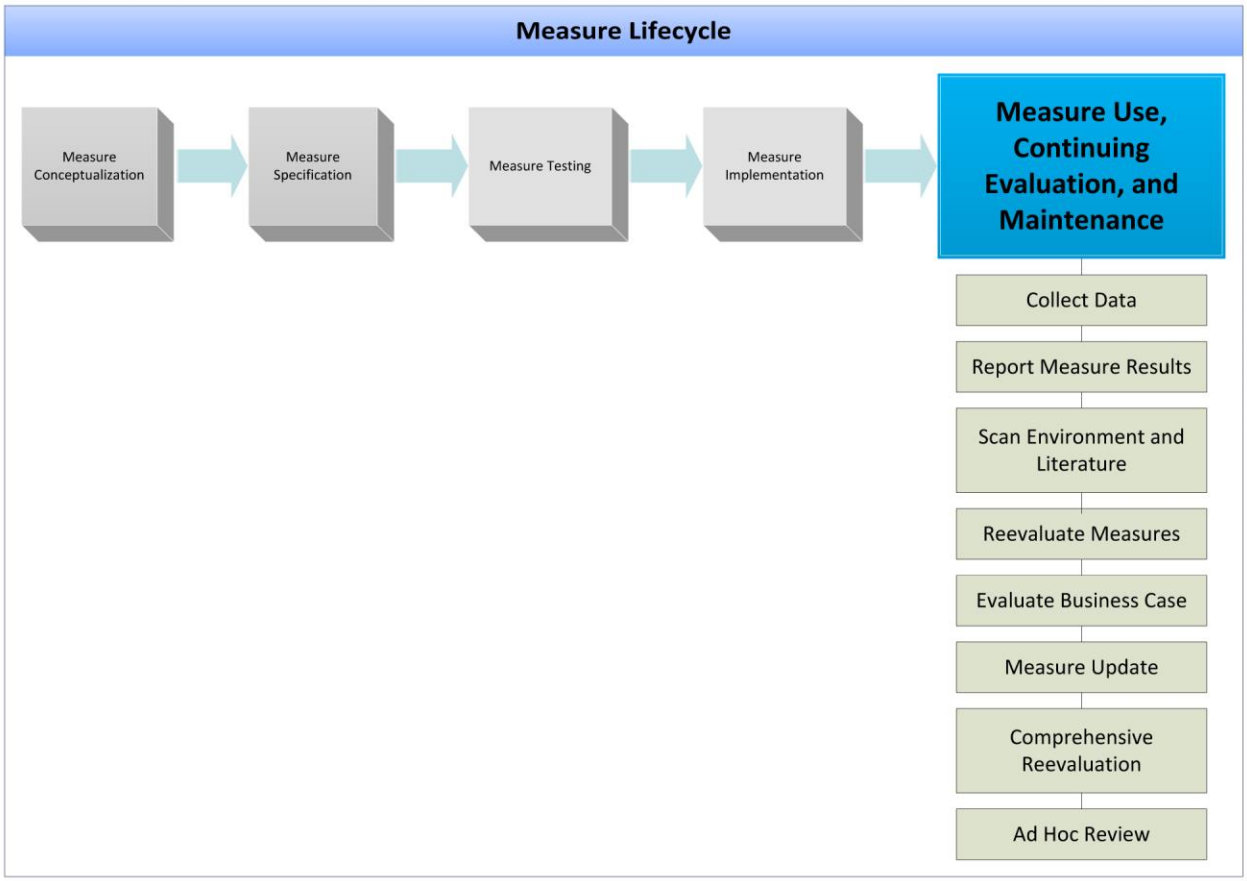
To help CMS ensure the continued soundness of the measures, the measure developer must provide strong evidence that the measure continues to add value to quality reporting measurement programs and that its construction continues to be sound. This work also helps CMS ensure that its measures maintain NQF endorsement.

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

- The aspects of care included in the specifications continued 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 the best 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 other competing measures.

As depicted in Figure 26: Flow of the Measure Lifecycle - Measure Use, Continuing Evaluation, and Maintenance, there are three basic types of measure reviews during maintenance: measure updates, comprehensive reevaluations, and ad hoc reviews.

Figure 26: Flow of the Measure Lifecycle - Measure Use, Continuing Evaluation, and Maintenance



The Blueprint describes the annual update, comprehensive reevaluation, and ad hoc review as distinct and separate activities. In practice, these activities sometimes overlap and are conducted concurrently. For example, when the measure is due for an annual update, there might be unforeseen reasons that might require an ad hoc review such as changes in scientific evidence supporting the measure. On the other hand, circumstances might occur that an ad hoc review is needed; but, since the measure is due for comprehensive reevaluation, the ad hoc review might be merged with the comprehensive reevaluation. Ideally, the measure maintenance schedule is aligned with the NQF endorsement maintenance cycle. However, in practice, for various reasons, these schedules may not align completely.

5.1—Measure Production and Monitoring

Measure production and monitoring includes the ongoing tasks necessary to use the measure over time. These tasks are described in this overview but refer to other chapters for more detailed instructions where applicable. The process of measure production and monitoring varies significantly from one measure set to another depending on a number of factors, which may include, but are not limited to:

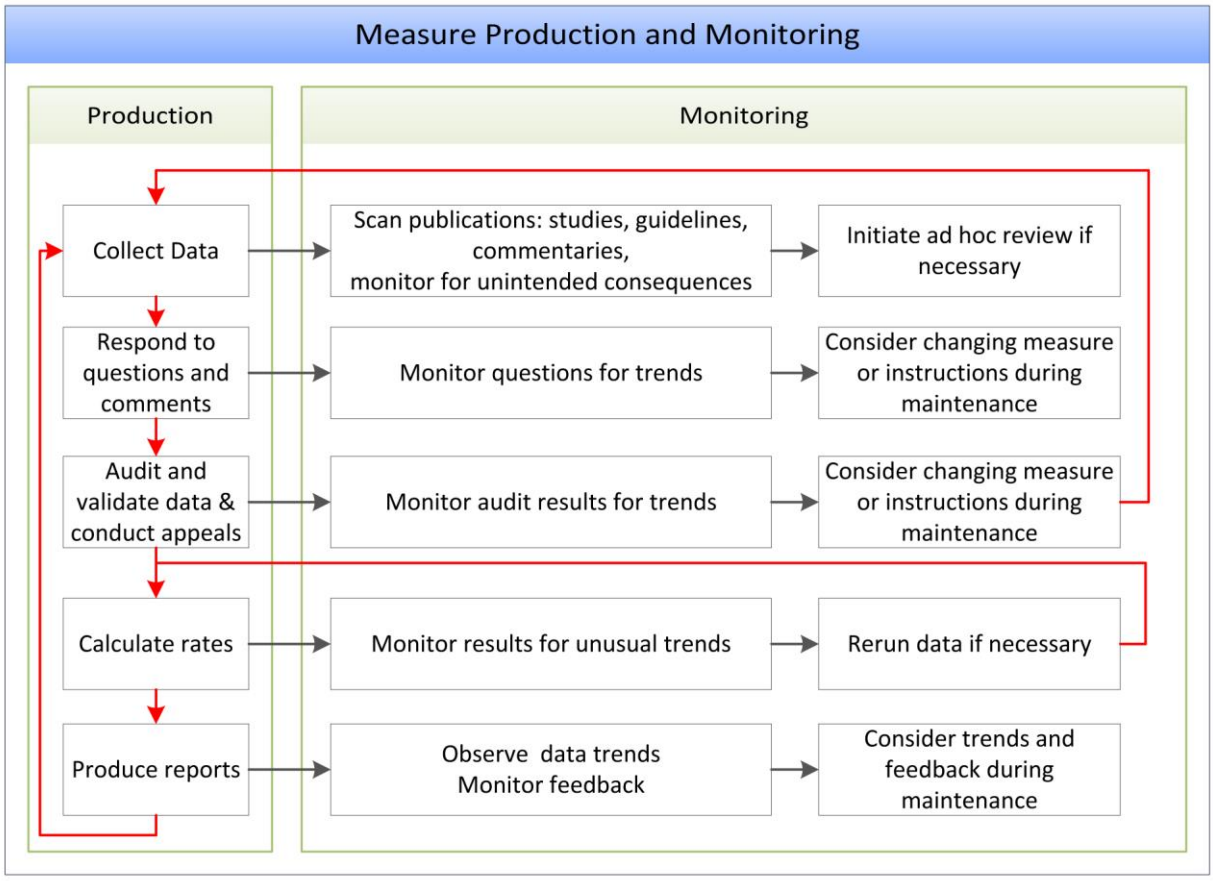
- Scope of measure implementation.

- Health care 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 27: Overview of the Measure Monitoring Process 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 7 steps listed below.

Figure 27: Overview of the Measure Monitoring Process



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
- Periodic environmental scans
- Data files of the measure rates and/or demographic information suitable for posting on CMS website

Procedure

Step 1 of 7: 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 three 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 Chapter 1.A—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.

Step 2 of 7: 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 two 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.

Step 3 of 7: 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:

- 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).

Step 4 of 7: 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.[211]

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.

Step 5 of 7: 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, socioeconomic status, income, region, gender, primary language, disability, or other classifications
- Frequency of use of exclusion or exception and how they influence rates
- 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
- Gaming or other unintended consequences
- 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.

[211] Department of Health and Human Services. *Section 508.* Available at: http://www.hhs.gov/web/508/index.html. Accessed on: June 1, 2015.

This information is reported during Comprehensive Reevaluation, as described later in this chapter.

Step 6 of 7: 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 measure update is described later in this chapter. Each measure is also fully reevaluated every three years to ensure that it still meets the measure evaluation criteria. The standardized process is described in the Comprehensive Reevaluation section later in this chapter.

As mentioned earlier, situations may also arise in which a measure must be reviewed before the scheduled measure update or comprehensive reevaluation. In this case, an ad hoc review is conducted. The standardized process is described later in this chapter under 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.[212] 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 subsection 5.5 of this chapter 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.

Step 7 of 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. The Priorities Planning chapter describes how CMS uses this input. CMS leadership may find information from measures monitoring valuable for setting priorities and planning future measurement projects. CMS may request an evaluation of current measures and sets used in the programs or initiatives and recommendations for ways to accommodate cross-setting use of the measures. The evaluation may also include options for alternative ways to interpret 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.

[212] 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: June 1, 2015.

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 2 steps help CMS ensure that its measures retain NQF endorsement.

Step 1 of 2: Apply measure evaluation criteria

Each measure undergoes an update at least annually and a rigorous comprehensive reevaluation every three years to assess its continued value based on the same set of standardized measure evaluation criteria used in measure development:

- Importance
- Scientific acceptability of the measure properties
- Feasibility
- Usability and use
- Harmonization

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 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.
- Data elements, codes, and parameters included in the specifications are the best ones to use to quantify the particular measure, and data collection still does not cause undue burden on resources.
- Calculations included in the specifications represent a clear and accurate reflection of the variation in the health outcome of interest, or the quality or efficiency of the care delivered.

Step 2 of 2: Report results of evaluation

Measure Evaluation Criteria and subcriteria are detailed in the Measure Evaluation Criteria and Instructions with a link to the 2013 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.

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 the Harmonization chapter in Section 1 of the Blueprint. The following 4 steps taken during measure maintenance will help ensure that measures continue to be harmonized after they are implemented.

Step 1 of 4: 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 11: Harmonization Decisions during Measure Maintenance below, describes harmonization issues and actions based on the numerator and denominator specifications.

Table 11: 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), or Justify differences, or 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), or Justify differences.
Numerator: Different measure focus Denominator: Different target population	Unique measures	Maintain existing measure.

Step 2 of 4: 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.

Step 3 of 4: Test scientific acceptability of measure properties

If harmonization results in changes to the measure specifications, testing is usually necessary. Further details about testing scientific acceptability are described in Chapter 3—Measure Testing.

Step 4 of 4: NQF evaluates for harmonization during measure maintenance

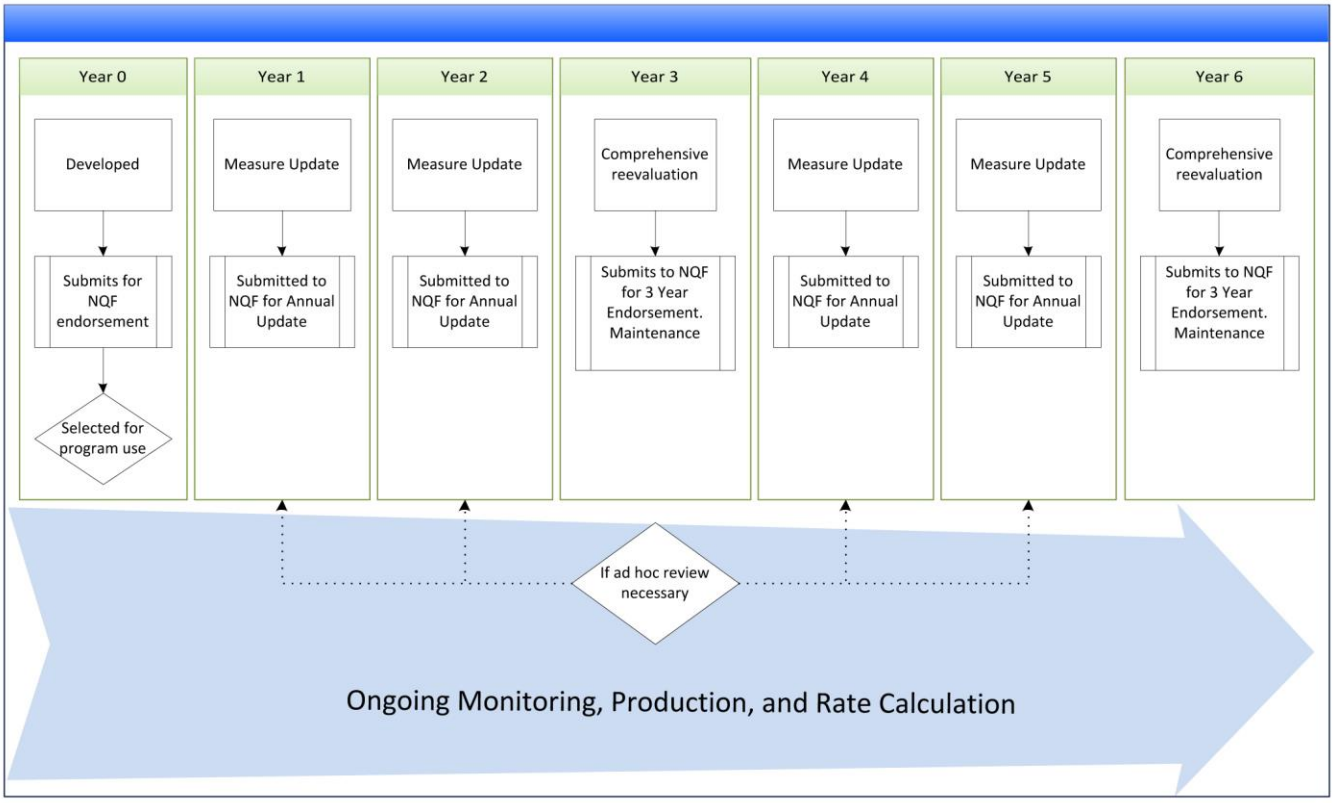
NQF will evaluate the measure for harmonization potential during the maintenance review of the measure. There may be instances where the measure developer may be unaware of newly developed similar or related measures until they have been submitted to NQF for review. If similar or related measures are 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.

Maintaining NQF endorsement

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. This section describes NQF's standardized measure endorsement maintenance processes. The measure developer is responsible for being familiar with NQF's current measure endorsement maintenance processes described on NQF's website.[213] NQF's endorsement maintenance processes are designed to ensure that NQF only continues to endorse measures that meet the criteria. NQF endorsement maintenance reviews are separate from the CMS Measure Management System maintenance reviews, but Figure 28: Measure Review Cycle Timeline 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.

[213] http://www.qualityforum.org/Measuring_Performance/Maintenance_of_NQF-Endorsed®_Performance_Measures.aspx

Figure 28: Measure Review Cycle Timeline



The following three types of maintenance reviews are described below, including deliverables and the steps required for each.

- Measure update
- Comprehensive reevaluation
- Ad hoc review

5.2—Measure Update

The first type of measure reevaluation is the measure update. This is usually a limited review of the precision of the measure's specifications—completed annually (or semiannually in some cases). Measure updates are scheduled at least annually to ensure the procedure, diagnostic, and other codes (CPT, ICD-10-CM, LOINC, etc.) used within the measure are updated when the code sets 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 two 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 three-year endorsement period, the measure steward is responsible for informing NQF immediately of the timing and purpose of the changes. An ad hoc review will be conducted if the changes materially affect the measure's original concept or logic.

The measure 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 measure updates to determine if revision is needed beyond updating the codes.

The measure update process involves three parts, below, divided into 8 steps (see Procedure, below):

- Gathering information that has been generated since the last review (the comprehensive reevaluation, measure update, or measure development—whichever occurred most recently).
- Recommending action.
- Approving and implementing the action(s).

Deliverables

- 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.[214]
- For measure developers maintaining eMeasures, the revised electronic specifications (eMeasure XML file, SimpleXML file, eMeasure 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 Measure Information Form.
- NQF Annual Update online submission regardless if any change was made to the measure.
- NQF submission documentation for any material changes to the measure.

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

[214] 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: June 1, 2015.

priority topics with the most potential to improve healthcare, and bring other benefits of parsimony. An annual measure 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.

Procedure

Step 1 of 8: Review the measure's code sets

Review the code sets used by the measure to determine:

- If new codes have been added to or deleted from the code set that may affect the measure.
- If 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, 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 eMeasure value sets, it is important to align with the vocabulary recommendations made by Health Information Technology Standards Committee Clinical Quality Workgroup and Vocabulary Task Force. The eMeasure Specification chapter of the eMeasure Lifecycle section provides more information on the procedure.

Step 2 of 8: 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 measure 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.

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 measure update. Details of the procedure are discussed later in this chapter under 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

Step 3 of 8: 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 Possible Outcomes of Maintenance Reviews.

The possible dispositions are:

- Retain
- Revise
- Remove
- Retire
- Suspend

Step 4 of 8: The COR reviews the recommendation for approval

Forward the recommendations to the COR, along with updated Measure Information Forms (or eSpecifications and value sets for eMeasures), and Summary of Changes/Release Notes. If significant changes were made, an updated Measure Justification Form and Measure Evaluation Report may be necessary.

The COR reviews the measure 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.

Step 5 of 8: 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, or re-education of providers. Notify the Measures Manager to ensure that the CMS Measures Inventory is updated appropriately regardless of the disposition decision.

Step 6 of 8: 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 three-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. Annual maintenance focuses on whether the measure remains current. 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 concept or logic. 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.[215] 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.

Step 7 of 8: Consider measures not owned/maintained by a CMS measure developer

If the measure developer responsible for implementing the measure for which CMS is not the measure owner (i.e., not ultimately responsible for maintaining the measure), then the measure developer will be responsible for monitoring the measure owner's maintenance of the measure. This includes ensuring that the measure is revised periodically in response to updates in the underlying code sets (e.g., CPT, ICD-9, LOINC) and that the measure is reevaluated in a manner consistent with (though not necessarily identical to) the requirements of 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.

[215] http://www.qualityforum.org/Measuring_Performance/Endorsed_Performance_Measures_Maintenance.aspx.

Step 8 of 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 measure 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 if necessary.

Confirm the deadline for each measure 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.

5.3—Comprehensive Reevaluation

Measure developers are required to conduct a thorough review of the measure every three 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.

Measures maintained by CMS need this comprehensive reevaluation regardless if they are endorsed. Generally, CMS and the measure developer can align the schedules of the reevaluations, so that the comprehensive reevaluation immediately precedes the NQF three-year maintenance review. This allows CMS time to review the findings and recommendations prior to submission to NQF.

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 9 steps, outlined in Procedure section below, 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.

Deliverables

- An updated Measure Information Form detailing all recommended changes to the measure.[216]
- For measure developers maintaining eMeasures, the revised electronic specifications (eMeasure XML file, SimpleXML file, eMeasure 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 Measure Information Form.
- 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.
- 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 three-year endorsement maintenance).
- 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.

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.

Harmonization

During comprehensive reevaluation, make full consideration to determine if there are related and 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

[216] 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: June 1, 2015.

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 earlier in this chapter in the section on Harmonization during measure maintenance. 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.

Procedure

Step 1 of 9: 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.

Step 2 of 9: 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:

- 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:

- 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, socioeconomic status, 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.

Step 3 of 9: Convene a Technical Expert Panel

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. Chapter 1.B—Technical Expert Panel 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. The Concepts of Measure Evaluation chapter of Section 1 describes the measure evaluation criteria. Measure selection criteria are discussed in Section 2—Measure Lifecycle, located in 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.

Step 4 of 9: 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 eMeasure value sets, it is important to align with the vocabulary recommendations made by Health Information Technology Standards Committee Clinical Quality Workgroup and Vocabulary Task Force. More information on these requirements is found in the eMeasure Specification chapter of Section 3—eMeasure Lifecycle.

Update the Measure Information Form (or eSpecifications and value sets for eMeasures) with any new measure specifications and coding. All changes to measure specifications should be described in the Measure Information Form 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.

Step 5 of 9: 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 Possible Outcomes of Maintenance Reviews.

The possible dispositions are:

- Retain
- Revise
- Remove
- Retire
- Suspend

Step 6 of 9: 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 is outlined in the middle column of Table 12, and the extent of testing required for measures not in use is covered in the right column.

Table 12: 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 Chapter 3—Measure Testing.

Step 7 of 9: 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 Chapter 1.C.—Public Comment.

Analyze the comments received and refine the measure as indicated. Document any changes in the Measure Information Form (or eSpecifications and value sets for eMeasures). 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 re-test 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.

Step 8 of 9: 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.

Step 9 of 9: Maintain NQF endorsement

NQF requires comprehensive review every three 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.[217]

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 six months 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 three-year maintenance reevaluation follows the NQF assignment schedule and not necessarily the measures' dates of endorsement. Depending on the volume of measures being reevaluated during a cycle, committees and topics from one cycle may need to be extended and scheduled to convene in the following year. As a result, measures may be subject to early or late three-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 cycle. These measures still require annual updates to be submitted to NQF.

NQF will send a standardized online submission template for the three-year endorsement maintenance review to the measure steward of record. The form will be prepopulated with information from the original or the most recent measure update submission. CMS notifies NQF regarding the appropriate measure developer contact for the three-year endorsement maintenance review.

The three-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 measure updates, and ongoing surveillance to complete the NQF three-

[217] http://www.qualityforum.org/Measuring_Performance/Endorsed_Performance_Measures_Maintenance.aspx.

year maintenance form. Following COR approval, the measure developer submits the report to NQF.

The comprehensive reevaluation for eMeasures (where it differs from the standard process) is described in Chapter 6—eMeasure Use, Continuing Evaluation, and Maintenance in the eMeasure Lifecycle section.

5.4—Ad Hoc Review

An 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 three-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; but 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.

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 it 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.

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 measure 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.

Deliverables

- Updated Measure Information Form, if the ad hoc review results in changes to the measure specifications.[218]
- For measure developers maintaining eMeasures, the revised electronic specifications (eMeasure XML file, SimpleXML file, eMeasure human-readable rendition [HTML] file) and value sets, if the ad hoc review results in changes to the measure specifications.
- 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.
- Updated Measure Evaluation Report, if the review resulted in a change to the measure's strengths and/or weaknesses.

Procedure

The CMS measure developer remains responsible to monitor the maintenance performed by the owner even if the measure developer is not the measure owner (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 sets (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

[218] 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: June 1, 2015.

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 three primary subparts, subdivided into 7 steps:

- Determining if an ad hoc review should be conducted.
- Conducting the review and recommending an outcome.
- Approving and implementing the approved outcome.

Step 1 of 7: Determine if 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 below). 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.

Step 2 of 7: 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 1.A.—Information Gathering.

Step 3 of 7: 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.

Step 4 of 7: Determine if 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.

Step 5 of 7: 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 at the end of this chapter under Possible Outcomes of Maintenance Reviews.

Depending on the findings from the previous steps, the recommendation may be:

- Retain
- Revise
- Remove
- Retire
- Suspend

Submit the recommendation along with supporting documentation and the updated Measure Information Form and Measure Evaluation Report (if recommending immediate revision or suspension until revision is possible) to the COR.

Step 6 of 7: 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.[219]

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.

Step 7 of 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.

National Quality Forum 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.[220]

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.

[219] 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: June 1, 2015.

[220] National Quality Forum. *Measuring Performance:* Maintenance of NQF-Endorsed® Performance Measures. Available at: http://www.qualityforum.org/Measuring_Performance/Endorsed_Performance_Measures_Maintenance.aspx. Accessed on: June 1, 2015.

- 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 may conduct an ad hoc endorsement review on an endorsed measure at any time if any of the following occur:

- Evidence supporting the measure has changed.
- Implementation of the measure results in unintended consequences.
- Material changes have been made to the measure which may affect the original measure's concept or logic.

Ad hoc endorsement reviews may be requested at any time by any party. Adequate evidence to justify the review and under which criterion the review is requested must be submitted when seeking an ad hoc maintenance review. NQF reviews the request and initiates an ad hoc review if there is adequate 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.

5.5—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 (described in Section 1, in the Concepts of Measure Evaluation chapter) and the measure selection criteria (described in Section 2, in the Measure Implementation chapter) 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.

Table 13: CMS Criteria for Measure Disposition is adapted from the "Standard CMS Measure Implementation Determination Criteria" and lists the criteria CMS uses to make decisions regarding the various dispositions described above.[221] Updated eMeasure-specific material added to the table is based on the "2014 Measures under Consideration List Program Specific Measure Priorities and Needs" document.[222]

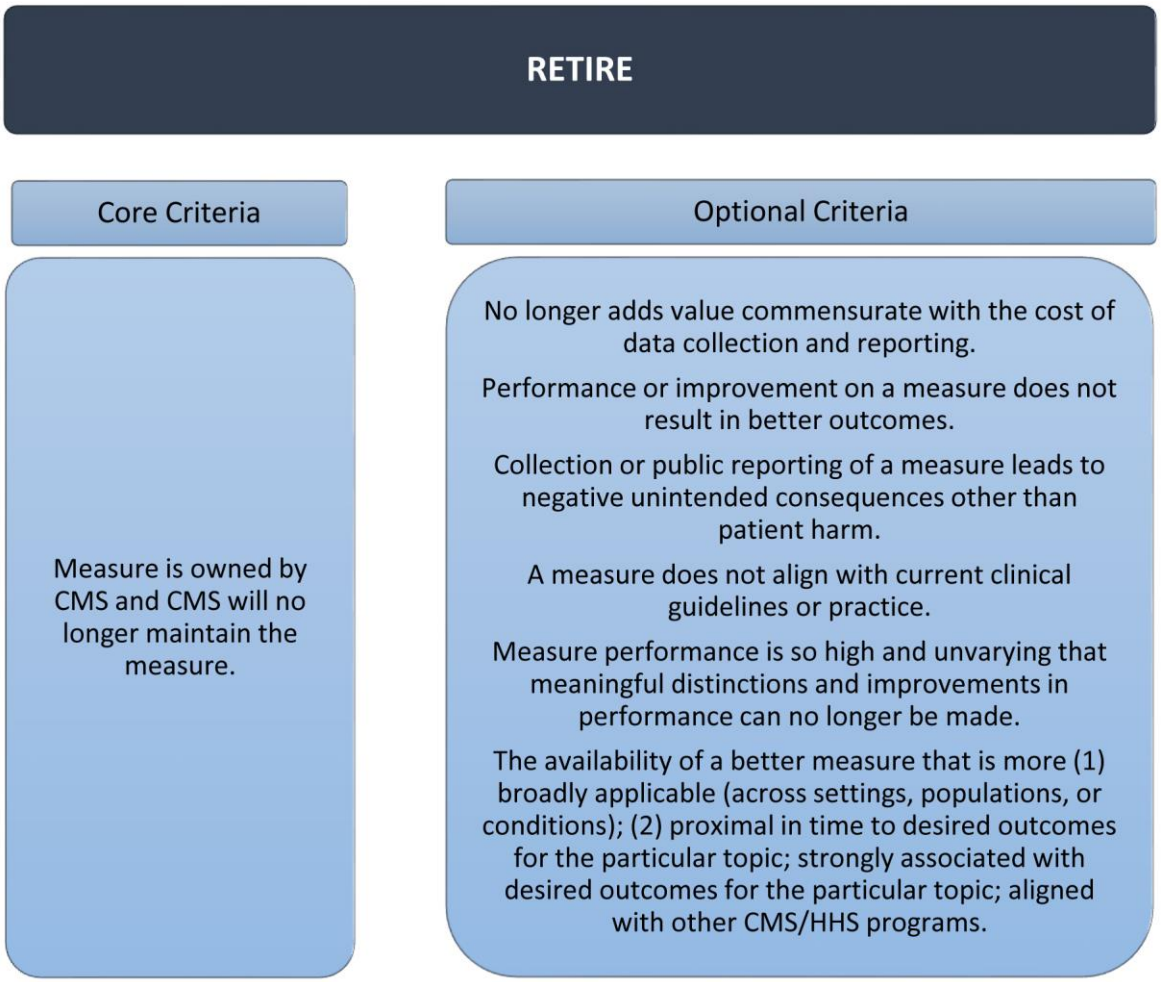
[221] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Standard CMS Measure Implementation Determination Criteria*. [unpublished] Prepared by E. Garcia and K. Goodrich for the CMS Quality Measures Task Force, March 26, 2012.

[222] 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.

Table 13: CMS Criteria for Measure Disposition

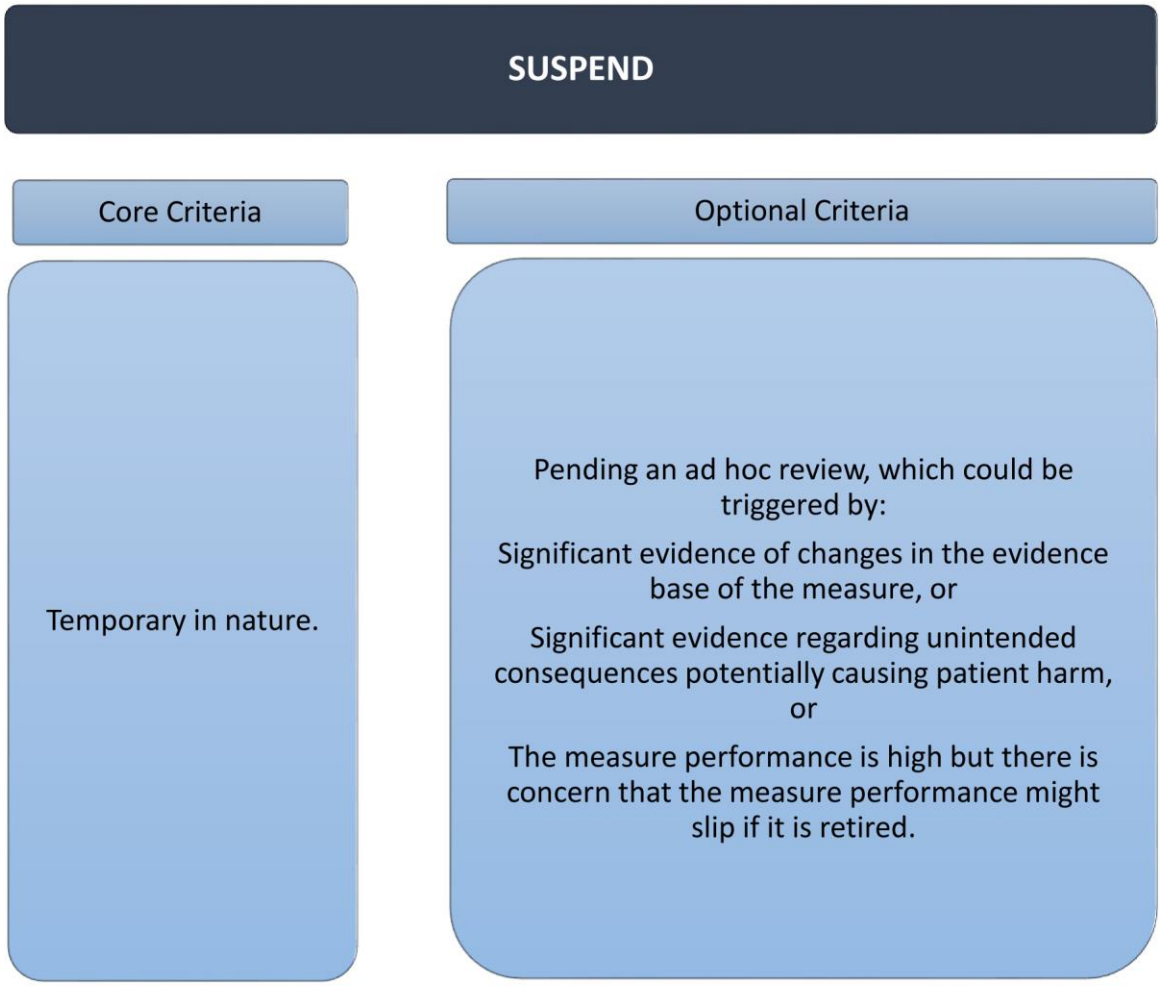
REVISE OR RETAIN

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 National Quality Strategy (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).	eMeasures must be fully developed and tested, entered in the MAT, created in the HQMF format. eMeasures must pass feasibility testing. eMeasures must undergo reliability and validity testing including review of the logic and value sets by the CMS partners, including, but not limited to, MITRE and the NLM. eMeasures must meet CMS EHR infrastructure requirements as defined by the EHR Incentive Program	Should be electronically specified whenever possible. Should be aligned with the EHR Incentive program were applicable.


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.
A measure 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; strongly associated with desired outcomes for the particular topic; aligned with other CMS/HHS programs.

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 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; strongly associated with desired outcomes for the particular topic; aligned with other CMS/HHS programs.


SUSPEND
Core Criteria
Optional Criteria
Temporary in nature.
Pending an ad hoc review, which could be triggered by:
Significant evidence of changes in the evidence base of the measure, or
Significant evidence regarding unintended consequences potentially causing patient harm, or
The measure performance is high but there is concern that the measure performance might slip if it is retired.

Section 3—eMeasure Lifecycle

Collecting and reporting accurate healthcare performance data has historically been a complex and time-consuming manual process. Performance measures most frequently use data that are routinely available. Claims data, laboratory results, and pharmaceutical usage data have traditionally been the source of information for these measures despite the fact that additional information valuable for performance measurement is now available in EHRs. 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.

eMeasures can promote greater consistency and improved uniformity in defining clinical concepts across measures, and also increased comparability of performance results. Through standardization of a measure's structure, metadata, definitions, and logic, the 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 will ideally be able to push a button and import these eMeasures into their EHRs. The HQMF Release 1 (R1) was published in 2009[223] and is the underlying structured representation used by the CMS MAT for eMeasures developed through June 2014. The representation of an eMeasure is simplified when measure developers author their eMeasures 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 level) or Category III (aggregate patient data) reports.[224] As the nation makes progress toward HIT adoption, much of the successes will rely on solid electronic representation of measurement and clinical support.

The purpose of the eMeasure Lifecycle Section is to guide measure developers through the procedures necessary to develop and maintain an eMeasure for either an adapted (retooled or reengineered) measure or a new (*de novo*) measure. This section does not repeat information from the rest of the Blueprint but documents where the processes deviate from the way other types of measures are developed.

The chapters in this section will reference general concepts and processes of measure development by linking to the appropriate chapter in Section 1 or Section 2 of the Blueprint for further information.

Measure developers need to be knowledgeable of the following:

- The Blueprint—in particular, this section, eMeasure Lifecycle
- The Measure Authoring Tool's (MAT's) user guide

[223] Announcement press release available at: http://www.hl7.org/documentcenter/public_temp_116FD0C6-1C23-BA17-0C50779726E5BD29/pressreleases/HL7_PRESS_20090827.pdf. Accessed June 1, 2015.

[224] Issued by Centers for Medicare & Medicaid Services. Guide for reading electronic clinical quality measures (eCQMs)-Version 5. Washington, DC: Centers for Medicare & Medicaid Services; Mar 2014. Available at: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/eMeasures_GuidetoReading.pdf. Accessed on: June 1, 2015.

- The eMeasures Issues Group (eMIG) recommendations
- VSAC

MAT. 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.

eMIG. To ensure consistency in the eMeasure development process, CMS convenes ongoing eMIG meetings. The eMIG serves as a forum to discuss and propose common solutions for issues encountered during the development and testing of eMeasures. A fundamental activity of the eMIG is to develop additional guidance for issues encountered by measure developers. Solutions proposed and presented at eMIG meetings address areas where common standards or approaches have not yet been specified, where different approaches can be harmonized, or where the Blueprint for the CMS Measures Management System does not supply guidance. Measure developers should contact a member of the CMS Measures Management System team for inclusion in these regularly scheduled eMIG meetings.

VSAC. The VSAC is provided by the NLM, in collaboration with the ONC and CMS. With a free Unified Medical Language System (UMLS) license, the VSAC provides searchable and downloadable access to all official versions of vocabulary sets contained in the eCQM currently used in CMS and other quality reporting programs (e.g., The Joint Commission). Each value set consists of the numerical values (codes) and human-readable names (terms), drawn from standard vocabularies such as SNOMED CT®, RxNorm, LOINC and ICD-10-CM, which are used to define clinical concepts used in clinical quality measures (e.g., patients with diabetes, clinical visit).[225] 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.

eMeasure Components

A health quality measure encoded in HQMF is referred to as an “eMeasure” or “eCQM”. As further described in this section, eMeasures are written to conform to the Health Level Seven (HL7) standard for representing a health quality measure as an electronic XML document that can be viewed in a standard Web browser, when associated with an eMeasure rendering style sheet. HQMF was originally published in 2010 as a HL7 Draft Standard for Trial Use (DSTU). A DSTU is issued at a point in the standards development lifecycle when many but not all of the guiding requirements have been clarified. A DSTU is intended to be tested and ultimately submitted to the HL7 ballot process, to be formalized into an American National Standards Institute (ANSI)-accredited standard.

Note: HQMF Release 2 (R2) was balloted by HL7 in September 2013 and published in December 2013. Updates were further approved in Spring 2014 at which time HQMF Release 2.1 (R2.1) was

[225] https://www.nlm.nih.gov/vsac/support/index.html. Accessed on: October 23, 2015.

published by HL7. The MAT July 2014 release addresses the content changes required for HQMF R2.1 allowing the HQMF data model and expression logic to evolve independently as necessary and to support composite measure metadata. Full MAT export of HQMF R2.1 is expected for Q4 2014. In the meantime, measures entered into the MAT July 2014 update will include the new capabilities allowed by HQMF R2.1; exporting an interim "simple XML" until the Q4 2014 MAT update is completed. A future version of the Blueprint will address these new changes.

eMeasures are specified using patient-level information coded in a format intended to be extracted to used data captured in EHRs. The process of creating an eMeasure is similar to 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, eMeasures require additional steps to map measure data elements to corresponding QDM components and standard vocabularies to assemble the data criteria. An eMeasure is composed of 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 includes the underlying HQMF syntax.
- Its major components include a Header and a Body. The Header identifies and classifies the document and provides important metadata about the measure. The HQMF Body contains eMeasure sections, e.g., data criteria, population criteria, and supplemental data elements.
- Value sets—Convey information about how different data elements within the CQM are defined based on code groupings. These sets include an OID, a list of codes, descriptions of those codes, and the code system from which the codes are derived.
- Human readable—A HyperText Markup Language (HTML) file (.html) that displays the eMeasure content in a human-readable format directly in a Web browser. This file does not include the underlying HQMF syntax.

The MAT produces the following items for an eMeasure:

- eMeasure XML
- SimpleXML file
- eMeasure HTML file
- eMeasure style sheet
- Standalone spreadsheet that contains all of the value sets referenced by the measure

Overview of the eMeasure Lifecycle

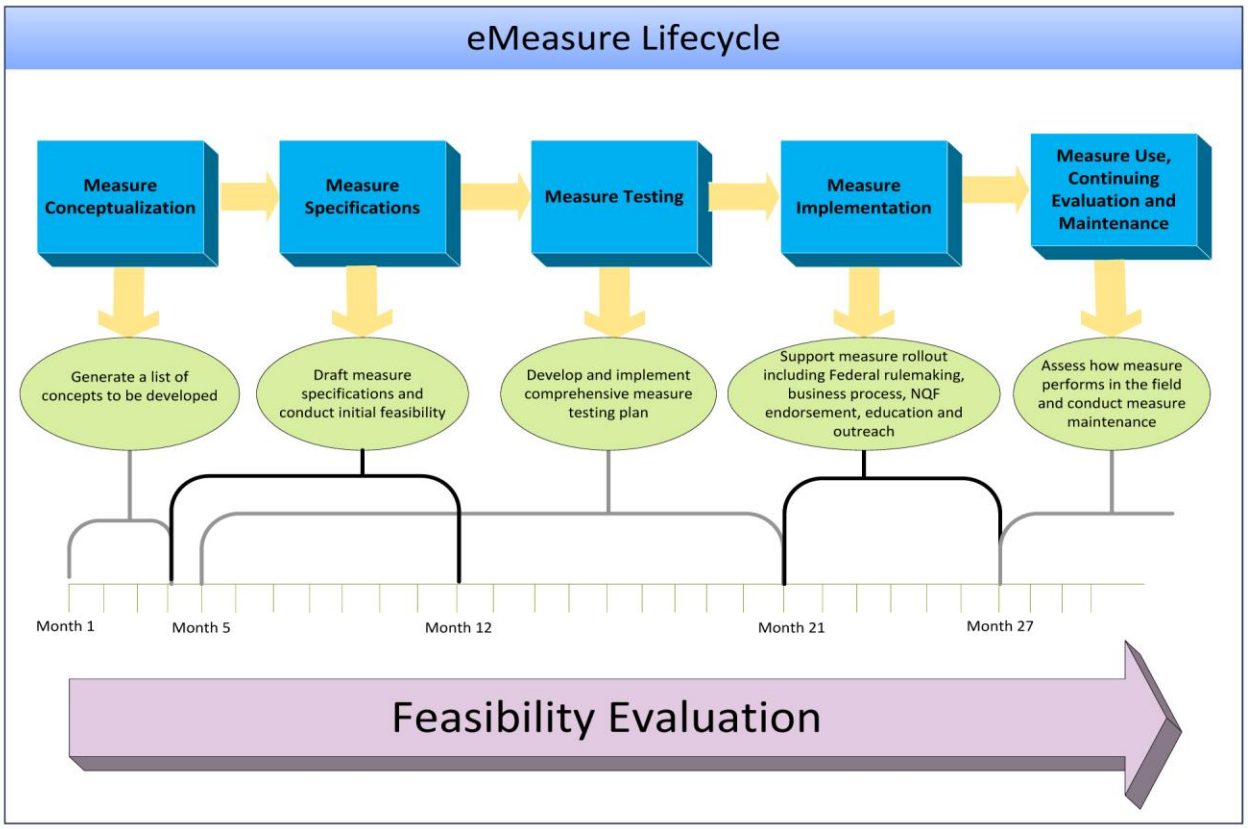
The overview of the eMeasure development process below provides a high-level summary of the steps used to develop and implement an eMeasure under CMS measure development contracts. These steps will be discussed in more detail throughout the section.

1. Perform environmental scan/information gathering
2. Review measure concepts with TEP
3. Create record in the ONC JIRA system CQM Issue Tracker project

4. Complete initial Measure Evaluation Report (measure developer's self-evaluation report)
5. Review by CMS of concept with the Self-Evaluation Report
6. Conduct feasibility testing of data source
7. Review by TEP of feasibility testing results
8. Draft eMeasure specifications
9. Review by TEP of specifications and modeling
10. Review by CMS of draft specifications and testing results
11. Review by CMS of first eMeasure evaluation and results
12. Perform quality assurance check - preliminary logic and value set review
13. Post measure for public comment
14. Develop eMeasure output using MAT
15. Conduct eMeasure testing
16. Revise specifications based on feedback
17. Perform quality assurance check – logic and value set review
18. Update eMeasure output using MAT
19. Run eMeasure test cases through Bonnie tool
20. Publish final eMeasure
21. Update eMeasure Evaluation Report
22. Submit for NQF endorsement

Figure 29: eMeasure Lifecycle illustrates the eMeasure development lifecycle that will be discussed in further detail in this section. Timelines shown in the figure are for illustrative purposes only, as actual measure development timelines will vary based on the measure. Though feasibility initially appears in the figure during Measure Specification, feasibility of the measure should be evaluated throughout the entire development process.

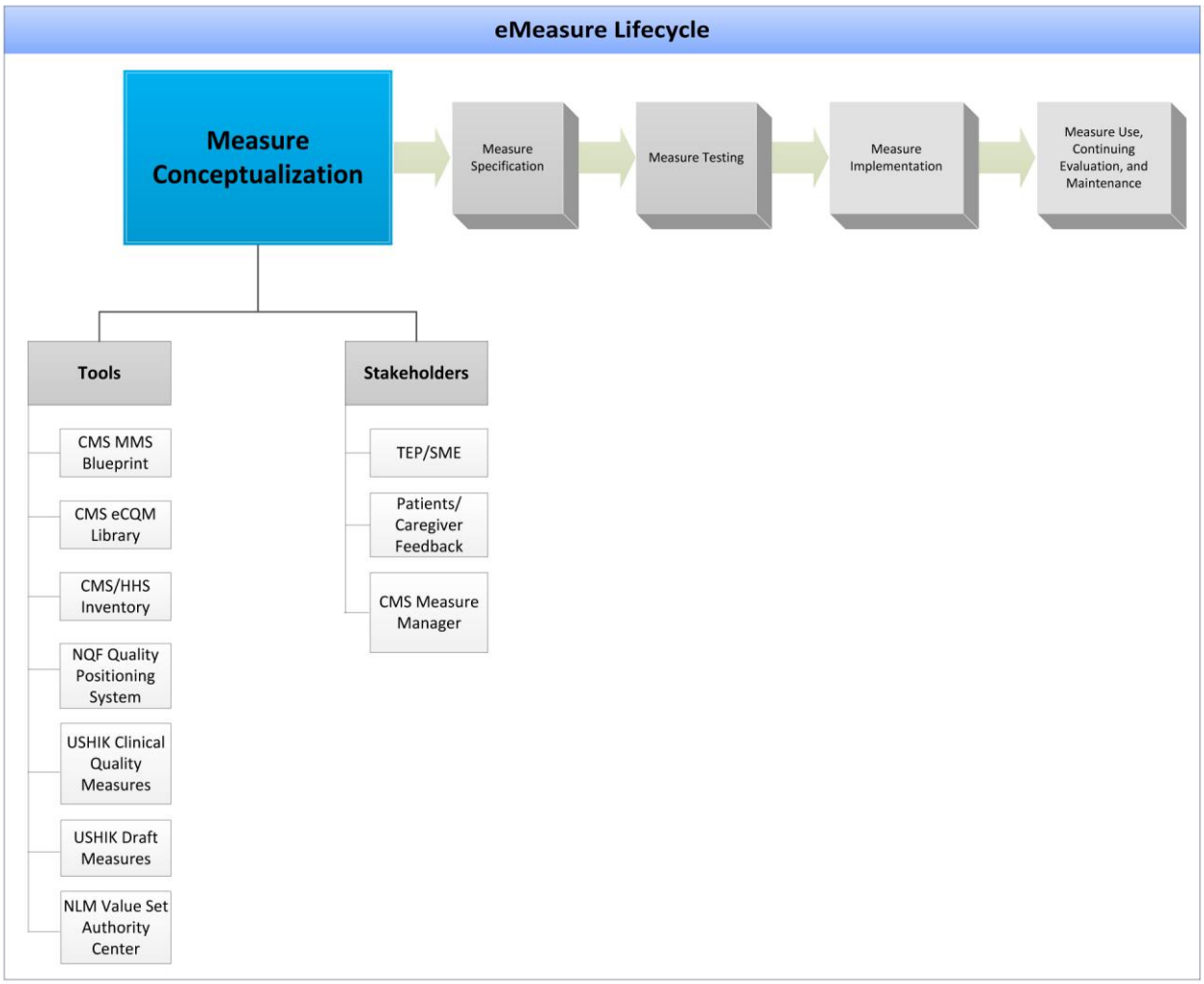
Figure 29: eMeasure Lifecycle



Chapter 1—eMeasure Conceptualization

In the measure conceptualization stage, the measure developer identifies whether existing measures may be adapted or retooled to fit the desired purpose. If no measures are identified that match the desired purpose, the measure developer works with a TEP to develop new measures. Depending on the findings during information gathering 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 eMeasure(s). The detailed process for development of a new (*de novo*) measure is outlined in Section 2 of the Blueprint. Figure 30: eMeasure Conceptualization Tools and Stakeholders depicts the tools and key stakeholders needed to develop a measure concept.

Figure 30: eMeasure Conceptualization Tools and Stakeholders



It is important to consider early in measure conceptualization what has been used before 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 re-work later in the process. Tools that help consider what concepts are feasible include those that allow the measure developer to evaluate the ability of codes within the terminologies to express the measure concepts (e.g., NLM VSAC).

1.1—Deliverables

- List of potential candidate measures or measure concepts
- Call for Measures (if necessary)
- Information Gathering Report—Summary Report of Environmental Scan and Empirical Analysis
- Measure testing report (if completed here)
- Measure Evaluation Report—includes the Feasibility Assessment and assessment of measure logic as part of the evaluation
- Draft Measure Justification Form (required for new eMeasures only)
- Draft Business Case
- TEP/SME Summary Report

1.2—eMeasure Feasibility Assessment

For eMeasures, feasibility assessment is necessary during the early stages of measure development (alpha testing). Data availability (and the quality of such data) and the impact to clinician workflow should also be assessed throughout the entire measure development process. The 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 eMeasure specifications to ensure that the data elements are available in a usable structured format 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. Identification of feasibility concerns early in the development of the measure allows for sufficient time to:

- Replace or revise data elements
- Stop further development, or
- Determine a plan for addressing the concerns.

In addition to information obtained from SMEs, empirical analysis can also be used to test the feasibility of data elements required for a measure. Feasibility considerations include:

- Data availability (including standardization).
- Accuracy of the information in the data.
- Maturity of standards.
- Standard vocabularies.

- Extent to which the data are collected 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. If the eMeasures in development are going to require risk-adjustment, note that a preliminary feasibility assessment should also be applied to the risk variables. General information on feasibility assessment is provided in the feasibility subsection of the Measure Testing chapter in Section 2. Refer to the NQF eMeasure Feasibility Assessment Report for more information and guidance.[226]

1.3—Information Gathering for eMeasures

For the most part, the information gathering process is the same for eMeasures as for measures developed using other data sources. However, eMeasures are based on information that should exist in a structured format in EHR systems. 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. Therefore, evaluation of the scientific acceptability (validity and reliability) of eMeasures is based on some unique assumptions and special considerations, as follows:

- eMeasure evaluation is based on use of only data elements that can be expressed using the QDM.
- Quality measures that are based on EHRs should significantly reduce measurement errors due to manual abstraction, coding issues, and inaccurate transcription errors.
- eMeasures are subject to some of the same potential sources of error as non-eMeasures 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. Other examples of potential sources of error include:
- Incorrect or incomplete measure specifications, including code lists (or 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 being used in different ways or entries made into the wrong EHR field.
- eMeasures are subject to an additional potential source of error from incorrect parsing of data by natural language processing software used to analyze information from text fields.
- Although data element reliability (repeatability) is assumed with computer programming of an eMeasure, empirical evidence is required to evaluate the reliability of the measure score.

[226] National Quality Forum. Report from the National Quality Forum: eMeasure Feasibility Assessment TECHNICAL REPORT. April, 2013. Available at: http://www.qualityforum.org/Publications/2013/04/eMeasure_Feasibility_Assessment.aspx. Accessed on: June 1, 2015.

- Initial data element validity for an eMeasure can be evaluated based on complete agreement between data elements and computed measure scores obtained by applying the eMeasure specifications to a simulated test EHR data set with known values for the critical data elements and computed measure score.
- A crosswalk of the eMeasure specifications (QDM quality data elements, code lists, and measure logic) is needed for a retooled measure. Analysis of actual data usage in a clinical setting is essential to determine if the measure concept is realistic. For example, a laboratory test result may indicate the available laboratory tests using a LOINC code set (value set) and the results using a SNOMED value set such as (a) reactive, (b) positive, (c) non-reactive, and (d) negative. Simulated data will provide results with expected SNOMED values. However, actual results may include such items as "NR," "non-R," "neg," "pos," "N," "P," or "inadequate specimen," "specimen not received," etc. If the analysis shows such results, the frequency should be considered before deciding what to include as a measure component.

The general procedure for gathering information is found in Chapter 1.A.—Information Gathering of Section 2.

1.4—Additional or Adapted Evaluation Subcriteria for Scientific Acceptability of the eMeasure Properties

Measure Evaluation criteria should be applied to measure concepts from the earliest phases of measure conceptualization. This principle applies to eMeasures the same way it applies to all other measures. The concepts are detailed in Chapter 5—Concepts of Measure Evaluation in Section 1. In addition to the standard measure evaluation criteria and related subcriteria, the following additional or adapted subcriteria are used to evaluate eMeasures.

The measure should be well defined and precisely specified so it can be implemented consistently within and across organizations and allow for comparability, and eMeasure specifications are based on the HQMF.

Data element validity should be demonstrated by comparing data elements exported electronically and data elements manually abstracted from the entire EHR and a statistical analysis of results; OR, complete agreement between data elements and computed measure scores obtained by applying the eMeasure specifications to a simulated test EHR data set with known values for the critical data elements. Note that the latter method is less robust as it assumes actual clinical documentation will parallel the method for incorporating the simulated data.

Documentation should demonstrate similarity (within empirically justified error limits) of scores produced by the retooled eMeasure specifications with scores produced by the original measure specifications.

- A crosswalk of the eMeasure specifications (QDM quality data elements, code lists, and measure logic) to the original measure specifications is needed for a retooled measure. Note that comparability is only an issue when maintaining two sets of specifications.

1.5—Technical Expert Panel/Subject Matter Expert

It is CMS's intention to increase focus on patient and caregiver input throughout the quality measure development process, including eMeasure development. Including a patient or caregiver representative on the TEP roster is an effective way to ensure input on the quality issues that are important to patients. Participation by consumer and patient advocacy organizations may be desirable but should not be used in lieu of actual patients. In addition, the TEP for an eMeasure should include recognized SMEs in relevant fields, such as:

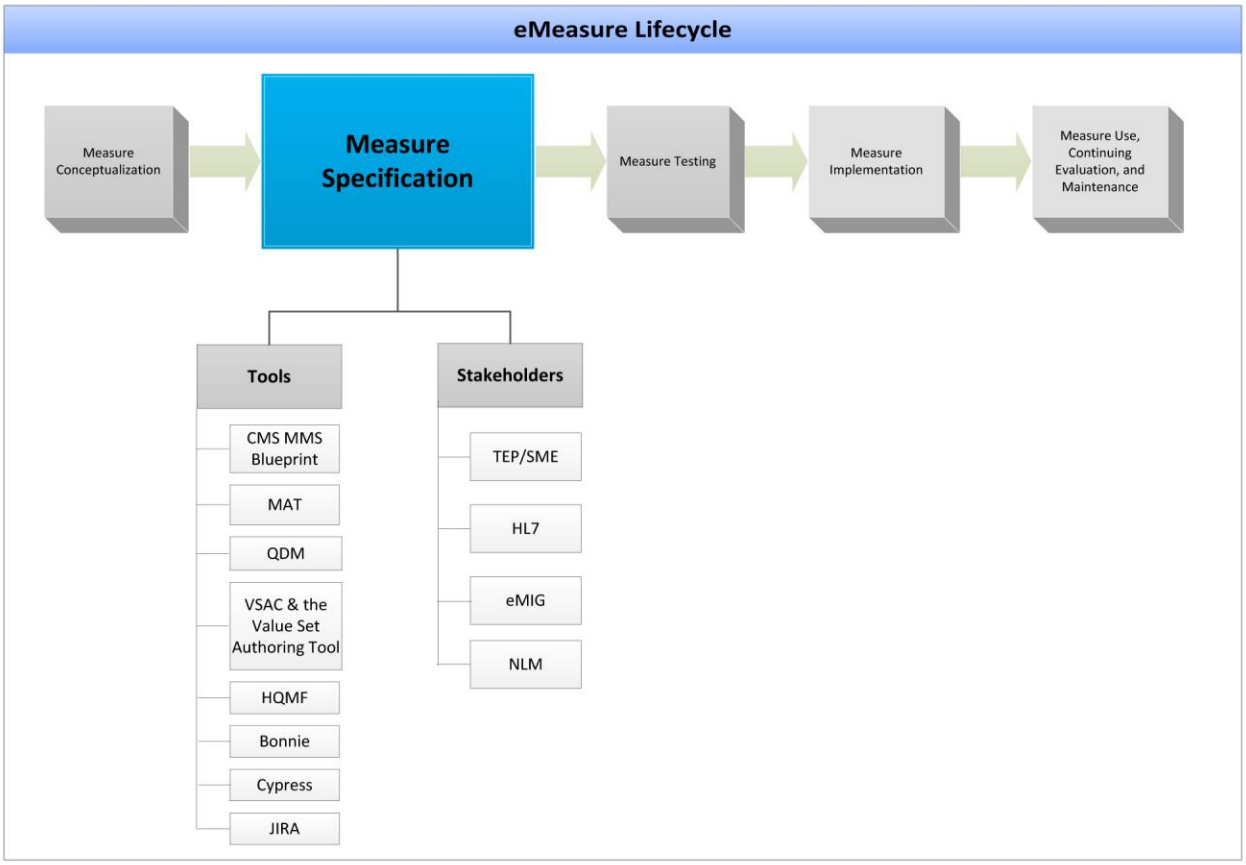
- EHR/health IT vendors.
- Implementers of electronic health record (EHR) systems.
- Medical informaticists.
- Programmers.
- Coding experts.
- Other measure developers—the collaborative process encouraged by the MIDS.
- Current EHR users (e.g., staff from measure testing sites).

Detailed information about the procedure for obtaining TEP input is found in Section 2, Chapter 1.B. —Technical Expert Panel.

Chapter 2—eMeasure Specification

eCQM (or eMeasure) 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 systems at various stages during development and maintenance. The systems used during eCQM development and maintenance include measure analysis and information gathering tool, MAT, VSAC, MITRE certification and testing tools (i.e., Cypress and Bonnie), as well as the ONC and CMS CQM Issue Tracker system (JIRA). The process described here applies to *de novo* eCQMs, retooled or reengineered eCQMs, and eCQM maintenance including annual update. Depending on the type of eCQM development, the appropriate process should be identified and applied. Figure 31: eMeasure Specification Tools and Stakeholders depicts the tools and key stakeholders needed for measure specification.

Figure 31: eMeasure Specification Tools and Stakeholders



2.1—Deliverables

- Draft eMeasure specifications
- Final eMeasure specifications
  - eMeasure XML File
  - Simple XML File

  - eMeasure human-readable rendition (HTML) File
  - eMeasure style sheet
  - Value sets—exported from the VSAC for delivery to CMS as the VSAC is the authoring center for value sets.
- Measure Justification Form (required for new eMeasures only)

2.2—JIRA Basics for eCQM Developers

ONC JIRA tracking system

The ONC JIRA project system is a platform that allows teams to create, capture, and organize issues, develop solutions, and follow team activities for multiple health IT projects. In addition, issue tracking in JIRA allows 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 and implementation 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 or recommendations on the eMIG.

Creating an account

If no JIRA account exists, set up an account on the JIRA website. Select “sign up for a new account” under the *Login* area of the Homepage. When prompted, fill in user-specific information and choose a password. Finish by clicking on the “Sign up” button to register.

Access to projects and measures

To access projects in JIRA, select the Projects menu. Here, users can view current and recently accessed projects, as well as browse a list of all projects available in JIRA. The two projects most relevant to eCQM development and maintenance are the CQM Issue Tracker and CQM Annual Update. The CQM Issue Tracker is used for tracking user- and vendor-reported issues related to published eCQMs and provides coordinators, developers, and stewards an opportunity to respond to these issues. JIRA assigns a specific ticket to each comment such that all tickets relevant to each eCQM can be tracked, related to other relevant tickets, and reported in the CQM Issue Tracker. The CQM Annual Update (CAU) is only accessible to the individuals who are involved in the measure annual update process. Generally, each CAU ticket links a significant number of CQM Issue Tracker tickets. Each is addressed in the subsequent annual update.

Innovative uses of JIRA

The function of JIRA has evolved and expanded from communication among various stakeholders on numerous health IT projects to a source of information as a central file-sharing repository for the 2014 Meaningful Use 2 (MU2) annual update, and aggregate public comments on eCQMs.

2.3—Develop eMeasure Specifications

HQMF

A clinical quality measure encoded in the Health Quality Measures Format is referred to as an "eMeasure" or "eCQM." HQMF is an HL7 standard for representing a health quality measure as an electronic XML document (refer to Appendix A: XML View of a Sample eMeasure for a detailed example of how an eMeasure will appear when rendered in a Web browser). 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 will ideally be able to push a button and import these eCQMs into their EHRs. The eCQMs ideally can be turned into queries that automatically gather data from the EHR's 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, descriptions of numerators and denominators, measure stewards, measure type and measure scoring, as well as NQF number. The HQMF Body contains eCQM chapters (e.g., data criteria, population criteria, and supplemental data elements). Each chapter can contain narrative descriptions and formally encoded HQMF entries.

The MAT was developed under contract with CMS to aid in the creation of eCQMs.[227] The authoring tool uses a graphical user interface to guide measure developers through the measure authoring process to create an eCQM. The tool hides much of the complexity of the underlying HQMF from the measure developer. Any eCQM intended to be submitted for NQF endorsement must be submitted in HQMF. This process is simplified when measure developers author their eCQMs in the MAT. This chapter of the Blueprint assumes that measure developers will be using the MAT and describes the process from that perspective.

The eCQM is the 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. The transmission format (e.g., the interoperability specification that governs how the individual or aggregate patient data are to be communicated to CMS) is another important component of the

[227] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Measure Authoring Tool.* Available at: https://www.emeasuretool.cms.gov/. Accessed on: June 1, 2015.

quality framework, known as Quality Reporting Document Architecture (QRDA). Aligning an eCQM with QRDA maximizes consistency and may lead to successful data submission to CMS.

Quality Data Model

The Health Information Technology Expert Panel (HITEP), a committee of content experts convened by NQF, created the QDM, formerly referred to as Quality Data Set or QDS.[228] The QDM continues to evolve and further information about the QDM including the QDM User Group can be found on HealthIT.gov. The latest version (QDM 4.1) was published July 2014 and is incorporated into the July 2014 MAT Update. The QDM is an information model that defines concepts that recur across quality measures and clinical care and is intended to enable automation of EHR use. The QDM contains six components:

- QDM element—an atomic (smallest possible) unit of information that has precise meaning to communicate the data required within a quality measure.
- Category—a particular group of information that can be addressed in a quality measure.
- Datatype—context expected for any given QDM element.
- Code set—standard vocabulary, taxonomy, or other classification system to define a QDM element's category (e.g., SNOMED-CT, LOINC, RxNorm).
- Value set—the list of codes, or values, used to define an instance of a category.
- Attribute—specific detail about a QDM element that further constrains a QDM element concept.

A QDM element is specified by selecting a category, the datatype in which the category is expected to be found with respect to electronic clinical data, a value set from an appropriate taxonomy (or vocabulary), and all necessary attributes. For example, defining a value set for pneumonia and applying the category (diagnosis) and the datatype "diagnosis, active" forms the QDM element, "diagnosis, active: pneumonia" as a specific instance for use in a measure.

Building Block Approach to eCQMs

Measure developers should develop eCQMs (or eMeasures) using the building block approach. This approach, built into the MAT, takes each datatype (e.g., diagnosis, active) in the QDM and represents it as a reusable pattern. Coupled with a value set (e.g., SNOMED CT, ICD 10-CM and ICD9-CM codes for pneumonia), a quality pattern becomes a QDM element representing an HQMF data criterion (e.g., "active diagnosis of pneumonia"). Data criteria are assembled (using Boolean and other logical, temporal, and numeric operators) into population criteria (e.g., "Diagnosis, Active: Pneumonia", AND "Patient Characteristic Birthdate: birth date" <= 17 year(s) starts before start of "Encounter, Performed: Encounter Inpatient"), thereby creating an unambiguous representation of a quality measure.

[228] National Quality Forum. *Quality Data Model.* Available at: https://www.qualityforum.org/QualityDataModel.aspx#t=2&s=&p=. Accessed on: June 1, 2015.

Thus, at a high level, the process of creating an eCQM is to map measure data elements to the correct datatypes in the QDM, associate each datatype with the correct value set(s) to create data criteria, and then assemble the data criteria into population criteria.

Measure Authoring Tool (MAT)

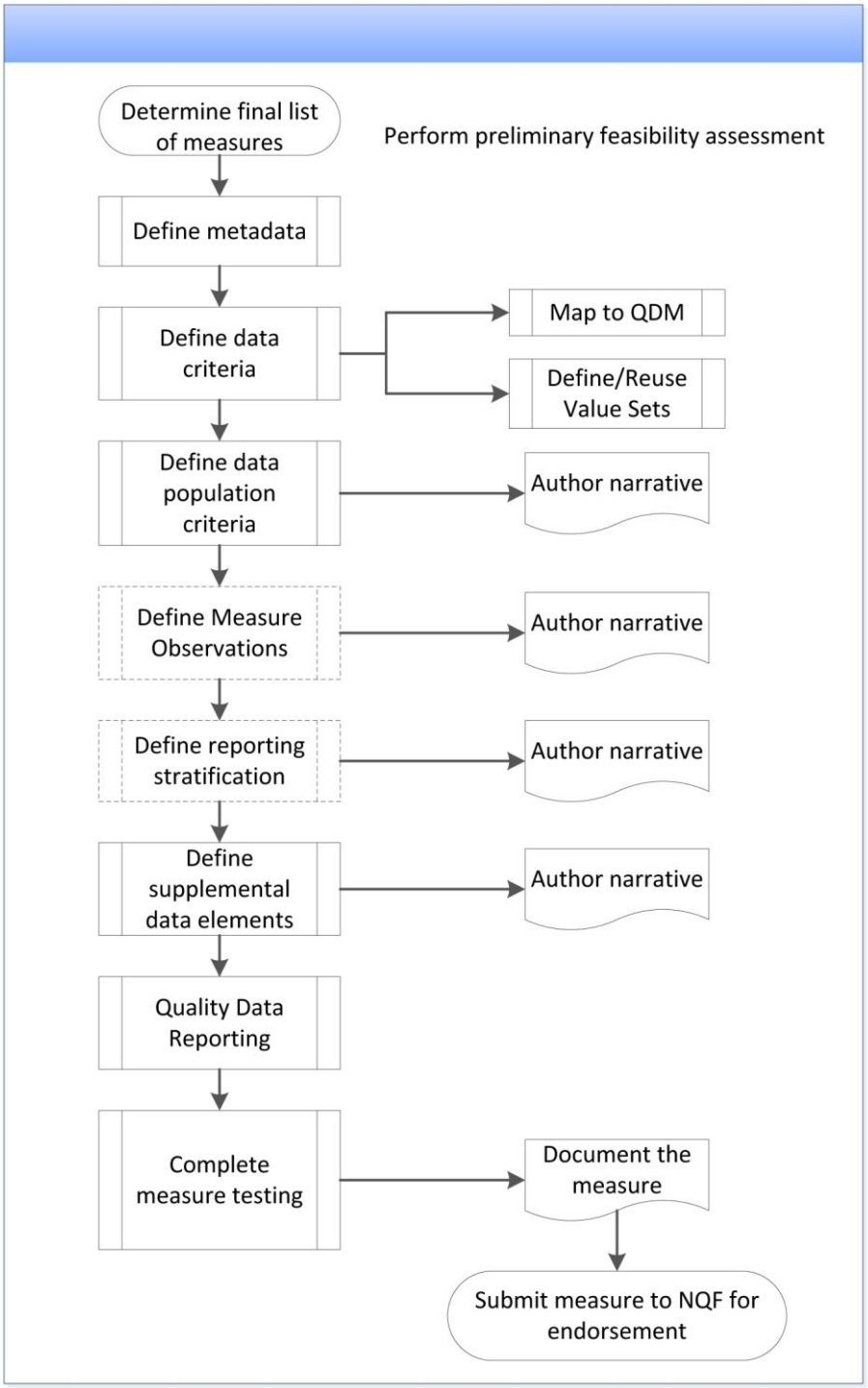
The MAT is a publicly available, web-based tool that is used by measure developers to create eCQMs. The tool allows measure developers to create their eCQMs in a highly structured format using the QDM and healthcare industry standard vocabularies. The MAT does not require its users to have an extensive knowledge of HQMF standards. The MAT is based on the QDM and the building block approach to creating eCQMs and supports common use cases and the existing patterns in the HQMF pattern library. The MAT requires ongoing maintenance and support to meet future measure authoring requirements. For instance, if a measure requires a new datatype and therefore a new corresponding quality pattern, the new datatype must be approved by the QDM user group and subsequently the pattern will be developed and then added to the MAT. Changes to the MAT are evaluated and prioritized by a Change Control Board (CCB) coordinated by CMS.

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 one week to be processed.

2.4—Procedure

Whether retooling or developing a *de novo* eCQM, the measure developer should use the process outlined in Figure 32: eMeasure Specification Development Process and detailed following the figure. The processes in dashed boxes are not required in all circumstances.

Figure 32: eMeasure Specification Development Process



Step 1: Determine final list of measures (and perform preliminary feasibility assessment)

Based on the environmental scan, measure gap analysis, and other information gathering activities, the measure developer submits the list of candidate measures—selected with TEP input—to the COR for approval. These measures may be retooled, reengineered, or *de novo*.

Before deciding to retool or reengineer a measure, a measure developer must consider the following issues:

- If the existing measure is NQF-endorsed, are the changes to the measure significant enough to require resubmission to NQF?
- 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?

These considerations must be discussed with the COR and the measure owner, and NQF endorsement status may need to be discussed with NQF. Upon approval from the COR, the measure developer proceeds with the development of detailed technical specifications for the measures using the MAT.

Prior to drafting initial eCQM specifications, the measure developer must determine the data elements necessary for the proposed measure and conduct preliminary feasibility assessments (alpha testing) to confirm availability of the information within a structured format in the electronic health record. Should feasibility concerns be identified early in the development of the measure, this will allow measure developers sufficient time to: (1) replace or revise data elements, (2) stop further development, or (3) maintain the data elements and determine a plan for addressing the concerns.

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

Step 2: Define metadata

The Header of an eCQM identifies and classifies the document and provides important metadata about the measure. eCQM metadata is summarized in the table below listing elements in the order that 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. Table 14 below lists all header (metadata) elements, definitions of header data elements, and guidance for measure developers to describe the element and conventions to use (e.g., "none" or "not applicable") if the element is optional. "Required" in the preferred term column indicates that the measure developer must populate the metadata element field as defined in column 2. 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" have been described for each metadata element and should be entered according to the Preferred Term column instructions (for example, 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.

Table 14: eCQM Metadata

Header Data Elements	Definition	Measure Developer Guidance	Preferred Term (Required, None, Not Applicable)
eCQM Title	The title of the quality eCQM.		Required
eCQM Identifier (Measure Authoring Tool)	Specifies the eCQM identifier generated by the MAT.	Field is auto-populated by the MAT.	Required
eCQM Version Number	A positive integer value used to indicate the version of the eCQM.	Displays the integer (whole number) the measure developer enters. The version number is a whole integer that should be increased each time the measure developer makes a substantive change to the content or intent of the eCQM or the logic or coding of the measure. Typos or word changes that do not affect the substantive content or intent of the measure would not require an increment in version number (e.g., changing from an abbreviation for a word to spelling out the word for a test value unit of measurement).	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 quality 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

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, continuous variable, ratio).		Required
Measure Type	Indicates whether the eCQM is used to examine a process or an outcome over time. (e.g., Structure, Process, Outcome).		Required
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 Emergency Department 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

Header Data Elements	Definition	Measure Developer Guidance	Preferred Term (Required, None, Not Applicable)
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
Rate Aggregation	Describes how to combine information calculated based on logic in each of several populations into one summarized result. (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).		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
Measurement Duration	Field will not be used.	Does not show in the style sheet.	None
Reference(s)	Identifies bibliographic citations or references to clinical practice guidelines, sources of evidence, or other relevant materials supporting		None

Header Data Elements	Definition	Measure Developer Guidance	Preferred Term (Required, None, Not Applicable)
	the intent and rationale of the eCQM.		
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
Transmission Format	May be a URL or hyperlinks that link to the transmission formats that are specified for a particular reporting program.	For example, it could be a hyperlink or URL that points to the Quality Reporting Document Architecture (QRDA) Category I implementation guide, or a URL or hyperlink that points to the Physician Quality Reporting System (PQRS) Registry XML specification. This is a free text field. Further guidance forthcoming.	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 quality assurance step, it is important to compare the human-readable description of the measure population (in the header) to the logic representation (in the body) for any discrepancies. This field will be the primary field to fully define the comprehensive eligible population for proportion/ratio eCQMs or the eligible measure population for continuous variable 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	For proportion/ratio measures, include the text “Equals Initial Population” where applicable.	Not Applicable (for continuous variable eCQMs)

Header Data Elements	Definition	Measure Developer Guidance	Preferred Term (Required, None, Not Applicable)
	eCQM. Different measures within an eCQM set may have different Denominators. Continuous variable eCQMs do not have a Denominator, but instead define a Measure Population.		
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 continuous variable eCQMs*)
Numerator	Numerators are used in proportion and ratio eCQMs. 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. 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 continuous variable 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. Note that the Numerator Exclusion population for a proportion measure is not yet supported by the current MAT.	None (*for ratio and proportion eCQMs*) Not Applicable (*for continuous variable 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. Denominator Exception are used only in proportion eCQMs. They are not appropriate for ratio or continuous variable 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 medical reasons.	None (*for proportion eCQMs*) Not Applicable (*for ratio or continuous variable eCQMs*)
Measure Population	Measure Population is used only in continuous variable eCQMs. It is a narrative description of the eCQM population. For example, all patients seen in the Emergency Department during the measurement period.	For continuous variable 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 emergency department, exclude those transferred directly to another acute care facility for tertiary treatment.	Note that Measure Population Exclusion is supported by the July 2014 MAT release.	None (*for continuous variable eCQMs*) Not Applicable (*for ratio or proportion eCQMs*)
Measure Observation	Measure Observation is used only in ratio and continuous variable eCQMs. They provide the description of how to evaluate performance.	Note that use of Measure Observation in ratio eCQMs is now supported by the July 2014 MAT release. Note that use of Measure Observation for risk variables in risk-	Not Applicable (for proportion eCQMs)

Header Data Elements	Definition	Measure Developer Guidance	Preferred Term (Required, None, Not Applicable)
	For example, the mean time from arrival to departure for all Emergency Department visits during the measurement period. Measure Observation is generally described using a statistical method such as count, median, mean, etc.	adjusted outcome measures is not yet supported by the current MAT. Note that risk variables in risk-adjusted outcome measures are currently represented as Supplemental Data Elements, but are represented as standalone Risk Adjustment Data Criteria in the MAT.	
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. Note that risk variables in risk-adjusted outcome measures are currently represented as Supplemental Data Elements, but are represented as standalone Risk Adjustment Data Criteria in the MAT.	Required

Additional Standardized Definitions and Conventions

Diagnosis Types

To more closely align measure intent and data workflow processes, four diagnosis types (principal, admission, discharge, and encounter) have standard definitions to allow for more granular representation. Once the definitions and concepts are added to the QDM, representation in HQMF, the MAT, and QRDA can also be finalized. Refer to Table 15 for the standardized diagnosis types available for use in eCQMs.

Table 15: Diagnoses Available for Use in eCQMs

Diagnosis Type	Definition
Principal	The condition established after study to be chiefly responsible for occasioning the admission of the patient to the hospital for care. *Uniform Hospital Discharge Data Set* (UHDDS). Federal Register (vol. 50, no. 147)
Admission	For an inpatient admission, the condition(s) identified by the clinician at the time of the patient's admission requiring hospitalization. Note: There can be multiple admission diagnoses.

Diagnosis Type	Definition
Discharge	For an inpatient setting—conditions identified during the hospital stay that either need to be monitored after discharge from the hospital and/or were resolved during the hospital course. *Healthcare Information Technology Standards Panel (HITSP) C83 Clinical Document Architecture (CDA) Content Modules Component* Note that discharge diagnoses do not need to be the same as "Principal" diagnoses.
Encounter	For an outpatient setting—conditions addressed during the encounter that were resolved or require ongoing monitoring.

Age Calculation Conventions & Logic Guidance

Measure developers must use the following conventions in conjunction with Appendix E: Time Unit and Time Interval Definitions and Appendix F: Time Interval Calculation Conventions when developing the age logic for eCQMs intended to be used in CMS programs. This will ensure standardization and allow for unambiguous interpretation. These conventions are intended to standardize the time calculation units for durations, which include age (e.g., the 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"). Refer to the MAT July 2014 release User Guide for updates to this guidance as some entry options have changed.

Table 16: Conventions for Constructing Age Logic

Measure Setting	Age Convention/Logic Guidance
Ambulatory	Convention: To be included in the eCQM, the patient's age must be >= IP inclusion criterion before the start of the Measurement Period. The Measurement Period is January 1–December 31, 20xx. Guidance: Note A—An example of a corresponding representation in HQMF for an IPP inclusion criterion for patient age range: • "Patient Characteristic Birthdate: birth date" >= 18 years starts before start of "Measurement Period"; AND • "Patient Characteristic Birthdate: birth date" < 75 years starts before start of "Measurement Period" Note B—Upper age bounds must use < (instead of <=)[229] because of the way "year" is calculated. If, for example, the upper age limit is "<= 75," then people who are slightly older than 75 (e.g., 75 years and 1 day) are not counted. Using "< 75" allows those that are 74 and 364 days old and turn 75 during the measurement period to be included. Using the example in Note A, if the upper age is expressed as "< 76," then people that are 75 and 364 days and turn 76 during the measurement year are included, and therefore the wrong age for inclusion.
Hospital	Conventions: In the context of hospital measures, patient age can be defined in multiple ways, all of which retain significance in the context of a single episode of care. While the most commonly used age calculation is patient age at admission (the age of the patient [in years] on the day the patient is admitted for inpatient care), there are situations where specific population characteristics or the measure's intent require distinct calculations of patient age: • Patient age at admission (in years) [refer to example in note A below] = Admission date - Birth date

[229] When considering the upper boundary age and the operator used in the measure, the primary consideration should be which age is *inappropriate* for inclusion in the measure.

Measure Setting	Age Convention/Logic Guidance
	• Newborn patient age at admission (in days) [refer to example in note B below] = Admission date - Birth date • Patient age at discharge (in years) [refer to example in note C below] = Discharge date - Birth date • Patient age at time of event (in years) [refer to example in note D below] = Event date - Birth date Guidance: In determining the operator used for the logic, consider the reference date that age is being calculated against. This will determine whether use of "less than equal to" or "less than" is more appropriate for the measure purposes. Note A—An example of a corresponding representation in HQMF for an IPP inclusion criterion for patient age range at admission: • "Patient Characteristic Birthdate: birth date" >= 18 years starts before start of x; AND • "Patient Characteristic Birthdate: birth date" <= 65 years starts before start of x Use of "<=" when the reference date is admission includes those patients who turn 65 years on the reference date, while using "<" operator would exclude them. Note B—An example of a corresponding representation in HQMF for newborn age: Patient is <= 20 days old at time of admission: • "Patient Characteristic Birthdate: birth date" <= 20 days starts before start of "Encounter Performed: hospital encounter (admission)" Note C—An example of a corresponding representation in HQMF for age at discharge: Patient is >= 10 years old at time of discharge: • "Patient Characteristic Birthdate: birth date" >= 10 years starts before start of "Encounter Performed: hospital encounter (discharge)" Note D—An example of corresponding representation in HQMF for age at time of an event: Patient is <21 at time of hepatitis vaccination: • "Patient Characteristic Birthdate: birth date" < 21 years starts before start of "Medication Administered: hepatitis vaccine (date time)" Use of "<" in this instance excludes patients who are 21 years of age at the reference date (vaccination), while "<=" would include them. The significance of < 21 years and <=20 years is equivalent, i.e., there is no age in years between 20 and 21.

Step 3: Define data criteria and key terms

Map to QDM

As previously indicated, the process of creating a new eCQM is to map measure data elements to the correct datatypes in the QDM, associate each datatype with the correct value set(s) to create data criteria, and then assemble the data criteria into population criteria. The process works somewhat differently for retooled measures.

Retooled measures

Measure developers need to identify data elements in the existing paper-based measure and map them to QDM datatype to define data criteria in a quality measure retooling or reengineering scenario (when transforming an existing paper measure into an eCQM). Measure developers will associate each QDM datatype with an existing value set or create a new value set if one does not

exist. Measure developers are expected to author a retooled measure directly in eCQM format using the MAT. The Health IT Standards Committee (HITSC) has developed a set of recommendations to report quality measure data using clinical vocabulary standards listed in When creating value sets, it is important to align with the vocabulary standards adopted for eCQM use as implemented by HITSC (refer to Table 17).

Recognizing that immediate use of clinical vocabularies may be a challenge, the HITSC also developed a transition plan that includes a list of acceptable transition vocabularies and associated time frames for use. The transitional vocabularies are listed in Table 188.

New measures

Measure developers are expected to author a new measure directly in eCQM format using the MAT. A data criterion will be constructed based on a QDM datatype. Measure developers will associate each QDM datatype with an existing value set (or create a new value set if one does not exist) and define additional attributes if applicable; all value sets authoring is done in VSAC, and the value sets in VSAC can be viewed in MAT. Measure developers should follow the HITSC recommendations to report quality measure data using clinical vocabulary standards listed in Table 17 and the transitional vocabularies listed in Table 18.

Clearly define any time windows

Time windows must be stated whenever they are used to describe the denominator, numerator, or exclusion. The measure developer must clearly indicate the index event used to determine the time window. Appendix E: Time Unit and Time Interval Definitions provides standardized definitions for time units and intervals. Measure developers must exactly state the interval units required to achieve the sensitivity necessary for the purpose of measurement. The selection of the time unit to be used should be made according to the level of granularity needed to meet the intent of the measure.

Example:

Medication reconciliation must be performed within 30 days following hospital discharge. Thirty days is the time window, and the hospital discharge date is the index event. If the minimum sensitivity was one month instead of 30 days, then the measure specification would state month instead of day as the unit of time.

Appendix F: Time Interval Calculation Conventions 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.”

Define/Reuse Value Sets

Value sets are specified and bound to QDM datatypes to create coded data elements in an eCQM. When creating value sets, it is important to align with the vocabulary standards adopted for eCQM use as implemented by HITSC (refer to Table 17).[230]

Table 17: ONC HIT Standards Committee Recommended Vocabulary Summary

Clinical Vocabulary Standards:	Others:
SNOMED CT	CVX—Vaccines Administered
LOINC	CDC-PHIN/VADS
RxNorm	UCUM—the Unified Code for Units of Measure
	ISO-639
	PHDC Payer Typology

On January 16, 2009, HHS released a final rule mandating that everyone covered by the HIPAA must implement ICD-10 for medical coding by October 1, 2013. However, on April 1, 2014, the "Protecting Access to Medicare Act of 2014" H.R. 4302 bill was signed, which delays the compliance date for ICD-10 from October 1, 2014, to October 1, 2015, at the earliest. Therefore, the time frame for including ICD-9-CM in eCQMs is extended.

The transition vocabulary standards summary and plan are listed below in Table 18.

Table 18: ONC HIT Standards Committee Transition Vocabulary Standards Summary and Plan

Transition Vocabulary	Transition period: Acceptable for reporting eCQM results	Final date for reporting eCQM results
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
ICD-10-CM	With dates of service on or after the implementation of ICD-10	Final Date: one year after MU-3 is effective
ICD-10-PCS	With dates of service on or after the implementation of ICD-10	Final Date: one year after MU-3 is effective
CPT	During MU 1,2,3 if unable to report using clinical vocabulary standards	Final Date: one year after MU-3 is effective
HCPCS	During MU 1,2,3 if unable to report using clinical vocabulary standards	Final Date: one year after MU-3 is effective

When specifying eCQMs using value sets, measure developers should note the following:

- ICD9-CM and ICD10-CM Group Codes—Codes that are not valid for clinical coding 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.

Some examples include:

[230] HITSC recommendations were developed by the HITSC Clinical Quality Workgroup and Vocabulary Task Force and accepted by the full HITSC in September 2011.

- Use the following fifth-digit sub-classification with category 948 to indicate the percent of body surface with third degree burn:
- 0—less than 10 percent or unspecified
- 1—10–19 percent
- 2—20–29 percent
- 3—30–39 percent
- 4—40–49 percent
- 5—50–59 percent
- 6—60–69 percent
- 7—70–79 percent
- 8—80–89 percent
- 9—90 percent or more of body surface
- 632 Missed abortion—This is a standalone code and does not require any additional digits to be valid.
- 633 Ectopic pregnancy—This is a non-billable code that must have additional digits to be valid (e.g., 633.1 Tubal pregnancy).

Allergy Value Sets

Allergy/Intolerance value sets, when drawn from RxNorm, should include RxNorm concepts having Term Types:

- Brand Name (BN)
- Ingredient (IN)
- Multiple Ingredient (MIN)
- Precise Ingredient (PIN)

Measure Developer Guidance:

- Create a substance-allergen value set in addition to a prescribe-able medication allergy value set (these two value sets will NOT be merged into a “super value set”).
- The QDM datatype for substance-allergen value sets will continue to be tied to the medication category value sets rather than substance category value sets.

Table 19: Allergy and Intolerance Value Set Naming Convention

Convention	Examples
End the value set name with the word “allergen”	“Antithrombotic Therapy Allergen”
	“Beta Blocker Therapy Allergen”

Medication Value Sets

Medication value sets, when drawn from RxNorm, should include RxNorm concepts having these Term Types:

- Brand Pack (BPCK)

- Generic Pack (GPCK)
- Semantic Brand Drug (SBD)
- Semantic Brand Drug Form (SBDF)
- Semantic Brand Drug Group (SBDG)
- Semantic Clinical Drug (SCD)
- Semantic Clinical Drug Form (SCDF)
- Semantic Clinical Drug Group (SCDG)

QDM Categories with Recommended Vocabularies

Table 20: Quality Data Model Categories With ONC HIT Standards Committee Recommended Vocabularies maps the recommendations to each QDM category. Measures that are retooled or developed must include ICD-10 codes where applicable.

Table 20: Quality Data Model Categories With ONC HIT Standards Committee Recommended Vocabularies

General Clinical Concept	Quality Data Model Category	Quality Data Model Attribute	Clinical Vocabulary Standards	Transition Vocabulary
Adverse Effect/ Allergy/Intolerance	Device	Reaction	SNOMED CT	N/A
	Diagnostic Study	Reaction	SNOMED CT	
	Intervention	Reaction	SNOMED CT	
	Laboratory Test	Reaction	SNOMED CT	
	Medication	Reaction	SNOMED CT	
	Procedure	Reaction	SNOMED CT	
	Substance	Reaction	SNOMED CT	
Substance	Substance	N/A	SNOMED CT	N/A
Condition/Diagnosis/ Problem	Condition/Diagnosis/ Problem	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	N/A	SNOMED CT	CPT, HCPCS, ICD-9-CM Procedures, ICD-10-PCS
Device	Device	N/A	SNOMED CT	N/A
Physical exam finding	Physical Exam	Result	SNOMED CT	N/A
Laboratory test (names)	Laboratory Test	N/A	LOINC	N/A
Laboratory test (results)	Laboratory Test	Result	SNOMED CT	N/A
Diagnostic study test names	Diagnostic Study	N/A	LOINC	HCPCS
Diagnostic study test results	Diagnostic Study	Result	SNOMED CT	N/A

General Clinical Concept	Quality Data Model Category	Quality Data Model Attribute	Clinical Vocabulary Standards	Transition Vocabulary
Units of Measure for results	N/A	N/A	UCUM-the Unified Code for Units of Measure	N/A
Intervention	Intervention	N/A	SNOMED CT	CPT, HCPCS, ICD-9-CM Procedures, ICD-10-PCS
Procedure	Procedure	N/A	SNOMED CT	CPT, HCPCS, ICD-9-CM Procedures, ICD-10-PCS
Patient characteristic, preference, experience (expected answers for questions related to patient characteristic, preference, experience)	Applies to All QDM Categories Except System Characteristic	Patient preference	SNOMED CT	N/A
Functional Status	Functional status	N/A	SNOMED CT	N/A
Categories of function	Functional status	Result	ICF-(International Classification of Functioning, Disability, and Health)	N/A
Communication	Communication	N/A	SNOMED CT	CPT, HCPCS
Assessment instrument questions (questions for patient preference, experience, characteristics)	Individual Characteristic	N/A	LOINC	N/A
Medications (administered, excluding vaccines)	Medication	N/A	RxNorm	N/A
Vaccines (administered)	Medication	N/A	CVX—Vaccines Administered	N/A
Patient characteristic (Administrative Gender, DOB)	Individual Characteristic	N/A	CDC-Public Health Information Network (PHIN)/Vocabulary Access and Distribution System (VADS) http://www.cdc.gov/phin/activities/vocabulary.html	N/A
Patient Characteristic (Ethnicity, Race)	Individual Characteristic	N/A	OMB Ethnicity/Race (scope) http://www.cdc.gov/phin/library/resources/vocabulary/CDC%20Race%20&%20Ethnicity%20Background%20and%20Purpose.pdf – expressed in PHIN VADS as value sets as the vocabulary	N/A

General Clinical Concept	Quality Data Model Category	Quality Data Model Attribute	Clinical Vocabulary Standards	Transition Vocabulary
Patient characteristic (Preferred language)	Individual Characteristic	N/A	ISO-639-1:2002	N/A
Patient characteristic (Payer)	Individual Characteristic	N/A	Payer Typology (Public Health Data Standards Consortium Payer Typology) (scope) 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

To the extent possible, use existing QDM value sets when developing new 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 vocabularies to build a new list.

CMS measure developers should refer to the periodic updates to these guidelines issued by eMIG for the most up-to-date vocabulary recommendations. Other measure developers not involved in eMIG may refer to the ONC HIT Standards Committee website.[231]

At times, there may be a need to request new SNOMED-CT concepts. The request should be submitted through the U.S. SNOMED CT Content Request System (USCRS) of the NLM.[232] Measure developers must sign up for a UMLS Terminology Services account to log into the USCRS.[233] Measure developers should consider contractual timelines when considering applying for new concepts.

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.[234] Measure developers should consider contractual timelines when considering applying for new concepts.

Retooling or creating a new measure may require reusing existing value sets or defining 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-9-CM codes, such as 250.0, 250.1, and 250.2 and SNOMED CT and ICD-10-CM codes as well. VSAC is the only authoritative tool to author value sets for Meaningful Use. Several tools exist to help build and

[231] Office of the National Coordinator for Health Information Technology. *Health Information Technology: Health IT Standards Committee.* Available at: https://www.healthit.gov/facas/health-it-standards-committee. Accessed on: June 1, 2015.

[232] National Library of Medicine. *U.S. SNOMED CT Content Request System (USCRS)*. Available at: https://uscrs.nlm.nih.gov/. Accessed on: June 1, 2015.

[233] National Library of Medicine, National Institutes of Health. *Unified Medical Language System®: UMLS Terminology Services.* Unified Medical Language System® is a registered trademark of the National Library of Medicine and the National Institutes of Health. Available at: https://uts.nlm.nih.gov/home.html. Accessed on: June 1, 2015.

[234] Regenstrief Institute, Inc. *Logical Observation Identifiers Names and Codes (LOINC®): Process for Submitting New Term Requests*. LOINC® is a registered trademark of Regenstrief Institute, Inc. Available at: http://loinc.org/submissions/new-terms. Accessed on: June 1, 2015.

maintain quality measure value sets. Some of these tools can take value set criteria (e.g., "all ICD9 codes beginning with 250") and expand them into an enumerated list. Where such value set criteria exist, they should be included as part of the value set definition.

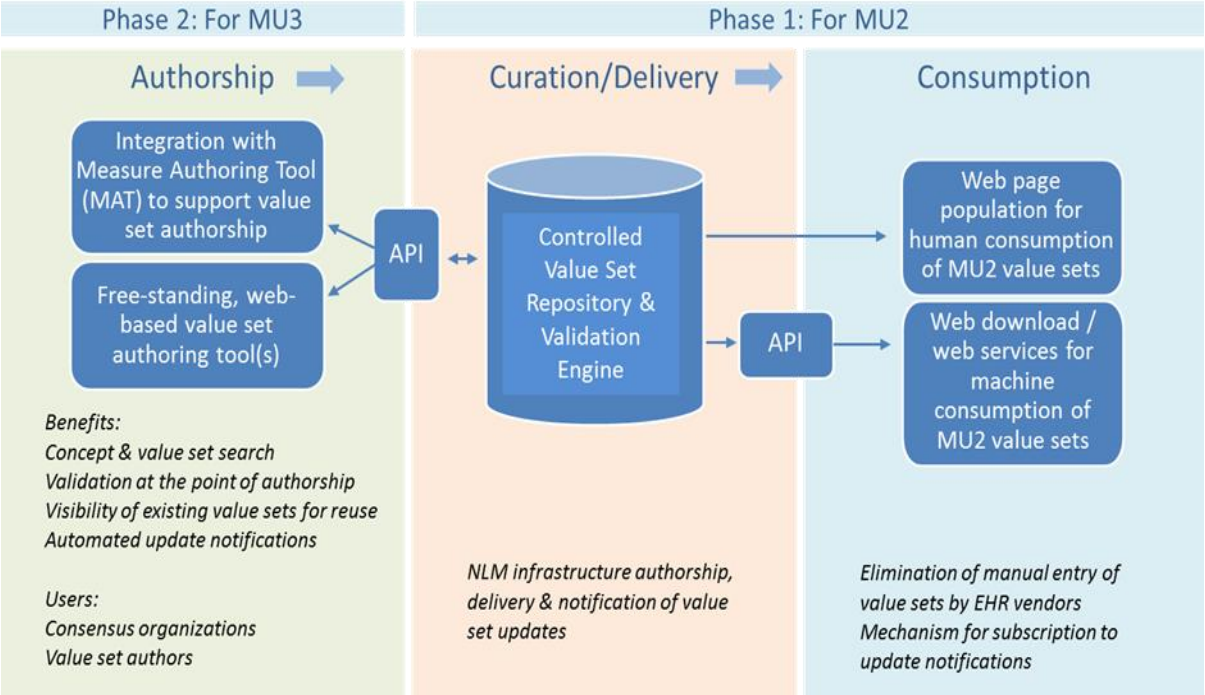
To identify the recommended vocabularies as defined by the HITSC, it may be necessary to identify multiple subsets for a measure data element (e.g., SNOMED-CT child value sets, ICD-9-CM child value sets, and ICD-10-CM child value sets). VSAC has defined a grouping mechanism for this scenario to create 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-9-CM child value set. They then define a Diabetes Mellitus Grouping value set that combines the two subsets and associate the grouped value set with the measure data element. The VSAC allows only one level of grouping. Therefore, the MAT measure logic clause is needed to combine two grouped value sets. As an example, to express all patients with hematologic malignancies may be one grouped value set combining SNOMED-CT, ICD-9-CM and ICD-10-CM values and patients with primary immunodeficiencies, and those with HIV infection should be similarly grouped. However, to indicate "all immunocompromising conditions for which live virus vaccines should be avoided" requires an eMeasure clause referring to active diagnoses for each of the related grouped value sets (hematologic malignancies, primary immunodeficiencies and HIV infection)."

When defining a value set, measure developers may need to include codes that are no longer active in the target terminology. For example, a measure developer may need to include retired ICD-9-CM codes in a value set so that historic patient data, captured when the ICD-9-CM codes were active, also satisfies a criterion. Measure developers need to carefully consider the context in which their value sets will be used to ensure that the full list of allowable codes is included. 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.

To improve value set authorship, curation, and delivery, NLM performed quality assurance checks to compare the validity of value set codes and terms with the latest source vocabularies. Figure 33: Vision for Robust Value Set Management below shows the NLM's implementation for value set management.[235]

[235] National Library of Medicine. *Value Set Authority Center.* Authoring Guidance. Available at: https://vsac.nlm.nih.gov/. Accessed on: June 1, 2015.

Figure 33: Vision for Robust Value Set Management



As value set authors and measure developers create their value sets within the VSAC Authoring Tool, the tool interactively assesses the code validity within a code system, as well as other quality assurance parameters. Measure developers should take proper actions as specified by NLM based on the analysis outcome. If the VSAC or NLM quality assurance teams identify value set deficiencies, measure developers should correct the value sets using the VSAC Authoring Tool.

Step 4: Define population criteria

Population criteria are assembled from underlying data criteria. A population for a particular measure depends on the type of measure scoring (proportion, ratio, continuous variable) that is planned, as shown in Table 21 below. Additionally, the definitions for these populations are provided in the glossary.

Table 21: 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=Option. and NP=Not Permitted.

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

For a proportion measure, there is a fixed mathematical relationship between population subsets. These mathematical relationships, and the precise method for calculating performance, are detailed in Appendix G: Proportion eCQM Calculations.

For a continuous variable measure, the population for a single measure is simply a subset of the Initial Population for the measure set representing patients meeting criteria for inclusion in the measure. Appendix H: Continuous Variable and Ratio eCQM Calculations provides the mathematical relationships and performance rate calculations for both ratio and continuous variable measures.

Denominator exclusion vs. denominator exception

Definitions of each population are provided in Table 21: Measure Populations Based on Type of Measure Scoring and in the glossary. Refer to the Technical Specifications subsection of Chapter 2—Measure Specification, in Section 2, for guidance regarding when to use Denominator Exclusion versus Denominator Exception. There is a significant amount of discussion on the use of exclusion and exception, particularly the ability to capture exception in EHRs. 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 eCQM 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.

Assemble data criteria

Measure developers will use Boolean operators (AND and OR) to assemble data criteria to form population criteria (e.g., “Numerator = DataCriterion1 AND NOT (DataCriterion2)”). In addition to the Boolean operators, measure developers can also apply appropriate temporal context and comparators such as “during,” relative comparators such as “FIRST” and “LAST,” EHR context (defined below), and more. Conceptual considerations are provided here. Refer to the MAT User Guide for authoring details.[236]

Temporal context and temporal comparators

The QDM recommends that all elements have a date/time stamp. These date/time stamps are required by an eCQM for any inferences of event timing (e.g., to determine whether

[236] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Measure Authoring Tool User Guide*-Updated 1/14/2014. Available at: https://www.emeasuretool.cms.gov/documents/10179/13108/MAT+User+Guide/c50cc2fb-01ee-4b97-9cd1-2e33feca8763. Accessed on: June 1, 2015.

DataCriterion1 occurred before DataCriterion2, to determine if a procedure occurred during a particular encounter).

Relative timings allow a measure developer to describe timing relationships among individual QDM elements to create clauses that add meaning to the individual QDM elements. Relative timings are described in detail in the QDM technical specifications.[237] For instance, "starts before end" is a relative timing statement that specifies a relationship in which the source act's effective time starts before the start of the target or starts during the target's effective time. An example of this is "A pacemaker is present at any time starts before end of the measurement period." QDM documentation also specifies a list of functions. Functions specify sequencing (ordinality) and provide the ability to specify a calculation with respect to QDM elements and clauses containing them. QDM documentation includes functions such as "FIRST," "SECOND," "LAST," and "RELATIVE FIRST." Measure developers should refer to the QDM technical specification document for descriptions and examples for a particular relative timing comparator and function.

Other data relationships

A data element in a measure can be associated with other data elements to provide more clarity. These relationships include "Is Authorized By" (used to express that a patient has provided consent); "Is Derived By" (used to indicate a result that is calculated from other values); "Has Goal Of" (used to relate a Care Goal to a procedure); "Causes" (used to relate causality); and "Has Outcome Of" (used to relate an outcome to a procedure as part of a care plan). The MAT user guide also shows how to add these relationships into an eCQM.[238]

Activities That Were "Not Done"

A negation attribute may be used to identify situations where an action did not occur or was not observed for a documented reason. Modeling of negation uses AND NOT to identify NULL values when an action was not performed for a medical, patient, or system reason. This approach does not change the expression of negation in the HQMF; however, it does change the code that is associated with these activities in the QRDA-I file. The intent of the null flavor in this context is to specify that ALL the activities in the value set were intentionally not done, not that a single activity was not done or that it is not known why that activity was not completed. It is not appropriate for developers to certify that an activity was not done using negation unless the provider intentionally did not order or perform the activity in question and documented a justification why that was the case. Developers using these concepts are expected to have a documented reason for these exclusions.

---

[237] Department of Health and Human Services. *Health Information Technology (Health IT): Quality Data Model*. Available at: http://www.healthit.gov/quality-data-model. Accessed on: June 1, 2015.

[238] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Measure Authoring Tool User Guide*-Updated 1/14/2014. Available at: https://www.emeasuretool.cms.gov/documents/10179/13108/MAT+User+Guide/c50cc2fb-01ee-4b97-9cd1-2e33feca8763. Accessed on: June 1, 2015.

Data Source

"Data Source," sometimes referred to as "EHR Context," represents the place where the data are expected to be found to resolve a criterion. The health record field indicates where to find data in a specified data store, for example, the location within an EHR such as a problem list.

As part of the measure development process, measure developers will need to evaluate whether a data source and its appropriate health record fields should be specified for a criterion. If they are explicitly asserted as part of a criterion, this data source and the health record field are considered prescriptive. For instance, if the data source is an EHR and the health record field is "problem list," and they are explicitly specified as part of a criterion such as "active problem of hypertension (health record field = problem list)," then one would expect that the information regarding whether a patient has an active problem of hypertension would be found in the EHR's problem list. And only if the active problem of hypertension is found in the EHR's problem list is it then considered meeting the criterion.

QDM dataflow attributes that will be added or redefined to support a prescriptive approach include:

- Data source (new attribute)—system or place where the information is captured (e.g., Lab system or EHR system); works directly with the health record field attribute.
- Health record field (unchanged)—specific location to find the data within the data source (e.g., a specific field or place in the system).
- Source (redefined attribute)—who provided the data (e.g., patient or patient proxy).
- Recorder (redefined attribute)—who entered the data (could have the same person named as the "source").

Author narrative

In addition to the brief narrative description of each population's criteria that are part of the measure metadata, a narrative representation of the HQMF logic is also automatically generated by the authoring tool. The computer-generated narrative is standardized and concise, and it can lack the richness of full text that sometimes helps in the understanding of a measure. This is especially true for measures that have complex criteria, where the computer-generated text may not be able to express the exact description that a measure developer would like to convey. As part of the quality assurance step, it is important to compare the human-readable description of the measure population (in the header) to the logic representation (in the body) for any discrepancies.

Step 5: Define measure observation

This step is only applicable to ratio and continuous variable measures. Measure observation is required for ratio and continuous variable measures, but they are precluded for other measures.

Measure observation is variables or calculations used to score a particular aspect of performance. For instance, a measure intends to assess the use of restraints. Population criteria for the measure include "patient is in a psychiatric inpatient setting" and "patient has been restrained." For this population, the measure defines a measure observation of "restraint time" as the total amount of time the patient has been restrained. Measure observation is not a set of criteria; rather, it is a definition of observations used to score a measure.

Note that for risk-adjusted measures, the risk variables are also to be included in this chapter as measure observation.

Author narrative

It is important for measure developers to write a corresponding brief narrative description of the measure observation as part of the metadata.

Step 6: Define reporting stratification

Measure developers 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 in a facility; stratify and report separately by age group [14–19, 20–25, and total 14–25]), based on discussions with their COR.

Reporting strata are optional. They can be used for proportion, ratio, or continuous variable measures. A defined value set is often a necessary component of the stratification variable so that all measures report in the same manner.

Reporting stratification vs. population criteria

A variable may appear both as a reporting stratum and as a population criterion. For example, a pediatric quality measure may need data aggregated by pediatric age strata (e.g., neonatal period, adolescent period) and may also define an Initial Population criterion that age be less than 18.

Author narrative

It is important for measure developers to write a corresponding brief narrative description of a measure's reporting stratification as part of the metadata.

Step 7: Define supplemental data elements

CMS defines supplemental data elements 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.

Supplemental data elements are similar to reporting stratification variables in that they allow for subgroup analysis. Whereas measure developers define reporting strata, CMS defines supplemental data elements.

CMS requires that transmission formats conveying single-patient-level data must also include the following supplemental data elements reported as structured data: sex, race, ethnicity, and payer. The specific identifier used for each supplemental element can be exported from the VSAC.[239] These supplemental data elements must be present for all CMS quality measures. Different supplemental data elements can be indicated as mandatory according to different program requirements. For example, Tax Identification Number (TIN), National Provider Identifier (NPI), CMS Certification Number (CCN), and Health Insurance Claim (HIC) number may be requested as supplemental data elements with additional guidance that will be provided by the program. The MAT automatically adds the supplemental data elements section to an eCQM when the eCQM is exported from the tool. The value sets referenced by the supplemental data elements are also automatically included in the value set spreadsheet exported by the authoring tool.

CMS is evaluating additional sets for inclusion in the future, and these may include preferred language and socioeconomic status, among others.

Author narrative

The narrative descriptions for supplemental data elements about sex, race, ethnicity, and payer are automatically added to the metadata by the MAT.

Step 8: Quality data reporting

Refer to the eMeasure Reporting chapter for details.

Step 9: Complete measure testing, document the measure, and obtain COR approval

Refer to the eMeasure Implementation chapter for details.

Step 10: Submit the eMeasure to NQF for endorsement

Refer to the eMeasure Implementation chapter for details.

2.5—eCQM Specifications Review Process

During eMeasure development or maintenance, an eMeasure is likely to be shared among various organizations (e.g., developers to stewards, clinical experts) for clinical business review, initial feasibility, validity, and reliability review, in order to discover any serious flaws early in the process. A bundle of essential files and data is necessary for proper review.

The review process has two phases: the Pre-MAT phase and the Post-MAT phase. Pre-MAT measures are draft measures that either have not been entered into the MAT or are being re-worked from a previous version of a measure in the MAT. Post-MAT measures are those that have been entered or updated in the MAT and require a final round of reviews before they are

[239] National Library of Medicine. *Value Set Authority Center*. Available at: http://vsac.nlm.nih.gov. Accessed on: June 1, 2015.

published. Once a measure has passed its Pre-MAT reviews, it is cleared for MAT entry. After the measure has been entered into the MAT, the measure is reviewed again in the Post-MAT phase.

Measure developers are responsible for creating eMeasure review package materials and uploading them to a centralized system. The ONC JIRA is the centralized system where the Pre-MAT package materials were uploaded in the appropriate JIRA project. Users of the JIRA eCQM review process should reference Appendix D: eMeasure Review Process Quick Reference for naming file conventions and entering comment boxes.

The measure developer initiates a Pre-MAT review cycle by (1) uploading a zip file with their documents to be reviewed and (2) placing a comment in the JIRA record that the measure is ready for review. Measure developers can include in the zip file any additional files they would like reviewed, such as clinical guidance documentation.

Pre-MAT eCQM Package Contents

Mockup human-readable file

eCQM measure developers may mock up an existing human-readable file or create a new human-readable file. The mockup file usually is a Microsoft (MS) Office Word document with loaded eCQM human-readable HTML file. However, file types other than MS Word may be submitted. Within the mockup file, developers are allowed to use typical Word features, such as commenting, tracking changes, color coding text, to present the new changes or describe the intent of changes or clarify intent of logic in population criteria. During the review process, reviewers are able to continue making changes, suggestions, or comments in the same mockup document. The mockup document becomes a shared vehicle where discussion of proposed changes is communicated and preserved.

Value set package

Value sets can be shared in several ways. The most straightforward way is to email the list of value set codes in an MS Excel spreadsheet. Note there are rules for value set development in the VSAC. Moving forward, all value sets will be required to have specific metadata that describe the purpose and content they contain. The value set metadata includes:

- Value set name—developed based on guidelines presented on the VSAC website.
- 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 relevancy, or the 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 defines the concepts to be selected for inclusion in the value set.
- Inclusion criteria—a free text statement that defines the concepts or codes should be included and why.

- Exclusion criteria—a free text statement that defines the concepts or codes excluded and why.

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 are described with care. Evaluation based on the value set name alone may be challenging. The advantage of providing an MS Excel spreadsheet for value set review is that it is simple and easy to browse codes. Value sets exported from the VSAC include the metadata described above. The disadvantage of providing an MS 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 would 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. Another disadvantage of the MS Excel spreadsheet approach is that reviewers may not have information regarding the code system versions or what codes have been retired or added. Also, if metadata fields in the VSAC are blank, reviewers do not know the scope of the value set, such as definitions of inclusion and exclusion, etc. Those who use MS Excel spreadsheets for review should be aware of these disadvantages.

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 will be described in a later section. Although access to the VSAC requires that all users apply for a UMLS Metathesaurus License, there is no charge for registration, and use of the value sets requires such a license. Some reviewers may find the registration process cumbersome for reviewing a small value set, but it is valuable.

Test patients

While development of test patients is not a requirement in the Pre-MAT bundle review stage when artifacts are reviewed and discussed before entering into the MAT, reviewers may find them helpful to evaluate the measure content if they are provided. It is recommended that measure developers develop test patients at the Pre-Mat stage. The test patients can be described in a few sentences in an MS Word document or MS Excel spreadsheet, or they can be captured in a structured way such as in QRDA files. As eCQMs are being implemented in EHRs and more tools such as Bonnie are developed and improved, it makes sense for stakeholders to start sharing test patients in QRDA files which can be exported from the Bonnie tool or from other tools as long as they conform to the QRDA standard.

Pre-MAT Core Review Processes

Logic review

The purpose of a logic review is to ensure the logic is valid, accurate, consistent, and efficient. The details of review criteria can be found in the eMeasure Testing chapter and Appendix C: eMeasure QA Checklists.

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 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 22 and take the appropriate remedial action.

Table 22: 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.

It is acceptable to use an MS Excel spreadsheet to capture and distribute the value set and codes. It is recommended, however, to use the NLM VSAC to author and review value sets. The key benefits of using VSAC are summarized below, 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 Meaningful Use 2014 eCQMs.

- VSAC provides search, retrieval, and download capabilities through a Web interface and application programming interfaces (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.

Clinical Review

The clinical reviewer is responsible for ensuring that the eMeasure uses appropriate clinical nomenclature and clinically appropriate value sets, and that the logic effectively captures the clinical intent of the measure. This complements the work of the value set reviewer who assesses the value set element alignment with specifications in the QDM, ensuring that the value sets chosen for the eMeasure align with the QDM and verified code systems. Clinical reviewers should reference the validity testing criteria in the eMeasure Testing chapter and Appendix C: eMeasure QA Checklists for details of review criteria.

The clinical review may be performed by measure stewards, clinical analysts, or clinical work groups. However, the ultimate responsibility for the quality of the clinical review lies with the measure steward. A clinical review may occur simultaneously with the logic review and value set reviews, or it may occur after the logic review and value set reviews have been completed. The reviewed mockup document should be uploaded to the ONC JIRA for the next step.

Test Case/Design Review

The purpose of test case review is to ensure that the test case scenarios adequately capture the elements of the measure that must be evaluated (attributes, controls, and variables). Pre-MAT review of the narrative test case design ensures that the measure developer test cases are sound before the logic coding is developed. For Post-MAT review, the purpose is to ensure that measure data are evaluated correctly using a test case scenario. The test should produce results that are expected, based on the test scenarios developed. Each population criteria should be accounted for by at least one planned test patient. Ideally, each logic phrase should be tested by one test patient.

Post-MAT Bundle and Core Review Processes

After Pre-MAT review, the eMeasure developers should have the necessary information to enter the eMeasure into the MAT. The MAT output package is exported, including a human-readable HTML file, an XML file for HQMF representation, an MS Excel spreadsheet for all value sets and

codes, and a simple XML file for simplified representation of the eMeasure generated by the MAT. The MAT output package becomes essentially the core of the Post-MAT material in addition to any other documents deemed useful for Post-MAT review. The main purpose of the Post-MAT review is to ensure the designs and resolutions proposed in the Pre-MAT review are correctly translated into the MAT output. Any new changes that occurred after the Pre-MAT review will also be reviewed.

Note: starting with the July 2014 release of the MAT, the output package will not include the HQMF XML file representation. The HQMF XML file is expected to be available in a late 2014 Q4 update to the MAT output. The remaining elements of the MAT output package will continue to be available for Post-MAT review including testing in Bonnie during the interim.

Logic Review

The MAT output package is the core content used for the Post-MAT logic review. If the reviewer needs to comment on the logic, the human-readable HTML should be converted into an MS Word document, and the reviewer's comments or suggestions should be captured. The Post-MAT review is not focused as much on alignment of logic with description as on whether the design or suggestions have been correctly captured in the MAT and subsequent MAT output. If there are additional comments or suggestions, a new mockup human-readable document will be created and uploaded to JIRA. The logic review cycle will be performed until all issues are addressed. Refer to Appendix C: eMeasure QA Checklists for details of logic review criteria.

Value Set Review

The spreadsheet in the MAT output package is the main source for the Post-MAT value set review. Similar to Post-MAT logic 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.

Clinical Review

Clinical reviewers, typically the stewards, will ensure the proposed changes have been successfully translated into the MAT output human-readable file and that the output value sets are correct. The clinical review process can be iterative and will not be closed until all issues are resolved. During each review cycle, the comments and suggestions are captured in documents and subsequently uploaded to JIRA. Refer to Appendix C: eMeasure QA Checklists for details of clinical review criteria.

Test Case/Design Review

Test case review in Post-MAT should not be different from the review process for Pre-MAT. Please note that the Test Case/Design Review has not been implemented in the MU2 2014 annual update.

MAT Version Identifiers and Functionality

The MAT offers the ability to create major or minor versions of a draft measure. When a major version is created, the version of the eMeasure is increased by 1 (for example the eMeasure version is increased from 3.0 to 4.0). When a minor version function is applied, the digit after the decimal point is increased; for instance, version 3.0 is increased to 3.1. When an eMeasure is updated, it is recommended to create a new version, whether major or minor, and then create a draft using the newly created version and make changes to the draft. During the annual update, the measure developers should check with their COR for program requirements regarding advancing versions. The detailed instructions of how to save a draft as a major or minor version, and how to create a draft of an existing measure can be found in the latest MAT user guide.

MAT-VSAC Integration

With the VSAC integration into MAT, the process of authoring value sets in the VSAC and using value sets to build eMeasures in the MAT becomes seamless. To retrieve value set data from the VSAC through the MAT, users must establish an active connection to the VSAC using their UMLS Metathesaurus License credentials. To request a UMLS license, submit a request to the NLM Unified Medical Language System website.[240] MAT users are not required to establish an active connection to the VSAC to work within the MAT; however, the VSAC value set data cannot be retrieved and applied to a measure without an active VSAC session. The detailed instructions to access VSAC within MAT can be found in the latest MAT user guide.

Enhancements

Creating or modifying QDM elements with/without reference to a VSAC value set

The MAT-VSAC integration allows users the ability to create QDM elements with or without a reference to a VSAC value set. The QDM Elements tab consists of two additional sub-tabs, Create Element and Applied Elements. Within the Create Element tab, MAT users are able to build QDM elements with or without a VSAC value set data, assign a category, assign a QDM datatype, and designate specific occurrence, if applicable. To create a QDM element with a reference to a VSAC value set, a user must be actively logged in to VSAC, enter the OID for the desired value set, and enter the value set version/effective date, if applicable. To create a QDM element without a VSAC value set, a user needs a temporary name for the QDM element and the desired category and QDM datatype. The Applied Elements tab allows users to view a list of elements applied to the measure, remove unused QDM elements, modify existing QDM elements, and update the VSAC

[240] http://www.nlm.nih.gov/databases/umls.html#access

value set data manually (e.g., apply a VSAC value set to a QDM element previously entered with a temporary name).

Update value sets used in measure creation with most recent VSAC data

The option to manually update applied elements with the VSAC value set data is available to allow MAT users to capture any updates made to the VSAC value set data after the data were applied to a measure within the MAT. An active connection to the VSAC is required to manually update applied QDM elements with the VSAC value set data. For applied data elements to which no specific version or date were applied at creation, the most recent version is retrieved for all the VSAC value sets when a MAT user manually requests an update from the VSAC. If a version or effective date was specified at the time a QDM element was created, the value set data in the MAT will be retained for that QDM element, and no update from VSAC is applied.

When the user packages an eCQM during an active VSAC session, the value sets associated with the applied data elements are automatically transferred from the VSAC to the MAT, and the value set spreadsheet in the MAT output package will contain the latest codes from the VSAC unless a version or effective date of a value set was specified at the time of QDM creation.

2.6—Innovations in eSpecifications

Improved eMeasure development process

The Lean Kaizen procedure has been introduced and applied to the eMeasure development process by CMS and ONC. Many discussions have taken place over the past year to map the current processes, improve the existing processes, and identify and remove inefficiencies in order to streamline eMeasure reviews and approvals.

The lessons learned from the Lean Kaizen discussions and/or events are incorporated into future releases of the Blueprint by the Measures Management System measure developer. In addition, measure developers are encouraged to be “agile” with the processes documented in the Blueprint in order to potentially identify a more efficient process that can be shared with other measure developers through the Kaizen Workgroups, eMIG meetings or MIDS Communication, Coordination and Collaboration (C3) Forum monthly meetings.

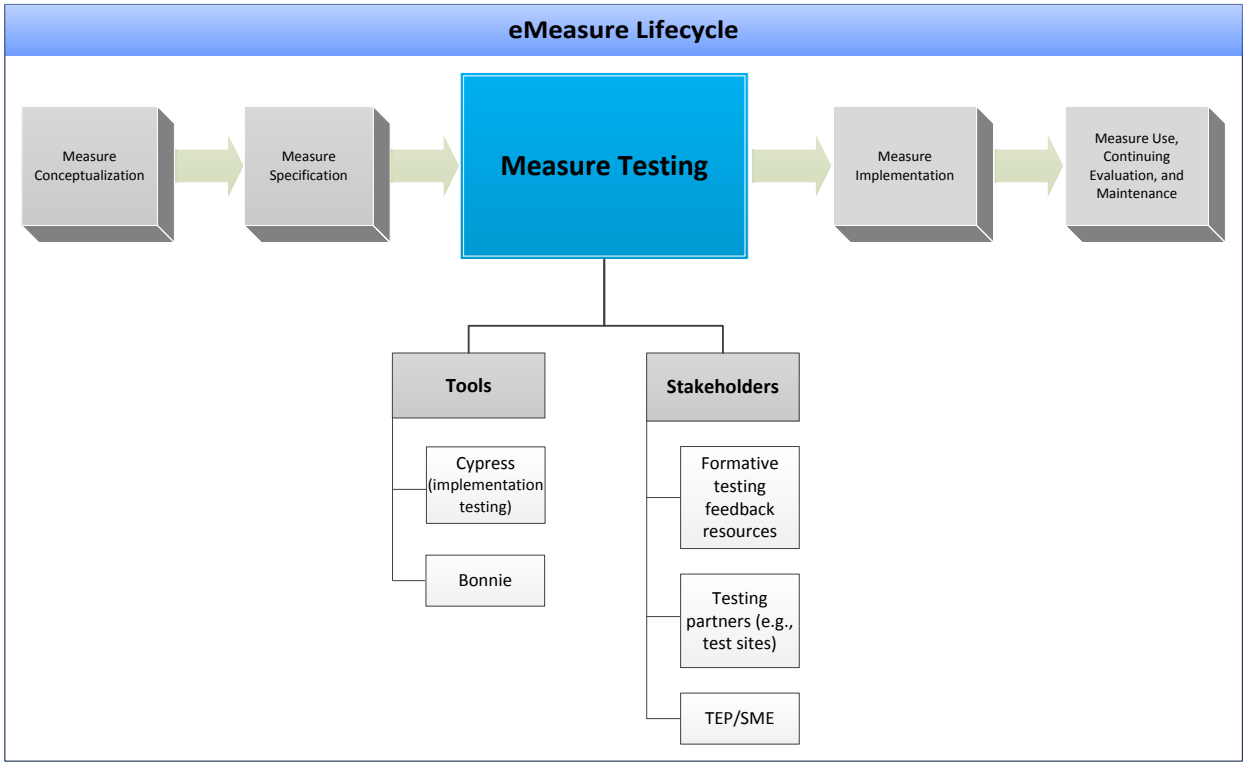
Chapter 3—eMeasure Testing

When evaluating an eMeasure's readiness for implementation and adoption, eMeasure testing assesses the extent to which an eMeasure meets the measure properties of feasibility, validity, and reliability. Testing measure properties is an iterative process with the purpose of refining and revising the eMeasure until all quality issues are resolved. The goal is to produce a reliable, valid eMeasure ready for implementation. eMeasure testing is possible once the eMeasure specification is completed in the MAT, and the eMeasure 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 electronic health record 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.

This chapter also includes recommendations for content to include in a testing plan and Testing Summary Report. The recommendations are intended to serve as a guide and are not prescriptive. Other testing methods and approaches may be appropriate for the measure being developed. Measure developers are encouraged to always select testing that is suitable for the measure. Properly conducting eMeasure testing and analysis is critical to approval of an eMeasure by CMS and endorsement by NQF. Figure 34: eMeasure Testing Tools and Stakeholders depicts the tools and key stakeholders needed for measure testing.

Figure 34: eMeasure Testing Tools and Stakeholders



3.1—Deliverables

- Measure Testing Plan
- Measure Testing Summary Report
- Updated eMeasure Specifications
- Updated Measure Justification Form
- Updated Measure Evaluation Report

3.2—Types of eMeasure Testing

As EHR systems become more widely available and more integrated, additional clinically documented 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 being retooled or reengineered for EHRs) to be specified using HQMF, which is a standard for representing a health quality measure (or clinical quality measure) as an electronic document.

In alignment with this format, measure developers are expected to author eMeasures 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, they also raise new considerations when testing measures that include EHR specification accuracy, EHR validity testing, measure score and element testing, testing of retooled 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.

Feasibility

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 eMeasure Feasibility Assessment Report for more information and guidance.[241]

Basic data element feasibility

Prior to drafting initial eMeasure specifications, the measure developer should consider the data elements necessary for the proposed measure and conduct preliminary feasibility assessments (alpha testing) to confirm availability of the information within a structured format within the EHR. Doing so ensures that a developed measure passes feasibility assessments during beta (field) testing. The test method is a survey of EHR vendors regarding the data types that can be captured in their individual EHR products in a structured format. The purpose of data element feasibility testing is to evaluate the reasonableness of collecting the expected data elements. The results of this testing provide information on the impact the data capture has on the typical clinical workflows. This supports the four areas specific to evaluating the feasibility of eMeasure data elements outlined in the NQF Data Element Feasibility Scorecard (Table 6 in the *Measure Evaluation Criteria and Guidance Summary Tables*).[242]

- **Data availability**—the extent to which the data are readily available in a structured format across EHR systems.
- **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.
- **Data standards**—the extent to which the data element is coded using a nationally accepted terminology/vocabulary standard. Standard data elements, associated definitions and code sets, and mapping to the QDM are expected. Refer to the eMeasure Specifications chapter for recommended vocabularies as they relate to the clinical concepts of the QDM.
- **Workflow**—the extent to which the effort of capturing the data element interferes with providers' typical workflow. For example, if capturing the information requires navigating multiple screens during a clinical encounter, the provider may skip entering the data.

[241] National Quality Forum. *Report from the National Quality Forum: eMeasure Feasibility Assessment TECHNICAL REPORT*. April, 2013. Available at: http://www.qualityforum.org/Publications/2013/04/eMeasure_Feasibility_Assessment.aspx. Accessed on: June 1, 2015.

[242] National Quality Forum. *Measure Evaluation Criteria and Guidance Summary Tables.* Effective July 2013, updated October 11, 2013. Available at: http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx. Accessed on: June 1, 2015.

3.2.1.2—Detailed data element feasibility

Additionally, measure developers should consider identifying any barriers to implementation related to technical constraints of EHRs and whether the data captured in the EHR are captured in a form that is semantically aligned with the expectations of the quality measure. Carefully consider the time and costs related to additional data entry by clinicians when replacing previous measures that rely on manual chart review.

A standardized score card or survey should be used as a screening tool to identify feasibility concerns during development, which assesses capture of data elements in current and future EHRs (for those scoring "low" on current feasibility). The following are the required characteristics for inclusion in a scorecard. Alternatively, developers may use the NQF data element feasibility scorecard example provided in the formal report:

- One scorecard per data element that assesses data availability, data accuracy, data standards, and workflow on a single EHR system.
- For each data element, the respective scorecard should address current and future capabilities (3–5 year capability) are rated on a scale of 1 through 3—"1" indicates a low score, while "3" indicates highly feasible.
- The scorecards use quantitative methods, yet permits consultation with the COR regarding comment documentation.[243]
- A survey assessment by quality measurement and/or IT experts about the detailed format and level of automatically extraction of eMeasure data element as captured in an EHR systems is a method used to test detailed data element feasibility.

When recruiting sites and vendors to test feasibility of a measure under development, measure developers should:

- Assess multiple EHR vendor systems.
- Use appropriate settings to test measure goals (hospitals, ambulatory, etc.).
- Consult with their COR regarding the PRA, which requires OMB approval before requesting most types of information from the public.

Validity

After the measure is electronically specified in the MAT, testing validates parts of the eMeasure package that is exported from the MAT such as the vocabulary file and the human rendering of the measure. Testing the different files from the 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 electronically specified measure confirms the intent of the measure, ensures the eMeasure logic is not ambiguous and expected test patients fall in the correct populations, data

[243] 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.

elements are aligned with national standards, and checks calculated scores from automated extraction for accuracy. Each testing approach is described below.

Ideally, certified EHRs 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 frequently into EHR fields other than those from which the vendor extracts data for measure reporting.
- Even small errors in the measure specifications, such as code lists or exclusion logic, may decrease measure validity. For example, omission of value set codes for commonly documented concepts can reduce the capture of appropriate patients in the measure's denominator.

Given the evolving potential for standardized data extraction of complex clinical information, different options should be considered during measure testing. These include:

- Comparison to abstracted records—eMeasure developers should consider comparing measure scores and measure elements across multiple test sites where EHR data practices may vary. This process generally requires extracting data from certified EHRs and comparing this information to manual review of the entire EHR record for the same patient sample. Sampling from diverse EHRs and comparison groups is critical to assess the measure's susceptibility to differences in data entry practices or EHR applications. Standard assessment of the agreement between the two methods using a reference strategy/criterion validity approach is often sufficient to demonstrate comparability.
- Comparison to simulated QDM data set—Because efforts are ongoing to ensure that the multitude of EHRs in use today are capable of exporting information suitable for a measure specified using HQMF and QDM standards, measure developers may have difficulty obtaining a sample that is adequately representative of the different practice patterns and certified EHR systems that will be in use when implementing a measure. To address this issue, validity testing can be augmented using a simulated data set (i.e., a test bed) that reflects standards for EHRs and includes sample patient data representing the elements used by the measure specifications. Provided the data set reflects likely patient scenarios and is constructed using QDM elements, the output can be used to evaluate the eMeasure logic and coding specifications. This approach is sometimes referred to as semantic validation, whereby the formal criteria in an eMeasure are compared to a manual computation of the measure from the same test database.

Measures originally specified using data sources other than EHR (i.e., chart abstraction or administrative claims data) can be modified or retooled/reengineered for use with EHRs. However, even if these measures were previously approved by CMS and show adequate reliability and validity, the eMeasure should be assessed for the following:

- Similarity to the originally approved specifications.
- Appropriately used QDM data types which represent the original measure specifications.
- Appropriately chosen codes and taxonomies.[244]

Consequently, at a minimum, a crosswalk of the eMeasure specifications should demonstrate that they represent the original measure. NQF indicates that this evidence warrants a "moderate" rating for the reliability and validity of a retooled measure. NQF guidance reserves a "high" rating for a measure when testing also demonstrates that the respecified measure produces similar results (within tolerable error limits) when compared to the previously approved measure that uses data sources other than EHR. Ideally, this may be achieved by obtaining representative samples of patients across EHRs and providers that allow calculation of the measure directly from EHRs, as well as calculation of the same patients using previously implemented methods (e.g., administrative claims or chart abstraction).

The crosswalk allows comparison across the data sources to determine if the EHR implementation produces similar (or superior) findings relative to data sources and specifications already in use. Statistics indicating agreement between the methods and data sources often provide a succinct summary of the similarity of the retooled eMeasure relative to prior implementation of the measure that does not use EHRs. Further investigation of patients where the data sources or methods do not match may also provide evidence for the adequacy of the retooled measure when a review of the patient record allows a definitive judgment regarding the appropriate disposition of the patient with respect to the measure.

A subjective evaluation of the human-readable rendition of the eMeasure 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 retooled or re-engineered measure that the eMeasure preserves the intent of the original paper-based measure equivalent "at face value." A subjective evaluation for a *de novo* measure includes confirmation by a clinical workgroup or TEP that eMeasure 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 eMeasure.

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 eMeasure. The test aims to ensure that the logic of the eMeasure 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 eMeasure.

Bonnie (described later in this chapter in Tools for Testing eMeasures) is a tool that measure developers can also use to evaluate eMeasure logic. Bonnie is able to consume the e-specified

[244] Ibid.

measure in HQMF Release 1, and measure developers can readily create test cases and confirm whether patients properly fall into the expected populations. Note that Bonnie is also programmed to consume the export of the July 2014 MAT update by using the interim simple XML output from the MAT export.

In conjunction with Bonnie testing, manual logic testing is performed to incorporate clinically relevant test cases. For example, clinical tests may include scenarios where the logic is "missing" data and/or a test case where an inappropriate code or taxonomy is chosen. The manual method of logic testing involves the comparison of answer keys from two analysts. Analyst One produces an answer key that indicates, for each record, whether it should be included in or excluded from the numerator, denominator, and other relevant measure populations. Analyst Two then performs a similar analysis of the testing patient records created by Analyst One but based on the authoritative measure specifications supplied by the clinical workgroup or measure steward. Analyst Two then independently produces another answer key based on the results of the second analysis. Results achieved by each analyst are then compared, to identify inadvertent error or logic difference between two measure specifications. This comparison identifies inadvertent error or logic difference between the authoritative measure logic of the paper-based measure and the specified measure.

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 retooled, re-engineered 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 eMeasure 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 eMeasure 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 entire EHR; or

- Complete agreement is observed between the known values from a simulated QDM-compliant data set and the elements obtained when the eMeasure specifications are applied to the data set. NQF guidance further clarifies that reliability testing of measure elements may be supplanted by evidence of measure element validity.

Measure score validity

An evaluation of the measure score of an eMeasure calculated through automatic extraction is compared to a measure score calculated using abstracted data. Using the EHR data captured through the electronic extraction and visual abstraction processes, electronically extracted and visually abstracted measure outcomes will be derived for the eMeasure. A comparative analysis is conducted to assess and examine whether the electronically extracted eMeasure outcome differs significantly from the visually abstracted outcome using the same eMeasure specifications and the same patient records. A sensitivity analysis is performed to determine the impact of missing or "incorrect" data on resulting measure scores.

Reliability

Testing for reliability involves experts assessing the human-readable format of the eMeasure for clarity and conformance 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 eMeasure is to evaluate the measure for clarity, logic ambiguity, and data element conformance to standard specifications that support consistent implementations.

Measure level ("face") reliability

Quality measurement and/or IT experts should assess whether the eMeasure, in its human-readable format, is clear and understandable for the purpose of writing queries, or mapping such queries to IT systems, in order to accurately extract the appropriate data from the EHRs.

Measure logic reliability

Like measure logic validity testing for a retooled or re-engineered eMeasure, a *de novo* measure is tested for measure logic reliability. *De novo* measures are tested for "reliability" since there is no authoritative source that can be used to evaluate the logic. The objective of the evaluation and method of testing are the same. Instead, in a *de novo* measure, one of the answer keys is created based on the evidence-based literature supplied by the clinical workgroup. Results by each analyst will then be compared and analyzed to identify inadvertent error or logic difference between the measure specification and logic intent based on the literature and clinical workgroup feedback.

Data element reliability

An objective evaluation of the eMeasure (HQMF) XML file and vocabulary file from the MAT should be performed to check conformance of individual data elements to the standardized formats established by the HQMF, QDM, XML, and clinical vocabulary standards. This test may

detect measure vocabulary inconsistencies (e.g., if a given clinical concept has multiple values sets with different code lists) or if a clinical concept has multiple value sets with the same code list, but with different value set names and identifiers. The measure developer should also identify conflicts that may exist among the various quality measurement and quality report standards and specifications defined by HL7, ONC, and the CMS Blueprint (e.g., different vocabulary requirements for a given data element under multiple standards). The eMeasure specifications should be validated against the Schematron and XML Schema encoded definitions to ensure technical compliance with the HQMF standard.

To help ensure the accuracy of these data elements, measure developers are expected to validate the content of the XML. This is often achieved using the following two methods:

- Syntactic validation—This method of accuracy validation ensures that the XML content follows (i.e., conforms to) specific constraints required by the HL7 HQMF Draft Standard for Trial Use (DSTU) and the XML patterns based on the QDM. These quality-checking processes are built into the MAT application (refer to the eMeasure Specifications chapter). Alternatively, other methods to confirm XML conformance can be used (i.e., HL7 ISO Schematron).
- The HL7 ISO Schematron is a possible mechanism for validating XML that is written outside the MAT; however, it may not include all of the components that are now built into the MAT. Additional resources for information—including technical specifications—on the HL7 ISO Schematron may be found at the ISO website.[245]
- Narrative validation—A fully constructed eMeasure is an XML document that can be viewed in a standard Web browser, when associated with an eMeasure rendering style sheet supplied by CMS. When rendered in a Web browser, the eMeasure is in a human-readable format which allows the measure author to assess the extent to which the machine-generated criteria correctly reflect the original measure criteria under development. When the measure author validates correctness of the human-readable format, this is considered narrative validation.

Testing Multiple Sites

Testing multiple sites for feasibility, validity and reliability is important to address potential variability in reporting based on differences in local workflow process. Each site should be provided with the same test cases to determine if each site's result is comparable with the others. 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.

[245] http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html. Accessed on: June 1, 2015.

3.3—Phases of eMeasure Testing

Testing of these measure properties is done in parallel with development during the two phases of testing: alpha and beta. Basic data element feasibility testing, measure level validity, measure logic validity, measure score validity, and data element reliability testing occur during the alpha phase. Conversely, detailed data element feasibility, measure level reliability, and data element validity occur during the beta phase.

Alpha

Alpha testing is an early, iterative form of internal testing that occurs prior to drafting the initial eMeasure specifications. The measure steward and measure developer should use results of alpha testing to determine and refine measure concepts for use in an eMeasure before completing the initial specifications and subsequent field testing. Alpha testing evaluates the feasibility and availability of data in a structured format within the EHR. Testing also focuses on identifying early any logic ambiguities and vocabulary inconsistencies in the electronically specified measure. Formative testing of the measure properties starts before the measure is entered into the MAT. Once the measure is specified in the MAT tools such as Bonnie assist with measure logic testing. (See the Tools for Testing eMeasures section below.)

Additionally, measure developers should consider identifying any barriers to implementation related to technical constraints of EHRs (e.g., lack of data provenance to support the data source required by the measure). Measure developers should also determine whether the data are captured in the EHR in a form that is semantically aligned with the expectations of the quality measure (e.g., radiology/imaging reports in unstructured rather than structured form, or laboratory results reported with standard units of measure). Note that when replacing previous measures that rely on manual chart review, the time and costs related to additional data entry by clinicians needs to be carefully considered.

Beta

Once the eMeasure is developed, it is tested in the “field” or beta tested. Beta testing involves extraction and abstraction of patient records from EHRs of participating providers. The number of beta testers for a measure may vary based on the nature of the testing required. Beta testing provides evidence of measure reliability across providers and uncovers variability in provider performance or relationship of the measure results to patient outcomes. In addition, data element validity testing and measure score validity testing occurs during this phase (as discussed above). Beta feasibility testing evaluates the ability to capture and report data from EHR systems.

3.4—Tools for Testing eMeasures

Bonnie Tool

Bonnie was released on April 3, 2014, and designed for testing eCQM.[246] With Bonnie, measure developers are able to 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 on whether or not the logic behaved as expected. This tool allows the user to readily identify criteria each test case (patient) must satisfy to qualify for a specific population (e.g., denominator, numerator, etc.). 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 testing patients and identifies errors missed with manual review. It is especially helpful identifying recursive logic. Refer to the Bonnie user guide for more information.

Cypress Tool

Cypress is an open source testing tool available to EHRs and EHR modules for calculating eMeasures. Cypress is the official testing tool for the 2014 EHR Certification program and is supported by ONC. Testing involves importing an eMeasure into Cypress and applying QRDA Category I test patient records in order to test the measure logic and process calculations against the eMeasure. Answer keys produced by the Cypress tool are then compared with manually generated answer keys.[247] Also, the Cypress tool assesses an eMeasure's logic, expression, and relevant value sets to identify potential gaps and defects. Like Bonnie, this tool supports efficiency in measure development when integrated early into the process to identify additional measure logic discrepancies and calculation errors that potentially could occur during adoption and implementation.

3.5—Procedure

Test Plan

Similar to the test plan that is described in Section 2, Chapter 3—Measure Testing, a test plan is created for eMeasure testing. The intent of the test plan is to describe the objectives and methods for testing a single or set of measures. The test plan is created early in the process and outlines the scope for both alpha and beta testing. Sufficient information is provided to the COR to ensure the measure is tested for validity, reliability, and feasibility.

The test plan should contain the following:

- Scope of testing which includes the name(s) of measures(s) and identifies who is conducting each part of the testing if collaboration with another organization is involved.

[246] https://bonnie.healthit.gov/users/sign_in

[247] Office of the National Coordinator for Health Information Technology. Cypress: Meaningful Use Stage 2 Testing and Certification Tool. Available at: http://projectcypress.org/about.html. Accessed on: June 1, 2015.

- Statistical testing plan for risk adjustment models as applicable.
- Statement of the objective(s)/research question.
- Testing methods and approach. Descriptions of the testing methods and approach should address feasibility, validity, and reliability properties of the measure.
- Description of the phases of testing: alpha and beta.
- Inclusion of any institutional review board (e.g., IRB) compliance process followed as applicable.
- Plan for analyzing data including a description of test statistics that support assessments.
- Timeline or schedule for the testing and report completion.

In addition to the elements above, the testing plan should include a description of the field data collection methods and procedures. When describing the test population, include in the test plan:

- Recruitment strategy—Describes the number and type of participants recruited and what part of the population they represent. Explain the process for recruitment (e.g., solicitation emails, correspondence letters).
- Data collection methods—Describes the process for data elements collection (e.g., electronically abstracted or visually abstracted) and who would perform this task.

Testing Summary Report

The Testing Summary Report is a description of the test(s) performed on an eMeasure. This testing report is a summary of the types of testing completed or in progress and is produced annually for the CMS COR. At minimum, the Testing Summary Report should include:

- List of eMeasure(s) tested.
- Types of testing performed.
- Methods/approaches employed, including tools.
- Validity and reliability testing findings and analysis.
- The number of test sites reporting each measure data element and the number of sites indicating the feasibility of each measure (a) overall, (b) with workflow changes, or (c) not feasible to implement at this time, and relevant comments.
- The coding systems used by each site for the specific data element (SNOMED, LOINC, etc.).
- The number of test sites using a specific data capture type for each data element (discrete/non-discrete, numeric, Boolean, etc.).
- Percentage of feasible elements for each measure per test site (reported as a range, the high and low sites).
- Percentage of test sites reporting that the measure as respecified retains the originally stated intention of the measure.
- Percentage of test sites reporting an acceptable face validity rating. In other words, the extent to which the measure appears to capture the single aspect of care or healthcare quality as intended, and the measure as specified is able to differentiate quality performance across providers.
- Outcome of testing.

- Analysis and conclusion.
- Recommendations.

In addition to the items listed above for inclusion in the measure Testing Summary Report, the feasibility component to the report should also contain the reference point or threshold against which the feasibility is being assessed. Measure feasibility should not be limited to the individual and aggregate data elements but should also address the feasibility of the overall specifications and within the context of the calculation logic.

When reporting the feasibility results to NQF during the measure submission process as described in the subsection on National Quality Forum Endorsement (Section 2, Chapter 4—Measure Implementation), the feasibility assessment information must, at a minimum, include:

- A description of the feasibility assessment.
- Data element score card results. This score card contains the resulting scores for all the data elements with respect to data availability, data accuracy, data standards, and workflow.
- Notes explaining all data elements with "low" scores, including the rationale and plan for addressing the concerns.

3.6—Testing Challenges

While testing an eMeasure, inherent challenges exist both in the alpha and beta phases. In the alpha phase, the iterative communication process to resolve issues occurs mainly between the measure developer, quality analyst, and measure steward. Keeping track of the issues requires a plan and close communication to ensure that the measure is ready for the beta phase. This may involve calls to discuss issues which can lengthen the time of this testing phase. In addition, the creation of sample test cases is a manual process and is time consuming. However, the development of innovative testing tools such as Bonnie has improved the process.

In the beta testing phase, a different set of obstacles exist. Only a small number of test sites may be available to validate data for some measures. When testing at a site, limited coded data may be available and what does exist may not use the standard vocabulary recommended by the ONC HIT Standards Committee. If the data capture criteria do not meet the eCQM criteria, the certified EHR technology may still fall short of capturing the needed data for the eCQM.

Chapter 4—eMeasure Implementation

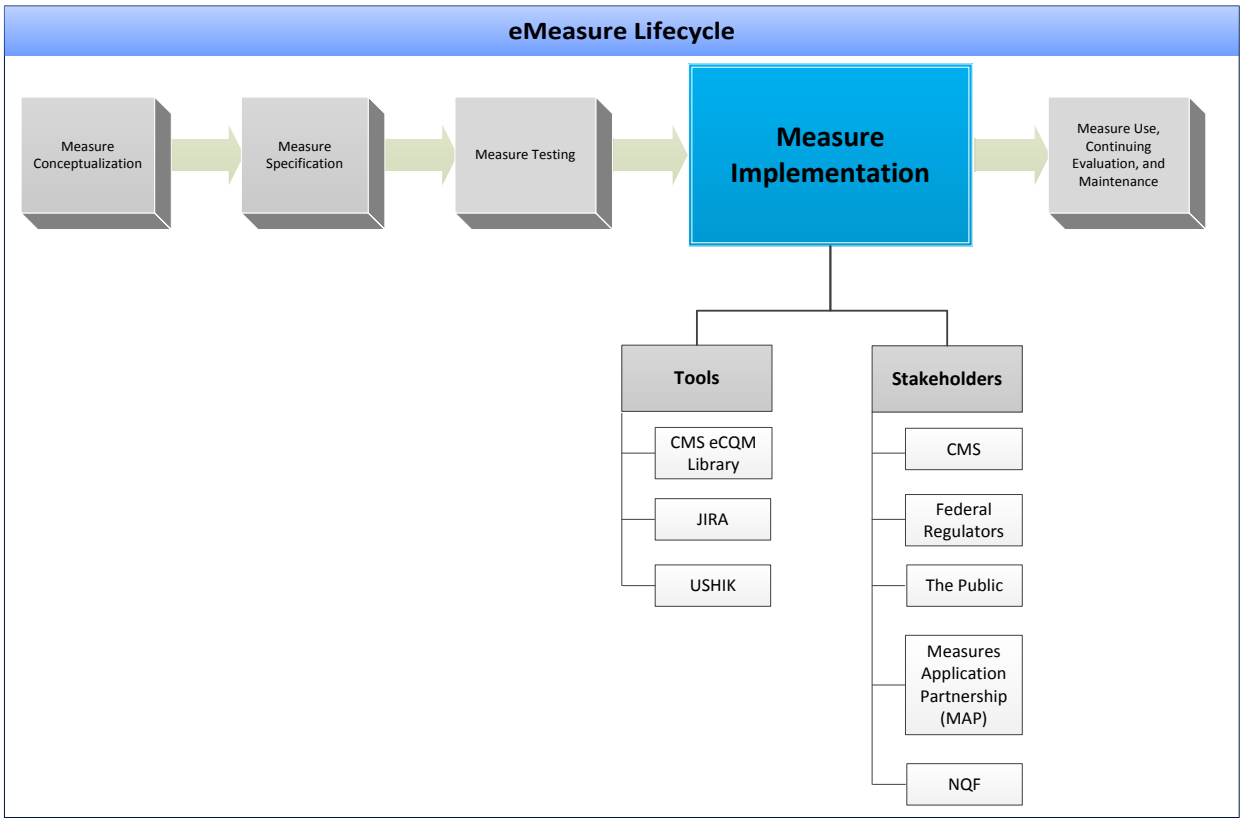
During this step, the measure developer prepares the measure to go through the federal rulemaking process and public comment period. The measure developer works with CMS to make sure that the public stakeholders have time to review and comment on the measure, and then revise the measure based on necessary feedback. Once the measure is final, it is submitted to NQF via an online measure submission available on the NQF website.[248] The measure developer supports the measure as it goes through the endorsement process. After the measure has been endorsed by NQF, the measure developer supports the implementation of the measure by developing a:

- Roll-out plan.
- Business process to support measure reporting.
- Data management plan.
- Auditing and appeals processes.
- Education and outreach efforts.

Figure 35: eMeasure Implementation Tools and Stakeholders depicts the tools and key stakeholders needed for measure implementation.

[248]National Quality Forum, Measuring Performance: Submitting Standards. http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx

Figure 35: eMeasure Implementation Tools and Stakeholders



4.1—Deliverables

- Project plan
- Summary report
- NQF endorsement submission documents
- Updated eMeasure specifications
- Updated Measure Information Form (if necessary) and Measure Justification Form[249]
- Provide justification for using value sets not in the VSAC and vetted by NLM (if needed)

Development of the complete electronic specifications is an iterative process including input from the measure steward (for retooled measures), information gathering, TEP, public comments, and measure testing conducted and results. (Refer to the eMeasure Testing chapter for guidance on types of testing methods during measure development.) Information from measure testing, the public comment period, or other stakeholder input may result in the need to make changes to the technical specifications prior to NQF submission. The measure developer will communicate and collaborate with the measure steward and/or TEP to incorporate changes before submitting the final measure to the COR for approval. A complete electronically specified measure submitted to NQF will include:

[249] 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: June 1, 2015.

- An updated eMeasure specification from the MAT Tool
- Measure Information Form (if necessary)
- Measure Justification Form (for *de novo* measures only)
- Any additional documents required to produce the measure as it is intended (such as risk adjustment methodology, business case, etc.)

4.2—Preparing eMeasure Specifications for Public Release and Public Comment

At the discretion of the COR, when eMeasure testing is complete, the measure developer may also obtain public comments on the draft eSpecification which is generated from the MAT. The ONC JIRA system process and the federal rulemaking process are two ways to gather public feedback on a draft eMeasure. The Measure Rollout information in Chapter 4—Measure Implementation of Section 2 provides more details on this phase. Because the public comment process reaches a broader group of people, it can provide information to improve measure feasibility, validity, and reliability by correcting the technical specifications. Also, the ONC JIRA system allows for measure developers and community users to submit feedback on eMeasures. Below are the steps to set up a JIRA account in order for measure developers and community users to submit feedback for measures that have been posted for public comment.

To set up a JIRA account, go to the JIRA website. Select “sign up for a new account” under the *Login* area of the Homepage to set up a new account and log into the JIRA tool. When prompted, fill in user-specific information and choose a password. Finish by clicking on the Sign up button to be registered. After logging into the JIRA tool, follow these steps to post a comment or question.

1. Select “Projects” at the top middle of the home screen.
2. Select “Comments on eCQMs under development” project.
3. Select “Create issue” (orange button) at the top/middle of the screen to enter comments.
4. Select the type of issue from the “Issue Type” dropdown menu.
5. Fill out the following fields:
    - Summary
    - Contact Name
    - Contact Email
    - Contact Phone
6. Enter any comments in the “Description” field.
7. Select the measure name related to the comments from the “Draft measures” dropdown box.
8. Select “Create” at the bottom left to submit the comments. To enter more comments, select “Create another” and then select “Create.”

The JIRA user guide provides more information on using JIRA. Further discussion of the role of public comment is found in Chapter 3—Stakeholder Input of Section 1. A description of the general process for obtaining public comment is found in Section 2, Chapter 1.C.—Public Comment.

4.3—NQF eMeasure Endorsement

Any eMeasure intended to be submitted for NQF endorsement must be submitted in HQMF. This process is supported when measure developers author their eMeasure in the MAT. Measure developers should consult with the NQF website for any current policy decisions related to the endorsement of eMeasures.

NQF endorses measures only as a part of a larger project to seek standards (measures) for a given healthcare condition or topic. If CMS decides that the measures developed under a project are to be submitted to the NQF for endorsement, and NQF is conducting a project for which the measure is applicable, the measure developer will support CMS in the submission process. Measures are submitted to NQF using the web-based electronic submission form. Upon the direction of the COR, the measure developer will initiate and complete the submission form. NQF makes periodic updates to the measure submission process. Consult the NQF website for the current process.[250]

If the measure is submitted to NQF for endorsement, the measure developer is required to provide technical support throughout the review process. This may include presenting the measure to the Steering Committee that is evaluating the measure and answering questions from NQF staff or the Steering Committee about the specifications, testing, or evidence.

During the course of the review, NQF may recommend revisions to the measure. If changes are recommended by NQF, all recommendations must be reported to and approved by the COR. The measure developer will then update the eMeasure in the MAT with any changes agreed upon during the endorsement process.

Any measure developed under contract with CMS should identify CMS as the measure steward unless special arrangements have been made in advance. Measure developers should consult with the COR if unclear or there are questions. Barring special arrangements, CMS should be identified on the NQF submission form as “Centers for Medicare & Medicaid Services (CMS).”

The MAT output and Measure Justification Form are designed to help measure developers complete an eMeasure for submission to NQF for endorsement. The Measure Justification Form, which is aligned with the NQF measure submission form, was designed to guide the measure developer throughout the measure development process in gathering the information in a standardized manner. The form also provides a crosswalk to the fields in the NQF measure submission to facilitate online information entry, should CMS decide to submit the measure for consideration.

Refer to the National Quality Forum Endorsement information in the Measure Implementation chapter of Section 2 for further discussion of this process.

[250] National Quality Forum. *Measuring Performance: Submitting Standards.* Available at: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx. Accessed on: June 1, 2015.

4.4—eMeasure Rollout, Implementation, and Publication

Publication and Packaging

Naming Conventions

CMS measure developers must follow the conventions detailed below when submitting final measure specifications for public release on the CMS website.

CMS eMeasure Identifier and eCQM Naming Convention

CMS created a unique "CMS eMeasure Identifier" to clearly and consistently identify eCQM files. The naming convention combines the eMeasure identifier assigned to the eCQM in the MAT with the "eMeasure Version Number", which is prepended by "CMS". The eMeasure Version Number is a numeric value used to indicate the published version of the eMeasure. Based on this universal naming convention, Eligible Professional measure (NQF0056-Diabetes: Foot Exam) would display the following for the first version of the measure: **CMS123v1**.

Table 23: Sample eMeasure Identifier

eMeasure Information	Value
eMeasure Identifier (from MAT)	123
eMeasure Version Number	1
CMS eCQM identifier	CMS123v1
Minor MAT version and Revision number*	x.1.xxx

*In 2015, the Measure Authoring Tool (MAT) released new functionality that allows minor versioning. This minor version will be visible in the human-readable HQMF under "eMeasure Version number", but only a single measure version will be released in the 2015 measure packages and that version will be locked for the 2016 measurement period. The measure versions will continue to be referred to by CMS as an integer that represents the major version number.

Individual eCQM Measure Package Components

The file type (.xml or .html) is added to the CMS eMeasure ID to complete the naming convention for the components of the eCQM package. Examples below:

Type of Artifact File Name

HQMF (XML file) **CMS123v1.xml** – this is the machine-readable

HQMF (HTML file) **CMS123v1.html** – this is the human-readable

If downloaded from a site offering UMLS authentication, you may also find:

Value Sets (Excel file) CMS123v1.xls – this contains the codes and value sets in the measure

Individual Measure Zip File and Folder Names

The naming conventions for the individual eMeasure packages (zip files and measure folder) that will contain the eMeasure XML file and human-readable rendition is described below in the order in which they must appear.

1. Setting for which the measure applies—"Eligible Professional" or "Eligible Hospital" measures. Use 2-letter abbreviation, EP for eligible professional or EH for eligible hospital.
2. CMS eMeasure ID.
3. NQF identifier—if not endorsed by NQF, the file will contain "NQFXXXX."
4. Abbreviated name for the clinical quality measure (example: "Colorectal_Cancer_Screen").

Zip file name structure: <EP|EH>_<CMSeMeasureID>_<NQFID>_<shortDescription>

Example: EP_CMS130v1_NQF0034_Colorectal_Cancer_Screen

Note: The NQF ID reads NQF NOT APPLICABLE in the HQMF when the measure is not endorsed by the NQF. All measures have been recommended by CMS but not all have been/are endorsed by the NQF.

We recommend use of the CMS eCQM ID to identify measures and their versions.

Measure Set Package Naming

The file names combine attributes that identify the:

1. Setting for which the measure applies—"EH_Hospital" or "EP_Professional"

2. Publication date—format: YYYY_MM_DD

Measure set package name structure: <EP|EH>_<Hospital|Professional>_eMeasures_<date>

Example: EH_Hospital_eMeasures_2012_10_17

(assuming the release date for these EH measures is 10/17/2012)

Versioning

Once CMS has released an eMeasure for public use, maintaining the measure version is critical. If any changes are made to the measure's specifications (e.g., major, minor, technical corrections, annual update), the version number will be advanced based on the versioning requirements set by the program and the COR prior to posting.

Measure Packaging for Distribution (by Setting)

The naming convention for the "all measures Zip files" (containing all of the individual eMeasure files grouped by setting) is described below; the components are listed in the order in which they must appear. The file names combine attributes that identify these items:

- CMS Rule year with which the eMeasures are associated (example: 2014).
- Files included—"eCQM" for Eligible Professional, "eCQM_Spec_for" for Eligible Hospital.
- Setting for which the measure applies—"EH" and "EP."
- Publication date—format: "MonthYYYY" for Eligible Professional, "Release _MonthYYYY" for Eligible Hospital.

Table 24: Sample File Naming provides examples of Zip file names using the convention for measure packaging by setting:

Table 24: Sample File Naming

File Contents/Setting	File Name
Eligible Hospital Zip file	2014_eCQM_Spec_for_EH_Release_April2013.zip
Eligible Professional Zip file	2014_eCQM_EP_June2013.zip

Rollout Process

Posting to eCQM Library

After the measure developer completes an eMeasure, the final eMeasure specifications are submitted to CMS for public release. CMS posts the following information on its online eCQM Library. [251]

- eCQM Specifications
- Technical Release Notes
- eCQM Logic and Implementation Guidance
- Guide for Reading eCQMs
- Resource table for Eligible Professional and Eligible Hospital measures

eCQM Specifications

The eCQM specifications are posted by release date and per setting (Eligible Hospital and Eligible Professional) and contain the individual zip files for the eMeasure (e.g., 2014 eCQM Specifications for Eligible Hospitals Release December 2012).

[251] Department of Health and Human Services, Centers for Medicare & Medicaid Services. *Electronic Clinical Quality Measures (eCQMs) Library*. Available at: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/eCQM_Library.html. Accessed on: June 1, 2015.

Technical Release Notes

This document lists the specific logic, header, and value set changes made to each eMeasure in each release.

eCQM Logic and Implementation Guidance

This document provides guidance for those implementing the eCQM electronic specifications. The content covers general implementation, defines specific logic and data element conceptualization, and provides details on versioning and time interval calculations. This guide also provides specific guidance for Eligible Hospital and Eligible Provider measures.

Guide for Reading eCQMs

CMS publishes a document titled *Guide for Reading eCQMs* to provide guidance in understanding and interpreting the 2014 eCQMs. It contains useful information for understanding the human-readable rendition (HTML) file of an eMeasure. A new version of the document is published when CMS releases updates to the Eligible Professional and Eligible Hospital measure specifications. The latest version of the document can be downloaded from the CMS eCQM Library website.[252] This guide should be used by developers to interpret and understand eMeasures.

Resource Tables

The resource tables list the measures with their CMS eMeasure ID, NQF number, version number, measure title, measure description, numerator statement, denominator statement, measure steward, PQRS number, and National Quality Strategy Domain.

Implementation

Posting to US Health Information Knowledgebase (USHIK)[253]

Approved eMeasures and their value sets may be downloaded, extracted, and accessed in the USHIK Meaningful Use Portal. This portal allows files to be downloaded in various formats: XML, Adobe (PDF), and comma separated value (CSV), including a single MS Excel (.xls) file that contains all the meaningful use quality measures and their value sets. Note that access to the value sets requires a free UMLS Metathesaurus License.

Innovations

[252] Ibid.

[253] Unites States Health Information Knowledgebase . Agency for Healthcare Research and Quality (AHRQ). http://ushik.org/mdr/portals. Accessed on: June 1, 2015.

NQF Approval for Trial Implementation and Testing Pilot

The following material is quoted from the April 8, 2014, NQF CSAC meeting materials.[254]

The purpose of the NQF approval for trial implementation and testing pilot is two-fold:

- Document the process for evaluating eMeasures submitted to NQF for evaluation against the Approval for Trial Implementation criteria instead of evaluation against the full NQF Endorsement criteria.
- Accept eMeasures submitted for Approval for Trial Implementation on a limited capacity before widely offering this option to all eMeasures submitted to NQF for evaluation.

The approval for trial implementation and testing is intended for eMeasures that are ready for implementation, but cannot yet be adequately tested to meet NQF endorsement criteria. For such eMeasures, NQF proposes to use the multi-stakeholder consensus process to evaluate and approve eMeasures for trial use that address important areas for performance measurement and quality improvement, though they may not have the requisite testing needed for NQF endorsement. These eMeasures must be assessed to be technically acceptable for implementation. The goal of approving eMeasures for trial use is to promote implementation and the ability to conduct more robust reliability and validity testing that can take advantage of the clinical data in EHRs.

Approval for trial use is NOT time-limited endorsement as it carries no endorsement label. Also, this is not a required two-stage review process: eMeasures that meet endorsement criteria do not need to first go through an approval for trial use.

eMeasures that are approved by NQF for trial use have been judged to meet criteria that indicate its readiness for implementation in real-world settings in order to generate the data required to assess reliability and validity. Such measures also could be used for internal performance improvement. However, such measures would not have been judged to meet all the criteria indicating it is suitable for use in accountability applications.

- Such measures will be considered Approved as Trial Measures for Implementation and Testing, NOT endorsed.
- When sufficient data have been accumulated for adequate reliability and validity testing, the eMeasure can be submitted to NQF for potential endorsement (not all may progress to endorsement).

NQF Criteria for Approval of Trial Measures for Implementation and Testing:

- The measure must be an eMeasure meaning the measure is specified in HQMF and must use the QDM. Output from the MAT ensures that an eMeasure is specified in HQMF and uses the QDM. However, the MAT is not required to produce HQMF. Alternate forms of "e-

[254] National Quality Forum. Consensus Standards Approval Committee. *Trial Implementation Discussion*. Available at: http://www.qualityforum.org/About_NQF/CSAC/Meetings/2014_CSAC_Meetings.aspx. Accessed on: June 1, 2015.

specifications" other than HQMF are not considered eMeasures. If HQMF or QDM cannot support all aspects of a particular measure construct, those may be specified outside HQMF. Please contact NQF staff to discuss format for measure specifications.

- Must use value sets vetted through the NLM's VSAC. This will help ensure appropriate use of codes and code systems and will help minimize value set harmonization issues in submitted eMeasures. If particular value sets are not vetted by VSAC, explain why they are used in the measure and describe plans to submit them to VSAC for approval.
- Must meet all criteria under Importance to Measure and Report (clinical evidence, performance gap, priority).
- The feasibility assessment must be completed.
- Results from testing with a simulated (or test) data set demonstrate that the QDM and HQMF are used appropriately and that the measure logic performs as expected.
- There is a plan for use and discussion of how the measure will be useful or accountability and improvement.
- Related and competing measures are identified with plan for harmonization or justification why the new measure is best in class.

Refer to the April 8, 2014 CSAC Meeting materials on the NQF website for more information on this topic.[255]

[255] National Quality Forum. Consensus Standards Approval Committee. *Trial Implementation Discussion*. Available at: http://www.qualityforum.org/About_NQF/CSAC/Meetings/2014_CSAC_Meetings.aspx. Accessed on: June 1, 2015.

Chapter 5—eMeasure Reporting

Once eMeasures are specified, tested, and implemented, the EHR systems turn the eMeasures into queries that retrieve the necessary information from the EHR's data repositories and generate quality data reports. As mentioned in earlier chapters, eMeasure reporting (the transmission format) is another important component of the quality reporting end-to-end framework. This chapter describes how individual and aggregate patient quality data can be transmitted to the appropriate agency using Quality Reporting Document Architecture (QRDA) Category I (individual patient level) and Category III (aggregate patient data) reports, respectively. Both QRDA Category I and Category III are DSTU standards for reporting quality measures.

5.1—Deliverables

- QRDA Category I XML File for individual patient report
- QRDA Category III XML File for aggregate report

5.2—Quality Reporting Document Architecture (QRDA) Category I

Individual patient level quality reports are transmitted using QRDA Category I. Each QRDA Category I report contains quality data for one patient for one or more quality measures, where the data elements in the report are defined by the particular measure(s) being reported on. The QRDA Category I, Release 2, is a DSTU standard that was published in July 2012. An errata package to this standard was released in December 2012 to fix errors that were identified since the July 2012 release. Another errata was published in June 2014. QRDA Category I Release 3 was published in June 2015.

The HL7 QRDA Category I is specifically designed with the building block approach. Developed with eMeasures in mind, it is based on the QDM. 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 eMeasure specification to eMeasure reporting. As specified in the Blueprint, measure developers use the MAT to develop eMeasures. The MAT exports eMeasures specified using QDM-based HQMF templates. Therefore, for any (or any set of) MAT-produced eMeasures, implementers should be able to follow the framework and construction rules specified in the QRDA Category I standard to create QRDA Category I reports.

In general, a QRDA Category I report contains a metadata header that has the information about a specific patient, the Measure Section lists the eMeasure(s) that is being reported, and the Reporting Parameters Section provides the information about the reporting period. Patient data elements as defined by the eMeasure(s) are reported through the Patient Data Section. A sample QRDA Category I report is shown in Figure 366.

Figure 36: QRDA Category I Sample Report

QRDA Incidence Report

Patient	Eve Everygirl		
Date of birth	February 1, 2002	**Sex**	Female
Race	White	**Ethnicity**	Not Hispanic or Latino
Contact info	2222 Home Street Burlington, MA 02368, US Tel: (781)555-1212	**Patient IDs**	111223333A 2.16.840.1.113883.4.572
Document Id	5b010313-eff2-432c-9909-6193d8416fac		
Document Created:	December 31, 2011		
Performer			
Author	Ann Quality, RN		
Contact info	1020 Healthcare Drive Burlington, MA 02368, US Tel: (555)555-1003		
Author	Good Health Report Generator		
Contact info	21 North Ave. Burlington, MA 02368, US Tel: (555)555-1003		
Legal authenticator	Virgil Verify, MD of Good Health Hospital signed at December 31, 2011		
Contact info	21 North Ave. Burlington, MA 02368, US Tel: (555)555-1003		
Document maintained by	Good Health Hospital		
Contact info	21 North Ave. Burlington, MA 02368, US Tel: (555)555-1003		

Table of Contents

- Measure Section
- Reporting Parameters
- Patient Data

Measure Section

eMeasure Title	Version neutral identifier	eMeasure Version Number	NQF eMeasure Number	eMeasure Identifier (MAT)	Version specific identifier
Children's Asthma Care (CAC-1) Relievers for Inpatient Asthma	dc78ee5d-1487-4d79-84c3-1dfdaff0781c	1	0143	93	8a4d92b2-373f-82e2-0137-7b9e21cc5c8f

Reporting Parameters

- Reporting period: 01 Jan 2011 - 31 Dec 2011

Patient Data

Data Element	Value	Date/Time
Encounter, Performed: Emergency Department Visit	Emergency Department visit	03/01/2011 4:00 - 03/01/2011 8:30
Encounter, Performed: Encounter Inpatient	Hospital admission	03/01/2011 9:00 - 03/03/2011 10:30
Diagnosis, Active: Asthma	Asthma	01/01/2011
Medication, Administered: Asthma Reliever	Albuterol 1.25 MG (albuterol sulfate 1.5 MG) per 3 ML Inhalant Solution	03/02/2011 9:00
Patient Characteristic Clinical Trial Participant	True	03/01/2011
Patient Characteristic Payer	Medicare	03/01/2011

5.3—Quality Reporting Document Architecture (QRDA) Category III

Whereas a QRDA Category I report carries single patient data, a QRDA Category III report carries aggregate data to summarize a provider's or organization's performance for one or more quality measures. Aggregate quality reports are transmitted based on the HL7 QRDA Category III, Release 1 standard, which was published in November 2012. An errata for QRDA III was published in July 2014.

Like a QRDA Category I report, a QRDA Category III report also contains a Measure Section that lists the eMeasure(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 particular health system over a specific period of time). A sample QRDA Category III report is shown in Figure 37.

Figure 37: QRDA Category III Sample Report

Legal authenticator	Good Health Hospital signed at August 11, 2012
Document maintained by	Good Health Hospital

Table of Contents

- Reporting Parameters
- Measure Section

Reporting Parameters

- Reporting period: 01 January 2012 - 31 March 2012
- First encounter: 05 January 2012
- Last encounter: 24 March 2012

Measure Section

eMeasure Title	Version neutral identifier	eMeasure Version Number	NQF eMeasure Number	eMeasure Identifier (MAT)	Version specific identifier
Anticoagulation Therapy for Atrial Fibrillation/Flutter	03876d69-085b-415c-ae9d-9924171040c2	1	0436	71	8a4d92b2-3887-5df3-0139-013b0c87524a

Member of Measure Set: Clinical Quality Measure Set 2011-2012 - b6ac13e2-beb8-4e4f-94ed-fcc397406cd8

- **Performance Rate:** 83% (Predicted = 62%)
- **Reporting Rate:** 84%
- **Initial Population:** 1000
  - **Male:** 400
  - **Female:** 600
  - **Not Hispanic or Latino:** 350
  - **Hispanic or Latino:** 650
  - **Black:** 300
  - **White:** 350
  - **Asian:** 350
  - **Payer - Medicare:** 250
  - **Payer - Medicaid:** 550
  - **Zipcode 92543:** 15
- **Denominator:** 500
  - **Male:** 200
  - **Female:** 300
  - **Not Hispanic or Latino:** 175
  - **Hispanic or Latino:** 325
  - **Black:** 150
  - **White:** 175
  - **Asian:** 175
  - **Payer - Medicare:** 125
  - **Payer - Medicaid:** 275
  - **Zipcode 92543:** 15
- **Numerator:** 400 (predicted=300)
  - **Male:** 100
  - **Female:** 300
  - **Not Hispanic or Latino:** 140
  - **Hispanic or Latino:** 260
  - **Black:** 120
  - **White:** 140
  - **Asian:** 140
  - **Payer - Medicare:** 100
  - **Payer - Medicaid:** 220
  - **Zipcode 92543:** 6
- **Denominator Exclusions:** 20
  - **Male:** 8

5.4—Procedure

CMS QRDA Category I and Category III Implementation Guide

CMS developed and published the CMS QRDA Category I and Category III Implementation Guides for Eligible Professionals and Eligible Hospitals for eCQM reporting.[256] In 2015, CMS moved to a combined approach for its implementation guides. The 2015 Guide is based on QRDA Category I,

[256] The CMS QRDA Implementation Guides are downloadable from the CMS eCQM Library web site. http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/eCQM_Library.html. Accessed on: July 27, 2015.

DSTU Release 2 and its errata updates (December 2012 and June 2014). The guide provides CMS-specific requirements for Eligible Professionals and Eligible Hospitals, such as requiring the CMS Certification Number for hospitals when submitting QRDA Category I reports, by further constraining the base HL7 standard.

5.5—Continued Improvement and Maintenance

For continued improvement and maintenance of the HL7 standards for QRDA Category I (R3), the general public has been able to submit errata and new feature requests through the HL7 DSTU Comment page.[257] The two-year DSTU period ended in July 2014 and was extended through July 2015 for the QRDA Category I standard. QRDA I had a new release published (Release 3) in June 2015. The two-year DSTU period for the QRDA Category III standard ended in November 2014 and was extended until November 2015. The DSTU comments have been and will be continually evaluated and discussed; the dispositions to the comments will be approved following the HL7 formal voting process. After the comments have been discussed and approved, an errata package may be published to address the identified technical errors. New feature requests were published in Release 3. Once the DSTU period has expired, the QRDA Category I Release 3 and QRDA Category III Release 1 may go through the HL7 ballot process as a normative standard.

ONC JIRA QRDA Project

The ONC JIRA system maintains a project specifically for QRDA. Issues and questions related to the CMS implementation guides, QRDA standards in general, and their implementations can be submitted to the QRDA JIRA project. CMS and ONC monitor the QRDA project closely in order to triage and address issues. The JIRA QRDA project is a valuable resource for CMS and ONC to continuously improve QRDA to report eMeasures, and to provide support to the measure implementers.

Synchronized Versioning of Quality Standards and Artifacts

The QRDA Category I standard is tightly coupled with the QDM, QDM-based HQMF Implementation Guide, and, therefore, the MAT. This intentional design effectively connects the end-to-end process from eMeasure specifications to eMeasure reporting. This coupling also means that careful planning and coordination among the multiple stakeholders should be put in place to maintain synchronized updates and maintenance of these standards and artifacts.

The QDM is currently being harmonized with other relevant clinical decision support standards. It will continue to evolve based on stakeholder input and feedback from the QDM User Group to meet new eMeasure development and reporting needs. Existing QDM datatypes and attributes may be modified and new QDM datatypes or attributes may be introduced as the QDM evolves. As a result, the QDM-based HQMF Implementation Guide and the QRDA Category I will need to be updated to support the QDM updates and to maintain the corresponding mappings between HQMF and QRDA representations for the QDM datatypes and attributes. The MAT should also be

[257] Health Level Seven DSTU Comment page. http://www.hl7.org/dstucomments/

updated accordingly. This synchronized versioning across these quality standards and artifacts is important to ensure success in continued support and improvement for both eMeasure development and eMeasure reporting. For better support and coordination, the ownership of the QRDA and QDM-based HQMF standards has been transitioned to the Clinical Quality Information Workgroup. The workgroup helps to organize and coordinate releases among these standards and artifacts.

The QRDA Category III is evolving based on reporting requirements. As it is a standard for aggregated reporting, the QRDA Category III does not have the tight coupling with the QDM, QDM-based HQMF, and MAT similar to the QRDA Category I standard. Therefore, the QRDA Category III does not have the synchronized versioning issue as described above.

Chapter 6—eMeasure Use, Continuing Evaluation, and Maintenance

Similar to the use and continuing evaluation for other types of measures, to help CMS ensure continued soundness of eMeasure specifications, measure developers will also be responsible for maintaining the eMeasures once they are implemented. There are three basic types of measure reviews during maintenance: measure annual updates, comprehensive reevaluations, and ad hoc reviews. For retooled or reengineered measures, updates and reevaluation are also necessary to align with changes to the corresponding paper-based measure specifications. Refer to the Measure Use, Continuing Evaluation, and Maintenance chapter in Section 2 for more information on this process.

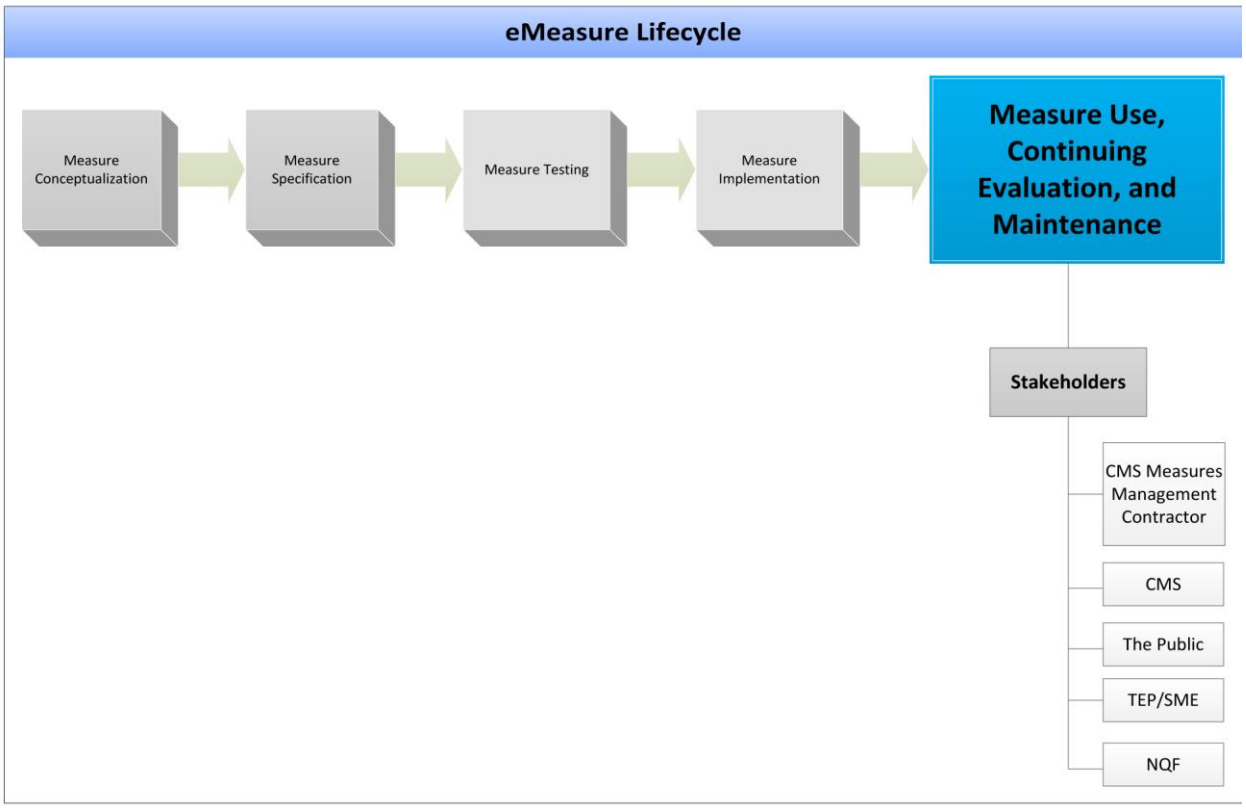
eMeasure updates or revisions may be required annually or every three years during the comprehensive reevaluation. Ideally, the measure maintenance schedule follows the NQF endorsement maintenance cycle. However, in practice, for various reasons, these schedules may not align completely. For example, as the HQMF and QDM standards evolve, eMeasures may need to be updated and reevaluated to conform to the latest standards.

The eMeasure updates and reevaluation are based on several sources:

- Code system updates from NLM
- Feedback from public comments through JIRA or other forums
- Proposed changes from CMS, the measure steward, and measure developers
- Suggestions from QA team regarding harmonizing data elements, value sets, and logic

Figure 38: eMeasure Use, Continuing Evaluation, and Maintenance Key Stakeholders below depicts the key stakeholders needed for measure use, continuing evaluation, and maintenance.

Figure 38: eMeasure Use, Continuing Evaluation, and Maintenance Key Stakeholders



6.1—Deliverables

- Updated eSpecifications (eMeasure XML file, SimpleXML file, eMeasure human-readable rendition [HTML] file) and value sets.
- Feedback review of JIRA comments.
- Release notes detailing the changes to the eSpecifications and value sets.
- 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.
- 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 three-year endorsement maintenance).
- 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.

6.2—eMeasure Evaluation and Testing After Implementation

eMeasure evaluation after implementation offers the opportunity to assess how diverse provider practice patterns affect the measure results. A comparison of the different types of evaluation activities discussed in this chapter is illustrated in Table 25: Application of Evaluation Methods. [258]

Table 25: Application of Evaluation Methods

Evaluation Activity	New or Newly Retooled eMeasures	Maintaining Implemented eMeasures
Validity		
Measure Level ("Face") Validity	Required	Required if any changes are made during maintenance
Measure Logic Validity	Required	Required if relevant changes are made during maintenance
Data Element Validity	Required	
Measure Score Validity	Required	
Reliability		
Measure Level ("Face") Reliability	Required	
Data Element Reliability	Required	Required if relevant changes are made during maintenance
Feasibility		
Basic Feasibility	Required	Required if relevant changes are made during maintenance
Data Element Feasibility	Required	

Once an eMeasure has been implemented, opportunities for ongoing evaluation and testing of the eMeasure arise through the measure maintenance process. This is particularly true given that:

- Some EHR systems may not have been available at the time of initial testing (prior to measure endorsement).
- eMeasure testing can only provide a snapshot of EHR capabilities.
- Advancements to vendor systems continually occur.

Validity and Reliability Reevaluation

Measures may need to be reevaluated because of the complexity of the specifications. For example, a specification that is complex to implement is susceptible to differences in data field usage by different users, or it may incorporate elements that are not consistently entered into the EHR fields created by vendors to report the measure results. Both of these events would impact measure logic validity. If the clinical vocabulary code sets in the measure undergo small changes, this impacts data element reliability. Given the difficulty of ensuring standardized data extraction across diverse EHRs and EHR users, different options should be considered when testing an eMeasure during ongoing maintenance periods. NQF recommendations for retooled eMeasures

[258] eMeasure evaluation addresses three key measure properties described in Chapter 3—eMeasure Testing: validity, reliability, and feasibility. Different validity, reliability, and feasibility testing methods are employed to test each of the different components in a measure to confirm an eMeasure's scientific acceptability and adoptability. The different types of eMeasure testing methods for each measure property are described in this section.

(and time-limited measures) previously tested on the original data source include testing of reliability and validity by the next endorsement maintenance review. These same methods should be considered for new measures undergoing maintenance for the first time. Refer to the eMeasure Testing chapter for more information on testing methods.

NQF guidance further clarifies that reliability testing of measure elements may be supplanted by evidence of measure element validity. Consequently, eMeasures undergoing a comprehensive reevaluation do not necessarily require testing of the measure data elements if prior evidence demonstrates sufficient validity. However, when an eMeasure has been widely implemented and element-level data are available from a large, diverse number of providers, examining differences across groups may help determine if provider practice patterns or specific EHR systems result in dramatic differences in availability of data element level information specified in the measure. This examination may also help inform recommendations for changing the measure specifications. For example, dramatic differences in the storage location of specific elements based on a particular EHR system may suggest areas that should be reviewed if measure specifications are changed.

Feasibility of eMeasures

Though usability of an eMeasure generally is demonstrated in a manner equivalent to a non-eMeasure, feasibility will require, at a minimum:

- Determination of which measure-specified data elements are typically captured in EHRs.
- Whether the EHRs can reliably extract the specified data.
- The ability of the EHRs to capture these data through customary workflow.
- Identification of any barriers to implementation related to technical constraints of EHRs.
- Whether data captured in the EHRs are captured in a form that is semantically aligned with the expectations of the quality measure.

For measures that have not yet been implemented at the time of the comprehensive reevaluation, the prior review may be sufficient, or it may be augmented by evaluation across a larger set of EHR systems. For measures that have already been implemented, feedback from eMeasure implementers may provide additional information suitable for the reevaluation of feasibility.

Adapted measures

As a measure undergoes continued evaluation, the measure will evolve and a may need to be adapted for wider use. The changes to the measures will require refinement and testing before implementing publicly. Adapted measures with material changes must be resubmitted for NQF endorsement consideration.

Testing

If there are material changes to specifications following measure evaluation, it may be necessary to retest the eMeasures. Examples of material changes include those affecting the original measure's concept, logic, intended meaning, or strength of the measure relative to the measure

evaluation criteria. All of these changes should be tested using the techniques described in the eMeasure Testing chapter, and a summary report documenting the results should be submitted.

6.3—Annual Update Process for eMeasures

Electronically specified measures are updated annually based on public feedback and response to JIRA tickets. Updates to the measure may impact the clinical intent of the measure, measure logic, and/or value set composition if terms are changed or retired. In response, the annual update process includes a clinical, measure logic, and value set review to ensure the scientific acceptability of the measure. The goal of the annual update process is to produce update and maintain measures that are of high quality. This process involves measure stewards, measure developers, measure testers, CMS, and other stakeholders.

Similar to eMeasure testing, each part of the review process is iterative and may involve several rounds of feedback and review with the intent of resolving all outstanding issues.

Clinical Review

During clinical review, measure concepts that are changed or altered are reviewed by the measure steward. Measure stewards check to see if the translations of the human-readable format to the logic statements correctly represent the intent of the change. Clinical concepts are also validated for accuracy. If value sets are changed, the terms in the value set are reviewed for appropriate representation and coverage of the concept intended by the measure. The clinical review should occur prior to re-entering the specifications into the MAT.

Logic Review

When a change to the measure logic is made, measure logic validity testing should be conducted as described in Types of eMeasure Testing. New test cases are created to test whether the logic is expressed accurately and places a patient in the expected numerator, denominator, or exception population. A preliminary logic review is often conducted prior to MAT entry to assess the complexity of the logic and impact implementation. Once the updated logic is in the MAT, a Post-MAT version of the measure may be imported into a testing tool such as Bonnie to test the changes in the measure.

Value set review

A value set review is conducted on all measures annually. At minimum, a value set is reviewed to see if there is an existing value set and to update the value set terms to the latest version of its code system. This process is referred to as value set harmonization. If a member of a value set is changed or altered, the value set will be reviewed for data element reliability testing. Details on this method of testing are covered in Types of eMeasure Testing. As mentioned above, a clinical value set review is conducted by the measure steward and the review takes place in the VSAC. In addition, a technical value set review should be performed to validate the codes with the latest version of the standard code system.

Value set harmonization

The development of eMeasures by various organizations, measure developers, and measure stewards, under different federal contracts, has led to different value sets being developed for the same or similar concepts. In the EHRs, hospitals and providers may have variations in the code list to represent the same clinical content. As a result, value sets from different reporting systems may have differences in representations of the same concept.

How they are identified

Because each measure developer is responsible for a small set of eMeasures, identifying duplications of value sets for data elements is difficult at the eMeasure development stage. However, end users of eMeasures, such as vendors, tend to more readily identify duplications as they evaluate and compare eMeasures side by side. The duplicated value sets are then reported by the vendors in the ONC JIRA system. Another approach to duplications is having the QA team or developers conduct an evaluation of overlap between codes among value sets. Once the potentially duplicated value sets are identified, a code-level comparison is followed and harmonization of codes or even value sets is determined. The harmonization process often requires a final decision by the measure steward after an evaluation of discrepancies in codes has been conducted. Sometimes identical value sets are found and the solution is relatively simple—a matter of keeping one value set and making the others obsolete. However, the level of impact on changing eMeasures should be considered as the purpose is to achieve optimum efficiency.

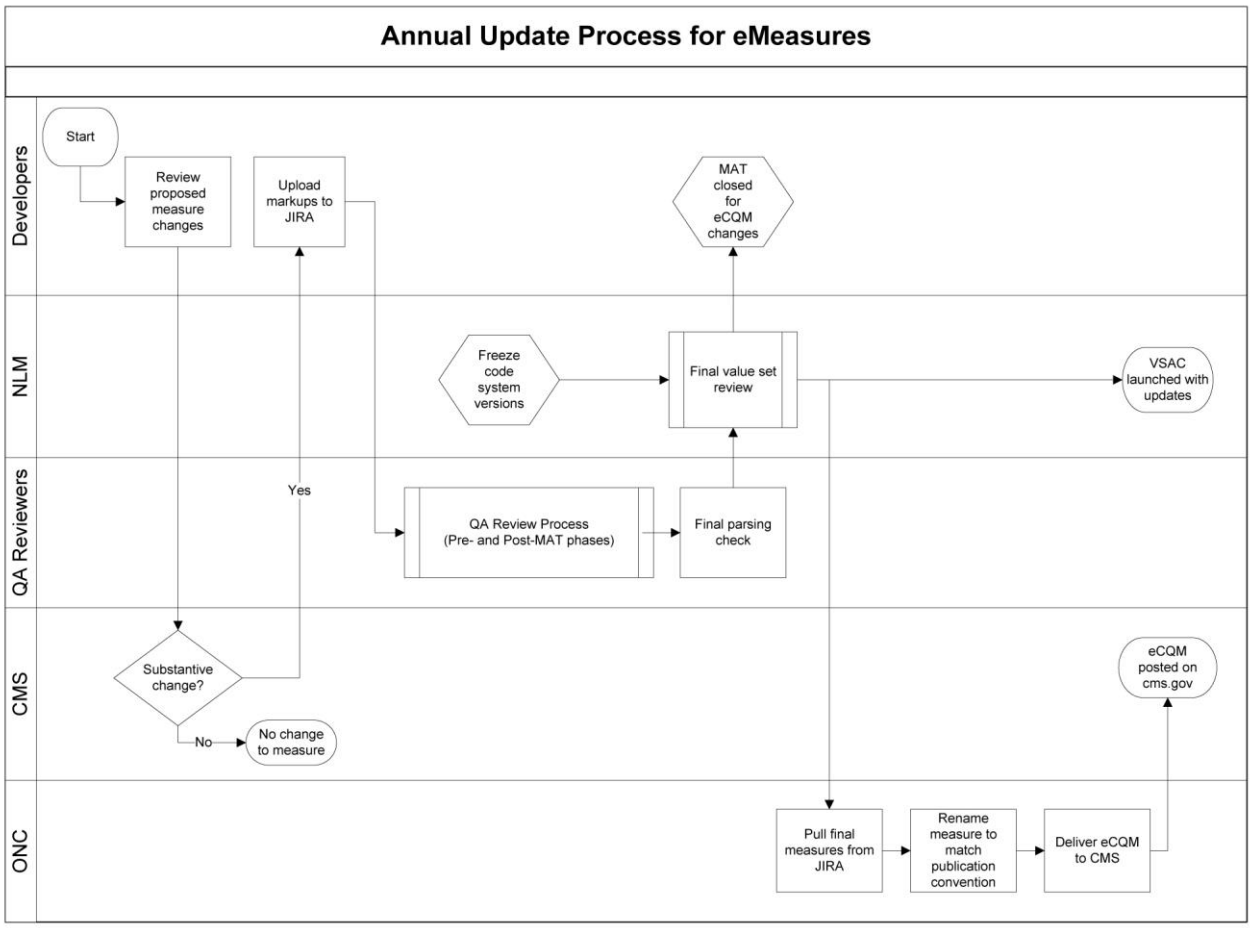
Harmonization process

An inpatient encounter can be captured using multiple value sets (shown in the table below). The detailed comparison at a code level is presented to the measure steward for evaluation of any discrepancies. The measure steward prefers certain codes based on their experience of processing the data. The value sets shown in Appendix I were harmonized into a single value set with OID 2.16.840.1.113883.3.666.5.307 during the 2014 Eligible Hospital annual update. Once a harmonized value set is finalized, all the other value sets will be replaced in the eMeasures. Often the harmonization is achieved via group discussions between eMeasure developers, the measure steward, and terminologists.

Once the clinical, logic, and value set reviews are complete, and the measure is signed off by the measure steward, the measure is ready for packaging and publication.

Figure 39: Annual Update Process for eMeasures is a high-level flow diagram showing the overall annual update process.

Figure 39: Annual Update Process for eMeasures



6.4—Comprehensive Reevaluation for eMeasures

As eMeasures are still relatively new, a comprehensive reevaluation has not currently been performed on existing eMeasures. However, the reevaluation process includes all of the elements of the annual update with the addition of a full measure evaluation. Measure developers/developers must demonstrate that the eMeasures continue to meet the following criteria:

- 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

Further details of comprehensive reevaluation applicable to all measures including eMeasures are found in Chapter 5—Measure Use, Continuing Evaluation, and Maintenance in Section 2. Deliverables for the eMeasure Comprehensive Reevaluation are the same as listed in Section 2 with the addition of the eSpecifications update.

6.5—Ad Hoc Review for eMeasures

The ad hoc review for eMeasures is similar to the annual update process and may be completed on a smaller or larger scale depending on the needed updates to the measure. Further details about ad hoc reviews, including when they might be triggered and how to conduct them, are also found in Chapter 5—Measure Use, Continuing Evaluation, and Maintenance in Section 2. Deliverables for the eMeasure Ad Hoc Review are the same as listed in Section 2, with the addition of the eSpecifications update.

Section 4—Tools, Appendices, and Forms

Acronyms and Abbreviations

ACA—Affordable Care Act
AHRQ—Agency for Healthcare Research and Quality
AMA—American Medical Association
CDA—Clinical Document Architecture
CDC—Centers for Disease Control and Prevention
CMS—Centers for Medicare & Medicaid Services
COR—Contracting Officer Representative
CPT—Current Procedural Terminology
CQM—Clinical Quality Measure
CSAC—NQF Consensus Standards Approval Committee
DENOM—Denominator
DENEX—Denominator Exclusion
DENEXCEP—Denominator Exception
DSTU—Draft Standard for Trial Use
eMIG—electronic Measures Issues Group
eCQM—electronic Clinical Quality Measure
EHR—electronic health record
HAC—Hospital-Acquired Conditions
HCPCS—Healthcare Common Procedure Coding System
HHS—Department of Health and Human Services
HIT—Health Information Technology
HITSC—Healthcare Information Technology Standards Committee
HL7—Health Level Seven
HQMF—Health Quality Measures Format
ICD-10-CM—International Classification of Diseases, Tenth Revision, Clinical Modification
ICD-10-PCS—International Classification of Diseases, Tenth Revision, Procedural Coding System
ICD-9-CM—International Classification of Diseases, Ninth Revision, Clinical Modification
IOM—Institute of Medicine
IPOP—Initial Population
LOINC—Logical Observation Identifiers Names and Codes
MACRA—Medicare Access and CHIP Reauthorization Act
MAP—Measure Applications Partnership
MAT—Measure Authoring Tool
MCC—Multiple Chronic Conditions
MIDS—Measure and Instrument Development and Support
MMS—Measures Management System
MUC—Measures under Consideration
NLM—National Library of Medicine
NQF—National Quality Forum

NUMER—Numerator
NUMEX—Numerator Exclusion
OID—Object ID
OMB—Office of Management and Budget
ONC—Office of the National Coordinator for Health Information Technology
PCORI—Patient-Centered Outcomes Research Institute
PHIN VADS—CDC Public Health Information Network (PHIN) Vocabulary Access and Distribution System (VADS)
PQRS—Physician Quality Reporting System
PRA—Paperwork Reduction Act
PRO—Patient-Reported Outcomes
PROM—Patient-Reported Outcome Measure
PRO-PM—Patient-Reported Outcome Performance Measurements
QDM—Quality Data Model (QDM, previously known as Quality Data Set)
QIN-QIO—Quality Innovation Network-Quality Improvement Organization
QRDA—Quality Reporting Document Architecture
RCT—Randomized Controlled Trial
SME—Subject Matter Expert
SNOMED CT—Systematized Nomenclature of Medicine—Clinical Terms
TEP—Technical Expert Panel
USCRS—SNOMED CT Content Request System
USHIK—US Health Information Knowledgebase
VSAC—Value Set Authority Center
XML—Extensible Markup Language

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."[259] 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.[260]

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

[259] 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: June 1, 2015.

[260] Krumholz, H., et al. Standard for measures: efficiency in health care. *Journal of the American College of Cardiology*. 2008;52 (number 18):1522.

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.

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.[261] 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 continuous variable values are identified by a measure. Also referred to as the performance calculation.

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 sets

A code system is a way to turn health information like a diagnosis or procedure name into numbers or code to make sharing information easier and faster. A code set is a specific version of that system's rules.[262]

Common use case

A use case describes the flow of quality information through an electronic health record (EHR) system for the purpose of quality measurement, feedback and reporting.[263]

---

[261] 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: June 1, 2015.

[262] National Quality Forum. *Glossary of Terms.* Available at: https://www.qualityforum.org/Measuring_Performance/Submitting_Standards/NQF_Glossary.aspx. Accessed on: June 25, 2015.

[263] 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.

Comparable data

The accuracy, reproducibility, risk-adjustability, and validity of the measure should not be affected if different systems use different sources for measures.[264] Data applied to a specific measure must be collected using similar methods and with a common definition throughout the population of interest.

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.[265] Also called composite measures.

Concatenation

Combined, joined, or linked together; interconnected, as in a chain.

Concordance rate

The proportion (and probability of that proportion) in a random sample of pairs that both individuals will have a certain characteristic, given that one of the pair has the characteristic.

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.

---

[264] Institute for Medicine, National Committee for Quality Assurance. *IOM's Envisioning the National Health Care Quality Report.* Washington, DC: The National Academies Press, 2001. Appendix E.

[265] National Quality Forum. *Composite Performance Measure Evaluation Guidance*. Washington, DC: National Quality Forum; Apr 2013; pg. 27; Contract No. HHSM-500-2009-00010C. Available at: http://www.qualityforum.org/Publications/2013/04/Composite_Performance_Measure_Evaluation_Guidance.aspx. Accessed on: June 1, 2015.

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.

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. Different measures within an eMeasure set may have different denominators. 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 is used only in proportion measures. Denominator exception allows for the exercise of clinical judgment and should be specifically defined where capturing the information in a structured manner fits the clinical workflow.

These cases are removed from the denominator; however the number of patients with valid exception may still be reported. Exception allows for the exercise of clinical judgment. 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.

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.[266]

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 Agency for Healthcare Research and Quality has developed a typology or analytical framework of efficiency measures: perspective, outputs, and inputs.[267]

The IOM defines efficiency as avoiding waste, including waste of equipment, supplies, and energy. To measure or assess efficiency, and ultimately value, associated with the care over the course of an episode of illness, the NQF has developed a framework to guide future and ongoing efforts in measuring efficiency in healthcare. The following constructs are essential to adequately assess the overall efficiency of the healthcare delivery system.[268]

- *Quality of care* is a measure of performance on the six IOM-specified healthcare aims: safety, timeliness, effectiveness, efficiency, equity, and patient centeredness.
- *Cost of care* is a measure of total healthcare spending, which includes 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.
- *Efficiency of care* is a measure of cost of care associated with a specific level of performance measured with respect to the other five IOM aims of quality.
- *Value of care* is a measure of a specified stakeholder's preference-weighted assessment of a particular combination of quality and cost of care performance. Examples of stakeholders

[266] Truman B., Smith CK, Roy K, et al. *CDC Health Disparities and Inequalities Report – United States, 2011*. Centers for Disease Control and Prevention, 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: June 1, 2015.

[267] 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: June 1, 2015.

[268] National Quality Forum . Measurement framework: evaluating efficiency across patient-focused episodes of care. Washington, DC: National Quality Forum; 2009; Report No. ISBN: 978-1-933875-42-2.

include individual patients, consumer organizations, payers, providers, governments, or societies.

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."[269]

Electronic specifications

Refers to measure specifications derived from EHRs. Electronic specifications contain four main components:

- Measure Overview/Description—contains the measure title, description, number, measurement period, measure steward, and other relevant information to the measure.
- Measure Logic—contains population criteria and measure logic for numerator, denominator, and exclusion categories. The measure logic contains the algorithm used to calculate performance.
- Measure Code Lists (Value sets)—contains all of the codes pertaining to the measure.
- Quality Data Model (QDM) elements—lists and describes each QDM data element with the measure. See Quality Data Model.

eMeasure

An eMeasure is a health quality measure encoded in a Health Quality Measures Format (HQMF). Also called eCQM, eMeasures pull the information needed to evaluate performance directly from the electronic record. They can be far more efficient than traditional approaches of extracting data from paper charts or claims databases.[270]

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.

---

[269] Healthcare Information and Management and Systems Society. *Resource Library: Electronic Health Records.* Available at: http://www.himss.org/library/ehr/?navItemNumber=13261. Accessed on: June 1, 2015.

[270] 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: June 1, 2015.

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.

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.

Formal

Officially recognized or final version.

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 eMeasures) should be harmonized when the intended meaning is the same. Harmonization of logic in eMeasures 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.[271]

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.

[271] Wikipedia. *Health information technology*. Available at: http://en.wikipedia.org/wiki/Health_information_technology. Accessed on: May 31, 2015.

Health Quality Measures Format (HQMF)

A standards-based representation of quality measures as electronic documents. A quality measure expressed in this way is also referred to as an eMeasure.

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 (patient) population

Refers to all events to be evaluated by a specific performance eMeasure 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.

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 observation from two or more human observers are congruent with each other.

Inverse measures

Inverse measures are measures where a lower performance rate is better. Therefore a zero performance rate for these measures is a good score. For example, The National Healthcare Safety

Network calculates most Healthcare-Associated Infections 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 program 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.[272]

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.[273]

Leverage point

A key state of health, clinical process, or event that has demonstrable effect on the health outcome. "Three considerations arise when evaluating leverage points: (1) the area being measured is an important contributing factor to the clinical or contextual process for the goal, (2) the area is one in which measurement and reporting is likely to stimulate improvement (through either selection or change), and (3) the purpose is to be selective rather than comprehensive."[274]

Logic

The criteria used to define a quality measure and its key components.[275]

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.

---

[272] http://en.wikipedia.org/wiki/JIRA. Accessed on: May 31, 2015.
[273] http://en.wikipedia.org/wiki/Kaizen. Accessed on: May 31, 2015.
[274] McGlynn EA. Selecting common measures of quality and system performance. *Medical Care.* 2003;41(suppl) 39-47.
[275] 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 down. 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 Authoring Tool (MAT)

A publicly available, web-based tool for measure developers to create eMeasures; it should also reduce the time required to create new quality measures, and to convert existing paper-based measures into EHR-readable format.[276]

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 score 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.[277]

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

---

[276] 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.

[277] 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.

exclusion, risk adjustment/stratification for outcome and resource use measures, methods to identify differences in performance, and comparability of data sources/methods.

Measure (performance measure or quality measure)

A mechanism to assign a quantity to an attribute by comparison to a criterion.[278] 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.[279]

Measures under Consideration

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.[280]

Medical record (data source)

Data obtained from the records or documentation maintained on a patient in any healthcare setting (for example, hospital, home care, long-term care, practitioner office). Includes automated and paper medical record systems.

Metadata

Precise, semantic definition of a data element. Provides additional descriptive information about an eCQM.

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.

---

[278] 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.

[279] 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.

[280] Section 3014 of the Affordable Care Act (ACA) (P.L. 111-148) created the new Section 1890A(a)(2) of the Social Security Act.

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.[281]

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 National Quality Strategy.

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

Define instances that should not be included in the numerator data. Numerator exclusion is 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.[282]

---

[281] 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.

[282] 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.

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.[283]

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 of:

- Health-related quality of life (including functional status).
- Symptoms and symptom burden (e.g. pain, fatigue).
- Experience with care.
- Health behaviors (e.g., smoking, diet, exercise).[284]

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.[285]

Patient-reported outcome-based performance measure (PRO-PM)

A performance measure that is based on PROM data aggregated for an accountable healthcare entity.[286]

Pattern

A named set of formal, computable constraints used to calculate a measure. Formerly referred to as *Quality Data Model Template*.[287]

---

[283] 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.

[284] 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.

[285] Ibid.

[286] Ibid.

[287] Adapted from *QDM December 2012 HQMF Templates***.** Available at: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=72518 (has examples which could be used for illustrative purposes). Accessed on: May 31, 2015.

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.[288]

Quality Data Model (QDM)

The QDM is part of NQF's work in health information technology. It is an "information model" that defines concepts used in quality measures and clinical care so that users can clearly and concisely locate and communicate pieces of electronic information.

[288] *Merriam-Webster's Online Dictionary, Public Domain.* Available at: http://www.merriam-webster.com/dictionary/public+domain. Accessed on: May 31, 2015.

The QDM can be used to help the designers of EHRs to improve consistency between different systems. This improves automation and the ability of different systems to exchange electronic information. [289]

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)

As defined by HL7, "a standard for communicating healthcare quality measurement information in support of calculation of eMeasures."

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.

Reengineered eMeasure

Based on the intent of a preexisting paper-based measures but does not necessarily include a one-to-one match of data elements; steward facilitates development. For example, AMI 30-day mortality is to eSpecify an existing measure construct without trying to exactly match the existing specifications.

[289] NQF Glossary: Available from http://www.qualityforum.org/QualityDataModel.aspx. Accessed May 12, 2015.

Re-expressed eMeasure

Starting with the initial measure concept and evidence and configuring the eMeasure de novo from the same evidence to address clinical data rather than data derived from claims-based or abstraction workflow.

Related measures

Related measures address either the same topic or the same population. This term is used when considering harmonization. See also competing measures.

Reliability (part of scientific acceptability)

Reflects that the measure is well defined and precisely specified so it can be implemented consistently within and across organizations and that it distinguishes differences in performance.

Reliability testing

Evaluates whether the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. Reliability 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).

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.

Retooling

Conversion of measures from paper-based format to an electronic (eMeasure) format.

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.[290]

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 eMeasure whereby the formal criteria in an eMeasure 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.

[290] National Quality Forum. *Glossary of Terms.* Available at: *https://www.qualityforum.org/Measuring_Performance/Submitting_Standards/NQF_Glossary.aspx*. Accessed on: June 25, 2015.

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.

Structure measure

A structure measure is one that assesses features of a healthcare organization or clinician relevant to its capacity to provide healthcare.

Structure (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)

Substitute variable

Also known as a proxy variable, it is used when direct assessment of the relevant variable is not possible. Relevance should be clinically and empirically apparent.

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 window

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.[291]

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:

- 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.

[291] 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.

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 Authority Center (VSAC)

The VSAC is a NIH public use tool for maintaining code set lists used in eCQM (called an object identifier or OID).

Value Sets

A value set is a list of specific clinical terms and the codes that correspond with them. A value set defines each of the clinical terms in the elements of a quality measure. Value sets support the calculation of eMeasures and the systematic exchange of health information. (Source NQF Glossary)

Zero Performance Rate

Individual measures with a 0% performance rate will not be counted as satisfactorily reporting for the 2012 Physician Quality Reporting System. The recommended clinical quality action must be performed on at least one patient for each individual measure reported by the eligible professional. A 0% performance rate could be due to the fact that none of the provider's eligible patients were in compliance for the measure or that the provider did not provide the correct quality action to the patient. Exception: When a lower rate indicates better performance, a 0% performance rate will be counted as satisfactorily reporting (100% performance rate would not be considered satisfactorily reporting). Performance exclusion quality-data codes are not counted in the performance denominator. If the eligible professional submits all performance exclusion quality-data codes, the performance rate would be 0/0 (null) and would be considered satisfactorily reporting.

Appendix A: XML View of a Sample eMeasure

This appendix[292] shows a detailed example of an eMeasure,[293] encoded per the HQMF standard and the additional rules stipulated in the eMeasure Lifecycle section, illustrating how it will appear when rendered in a Web browser. Following the example, a high-level view of the underlying XML is provided. Note, this sample is provided using HQMF Release 1 (R1). The HQMF Release 2.1 (R2.1) human readable version is available to export from the MAT with the July 2014 version. The full HQMF R2.1 is expected to be available from the MAT in late Q4 2014.

Table 26: Example of XML View

eMeasure Title	Asthma Assessment		
eMeasure Identifier (Measure Authoring Tool)	123	eMeasure Version Number	1
NQF Number: 0001	GUID	59B24505-B9D1-48B8-BEB2-AF35219AA689	
Measurement Period	January 1, 2011, through December 31, 2011		
Measure Steward	American Medical Association-Physician Consortium for Performance Improvement		
Measure Developer	American Medical Association-Physician Consortium for Performance Improvement		
Endorsed by	National Quality Forum		
Description	Percentage of patients aged 5 through 40 years with a diagnosis of asthma who were evaluated during at least one office visit within 12 months for the frequency (numeric) of daytime and nocturnal asthma symptoms.		
Copyright	©2010 American Medical Association. All Rights Reserved.		
Disclaimer	None		
Measure scoring	Proportion		
Measure type	Process		
Stratification	Population 1: DOB 14–23 years before start of measurement period Population 2: DOB 14–19 years before start of measurement period Population 3: DOB 20–23 years before start of measurement period		
Risk Adjustment	None		
Rate Aggregation	None		
Rationale	Appropriate treatment of asthma patients requires accurate classification of asthma severity. Physician assessment of the frequency of asthma symptoms is the first step in classifying asthma severity.		
Clinical Recommendation Statement	To determine whether the goals of therapy are being met, monitoring is recommended in the 6 areas listed below: • Signs and symptoms (daytime; nocturnal awakening) of asthma • Pulmonary function (spirometry; peak flow monitoring) • Quality of life/functional status • History of asthma exacerbations • Pharmacotherapy (as-needed use of inhaled short-acting beta2-agonist, adherence to regimen of long-term-control medications) • Patient-provider communication and patient satisfaction (NAEPP/NHLBI)		
Improvement Notation	Higher score indicates better quality		
Reference	National Asthma Education and Prevention Program Expert Panel Report 2: Guidelines for the diagnosis and management of asthma. National Heart, Lung, and Blood Institute, National		

[292] Note that though value sets are currently authored in the measure authoring tool, they will be authored directly in VSAC in the near future, and they are not part of the published eMeasure. Rather, the published eMeasure contains value set references.

[293] This is a sample measure, solely for illustration purposes. It is not a real measure.

	Institutes of Health; July 1997. NIH Publication No. 97-4051. Available at: http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.htm. Accessed on: August 2002.
Reference	National Asthma Education and Prevention Program Expert Panel Report 2 Update: Guidelines for the diagnosis and management of asthma – update on selected topics 2002. National Heart, Lung, and Blood Institute, National Institutes of Health; August 2002. NIH Publication No. 97-4051. Available at: http://www.nhlbi.nih.gov/guidelines/asthma/asthsumm.htm. Accessed on: September 2002.
Definition	None
Guidance	None
Transmission Format	None
Initial Population	Brief narrative description of population. Must be consistent with the computer-generated narrative logic in the body of the eMeasure. The computer-generated narrative is standardized and concise, and it can lack the richness of full text that sometimes helps in the understanding of a measure. This is especially true for measures that have complex criteria, where the computer-generated text may not be able to express the exact description that a measure developer would like to convey. As part of the quality assurance step, it is important to compare the human-readable description of the measure population (in the header) to the logic representation (in the body) for any discrepancies.
Denominator	Brief narrative description of population. If population isn't applicable, then say “Not Applicable.” If population isn't defined, then say “None”.
Denominator Exclusion	Brief narrative description of population. If population isn't applicable, then say “Not Applicable.” If population isn't defined, then say “None”.
Numerator	Brief narrative description of population. If population isn't applicable, then say “Not Applicable.” If population isn't defined, then say “None”.
Numerator Exclusion	None
Denominator Exception	Brief narrative description of population. If population isn't applicable, then say “Not Applicable.” If population isn't defined, then say “None”.
Measure Population	Not Applicable
Measure Observation	Brief narrative description of measure observation. If not a Continuous Variable measure, then should say “Not Applicable”.
Supplemental Data Elements	For every patient evaluated by this measure also identify payer, race, ethnicity, and sex.

Table of Contents

- Population Criteria
- Data Criteria
- Measure Observation (only shown if present)
- Reporting Stratification (only shown if present)
- Supplemental Data Elements

Population criteria

Populations will vary based on type of measure scoring

Initial Population =

- AND:

- AND: "Patient Characteristic Birthdate: birth date" >= 5 year(s) starts before start of "Measurement Period"
- AND: "Patient Characteristic Birthdate: birth date" < 40 year(s) starts before start of "Measurement Period"
- AND: "Diagnosis Active: Asthma" starts before or during ("Encounter, Performed: Encounter Office & Outpatient Consult" during "Measurement Period")
- AND: >= 2 count(s) of
- AND: "Encounter, Performed: Encounter Office & Outpatient Consult" during "Measurement Period"

Denominator =

- AND: "Initial Population"

Denominator exclusion =

- None

Numerator =

- AND:
- OR:
  - AND: "Symptom Assessed: Asthma Daytime Symptoms Quantified"
  - AND: "Symptom Assessed: Asthma Nighttime Symptoms Quantified"
  - Starts before or during ("Encounter, Performed: Encounter Office & Outpatient Consult" during "Measurement Period")
- OR:
  - AND: "Symptom Active: Asthma Daytime Symptoms"
  - AND: "Symptom Active: Asthma Nighttime Symptoms"
  - Starts before or during ("Encounter, Performed: Encounter Office & Outpatient Consult" during "Measurement Period")
- OR: "Risk category/assessment: Asthma Symptom Assessment Tool" starts before or during ("Encounter: Encounter Office & Outpatient Consult" during "Measurement period")

Denominator exception =

- None

Data criteria

- "Diagnosis, Active: Asthma" using "Asthma GROUPING Value Set (2.16.840.1.113883.3.526.03.362)"
- "Encounter, Performed: Encounter Office & Outpatient Consult" using "Encounter Office & Outpatient Consult CPT Value Set (2.16.840.1.113883.3.526.02.99)"

- "Patient Characteristic Birthdate: birth date" (age) using "birth date LOINC Value Set (2.16.840.1.113883.3.560.100.4)"
- "Risk Category/assessment: Asthma Symptom Assessment Tool" using "Asthma Symptom Assessment Tool SNOMED-CT Value Set (2.16.840.1.113883.3.526.2.143)"
- "Symptom, Active: Asthma Daytime Symptoms" using "Asthma Daytime Symptoms SNOMED-CT Value Set (2.16.840.1.113883.3.526.02.440)"
- "Symptom, Active: Asthma Nighttime Symptoms" using "Asthma Nighttime Symptoms SNOMED-CT Value Set (2.16.840.1.113883.3.526.2.441)"
- "Symptom, Assessed: Asthma Daytime Symptoms Quantified" using "Asthma Daytime Symptoms Quantified SNOMED-CT Value Set (2.16.840.1.113883.3.526.2.141)"
- "Symptom, Assessed: Asthma Nighttime Symptoms Quantified" using "Asthma Nighttime Symptoms Quantified SNOMED-CT Value Set (2.16.840.1.113883.3.526.2.142)"

Reporting stratification

Reporting Stratum 1 =

- AND: "Patient Characteristic Birthdate: birth date" >= 14 year(s) starts before start of "Measurement Period"
- AND: "Patient Characteristic Birthdate: birth date" < 23 year(s) starts before start of "Measurement Period"

Reporting Stratum 2 =

- AND: "Patient Characteristic Birthdate: birth date" >= 14 year(s) starts before start of "Measurement Period"
- AND: "Patient Characteristic Birthdate: birth date" < 19 year(s) starts before start of "Measurement Period"

Reporting Stratum 3 =

- AND: "Patient Characteristic Birthdate: birth date" >= 20 year(s) starts before start of "Measurement Period"
- AND: "Patient Characteristic Birthdate: birth date" < 23 year(s) starts before start of "Measurement Period"

Supplemental data elements

- "Patient Characteristic Ethnicity: Ethnicity" using "Ethnicity CDC Value Set (2.16.840.1.114222.4.11.837)"
- "Patient Characteristic Sex: ONC Administrative Sex" using "ONC Administrative Sex Value Set (2.16.840.1.113762.1.4.1)"
- "Patient Characteristic Payer: Payer" using "Payer Source of Payment Typology Value Set (2.16.840.1.114222.4.11.3591)"

- "Patient Characteristic Race: Race" using "Race CDC Value Set (2.16.840.1.114222.4.11.836)"
- Measure set CLINICAL QUALITY MEASURE SET 2011–2012

Appendix B: Electronic Specification (eMeasure)

The prior appendix shows a detailed example of an eMeasure, encoded per the HQMF standard and the additional rules stipulated in the eMeasure Lifecycle section, illustrating how it will appear when rendered in a Web browser. This appendix provides a high-level view of the underlying XML.

The MAT will generate a much more detailed XML document than is shown here, but measure developers will require some knowledge of XML and the HQMF standard in order to make any subsequent manual edits to eMeasures generated by the tool. The skeletal example below shows the major components of a prototypic eMeasure document. Note, this sample is provided using HQMF Release 1 (R1). The HQMF Release 2.1 (R2.1) human readable version is available to export from the MAT with the July 2014 version. The full HQMF R2.1 is expected to be available from the MAT in late Q4 2014.

Table 27: Sample Electronic Specification

```
<QualityMeasureDocument>
 ... eMeasure Header ...
 <section>
 <title>Data criteria</title>
 <text>... narrative data criteria descriptions ...</text>
 <entry>... Formal data criteria definition ...</entry>
 <entry>... Formal data criteria definition ...</entry>
 ...
 </section>
 <section>
 <title>Population criteria</title>
 <text>... narrative population criteria descriptions ...</text>
 <entry>... Formal population criteria definition ...</entry>
 <entry>... Formal population criteria definition ...</entry>
 ...
 </section>
 <section>
 <title>Measure observation</title>
 <text>... narrative measure observation descriptions ...</text>
 <entry>... Formal measure observation definition ...</entry>
 <entry>... Formal measure observation definition ...</entry>
 ...
 </section>
 <section>
 <title>Reporting stratification</title>
 <text>... narrative reporting stratification descriptions ...</text>
 <entry>... Formal reporting stratification definition ...</entry>
 <entry>... Formal reporting stratification definition ...</entry>
 ...
 </section>
 <section>
 <title>Supplemental data elements</title>
 <text>... narrative supplemental data elements descriptions ...</text>
 <entry>... Formal supplemental data elements definition ...</entry>
 <entry>... Formal supplemental data elements definition ...</entry>
```

```
 ...
 </section>
 <section>
 <section>...</section>
 </section>
</QualityMeasureDocument>
```

Appendix C: eMeasure QA Checklists

Each checklist below corresponds to a particular type of review within the eCQM QA Review Process. Using a standard checklist for each of the review types helps to ensure consistency across different reviewers and allows metrics to be collected about the process and about the quality assurance pass/fail rates.[294]

Table 28: 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 patters 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			Are the mathematic inequalities reflective of the measure intent and represent the intended populations (for example: when intended the inequality represents less than rather than less and or equal to)?	
L-8	Y			Are operator precedence rules followed as specified in the Clinical Quality eMeasure Logic and Implementation Guidance v1.5?	
L-9	Y			Are QDM statement syntax rules followed in the logic?	
L-10	Y			Have subset operators been applied correctly as defined in the Clinical Quality eMeasure Logic and Implementation Guidance v1.5?	
L-11	Y			Are time intervals represented in similar units (e.g. hours)?	
L-12	Y			Are temporal logic operators used correctly as defined in the HL7 v3 Vocabulary Standard?	
L-13	Y			Are common populations and population criteria across similar measures harmonized (e.g., inpatient or medication)?	
L-14	Y			Are annotations included in the logic for sections of the measure that have been updated/changed?	

[294] Technical Authority for the Unified Clinical Quality Improvement Framework (Tacoma). *eCQM QA Review process 2.0* . Centers for Medicare & Medicaid Services, Office of the National Coordinator for Health Information Technology.

#	Mandatory?	Reviewed?	Passed?	Item	Comment
L-15	Y			(Post-MAT) Examine the .xml file. Is there any empty exception or empty exclusion?	
L-16	N			Have value set reviewers been alerted when the inappropriate use of value sets has been identified?	
L-17	N			Additional comments/issues/suggestions.	

Table 29: Clinical Review Checklist

#	Mandatory?	Reviewed	Passed	Item	Comments
C-1				Does each clinical element included in the measure narrative appear in the logic code (no errors of omission)?	
C-2				Are there errors when comparing the human-readable narrative to the logic narrative?	
C-3				Are there any incorrect translations of human-readable narrative to logic portion?	
C-4				Has the clinical algorithm been applied correctly in the logic code?	
C-5				Does the eCQM depart from recommended code systems (from a clinical context)?	
C-6				Are the medical terms used accurately and in the correct context (overall clinical validity and accuracy)?	
C-7				Are the medical terms clearly defined in the narrative (metadata) to provide enough guidance to support logic in the body?	
C-8				Are the medical terms used in the eCQM consistent with their definition in the Data Dictionary?	
C-9				Do the value sets chosen correspond to the value set selection intended by the measure steward?	
C-10				Are the value sets clinically appropriate?	
C-11				Do the value sets provide complete and appropriate coverage of and alignment with the data elements in the measure?	
C-12				Additional comments/issues/suggestions.	

Appendix D: eMeasure Review Process Quick Reference

Table 30: Naming File Attachments is intended to provide a quick reference for users of the JIRA eCQM review process for naming files and entering comments.[295]

Table 30: Naming File Attachments

Item	Naming Convention	Example
Pre-MAT review bundle	<eCQM ID #>_<Pre or Post> MAT_Cycle<#>.zip	CMS123_PreMAT_Cycle1.zip
Post-MAT review bundle	<eCQM ID #>_<Pre or Post> MAT_Cycle<#>.zip	CMS123_PostMAT_Cycle1.zip
Logic review	<eCQMID #>_LReview<Pre or Post>MAT_Cycle< #>.xls	CMS123_LReview_PreMat_Cycle1.xls
Word document with logic reviewer comments	< eCQM ID #>_<>_<Pre- or Post>MAT_Cycle<#>_LogicComments.doc	CMS123_PreMAT_Cycle2_LogicComments.doc
Pre-MAT completed logic review bundle. (This zip file would contain the completed QA checklist and any markup document.)	< eCQM ID #>_LReviewPreMAT_Cycle< #>.zip	CMS123_LReview_PreMat_Cycle1.zip
Post-MAT completed logic review bundle. (This zip file would contain the completed QA checklist and any markup document.)	< eCQM ID #>_LReviewPostMAT_Cycle< #>.zip	CMS123_LReview_PostMat_Cycle1.zip

Table 31: Comment Boxes is the quick reference for entering comments.

Table 31: Comment Boxes

Action	Convention	Example
Attaching a bundle for review (Pre- or Post-MAT)	<Phase>, <Cycle #> bundle ready for review.	Pre-MAT, Cycle 2 bundle ready for review.
Completed run through Bonnie	Measure and test cases have been loaded and tested in Bonnie.	Measure and test cases have been loaded and tested in Bonnie
Completed QA Review	<Passed/failed> <Phase>, <Cycle #> <review type> review. <Any additional comments>	Failed Post-MAT, Cycle 1 review.
Closing a review cycle	<Phase>, <Cycle #> closed. <Any additional comments>	Pre-MAT, Cycle 2 closed. Returning to developer for remediation.

[295] Technical Authority for the Unified Clinical Quality Improvement Framework (Tacoma). *"Cheat Sheet"— eCQM QA Review process 2.0.* Centers for Medicare & Medicaid Services, Office of the National Coordinator for Health Information Technology.

Appendix E: Time Unit and Time Interval Definitions

ISO 8601:2004 defines data elements and interchange formats for the representation of dates and times, including time intervals. Table 32: Time Unit and Time Interval Definitions is 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 eMeasures.

NOTE: Also reference the section on Age Calculation Conventions and Logic Guidance to understand how these time interval definitions relate to the development of age logic.

Table 32: Time Unit and Time Interval Definitions

Term	ISO Definition	ISO Notes
Time interval	Part of the time axis limited by two instants.	A time interval comprises all instants between the two limiting instants and, unless otherwise stated, the limiting instants themselves.
Duration	Non-negative quantity attributed to a time interval, the value of which is equal to the difference between the time points of the final instant and the initial instant of the time interval, when the time points are quantitative marks.	The SI unit of duration is the second. ▪ In case of discontinuities in the time scale, such as a leap second or the change from winter time to summer time and back, the computation of the duration requires the subtraction or addition of the change of duration of the discontinuity.
Nominal duration	Duration expressed among others in years, months, weeks, or days.	The duration of a calendar year, a calendar month, a calendar week or a calendar day depends on its position in the calendar. Therefore, the exact duration of a nominal duration can only be evaluated if the duration of the calendar years, calendar months, calendar weeks, or calendar days used are known.
Second	Base unit of measurement of time in the International System of Units (SI) as defined by the International Committee of Weights and Measures.	
Minute	Unit of time, equal to 60 seconds.	
Hour	Unit of time, equal to 60 minutes.	
Day <unit of time>	Unit of time, equal to 24 hours.	
Calendar day	Time interval starting at midnight and ending at the next midnight, the latter being also the starting instant of the next calendar day.	A calendar day is often also referred to as a day. The duration of a calendar day is 24 hours, except if modified by: ▪ The insertion or deletion of leap seconds, by decision of the International Earth Rotation Service (IERS), or ▪ The insertion or deletion of other time intervals, as may be prescribed by local authorities to alter the time scale of local time.
Day <duration>	Duration of a calendar day.	The term “day” applies also to the duration of any time interval which starts at a certain time of day at a certain calendar day and ends at the same time of day at the next calendar day.
Calendar week	Time interval of seven calendar days starting with a Monday.	A calendar week is also referred to as week.

Term	ISO Definition	ISO Notes
Week	Duration of a calendar week.	The term “week” also applies to the duration of any time interval which starts at a certain time of day at a certain calendar day and ends at the same time of day at the same calendar day of the next calendar week.
Calendar month	Time interval resulting from the division of a calendar year in 12 time intervals, each with a specific name and containing a specific number of calendar days.	A calendar month is often referred to as month.
Month	Duration of 28, 29, 30, or 31 calendar days depending on the start and/or the end of the corresponding time interval within the specific calendar month.	The term “month” applies also to the duration of any time interval which starts at a certain time of day at a certain calendar day of the calendar month and ends at the same time of day at the same calendar day of the next calendar month, if it exists. In other cases, the ending calendar day has to be agreed on. In certain applications, a month is considered as a duration of 30 calendar days.
Calendar year	Cyclic time interval in a calendar which is required for one revolution of the Earth around the Sun and approximated to an integral number of calendar days.	A calendar year is also referred to as year. Unless otherwise specified, the term designates in this International Standard a calendar year in the Gregorian calendar.
Year	Duration of 365 or 366 calendar days depending on the start and/or the end of the corresponding time interval within the specific calendar year.	The term “year” applies also to the duration of any time interval which starts at a certain time of day at a certain calendar date of the calendar year and ends at the same time of day at the same calendar date of the next calendar year, if it exists. In other cases, the ending calendar day has to be agreed on.
Common year	Calendar year in the Gregorian calendar that has 365 calendar days.	
Leap year	Calendar year in the Gregorian calendar that has 366 days.	

The standard also defines multiple formats for the representation of date and time as well as time intervals and durations. ISO 8601 postulates that duration can be expressed by a combination of components with accurate duration (hour, minute, and second) and components with nominal duration (year, month, week, and day). The standard allows for the omission of lower-level components for “reduced accuracy” applications.

Appendix F: Time Interval Calculation Conventions

The unit in which a time interval (or its duration) is expressed may depend on the necessary level of accuracy for the purposes of measurement. As such, the selection of the time unit to be used should be made according to the level of granularity needed to meet the intent of the measure.

The conventions listed in Table 33: Time Interval Calculation Conventions are intended to standardize the interpretation of 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."

For the purposes of this document, date 2 is assumed to be more recent than date 1. All dates in this table are represented according to the extended formatted designated in ISO 8601: YYYY-MM-DD HH:MM:SS and a 24-hour timekeeping system. The results of these calculations are always an integer. If the result includes decimals, it should be truncated (rounded down) to the unit.

NOTE: Also reference the section on Age Calculation Conventions and Logic Guidance to understand how these time interval definitions relate to the development of age logic.

Table 33: Time Interval Calculation Conventions

Calculation Unit	Definition	Calculation
Year	Duration of any time interval which starts at a certain time of day at a certain calendar date of the calendar year and ends at: • The same time of day at the same calendar date of the next calendar year, if it exists. • The same time of day at the immediately following calendar date of the ending calendar year, if the same calendar date of the ending calendar year does not exist. • Notes: • When in the next calendar year the same calendar date does not exist, the ISO states that the ending calendar day has to be agreed on. The above convention is proposed as a resolution to this issue. • 2. The ISO permits the representation of dates and times with "reduced accuracy," when necessary. For the purposes of quality measures, duration expressed in years ignores the time of day associated with the date/time stamps used in the calculation, e.g., the number of years will not change until the month and day of date 2 reach (or surpass) the month and day of date 1; time of day should be	Month (date 2) < month (date 1): Duration (years) = year (date 2) – year (date 1) – 1 Example 1: • Date 1: 2012-03-10 22:05:09 • Date 2: 2013-02-18 19:10:03 • Duration = year (date 2) – year (date 1) – 1 = 2013 – 2012 – 1 = 0 years • Month (date 2) = month (date 1) and day (date 2) >= day (date 1) • Duration (years) = year (date 2) – year (date 1) Example 2.a: day (date 1) = day (date 2) • Date 1: 2012-03-10 22:05:09 • Date 2: 2013-03-10 08:01:59 • Duration = year (date 2) – year (date 1) = 2013 – 2012 = 1 year Note: Time of day is ignored in this calculation. If time of day were to be considered for the calculation, the duration of the time interval would be 0 years according to the definition. Example 2.b: day (date 2) > day (date 1) • Date 1: 2012-03-10 22:05:09 • Date 2: 2013-03-20 04:01:30 • Duration = year (date 2) – year (date 1) = 2013 – 2012 = 1 year

Calculation Unit	Definition	Calculation
	ignored or set to 00:00:00 for this calculation.	• Month (date 2) = month (date 1) and day (date 2) < day (date 1) • Duration (years) = year (date 2) – year (date 1) – 1 Example 3.a: • Date 1: 2012-02-29 • Date 2: 2014-02-28 • Duration = year (date 2) – year (date 1) – 1 = 2014 – 2012 – 1 = 1 year • Month (date 2) > month (date 1) • Duration (years) = year (date 2) – year (date 1) Example 4.a: • Date 1: 2012-03-10 11:16:02 • Date 2: 2013-08-15 21:34:16 • Duration = year (date 2) – year (date 1) = 2013 – 2012 = 1 year Example 4.b: • Date 1: 2012-02-29 10:18:56 • Date 2: 2014-03-01 19:02:34 • Duration = year (date 2) – year (date 1) = 2014 – 2012 = 2 years Note: Because there is no February 29 in 2014, the number of years can only change when the date reaches March 1, the first date in 2014 that surpasses the month and day of date 1 (February 29).
Month	Duration of any time interval which starts at a certain time of day at a certain calendar day of the calendar month and ends at: • The same time of day at the same calendar day of the ending calendar month, if it exists. • The same time of day at the immediately following calendar date of the ending calendar month, if the same calendar date of the ending month year does not exist. Notes: • When in the next calendar year the same calendar date does not exist, the ISO states that the ending calendar day has to be agreed on. The above convention is proposed as a resolution to this issue. • The ISO permits the representation of dates and times with "reduced accuracy," when necessary. For the purposes of quality measures, duration expressed in months ignores the time of day associated	Day (date 2) >= day (date 1) Duration (months) = (year (date 2) – year (date 1))*12 + (month (date 2) – (month date 1)) Example 1.a: • Date 1: 2012-03-01 14:05:45 • Date 2: 2012-03-31 23:01:49 • Duration = (year (date 2) – year (date 1))*12 + (month (date 2) – (month date 1)) • = (2012 – 2012)*12 + (3 – 3) = 0 months Example 1.b: • Date 1: 2012-03-10 22:05:09 • Date 2: 2013-06-30 13:00:23 • Duration = (year (date 2) – year (date 1))*12 + (month (date 2) – (month date 1)) • = (2013 – 2012)*12 + (6 – 3) = 12 + 3 = 15 months • Day (date 2) < day (date 1)

Calculation Unit	Definition	Calculation
	with the date/time stamps used in the calculation, e.g., the number of months will not change until the month and day of date 2 reach (or surpass) the month and day of date 1; time of day should be ignored or set to 00:00:00 for this calculation.	• Duration (months) = (year (date 2) – year (date 1))*12 + (month (date 2) – month (date 1)) – 1 Example 2: • Date 1: 2012-03-10 22:05:09 • Date 2: 2013-01-09 07:19:33 • Duration = (year(date 2) – year(date 1))*12 + (month (date 2) – (month date 1)) – 1 • = (2013 – 2012)*12 + (1 – 3) - 1 = 12 – 2 - 1 = 9 months
Week	Duration of any time interval which starts at a certain time of day at a certain calendar day at a certain calendar week and ends at the same time of day at the same calendar day of the ending calendar week. Notes: • The ISO permits the representation of dates and times with "reduced accuracy," when necessary. For the purposes of quality measures, duration expressed in weeks ignores the time of day associated with the date/time stamps used in the calculation, e.g., a complete week is always seven days long: a week has passed when the same weekday in the following week is reached; time of day should be ignored or set to 00:00:00 for this calculation.	Duration[296] = [date 2 – date 1 (days)]/7 Example 1: • Date 1: 2012-03-10 22:05:09 • Date 2: 2012-03-20 07:19:33 • Duration = [# days (month (date 1)) – day (date 1) + # days (month (date 1) + 1) + # days (month (date 1) + 2) + … + # days (month (date 2) – 1) + day (date 2)]/7 • = (20 – 10)/7 = 10/7 = 1 week (result truncated to integer).
Day	Duration of any time interval which starts at a certain calendar day and ends at the next calendar day (1 second to 47 hours, 59 minutes and 59 seconds); the number of times midnight is crossed. Notes: • The ISO permits the representation of dates and times with "reduced accuracy," when necessary. For the purposes of quality measures, duration expressed in days ignores the time of day associated with the date/time stamps used in the calculation. Time of day should be ignored or set to 00:00:00 for this calculation.	The duration in days between two dates, will be generally given by subtracting the start calendar date to the end calendar date, regardless of the time of day between the two dates. Example 1: • Date 1: 2012-01-31 12:30:00 • Date 2: 2012-02-01 09:00:00 • Duration = 02-01 – 01-31 = 1 day Example 2: • Date 1: 2012-01-31 12:30:00 • Date 2: 2012-02-01 14:00:00 • Duration = 02-01 – 01-31 = 1 day
Hours	The number of 60-minute cycles between two given dates. Notes:	The duration in hours between two dates is the number of minutes between the two dates, divided by 60. The result is truncated to the unit.

[296] For information on how to calculate the duration, in days, of a time interval, please refer to the "day" calculation unit in this table.

Calculation Unit	Definition	Calculation
	• Use the date, hour, and minute of the date/time stamps to compute the interval. • Seconds are not used in the calculation.	Example 1: • Date/Time 1: 2012-03-01 03:10 • Date/Time 2: 2012-03-01 05:09 • Duration = 1 hour Example 2: • Date/Time 1: 2012-02-29 23:10 • Date/Time 2: 2012-03-01 00:10 • Duration = 1 hour Example 3: • Date/Time 1: 2012-03-01 03:10 • Date/Time 2: 2012-03-01 04:00 • Duration = 0 hours
Minutes	The number of minutes between two given dates. Notes: • Seconds are not used in the calculation.	Example 1: • Date/Time 1: 2012-03-01 03:10 • Date/Time 2: 2012-03-01 05:20 • Duration = 130 minutes Example 2: • Date/Time 1: 2012-02-29 23:10 • Date/Time 2: 2012-03-01 00:20 • Duration = 70 minutes

Timing relationships with no calculation unit defined

When no threshold is defined in a measure phrase that compares timing between two elements, a common level of granularity needs to be established for comparisons within and across EHRs. For instance, the criterion:

A starts before start of B

Depending on the level of granularity of the data (e.g., HH:MM:SS vs. HH:MM), the result of the timing comparison may vary if no comparison unit is defined:

Example 1

- A: 2012-01-01 11:00:01
- B: 2012-01-01 11:00:02
- A starts before B = TRUE

Example 2

- A: 2012-01-01 11:00
- B: 2012-01-01 11:00
- A starts before B = FALSE

To resolve this type of issue, calculations should be performed with dates represented with reduced accuracy to the minute (date/time truncated to minute prior to calculation, i.e., seconds not used in calculation):

Example 3

- A: 2012-01-01 11:00:01
- B: 2012-01-01 11:00:02
- A starts before B = FALSE

Note: Seconds are not used in the calculation.

Example 4

- A: 2012-01-01 11:00
- B: 2012-01-01 11:00
- A starts before B = FALSE

If no unit of comparison has been defined in the logic (e.g., A starts before start of B), minutes will be used as the default unit for the calculation. This convention effectively renders the following criteria equivalent:

- A starts before start of B
- A > 0 minute(s) starts before start of B
- A > 1 minute(s) starts before start of B

If another unit is intended, it should be specified in the criterion, for instance:

- If the intended unit of comparison is the second:
- A starts before start of B would “default” to comparison in minutes.
- A > 1 second(s) starts before start of B would explicitly refer the granularity intended for the calculation of duration between the two date/time stamps (criterion would be true if A started at least one second before B).
- If the intended unit of comparison is the day:
- A > 1 day(s) starts before start of B (criterion would be true if A started at least one day before B).

The best practice is to always explicitly define the unit of calculation, as this shows intent on the part of the measure developer and provides a completely unambiguous interpretation of the level of granularity with which the computation should be performed. However, if no unit is defined, the above convention determines that the computation unit should be minutes.

Other date/time-related calculations using QDM functions

The above definitions are focused on the calculation of durations. However, in certain applications, it may be necessary to compare portions of the date/time elements directly, e.g., comparing years or months. This can be achieved by using existing QDM/MAT functions that allow the extraction of specific portions of date/time elements. A complete list of these functions is available in the Measure Authoring Tool User Guide.[297]

[297] Department of Health and Human Services. Centers for Medicare & Medicaid Services. *Measure Authoring Tool: The Measure Authoring Tool User Guide*. User Guide can be accessed via the "Training and Resources" tab, by clicking the "User Guide" link. Available at: https://www.emeasuretool.cms.gov/. Accessed on: June 1, 2015.

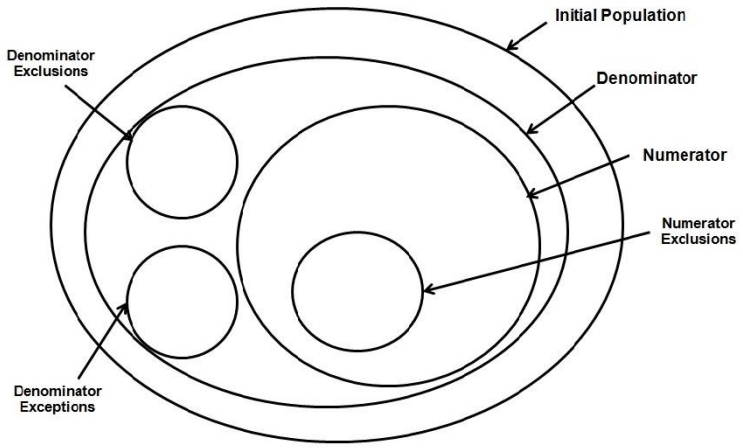
Appendix G: Proportion Measure Calculations (exclusion, exception, and aggregation)

This appendix provides further guidance on the precise mathematical relationships between populations in a proportion 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 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 40: Proportion Measure Populations shows a fixed mathematical relationship between the populations in a proportion measure. The definitions of the various populations in a proportion measure are listed in Table 14: eCQM Metadata. Note that the numerator exclusion population for a proportion measure is not supported by the current MAT.

Figure 40: Proportion Measure Populations



From these relationships and definitions, we define the sequential steps to be used when determining whether a patient falls into a given population:

1. Initial population (IPOP): Identify those patients who meet the IPOP criteria.
2. Denominator (DENOM): Identify that subset of the IPOP that meet the DENOM criteria.
3. Denominator exclusion (DENEX): Identify that subset of the DENOM that meet the DENEX criteria. There are patients who should be removed from the denominator as exclusion. Once these patients are removed, the subset remaining would reflect the denominator per criteria.
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. Numerator Exclusion is used only in ratio eMeasures 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 "LDL Cholesterol is less than 100" and there is no LDL Cholesterol result in the EHR, 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 EHR, then because there is positive evidence of an ejection fraction less than 40, the criterion is met.[298]

Supplemental Data Elements

CMS defines Supplemental Data Elements (variables) which are 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. CMS specifies that sex, race, ethnicity, and payer are the four supplemental data elements that must be defined in each of the Meaningful Use Stage 2 measures. Different supplemental data elements can be indicated as mandatory according to different program requirements. For example, TIN, NPI, CCN, and beneficiary claim number (HIC) may be requested as supplemental data elements with additional guidance that will be provided by the program.

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 in a facility; stratify and report separately by age group [14-19, 20-25, and total 14-25]). The Reporting Stratification section is always included in an eMeasure 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 its own heading.

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 3 numerators, and the patient is included in the first numerator, the patient should be scored for inclusion/exclusion from the populations related to the other numerators as well. When the measure definition includes stratification, each population in the measure definition should be reported both *without*

[298] Many measures will be more specific with respect to which observation to use when comparing against a criterion (such as "**LAST** ejection fraction is less than 40").

stratification, and *stratified* by each stratification criteria. Specific programs may require reporting of performance rates. The performance rate is defined as: Rate = N/(D – DExclusion – DException)

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).

- Null Scores: If a measure within a measures group is not applicable to a patient, the patient would not be counted in the performance denominator for that measure (e.g., Preventive Care Measures Group - Measure #39: Screening or Therapy for Osteoporosis for Women would not be applicable to male patients according to the patient sample criteria). 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 eligible professional submits all performance exclusion quality data codes, the performance rate would be 0/0 (null) and would be considered satisfactorily reporting.

Proportion measure example #1

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

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, with 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, with 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, with 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.

Proportion measure example #2

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

A fictitious 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, etc.

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)

From the example above, 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 also meet DENEX).
- Numerator: N=75 (meet DENOM, not in DENEX, and also meet NUMER criteria).
- Denominator exception: N=5 (meet DENOM, not in DENEX, not in NUMER, and also meet the DENEXCEP criteria).
- Performance rate = (NUMER – NUMEX) / (DENOM – DENEXCEP) = (75-0) / (150-20-5) = 0.6

Additional information about emeasures and proportion measure calculation can be found at the electronic clinical quality improvement (eCQI) resource center at https://ecqi.healthit.gov/.

Appendix H: Continuous Variable and Ratio Measure Calculations

This appendix 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.

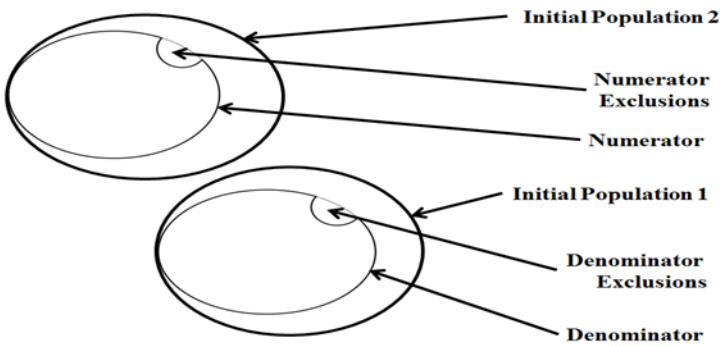
Measure Populations

There is a fixed mathematical relationship between the populations in a ratio or CV measure, as shown in Figure 41: Ratio Measure Populations and Ratio Measures are often two continuous variable calculations for related populations. For example the median Emergency Department waiting time for the index hospital as the numerator, and the median Emergency Department waiting time for the region in which the hospital is located. Thus, the ratio numerator may actually refer to a “Measure Observation” and a Measure Population. The same applies to the ratio denominator (but for a different population).

Figure 42: Continuous Variable Measure Populations. The definitions of the various populations in a ratio and a CV measure are listed in Table 14: eCQM Metadata.

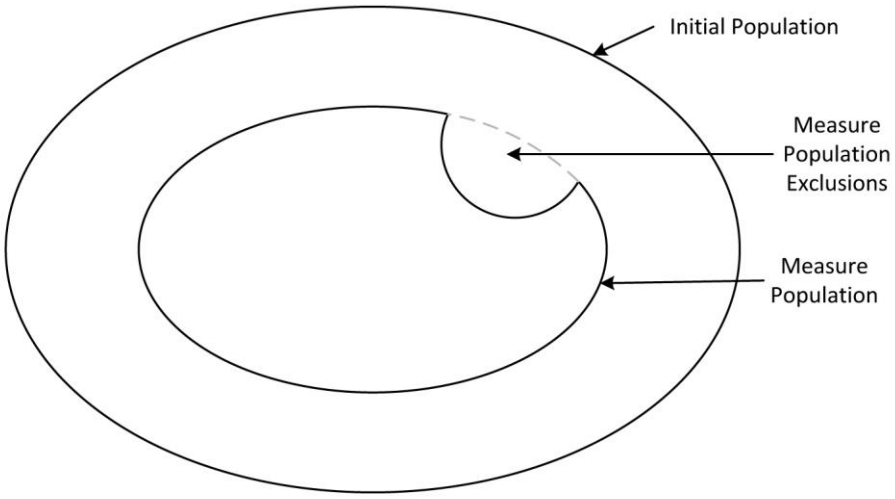
NOTE: The Measure Population Exclusion population for a CV measure is now supported by the July 2014 MAT release.

Figure 41: Ratio Measure Populations



Ratio Measures are often two continuous variable calculations for related populations. For example the median Emergency Department waiting time for the index hospital as the numerator, and the median Emergency Department waiting time for the region in which the hospital is located. Thus, the ratio numerator may actually refer to a "Measure Observation" and a Measure Population. The same applies to the ratio denominator (but for a different population).

Figure 42: Continuous Variable Measure Populations



From these relationships and definitions, we define the ratio measure query process:

- Initial population (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.)
- Denominator (DENOM): Identify that subset of the IPOP that meet the DENOM criteria.
- Denominator exclusion (DENEX): Identify that subset of the DENOM that meet the DENEX criteria.
- Numerator (NUMER): Identify that subset of the IPOP that meet the NUMER criteria.
- Numerator exclusion (NUMEX): Identify that subset of the NUMER that meet the NUMEX criteria.

And we define the continuous variable measure query process:

- Initial population (IPOP): Identify those cases that meet the IPOP criteria.
- Measure population (MSRPOPL): Identify that subset of the IPOP that meet the MSRPOPL criteria.
- Measure population exclusion (MSRPOPLEX): Identify that 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 EHR, 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 EHR, then because there is positive evidence of an ejection fraction less than 40, the criterion is met.[299]

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 be measured for all patients meeting the measure population criteria, and might require that for the measure population as a whole, average wait time be calculated. It is important to note that observations on the individual can be sent in a QRDA Category I, whereas aggregate calculations can be sent in a QRDA Category III.

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 continuous variable measure, an individual observation is determined for each case in the MSRPOPL and not in the MSRPOPLEX.

For a continuous variable 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 continuous variable 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:[300]

- 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.

[299] Many measures will be more specific with respect to which observation to use when comparing against a criterion (such as "**LAST** ejection fraction is less than 40").

[300] Individual calculations can be sent in a QRDA Category I, whereas aggregate calculations can be sent in a QRDA Category III.

Ratio = aggregate NUMER / aggregate DENOM

In continuous variable measures, an individual observation is made for each case falling into the measure population. These individual observations are then used to calculate the aggregate continuous variable, 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.

Ratio Measure Example

Note: This is an example of the mathematical relationships between populations in a proportion 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 (specifically, the ratio of number of central line infections per total central line days) defines the following population criteria:

- IPOP: All patients aged 65 years and older admitted to hospital.
- DENOM: Presence of central line.
- DENEX: Patient is immunosuppressed.
- NUMER: Central line blood stream infection.
- NUMEX: Infection is deemed to be a contaminant.

Individual observations:

- DENOM observation: # 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 # of hospital days with central line.
- NUMER aggregation: Count of NUMER – NUMEX.

Determination of Measure individual observations:

Mr. Jones is 75 years old, 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, 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. 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

From the example above, taking into account 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
- DENEX: N = 2
- NUMER: N = 6
- 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.

Continuous Variable Measure Example

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

A fictitious CV measure (based on NQF 0495) defines the following population criteria:

- IPOP: All patients aged 65 years and older admitted to hospital.
- MSRPOPL: Admitted to hospital from ED.
- MSRPOPLEX: None.
- Individual observations: None.
- MSRPOPL observation: # hours in ED.

Aggregation rules:

Aggregate MSRPOPL: Median hours in ED.

Determination of Measure individual observations:

Mr. Jones is 75 years old, admitted to the hospital from the ED. He spent 1.5 hours 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:
- MSRPOPL observation: 1.5 hours.

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:
- MSRPOPL observation: N/A (Mr. James does not meet the MSRPOPL criteria).

Measure aggregate calculations:

From the example above, taking into account 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: 1.6 hours (median of all individual observations of time spent in the ED).

Appendix I: Harmonization of Inpatient Encounter Value Sets

The following table lists value sets that were harmonized into a single value set with OID 2.16.840.1.113883.3.666.5.307 during the 2014 Eligible Hospital annual update.

Table 34: Harmonization of Inpatient Encounter Value Sets

Description	Encounter Inpatient 2.16.840.1.113883.3.666.5.307	Encounter Inpatient 2.16.840.1.113883.3.666.5.3001	Hospital measures-Acute care hospital encounter 2.16.840.1.113883.3.666.5.2289	Hospital Measures-Encounter Inpatient 2.16.840.1.113883.3.666.5.625	Inpatient Encounter 2.16.840.1.113883.3.117.1.7.1.23	Non-Elective Inpatient Encounter 2.16.840.1.113883.3.117.1.7.1.424
Hospital admission, emergency, from emergency room, accidental injury (procedure)					X	X
Hospital admission, elective, with complete pre-admission work-up (procedure)	X	X	X	X	X	
Hospital admission, boarder, for social reasons (procedure)				X		
Hospital admission for isolation (procedure)	X		X	X		
Hospital admission, elective, with partial pre-admission work-up (procedure)	X	X	X	X	X	
Hospital admission, emergency, indirect (procedure)					X	X
Admission to cardiac intensive care unit (procedure)			X			
Admission to respiratory intensive care unit (procedure)			X			
Admission to neurological intensive care unit (procedure)			X			
Admission to metabolic intensive care unit (procedure)			X			
Emergency hospital admission (procedure)					X	X
Non-urgent hospital admission (procedure)	X	X	X	X	X	X
Non-urgent trauma admission (procedure)	X		X	X		
Hospital admission, emergency, from emergency room, medical nature (procedure)					X	X
Hospital admission, urgent, 48 hours (procedure)	X	X	X	X	X	X
Hospital admission, for research investigation (procedure)	X	X	X	X		
Hospital admission, under police custody (procedure)	X	X	X	X	X	X
Hospital patient (finding)					X	X
Hospital admission, type unclassified, explain by report (procedure)	X	X	X	X	X	X
Admission for respite care (procedure)	X			X		
Admission to establishment (procedure)	X		X	X		
Admission to community hospital (procedure)	X	X	X	X	X	X
Admission to GP hospital (procedure)	X	X	X	X	X	X

Description	Encounter Inpatient 2.16.840.1.113883.3.666.5.307	Encounter Inpatient 2.16.840.1.113883.3.666.5.3001	Hospital measures-Acute care hospital encounter 2.16.840.1.113883.3.666.5.2289	Hospital Measures-Encounter Inpatient 2.16.840.1.113883.3.666.5.625	Inpatient Encounter 2.16.840.1.113883.3.117.1.7.1.23	Non-Elective Inpatient Encounter 2.16.840.1.113883.3.117.1.7.1.424
Admission to private hospital (procedure)	X	X	X	X	X	X
Admission to long stay hospital (procedure)	X			X	X	X
Admission to tertiary referral hospital (procedure)	X	X	X	X	X	X
Admission to ward (procedure)	X	X	X	X		
Admission to day ward (procedure)				X		
Admission to day hospital (procedure)				X		
Admission to psychiatric day hospital (procedure)				X		
Admission to psychogeriatric day hospital (procedure)				X		
Admission to elderly severely mentally ill day hospital (procedure)				X		
Admission to care of the elderly day hospital (procedure)				X		
Admission to intensive care unit (procedure)			X			
Admission to adult intensive care unit (procedure)			X			
Inpatient stay (finding)					X	X
Hospital admission (procedure)	X	X	X	X	X	X
Hospital admission, pre-nursing home placement (procedure)	X		X	X		
Unexpected admission to intensive care unit (procedure)			X			
General outpatient clinic admission (procedure)				X		
Unexpected hospital admission (procedure)	X	X	X	X	X	X
Hospital re-admission (procedure)	X	X	X	X	X	X
Hospital admission, transfer from other hospital or healthcare facility (procedure)	X	X		X	X	X
Hospital admission, pre-certified by medical audit action (procedure)	X	X	X	X		
Hospital admission, mother, for observation, delivered outside of hospital (procedure)	X		X	X		
Hospital admission, special (procedure)	X	X	X	X	X	X
Hospital admission, short-term (procedure)	X	X	X	X	X	X
Hospital admission, donor for transplant organ (procedure)	X			X		
Hospital admission, parent, for in-hospital child care (procedure)				X		
Specialty clinic admission (procedure)				X		
Hospital admission, involuntary (procedure)	X	X	X	X	X	X
Hospital admission, for laboratory work-up, radiography, etc. (procedure)	X			X		
Hospital admission, infant, for observation, delivered outside of hospital (procedure)				X		

Description	Encounter Inpatient 2.16.840.1.113883.3.666.5.307	Encounter Inpatient 2.16.840.1.113883.3.666.5.3001	Hospital measures-Acute care hospital encounter 2.16.840.1.113883.3.666.5.2289	Hospital Measures-Encounter Inpatient 2.16.840.1.113883.3.666.5.625	Inpatient Encounter 2.16.840.1.113883.3.117.1.7.1.23	Non-Elective Inpatient Encounter 2.16.840.1.113883.3.117.1.7.1.424
Hospital admission, from remote area, by means of special transportation (procedure)	X	X	X	X	X	X
Hospital admission, short-term, day care (procedure)				X		
Hospital admission, by legal authority (commitment) (procedure)	X			X	X	X
Hospital admission, limited to designated procedures (procedure)	X	X		X		
Hospital admission, emergency, from emergency room (procedure)					X	X
Hospital admission, short-term, 24 hours (procedure)	X	X		X	X	X
Hospital admission, for observation (procedure)	X	X		X		
Hospital admission, emergency, direct (procedure)					X	X
Hospital admission, elective, without pre-admission work-up (procedure)	X	X	X	X	X	
Hospital admission, blood donor (procedure)	X			X		
Hospital admission, elective (procedure)	X	X	X	X	X	

Appendix J: Summary of Changes to Blueprint, version 11.2

This appendix presents a high-level summary of the changes found in the CMS MMS Blueprint, version 11.2. These changes are arranged by Section, Chapter and page number.

Introduction and Overall

- Pages (throughout), Modified some of the terms that establish a measure's definition according to the semantics established in HQMF. Notably, "exclusions" became "exclusion"; "exceptions" became "exception"; and "observations" became "observation", when used in the context of describing measure specifications.

Section 1 Chapter 1

- Pages 25-27, Included communication protocol among CORs, measures manager and measure contractors.
- Pages 27, Further defined measure governance roles.

Section 1 Chapter 3

- Pages 36-46, Expanded language on person and family (caregiver) participation and engagement.
- Page 50, Added language about using Federal Register Notices for public comment.

Section 1 Chapter 5

- Page 63-68, Added language regarding the development of both an initial and final business case.

Section 1 Chapter 6

- Pages 73-77, Included discussion on measure classification by domain, NQS priority and setting.
- Pages 79-85, Added clarifying language on composite performance meaures, patient-reported outcome measures and multiple chronic condition measures.

Section 2 Overall

- Pages (throughout), Included descriptions for how many steps and substeps were included in each chapter

Section 2 Chapter 1

- Pages 92-93, Added timeline to existing measure lifecycle figure as well as additional measure timeline figure

Section 2 Chapter 1.A

- Page 95, Added table of related and competing measures.
- Page 97, Added expanded language on web posting timeframes and deadlines.
- Pages 97-103, Included extensive new detail and template for environmental scans.

Section 2 Chapter 1.C

- Page 130, Removed the sentence: "A template for this report is available but is not required."

Section 2 Chapter 2

- Pages (throughout), Standardized semantics for measure title lead-ins, descriptions, numerator and denominator statements, age description criteria and numerator and denominator inclusion and exclusion criteria.

Section 2 Chapter 4

- Page 209-210, Deleted discussion of NQF's now-defunct time-limited endorsement status and included discussion of NQF's trial use approval status for eMeasures.

Appendix J

- Page 402-404, Revised to include changes for version 11.2.

Templates Public Comment Summary Web Posting

- Page 475, Deleted reference to inclusion of documents for comment, since comments have already been received at the time of the posting of the summary.

Templates TEP Call for TEP Web Page Posting

- Pages 479, Added expiration notice.

Templates TEP Charter

- Page 488, Deleted expiration notice.

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 Name:

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.

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.[301]

Incidence and prevalence:

The purpose of this item 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

[301] Concepts incorporated from Business Case sample prepared by FMQAI for Glycemic Control measures in 2014.

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.

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.[302]

Major clinical, functional status and other benefits:

This discussion of anticipated benefits may come primarily from the literature. It should include estimates of cost savings to Medicare, but also to patients and caregivers. Costs of treatment may be offset by reduced morbidity and mortality, improved functional status, improved quality of life, and prevention of lost wages.[303]

Successful intervention is feasible:

Present the evidence that not only is there a quality gap but that improvement is possible and feasible. This is not limited to feasibility of data collection according to measure specifications, but that changes are feasible based on the actions of those being held responsible.[304]

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.[305]

Complications or untoward effects:

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.[306]

---

[302] Concepts incorporated from Business Case sample prepared by Yale for THA/TKA resource use measure in 2014.
[303] Concepts incorporated from Business Case sample prepared by QP Rhode Island for Antipsychotic Polypharmacy and Warfarin Monitoring in 2009.
[304] Ibid
[305] Concepts incorporated from Business Case sample prepared by FMQAI for Glycemic Control measures in 2014.
[306] Ibid.

Estimated size and trajectory of anticipated improvement:

Estimate the expected performance of the measure. 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.[307]

Model and formulas used:

Describe the assumptions, variables, and formulas used to construct the business case.[308]

Limitations:

Describe any limitations in the data or the assumptions used in the business case.[309]

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.[310]

[307] Concepts incorporated from Business Case sample prepared by QP Rhode Island for Antipsychotic Polypharmacy and Warfarin Monitoring in 2009.
[308] Concepts incorporated from Business Case sample prepared by FMQAI for Glycemic Control measures in 2014.
[309] Ibid.
[310] 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.

Blank Business Case 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:

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.

Summary:

Incidence and prevalence:

Current performance, including any disparities:

Major clinical, functional status and other benefits:

Successful intervention is feasible:

Costs of clinical care:

Complications or untoward effects:

Estimated size and trajectory of anticipated improvement:

Model and formulas used:

Limitations:

Net benefit:

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.

Measure Information Form Instructions

This form is a guide for your use when submitting measures to the National Quality Forum (NQF). 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 6.5, and references corresponding fields from that submission form. If you do not already have an NQF login, you will be prompted to enroll. The NQF Submitting Standards webpage contains a link to the submission form that will take you to the start of the NQF measure submission process.[311] Your submission can be saved, closed, and reopened for editing at any time prior to final submission. After you press the submit button on the NQF website form, the measure can no longer be edited.

The NQF online measure submission includes nine tabs that can be navigated in any order. The tabs are divided into two sections for Blueprint purposes—the Measure Information Form and the Measure Justification Form. All portions of the online submission are required for a measure to be considered for endorsement by NQF. Information from the Measure Information Form and the Measure Justification Form may be used for other purposes, so developers may be asked to complete the Measure Information Form 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 three years.
C. The intended use of the measure includes both accountability applications (including public reporting) and performance improvement to achieve high-quality, efficient healthcare.

[311] http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx

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. Patient-reported outcomes include health-related quality of life, functional status, symptom burden, experience with care, and health-related behavior.

- Composite
- Cost-resource use
- Efficiency
- Outcome
- Patient-reported outcome (PRO)
- Process
- Structure

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.

Subject/Topic Areas De.5.

Document all of the topic areas that apply from the following choices.

- Behavioral Health: Behavioral Health

- Behavioral Health: Alcohol, Substance Use/Abuse
- Behavioral Health: Attention Deficit Hyperactivity Disorder (ADHD)
- Behavioral Health: Depression
- Behavioral Health: Post-Traumatic Stress Disorder (PTSD)
- Behavioral Health: Screening
- Behavioral Health: Serious Mental Illness
- Behavioral Health: Suicide
- Behavioral Health: Tobacco Use
- Cancer: Cancer
- Cancer: Bladder
- Cancer: Breast
- Cancer: Colorectal
- Cancer: Gynecologic
- Cancer: Hematologic
- Cancer: Liver
- Cancer: Lung, Esophageal
- Cancer: Pancreatic
- Cancer: Prostate
- Cancer: Screening
- Cancer: Skin
- Cardiovascular: Cardiovascular
- Cardiovascular: Acute Myocardial Infarction
- Cardiovascular: Atrial Fibrillation
- Cardiovascular: Congestive Heart Failure
- Cardiovascular: Hyperlipidemia
- Cardiovascular: Hypertension
- Cardiovascular: Ischemic Heart Disease, Coronary Artery Disease
- Cardiovascular: Percutaneous Coronary Intervention (PCI)
- Cardiovascular: Screening
- Endocrine: Endocrine
- Endocrine: Diabetes
- Endocrine: Screening
- Endocrine: Thyroid Disorders
- Gastrointestinal (GI): Gastrointestinal (GI)
- Gastrointestinal (GI): Appendicitis
- Gastrointestinal (GI): Cirrhosis
- Gastrointestinal (GI): GI Bleeding
- Gastrointestinal (GI): Gall Bladder Disease
- Gastrointestinal (GI): Gastroenteritis
- Gastrointestinal (GI): Gastro-Esophageal Reflux Disease (GERD)
- Gastrointestinal (GI): Polyps
- Gastrointestinal (GI): Screening

- Gastrointestinal (GI): Peptic Ulcer
- GU/GYN: GU/GYN
- GU/GYN: Incontinence
- GU/GYN: Screening
- Head, Eyes, Ears, Nose, Throat (HEENT): Head, Eyes, Ears, Nose, Throat (HEENT)
- Head, Eyes, Ears, Nose, Throat (HEENT): Dental
- Head, Eyes, Ears, Nose, Throat (HEENT): Ear Infection
- Head, Eyes, Ears, Nose, Throat (HEENT): Hearing
- Head, Eyes, Ears, Nose, Throat (HEENT): Pharyngitis
- Head, Eyes, Ears, Nose, Throat (HEENT): Screening
- Head, Eyes, Ears, Nose, Throat (HEENT): Vision
- Infectious Diseases: Infectious Diseases
- Infectious Diseases: Hepatitis
- Infectious Diseases: Human Immunodeficiency Virus/Acquired Immune Deficiency Syndrome (HIV/AIDS)
- Infectious Diseases: Immunization
- Infectious Diseases: Respiratory
- Infectious Diseases: Sexually Transmitted
- Infectious Diseases: Tuberculosis
- Infectious Diseases: Screening
- Mental Health: Mental Health
- Mental Health: Alcohol, Substance Use/Abuse
- Mental Health: Depression
- Mental Health: Domestic Violence
- Mental Health: Serious Mental Illness
- Mental Health: Suicide
- Musculoskeletal: Musculoskeletal
- Musculoskeletal: Osteoarthritis
- Musculoskeletal: Rheumatoid Arthritis
- Musculoskeletal: Hip/Pelvic Fracture
- Musculoskeletal: Joint Surgery
- Musculoskeletal: Low Back Pain
- Musculoskeletal: Osteoporosis
- Neurology: Neurology
- Neurology: Brain Injury
- Neurology: Cognitive Impairment/Dementia
- Neurology: Delirium
- Neurology: Stroke/Transient Ischemic Attack (TIA)
- Perinatal and Reproductive Health: Perinatal and Reproductive Health
- Perinatal and Reproductive Health: Gynecology
- Perinatal and Reproductive Health: Newborn
- Perinatal and Reproductive Health: Perinatal

- Perinatal and Reproductive Health: Screening
- Prevention: Prevention
- Prevention: Development/Wellness
- Prevention: Immunization
- Prevention: Malnutrition
- Prevention: Obesity
- Prevention: Physical Activity
- Prevention: Screening
- Prevention: Tobacco Use
- Pulmonary/Critical Care: Pulmonary/Critical Care
- Pulmonary/Critical Care: Asthma
- Pulmonary/Critical Care: Chronic Obstructive Pulmonary Disease (COPD)
- Pulmonary/Critical Care: Critical Care
- Pulmonary/Critical Care: Dyspnea
- Pulmonary/Critical Care: Pneumonia
- Pulmonary/Critical Care: Sleep Apnea
- Renal: Renal
- Renal: Chronic Kidney Disease (CKD)
- Renal: End Stage Renal Disease (ESRD)
- Surgery: Surgery
- Surgery: Cardiac Surgery
- Surgery: General Surgery
- Surgery: Perioperative
- Surgery: Thoracic Surgery
- Surgery: Vascular Surgery

Crosscutting Areas De.6.

Again, document all that apply from the following choices:

- Health and Functional Status: Health and Functional Status
- Health and Functional Status: Development/Wellness
- Health and Functional Status: Functional Status
- Prevention: Prevention
- Prevention: Immunization
- Prevention: Nutrition
- Prevention: Obesity
- Prevention: Physical Activity
- Prevention: Screening
- Prevention: Social Determinants
- Access
- Care Coordination: Care Coordination
- Care Coordination: Readmissions

- Disparities
- Functional Status
- Overuse
- Palliative Care and End of Life Care
- Patient and Family (caregiver) Engagement
- Safety: Safety
- Safety: Complications
- Safety: Healthcare Associated Infections
- Safety: Medication Safety
- Safety: Venous Thromboembolism
- Safety: Readmissions
- Safety: Workforce

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 Measure Authoring Tool (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 Measure Information Form 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.

Time Period for Data S.5.

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. Note if there are different time periods for the numerator and denominator.

Numerator Details S.6.

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.

Denominator Statement S.7.

Give a narrative description of the broadest population (based on the evidence) for which the target process, condition, event, or outcome is applicable. For example:

> Patient's [age] with [condition] in [setting] during [time frame]

Target Population Category S.8.

Identify all that apply.

- Children's Health
- Maternal Health
- Populations at Risk: Populations at Risk
- Populations at Risk: Dual-eligible beneficiaries
- Populations at Risk: Individuals with multiple chronic conditions
- Populations at Risk: Veterans
- Senior Care

Denominator Details S.9.

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.10.

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.11.

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.12.

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.13.

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.)

Statistical Risk Model and Variables S.14.

Name the statistical method, and list all the risk factor variables. Risk model development should be addressed with measure testing under Scientific Acceptability.

Detailed Risk Model Specifications S.15.

Provide these in an attached data dictionary/code list in an .xls or .csv file in the required format listed in S.2b. Also indicate if they are available at the measure-specific URL identified in S.1.

Risk model details (including coefficients, equations, codes with descriptors, and definitions) should be provided on a separate worksheet in an .xls or .csv file (the required format listed in S.2b.) with the data dictionary/code lists mentioned at S.2b.

Type of Score S.16.

- Count
- Rate/proportion
- Ratio
- Categorical (e.g., yes or no)
- Continuous variable (e.g., an average)
- Other (specify)

Interpretation of Score S.17.

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.18.

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.

Calculation Algorithm/Measure Logic Diagram URL or Attachment S.19.

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.20.

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 patient-reported outcome-based performance measure (PRO-PM), identify whether (and how) proxy responses are allowed.

Survey/Patient-Reported Data S.21.

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.

Missing Data S.22.

Specify how missing data are handled (imputation, delete case, etc.). This item is required for composite measures and PRO-PMs.

Data Source S.23.

Indicate all sources for which the measure is specified and tested.

- Administrative Claims
- Electronic Clinical Data: Electronic Clinical Data
- Electronic Clinical Data: Electronic Health Record
- Electronic Clinical Data: Imaging/Diagnostic Study
- Electronic Clinical Data: Laboratory
- Electronic Clinical Data: Pharmacy
- Electronic Clinical Data: Registry
- Healthcare Provider Survey
- Management Data
- Paper Medical Records
- Patient Reported Data/Survey
- Other (describe)

Data Source or Collection Instrument S.24.

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.25.

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.26.

Indicate only the levels for which the measure is specified and tested.

- Clinician: Individual
- Clinician: Group/Practice
- Clinician: Team
- Facility
- Health Plan
- Integrated Delivery System
- Population: Community

- Population: County or City
- Population: National
- Population: Regional
- Population: State

Care Setting S.27.

Indicate only the settings for which the measure is specified and tested.

- Ambulatory Care: Ambulatory Surgery Center
- Ambulatory Care: Clinician Office/Clinic
- Ambulatory Care: Outpatient Rehabilitation
- Ambulatory Care: Urgent Care
- Behavioral Health/Psychiatric: Inpatient
- Behavioral Health/Psychiatric: Outpatient
- Dialysis Facility
- Emergency Medical Services/Ambulance
- Home Health
- Hospice
- Hospital/Acute Care Facility
- Imaging Facility
- Laboratory
- Pharmacy
- Post-Acute/Long Term Care Facility: Nursing Home/Skilled Nursing Facility
- Post-Acute/Long Term Care Facility: Inpatient Rehabilitation Facility
- Post-Acute/Long Term Care Facility: Long Term Acute Care Hospital
- Other (describe)

Composite Performance Measure S.28.

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.

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.
Subject/Topic Areas De.5.
Crosscutting Areas De 6.

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.
Numerator Statement S.4.
Time Period for Data S.5.
Numerator Details S.6.
Denominator Statement S.7.
Target Population Category S.8.
Denominator Details S.9.
Denominator Exclusion (NQF Includes "Exception" in the "Exclusion" Field) S.10.
Denominator Exclusion Details (NQF Includes "Exception" in the "Exclusion" Field) S.11.
Stratification Details/Variables S.12.
Risk Adjustment Type S.13.
Statistical Risk Model and Variables S.14.
Detailed Risk Model Specifications S.15.
Type of Score S.16.
Interpretation of Score S.17.
Calculation Algorithm/Measure Logic S.18.
Calculation Algorithm/Measure Logic Diagram URL or Attachment S.19.
Sampling S.20.
Survey/Patient-Reported Data S.21.
Missing Data S.22.
Data Source S.23.
Data Source or Collection Instrument S.24.
Data Source or Collection Instrument (Reference) S.25.
Level of Analysis S.26.
Care Setting S.27.
Composite Performance Measure S.28.

Measure Justification Form Instructions

This form is a guide for use when submitting measures to the National Quality Forum (NQF). Use it to draft responses for each section in preparation for entering them in the measure submission. The Measure Justification Form tracks very closely to the NQF online measure submission and references corresponding fields from that submission. When all preparatory work has been completed, click the NQF Login link to begin the actual submission process.[312] The submission can be saved, closed, and reopened for editing at any time prior to submission. It also can be printed at any time. Once the submit button on the NQF website form is pressed, however, the measure can no longer be edited.

The NQF online measure submission includes nine tabs that can be navigated in any order. The tabs are divided into two sections for Blueprint purposes—the Measure Information Form and the Measure Justification Form. All portions of the online submission are required for a measure to be considered for endorsement by NQF. The numbers used throughout this form correspond to the same numbered items on the NQF submission.

Measure Name

Give the measure name that was used for the Measure Information Form. 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 Measure Information Form.

- 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.

[312] https://login.qualityforum.org/login.aspx

Opportunity for Improvement (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.

- 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.
- Respond to all questions as instructed with answers immediately following the question. 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 Grading of Recommendations, Assessment, Development, and Evaluation (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; AQA Principles of Efficiency Measures).

1a.1.—This is a Measure of:

- Outcome.
- Health outcome: Name the health outcome. Health outcome includes patient-reported outcomes (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.

Health Outcome Performance Measure (if not a health outcome, skip to 1a.3.)

1a.2.—Linkage

Briefly state or diagram the linkage between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1—Rationale

State the rationale supporting the relationship between the health outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

For health outcome performance measures, no further information is required; however, evidence for any of the structures, processes, interventions, or service identified above may be provided.

Intermediate Outcome, Process, or Structure Performance Measure

1a.3.—Linkage

Briefly state or diagram the linkages between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1.—Source of Systematic Review

What is the source of the systematic review of the body of evidence that supports the performance measure?

- Clinical Practice Guideline recommendation—complete sections 1a.4. and 1a.7.
- U.S. Preventive Services Task Force Recommendation—complete sections 1a.5. and 1a.7.
- Other systematic review and grading of the body of evidence (e.g., Cochrane Collaboration, Agency for Healthcare Research and Quality [AHRQ] Evidence Practice Center)—complete sections 1a.6. and 1a.7.
- Other—complete section 1a.8.

Please complete the sections indicated above for the source of evidence. The sections that do not apply may be skipped.

1a.4.—Clinical Practice Guideline Recommendation

1a.4.1.—Guideline Citation

Guideline citation (including date) and URL for guideline (if available online).

1a.4.2.—Specific Guideline

Identify guideline recommendation number and/or page number and quote verbatim the specific guideline recommendation.

1a.4.3.—Grade

Grade assigned to the quoted recommendation with definition of the grade.

1a.4.4.—Grades and Associated Definitions

Provide all other grades and associated definitions for recommendations in the grading system. If separate grades for the strength of the evidence, report them in section 1a.7.

1a.4.5.—Methodology Citation

Citation and URL for methodology for grading recommendations (if different from 1a.4.1.).

1a.4.6.—Quantity, Quality, and Consistency

If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

- If yes, complete section 1a.7.
- If no, report on another systematic review of the evidence in sections 1a.6. and 1a.7.; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7.

1a.5.—United States Preventive Services Task Force Recommendation

1a.5.1.—Recommendation Citation

Recommendation citation (including date) and URL for recommendation (if available online).

1a.5.2.—Specific Recommendation

Identify recommendation number and/or page number and quote verbatim the specific recommendation.

1a.5.3.—Grade

Grade assigned to the quoted recommendation with definition of the grade.

1a.5.4.—Grades and Associated Definitions

Provide all other grades and associated definitions for recommendations in the grading system. The grading system for the evidence should be reported in section 1a.7.

1a.5.5.—Methodology Citation

Citation and URL for methodology for grading recommendations (if different from 1a.5.1.):

Complete section 1a.7.

1a.6. Other Systematic Review of the Body of Evidence

1a.6.1.—Review Citation

Citation (including date) and URL (if available online).

1a.6.2.—Methodology Citation

Citation and URL for methodology for evidence review and grading (if different from 1a.6.1.):

Complete section 1a.7.

1a.7.—Findings from Systematic Review of Body of the Evidence Supporting the Measure

1a.7.1.—Specifics Addressed in Evidence Review

What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2.—Grade

Grade assigned for the quality of the quoted evidence with definition of the grade.

1a.7.3.—Grades and Associated Definitions

Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4.—Time Period

What is the time period covered by the body of evidence? (Provide the date range, e.g., 1990–2010.)

Quantity and Quality of Body of Evidence

1a.7.5.—Number and Type of Study Designs

How many and what type of study designs are included in the body of evidence (e.g., three randomized controlled trials and one observational study)?

1a.7.6.—Overall Quality of Evidence

What is the overall quality of evidence across studies in the body of evidence? (Discuss the certainty or confidence in the estimates of effect, particularly in relation to study factors such as design flaws, imprecision due to small numbers, or indirectness of studies to the measure focus or target population).

Estimates of Benefit and Consistency Across Studies in Body of Evidence

1a.7.7.—Estimates of Benefit

What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence (e.g., ranges of percentages or odds ratios for improvement/decline across studies, results of meta-analysis, and statistical significance)?

1a.7.8.—Benefits Over Harms

What harms were studied and how do they affect the net benefit?

Update to the Systematic Review(s) of the Body of Evidence

1a.7.9.—Provide for Each New Study

If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: (1) citation, (2) description, (3) results, and (4) impact on conclusions of systematic review.

1a.8.—Other Source of Evidence

If the source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which the performance measure is based.

1a.8.1.—Process Used

What process was used to identify the evidence?

1a.8.2.—Citation

Provide the citation and summary for each piece of evidence.

Evidence to Support Measure Focus (1b.)

1b.1.—Rationale

Briefly explain the rationale for this measure (i.e., the improvements in quality envisioned by use of this measure).

If the measure is a composite, a combination of component measure scores, all-or-none, or any-or-none, skip this question and provide the rationale for the composite in question 1d.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. and 1b.5.—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, socioeconomic status, 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.1.) 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. Include citations of the relevant literature used.

High Priority (Measure Evaluation Criterion 1c.)

1c.1.—Demonstrated High-Priority Aspect of Health Care

- Affects large numbers
- A leading cause of morbidity/mortality
- Frequently performed procedure
- High resource use

- Patient/social consequences of poor quality
- Severity of illness

1c.2.—If "Other" please describe

1c.3.—Epidemiologic or Resource Use Data

Provide epidemiologic or resource use data that demonstrate the measure addresses a high-priority aspect of healthcare. List the citations in 1c.4.

1c.4.—Citations

List citations for data demonstrating high priority provided in 1c.3.

1c.5.—Patient-Reported Outcome Performance Measure (PRO-PM)

If the measure deals with patient-reported health-related quality of life, functional status, symptoms, burden, experience with care, or health-related behaviors, provide evidence that the target population values the measured PRO and finds it meaningful. Describe how and from whom the input was obtained.

1d.1.—Composite Performance Measure

Please identify the composite measure construction

- Two or more individual performance measure scores combined into one score
- All-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient)
- Any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

1d.2.—Composite quality construct

Describe

- The overall area of quality.
- Included component measures.
- The relationship of the component measures to the overall composite and to each other.

1d.3.—Rationale for constructing a composite measure

Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.

1d.4.—Aggregation and weighting

Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.

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 the Measure Evaluation Report. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; and has demonstrated adequate discrimination and calibration. Risk factors that influence outcomes should not be specified as exclusion. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for cardiovascular disease risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to risk adjust the differences away.

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.

Data Sample Description (1.)

This description should be the same as what is submitted with the Measure Information Form. 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 electronic health record
- eMeasure (HQMF) implemented in EHRs
- Other:

Measure Tested With Data From:

- Abstracted from paper record
- Administrative claims
- Clinical database/registry
- Abstracted from electronic health record
- 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?

Exclusion Analysis (2b3)

If there are no exclusion, 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

If the measure is not an intermediate, or health outcome, or patient-reported outcome 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 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).

2b4.4.—Statistical Results

What were the statistical results of the analyses used to select risk factors?

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.

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?

Comparability of Multiple Data Sources/Methods

If there is only one set of specifications, this section can be skipped. This criterion is directed 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). If the data sources or methods are not comparable, then consider submitting the different specifications 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?

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 patient-reported outcome-based performance measure (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 three years and publicly reported within six 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 three years and publicly reported within six 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.1. 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

4b.2. 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.

Unintended Negative Consequences

4c.1. Were any unintended negative consequences to individuals or populations identified during testing, OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.

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.

5.1a. The measure titles and NQF numbers are listed here.

5.1b. If the measures are not NQF-endorsed, indicate the measure title.

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

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. Patient-Reported Outcome Performance Measure (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

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. Figure 8: Applying Measure Evaluation Criteria depicts this process, and Chapter 5—Concepts of 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 National Quality Forum Measure Evaluation Criteria very closely.[313]

Additional instruction for evaluating the specific measure types—Composite Performance Measure Evaluation Guidance,[314] Cost-Resource Use Measures, eMeasures, and Patient-Reported Outcomes in Performance Measurement[315]—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

1. Evidence, Performance Gap, and Priority (Impact)—Importance to Measure and Report:

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-impact aspect of healthcare 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.

---

[313] https://login.qualityforum.org/login.aspx.

[314] 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.

[315] 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.

Patient-reported outcome-based performance measures

For patient-reported outcome (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 process or structure of care.[316] 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.

Note: 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.

Note: The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grade definitions[317] and methods,[318] or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.[319]

- 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

---

[316] 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.

[317] 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.

[318] 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.

[319] 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.

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.

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;[320] American Quality Alliance's Principles of Efficiency Measures).[321]

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.

AND

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

[320] 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.

[321] 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.

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.

AND

1c. High Priority (previously referred to as High Impact)

This criterion emphasizes that the specific measure focus (i.e., what is measured) should be considered important enough to expend resources for measurement and reporting according to CMS goals and priorities as well as recognized national health goals and priorities. The performance measure addresses:

- A specific national health goal/priority identified by the Department of Health and Human Services or the National Priorities Partnership convened by NQF.

OR

- A demonstrated high-priority (high-impact) aspect of healthcare (for example, affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use [current and/or future]; severity of illness; and severity of patient/societal consequences of poor quality).

Composite performance measures

The priority criterion (1c) applies to the composite performance measure as a whole.

Cost and resource use measures

Cost and resource use measures should also include a summary of evidence to support the high impact of measurement in this area.

1d. Explicit logic for composite measures

Composite performance measures

Subcriterion 1d 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.

2. 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.

2a. Reliability

2a1. Precisely specified

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 Health Quality Measures Format (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 must also meet the following subcriteria in order to pass 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 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.

eMeasures

eMeasure specifications include data type from the QDM, code lists, EHR field, measure logic, original source of the data (person or system) and recorder (person).

Patient-reported outcome-based performance measures

For patient-reported outcomes (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.

2a2. Data elements repeatable

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).

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 intraclass correlation coefficient.

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.

Patient-reported measures

For patient-reported outcomes, reliability and validity should be demonstrated for both the data (PROM) and the PRO-PM performance measure score.

2b. 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.

2b1. Specifications consistent with evidence

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 patient-reported outcomes, 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.

AND

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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) 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.

Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (for example, poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for cardiovascular disease risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

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. Scoring methods

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

Measures must pass these additional subcriteria to pass validity:

- 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 ESRD patients]).
- Exclusion.
- 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 ratio scores 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.)
- Significant Differences in Performance.
- Comparability of Multiple Data Sources.
- Validity testing.

2c. Disparities

If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (for example, by race, ethnicity, socioeconomic status, gender);

OR

Rationale/data justifies why stratification is not necessary or not feasible.

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.

3. 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 electronic health records 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 National Quality Forum: 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 HIT Standards Committee'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.

4. 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 three years after initial endorsement and are publicly reported within six 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.

AND

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.

AND

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.

AND

4d. 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.

5. 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. Either the measure specifications must be harmonized with related measures so that they are compatible or the differences must be justified. Evaluation of harmonization examines the extent to which measure specifications are either harmonized with related measures so they are uniform or compatible, or the differences must be justified (for example, dictated by evidence).

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.

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		

Subcriteria	Anticipated NQF Rating	Rating Improvement Plan (if Low/Moderate)
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 three to five years based on a projection of the maturation of the electronic health record, 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:

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. If there are separate files, put them together in a zip file.>

Expiration Notice:

This notice expires on < list the date>.

Project Specific Instructions:

- <Project specific instructions>
- Send your comments to <insert email address>.

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 < 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 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. If there are separate files, combine them into a zip folder.>

Expiration Notice:

This notice expires on < list the date>.

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 may be adjusted as needed.>

Date Posted	Measure Set or Measure	Text of Comments	Name, Credentials, and Organization of Commenter	E-Mail Address	Type of Organization	Recommendations/Actions Taken

Technical Expert Panel Call for TEP Web Page 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
- Health care 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>

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
- Health care 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: Click here to enter text.
Credentials: Click here to enter text.
Professional Role: Click here to enter text.
Organizational Affiliation: Click here to enter text.
City: Click here to enter text.
State: Click here to enter text.
Mailing address: Click here to enter text.
Telephone: Click here to enter text.
Email: Click here to enter text.

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: Click here to enter text.
Credentials: Click here to enter text.
Professional Role: Click here to enter text.
Organizational Affiliation: Click here to enter text.
City: Click here to enter text.
State: Click here to enter text.
Mailing address: Click here to enter text.
Telephone: Click here to enter text.
Email: Click here to enter text.

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. Click here to enter text.

- 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. Click here to enter text.

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

Technical Expert Panel 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

Technical Expert Panel Composition (Membership) List Web Page 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. The purpose of this project is to:

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
- Health care 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>.

Technical Expert Panel 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. This project *<provide background on the measure project, including overall goals. >*

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.>

Technical Expert Panel Summary Web Page 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. The purpose of this project is to:

Project Objectives:

- <list contract objectives>

Expiration Notice:

This notice expires on < list the date>.