



CMS Measures Management System Blueprint

Version 16.0 | September 2020



TABLE OF CONTENTS

	PAGE
PREFACE.....	V
Evolution of the Blueprint	v
Organization of the Blueprint materials.....	v
1 CMS QUALITY MEASURE DEVELOPMENT.....	1
1.1 Quality Across CMS.....	2
1.2 Meaningful Measures Initiative.....	2
1.3 Successes to Date.....	4
1.4 Critical Challenges.....	5
1.5 CMS Goals and Priorities.....	5
2 MEASURE PRIORITIZATION AND PLANNING.....	7
2.1 CMS Measure Planning Inputs	7
2.1.1 Meaningful Measures Initiative	7
2.1.2 Patients, Public, and Other Stakeholders.....	7
2.1.3 Legislative Mandates.....	8
2.1.4 Quality Measure Development Plan (MDP).....	8
2.1.5 Impact Assessment and Other Reports	8
2.2 Role of the Measure Developer in Priorities Planning	9
3 QUALITY MEASURE DEVELOPMENT.....	11
3.1 Structure-Process-Outcome	11
3.2 Overview.....	11
3.3 General Principles for Measure Development	13
3.4 Technical Principles for Measure Development.....	14
3.5 Measure Evaluation in the Measure Lifecycle.....	15
3.5.1 Applying Measure Evaluation Criteria	15
3.5.2 Timing of Measure Evaluation	16
3.6 Measure Categorization.....	17
3.7 Roles in Measure Development	18
3.7.1 Measure Stewards.....	18
3.7.2 Measure Developers.....	18
3.7.3 The Measures Management System (MMS)	18
3.7.4 Stakeholders.....	19
4 MEASURE CONCEPTUALIZATION	23
4.1 Information Gathering.....	24
4.1.1 Identify the Healthcare Quality Issues to be Addressed and Determine Their Priority Area	25
4.1.2 Conduct an Environmental Scan	25
4.1.3 Analyze Empirical Data, as Appropriate.....	26
4.1.4 Evaluate Information Collected During Environmental Scan and Empirical Data Analysis.....	27

4.1.5	Conduct a Measurement Gap Analysis to Identify Areas for New Measure Development.....	27
4.1.6	Justify the Creation of New Measures.....	27
4.1.7	Apply Measure Evaluation Criteria.....	28
4.1.8	Prepare an Initial List of Measures or Measure Topics.....	28
4.2	Business Case Development.....	28
4.2.1	Business Case Best Practices.....	30
4.3	Stakeholder Engagement.....	33
4.3.1	Technical Expert Panel.....	33
4.3.2	Person and Family Engagement.....	33
4.3.3	Public Comment.....	34
4.3.4	Other Stakeholder Engagement.....	35
5	MEASURE SPECIFICATION.....	36
5.1	Measure Specifications by Measure Category.....	38
5.2	Define the Data Source(s).....	39
5.3	Develop Specifications and Definitions.....	40
5.3.1	Define the Initial Population.....	41
5.3.2	Define the Denominator.....	41
5.3.3	Determine if there is a need for a Denominator Exclusion.....	42
5.3.4	Define the Numerator.....	42
5.3.5	Determine if there is a need for a Denominator Exception.....	43
5.3.6	Determine if there is a need for a Numerator Exclusion.....	44
5.3.7	Define Stratification Scheme.....	45
5.3.8	Use Positive Evidence.....	46
5.4	Specify the Code and/or Code Systems.....	46
5.5	Construct Data Protocol.....	46
5.5.1	Define Key Terms, Data Elements, Codes, and Code Systems.....	47
5.5.2	Describe the Level of Measurement/Analysis.....	48
5.5.3	Describe Sampling.....	48
5.5.4	Determine Risk Adjustment.....	48
5.5.5	Clearly Define Any Time Intervals.....	49
5.5.6	Describe How to Score and Report the Measure Results.....	50
5.5.7	Develop the Calculation Algorithm.....	50
5.6	Document the Measures.....	50
5.6.1	Finalize the Measure Name and Description.....	51
5.6.2	Making Changes.....	52
5.7	Stakeholder Engagement in Measure Specification.....	52
5.8	Special Considerations for Medicaid-Focused Measures.....	52
5.9	Harmonization.....	52
6	MEASURE TESTING.....	55
6.1	The Measure Testing Process.....	56
6.2	Testing and Measure Evaluation Criteria.....	58
6.2.1	Importance.....	58
6.2.2	Scientific Acceptability.....	59
6.2.3	Feasibility.....	64
6.2.4	Usability and Use.....	65

6.3	Develop the Testing Work Plan	65
6.4	Perform Sampling	66
6.5	Implement the Plan.....	67
6.6	Analyze the Test Results	67
6.7	Refine the Measure.....	67
6.8	Retest the Refined Measure	67
6.9	Measure Testing Summary.....	67
6.10	Stakeholder Engagement During Measure Testing.....	69
6.11	Testing and Evaluation for Special Measures.....	69
7	MEASURE IMPLEMENTATION.....	70
7.1	The Implementation Process	70
7.2	Measure Selection	71
7.2.1	Pre-Rulemaking Process	71
7.2.2	CMS Rulemaking Processes.....	73
7.2.3	Stakeholder Engagement in Measure Selection.....	73
7.3	Rollout	74
8	MEASURE USE, CONTINUING EVALUATION, AND MAINTENANCE.....	76
8.1	Measure Production and Monitoring	77
8.1.1	Conduct Data Collection and Ongoing Surveillance	78
8.1.2	Respond to Questions About the Measure	79
8.1.3	Monitor and Analyze the Measure Rates and Audit Findings	79
8.1.4	Perform Measure Maintenance or Ad Hoc Review, When Appropriate.....	79
8.2	Continuing Evaluation	80
8.3	Measure Maintenance.....	80
8.4	Evaluation during Measure Maintenance	82
8.4.1	Apply Measure Evaluation Criteria	83
8.4.2	Document Results of Evaluation	83
8.5	Measure Maintenance Reviews	83
8.5.1	Annual Update.....	84
8.5.2	Comprehensive Reevaluation.....	84
8.5.3	Ad Hoc Review.....	84
8.6	Stakeholder Engagement During Measure Maintenance Reviews	85
9	TOOLS AND RESOURCES FOR MEASURE DEVELOPERS	86
9.1	Bonnie 	86
9.2	Clinical Quality Language (CQL) Style Guide 	86
9.3	CMS Data Element Library (DEL).....	86
9.4	CMS Measures Inventory Tool (CMIT).....	86
9.5	Environmental Scanning Support Tool (ESST)	86
9.6	CQL-to-Expression Logical Model (ELM) Translator Reference Implementation	87
9.7	Cypress 	87
9.8	Electronic Clinical Quality Improvement Resource Center (eCQI Resource Center) 	87
9.9	Electronic Clinical Quality Measure (eCQM) Logic and Implementation Guidance	87
9.10	Guide for Reading eCQMs 	87

9.11 *Measure Authoring Tool (MAT)* 88

9.12 *Measures Manager*88

9.13 *Measures Management System Website*.....88

9.14 *National Quality Forum Incubator (NQF Incubator®)*88

9.15 *National Quality Forum Quality Positioning System (QPS)*88

9.16 *Value Set Authority Center (VSAC)*.....89

APPENDIX A: SUMMARY OF CHANGES TO BLUEPRINT90

APPENDIX B: REFERENCES91

LIST OF FIGURES

Figure 1. Meaningful Measures Framework..... 4

Figure 2. CMS Strategic Priorities 6

Figure 3. The Measure Lifecycle12

Figure 4. High-Level View of Major Measure Lifecycle Tasks.....13

Figure 5. Applying Measure Evaluation Criteria16

Figure 6. Stakeholder Engagement in the Measure Lifecycle.....20

Figure 7. Flow of the Measure Lifecycle—Measure Conceptualization24

Figure 8. Environmental Scan Data Sources26

Figure 9. Flow of Inputs into Business Case Development31

Figure 10. Inputs and Uses for the Business Case.....32

Figure 11. Flow of the Measure Lifecycle—Measure Specification.....36

Figure 12. Flow of the Measure Lifecycle—Measure Testing.....55

Figure 13. Flow of the Measure Lifecycle—Measure Implementation70

Figure 14. Sample Pre-rulemaking Process Timeline.....72

Figure 15. Flow of the Measure Lifecycle—Measure Use, Continuing Evaluation, and Maintenance.....77

Figure 16. Overview of the Measure Production and Monitoring Process.....78

Figure 17. CMS Criteria for Measure Disposition: Revise or Retain.....81

Figure 18. CMS Criteria for Measure Disposition: Retire.....82

Figure 19. CMS Criteria for Measure Disposition: Remove.....82

List of Tables

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

Table 2. Measure Definitions17

Table 3. Refinement of the business case is an ongoing process that occurs throughout the Measure Lifecycle.30

Table 4. Measure Specifications by Measure Category39

Table 5. Strengths and Limitations of Different Data Sources.....39

Table 6. Features of Alpha and Beta Testing.....56

PREFACE

EVOLUTION OF THE BLUEPRINT

From Version 1 through 16, Blueprint updates have incorporated changes in the regulatory environment and in healthcare quality measurement science to meet evolving needs of measure developers. Each update has reflected systematically gathered stakeholder input.

In response to feedback that the Blueprint in its current form was not optimal for a broad variety of audiences, the Centers for Medicare & Medicaid Services (CMS) elected to streamline Version 16 of the Blueprint and create a separate volume tailored to MIDS contractors and their government leads. The Blueprint Version 16 package contains sections addressing a) basic steps needed to complete the Measure Lifecycle; b) supplemental materials that provide more detail about the Measure Lifecycle and/or specific types of [measures](#) (i); c) templates; and d) a QuickStart Guide. The Blueprint editors removed duplicative detail and folded information that is relevant only to CMS-contracted measure developers and their government leads into a contractual edition available only to MIDS contractors and CMS. *The Blueprint for the CMS Measures Management System: Contractual Edition* includes checklists of essential milestones and deliverables for MIDS 2018 Indefinite Delivery/Indefinite Quantity (IDIQ) Task Orders (TOs) and the parties responsible for each. Hopefully, these changes will make the Blueprint more accessible to new and non-CMS-contracted measure developers and easier for all users to find the information they need. Note: Blueprint in block letters refers to the total Blueprint package. *Blueprint* in italics refers to the streamlined piece of the Blueprint Package. *The Blueprint: Contractual Edition* piece is also in italics.

A high-level list of changes in *Blueprint* Version 16 is available in [Appendix A](#).

Recommendations for changes to the content, structure, or organization of the Blueprint are welcome. Given the major shift in Version 16's structure, CMS and the Measures Manager are particularly interested in your comments about the changes. Should you choose to provide input, please submit your comments and suggestions through the [MMS Support mailbox](#)[↗]. General feedback, as well as specific suggestions for edits are welcome. If requesting specific edits, please include specifics about recommended change(s) in your message, including

- version of the Blueprint referenced
- relevant chapter number or document section and title
- page number
- relevant text to modify, if applicable
- new text to add, if applicable
- rationale for change
- point of contact information

Consideration of recommended changes is year-round and incorporated into the next review cycle of the document, if appropriate.

ORGANIZATION OF THE BLUEPRINT MATERIALS

The CMS Measures Management System (MMS) Blueprint package consists of two primary documents, two sets of ancillary documents, a QuickStart Guide, and an introductory guide:

- *The Blueprint for the CMS Measures Management System* (the *Blueprint*) covers the high-level quality measurement philosophy and guidance for the development and maintenance of quality measures . Information in the *Blueprint* is applicable to all measure developers, regardless of level of experience or whether the intent of the measure(s) they develop is for use in a CMS program or some other purpose.
- *The Blueprint for the CMS Measures Management System: Contractual Edition* (the *Blueprint: Contractual Edition*) contains identical material addressed in the *Blueprint* augmented with contractual guidance for CMS-contracted measure development and maintenance contractors (MDMs) and their government leads. Subsections titled Contractual Guidance and Considerations and Task Lists contain this additional guidance.
- Supplemental Materials are a set of standalone documents that provide an in-depth treatment of measure development topics.
- Templates are resources available to all measure developers to guide measure development processes and format the resulting products. There are contractual requirements for CMS-contracted MDMs to adhere to the guidance and specifications delineated in each of the Forms and Templates unless otherwise directed by their contracting officer’s representative (COR).
- Designed with both experienced and novice measure developers in mind, the *CMS MMS Blueprint QuickStart Guide* (*QuickStart*) is a 25 page, start-to-finish overview of measure development, implementation, and maintenance steps and processes. Each section includes information about important steps associated with a given stage of the Measure Lifecycle, along with links to additional resources, templates, and references to specific chapters and supplemental materials of the *Blueprint*.
- Quality Measures: How They Are Developed, Used, & Maintained (“Quality Measurement 101”) is a brief, easy-to-read introduction to quality measures and the measure lifecycle designed for non-developers.

Both the *Blueprint* and the *Blueprint: Contractual Edition* link to the Supplemental Materials and Templates. The Supplemental Materials also have links to the Templates.

Additionally, the *Blueprint* materials address single measures or a small set of closely related measures , not measures at the portfolio/measure set  level.

Blueprint editors have adopted a modified American Psychological Association (APA) format for the *Blueprint* documents. Journal articles and books with hyperlinks to the article or book have APA format inline citations, if available. Grey literature  uses document names with hyperlinks. The editors use APA format for punctuation, abbreviations, bullets, and the reference list.

 Note: A computer icon indicates eCOM -specific information throughout the *Blueprint* documents.

1 CMS QUALITY MEASURE DEVELOPMENT

Increased emphasis on performance measurement is driving a fundamental change in the United States (U.S.) healthcare system. In nearly every setting of care, CMS is moving from paying for services to paying for value. CMS’s goal is to foster value by promoting the highest quality, safety, and care experience with the most affordable, cost-efficient service possible for Americans. [Table 1](#) highlights four payment categories that represent the progression of payment reform for clinicians and facilities for their services. Initiated with the passage and implementation of the [Patient Protection and Affordable Care Act \(ACA\)](#)¹ and more recently driven by the [Medicare Access and Children’s Health Insurance Program \(CHIP\) Reauthorization Act \(MACRA\)](#)², CMS is well on its way to transitioning from a fee for service (FFS) system to a payment system based on quality and value. In the near term, few payments in the Medicare program will continue to be based on Category 1 criteria and there will be a rapid transition to the majority of payments falling under the criteria for Categories 3 and 4.

Table 1. [Framework for Progression of Payment to Clinicians and Organizations in Payment Reform \(adapted from Rajkumar et al., 2014\)](#)³

	Category 1: FFS No Link to Quality	Category 2: FFS Link to Quality	Category 3: Alternative Payment Models (APMs) on FFS Architecture	Category 4: Population Based Payment
Description	Payments are based on volume of services and not linked to quality or efficiency	At least a portion of payments vary based on the quality or efficiency of healthcare delivery	<ul style="list-style-type: none"> Some payment is linked to the effective management of a <u>population</u> or an episode of care Payments still triggered by delivery of services, but opportunities for shared savings or two-sided risk 	<ul style="list-style-type: none"> Service delivery is not directly triggering payment, so volume is not linked to payment Clinicians and organizations are paid and responsible for the care of a beneficiary for a long period (e.g., > one year)
Medicare	<ul style="list-style-type: none"> Limited in Medicare FFS Majority of Medicare payments now are linked to quality 	<ul style="list-style-type: none"> Hospital value-based purchasing Hospital readmissions/hospital-acquired condition reduction programs 	<ul style="list-style-type: none"> Accountable Care Organizations Medical homes Bundled payments 	<ul style="list-style-type: none"> Eligible Pioneer accountable care organizations in Years 3 through 5 Some Medicare Advantage plan payments to clinicians and organizations
Medicaid	Varies by state	<ul style="list-style-type: none"> Primary care case management Some managed care models 	<ul style="list-style-type: none"> Integrated care models under FFS Managed FFS models for Medicare–Medicaid beneficiaries Medicaid health homes Medicaid shared savings models Medicaid waivers for delivery reform incentive payments Episodic-based payments 	<ul style="list-style-type: none"> Some Medicaid managed care plan payments to clinicians and organizations Some Medicare–Medicaid (duals) plan payments to clinicians and organizations

The stakes are higher than ever for persons (patients)¹ and providers, as payment programs use quality measurement and are transparent in public reporting, which may impact both patient choice and accountability for healthcare providers. As such, clinical quality measures  (CQMs) must be meaningful, robust,² valid, feasible, based in scientific evidence, and well tested to ensure that the measures  do not lead to unintended negative consequences or burden for patients or providers. In striving to achieve the goals of the Meaningful Measures Initiative, developed measures must be meaningful to patients and the providers who serve them; represent opportunities for improvement in care quality and hence, actionable; and differentiate quality in a meaningful and valid way. To accomplish these goals, the strategies for measure development outlined in this document must be at the forefront.

1.1 QUALITY ACROSS CMS

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

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

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

1.2 MEANINGFUL MEASURES INITIATIVE

CMS launched the comprehensive Meaningful Measures Initiative in 2017 which identifies high priority areas for quality measurement and improvement. The purpose of this initiative is to improve outcomes for patients, their families, and providers while also reducing burden and moving payment toward value through focusing everyone’s efforts on the same quality areas. The Meaningful Measures Initiative also helps to identify and close important gap areas of measures, align measures across the continuum of

¹ The Blueprint uses the terms *persons* and *patients* interchangeably.

² Throughout the document, “robust” refers to measures with the most vigorous quality action or guidance or as a descriptor to describe strong, vigorous, or thoroughly vetted components of a measure.

care and across payors, and spur innovation in new types of [measures](#)^① such as patient-reported measures and electronic measures.

CMS is continuing to drive towards patient-centered, value-based care through the development, selection, and implementation of quality measurement. Specifically, the CMS quality measurement needs identified include

- providing rapid performance feedback to providers
- accelerating the move to fully digital measures
- unleashing voice of patient through use of [patient-reported outcome measures](#)^①
- using measures that will advance innovative payment structures
- increasing [alignment](#)^① of measures
- promoting use of all payer data (where feasible)
- focusing on major domain outcomes

Regulatory reform and reducing regulatory burden are high priorities. CMS continuously works to find ways to reduce burden on providers, while empowering patients. By identifying the highest priorities for quality measurement and improvement, the [Meaningful Measures Initiative](#)^② provides a framework, as shown in [Figure 1](#),³ for core issues that are most vital to improving patient outcomes and has led to evaluation of measures across many programs with a resulting 20% reduction of measures used in Medicare quality programs to date. The Meaningful Measures Initiative represents a new approach to [quality measures](#)^① that will reduce the collection and reporting burden, while producing quality measurement that is more focused on meaningful outcomes.

³ There is regular evaluation of the Meaningful Measures Framework. It is updated to reflect stakeholder feedback given to CMS and any shift in CMS priorities. For the most up to date information, visit the [Meaningful Measures Hub](#)^②.

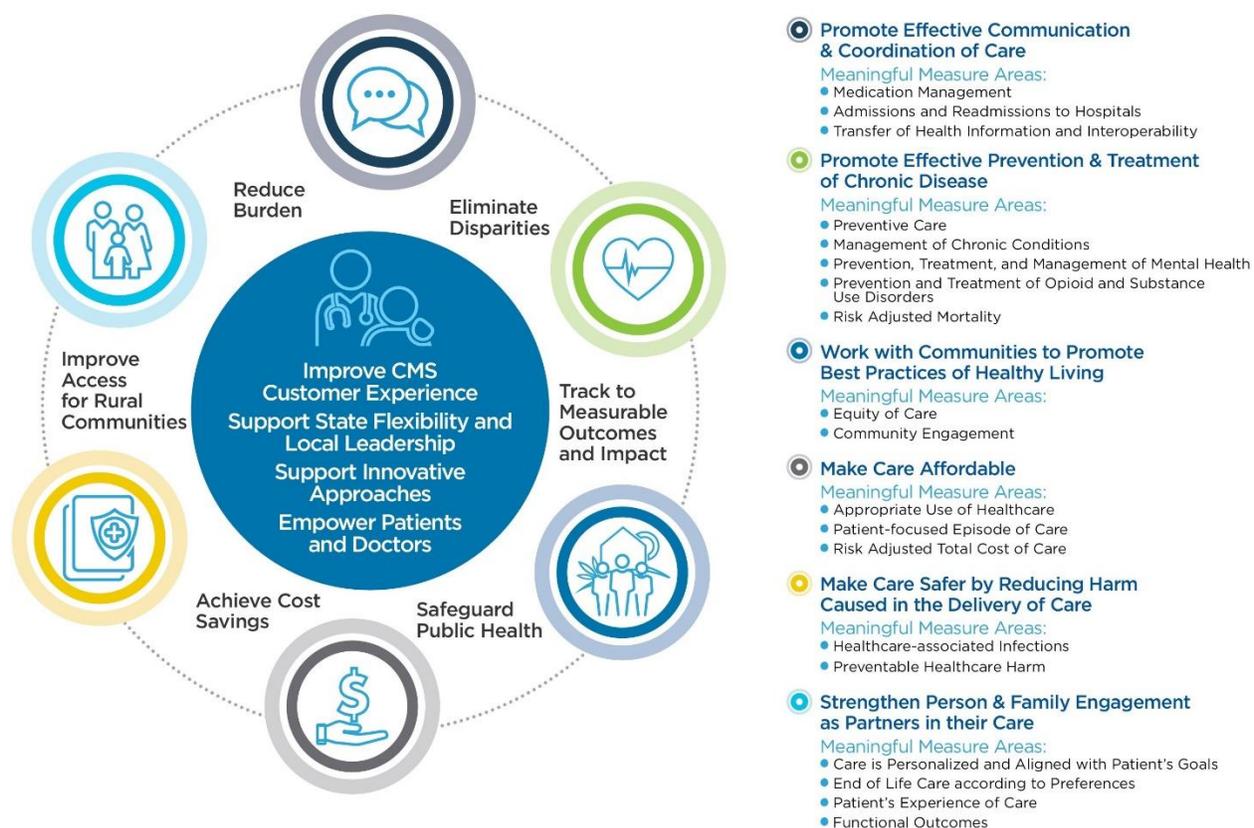


Figure 1. Meaningful Measures Framework

1.3 SUCCESSES TO DATE

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

Many CMS-developed measures ^① are National Quality Forum (NQF)-endorsed and/or recommended by the NQF-convened [Measure Applications Partnership \(MAP\)](#). ^① However, as performance on quality metrics is increasingly tied to provider payment, there is an increasing need for measure development to be a more flexible and efficient process with a shorter development time frame. CMS continually seeks innovations and process improvement to meet these challenges. CMS has also started to remove measures from programs that have minimal room for improvement, no longer have supporting evidence, are duplicative of other measures, or are of low value from the patient or clinical workflow perspective.

CMS is rebalancing the portfolio of measures to contain more outcome measures ^① and fewer process measures ^①, with the goal of better addressing performance gaps in the Meaningful Measures Initiative. As part of this effort, the Measures Manager maintains the [CMS Measures Inventory Tool](#) ^① (CMIT). The Tool includes measures throughout the Measure Lifecycle for dozens of CMS programs and initiatives including measures under development (MUD), measures under consideration (MUC) ^①, measures that

have entered the rulemaking process, measures ① actively in a program, and measures that have been removed from programs. The Inventory includes measure title, description, NQF endorsement status, and measure type with the goal of providing users with a complete picture of how the measure quantifies performance quality within the various CMS programs.

1.4 CRITICAL CHALLENGES

The challenges to developing measures that are meaningful and appropriate for the Quality Payment Program, which may be applicable to other programs, are described in detail in the [CMS Quality Measurement Development Plan](#) ² (MDP) and cannot all be enumerated here. However, some of the key opportunities include

- partnering with patients in the measure development process
- partnering with frontline clinicians and professional societies
- aligning ① measures across programs, payors, and payment systems
- reducing clinician burden of data collection for measure reporting
- shortening the time frame for measure development
- streamlining data acquisition for measure testing ①
- identifying and developing meaningful outcome measures ①
- developing patient-reported outcome measurement (PROM) ① tools and appropriate use measures
- developing measures that promote shared accountability across settings and providers

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

1.5 CMS GOALS AND PRIORITIES

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

- improving the CMS customer experience
- ushering in an era of state flexibility and local leadership
- supporting innovative approaches to improve quality, accessibility, and affordability
- empowering patients and clinicians to make decisions about their healthcare.

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

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

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

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

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

These goals are framed into a strategic wheel (Figure 2) reflecting the strategic initiatives across the agency.

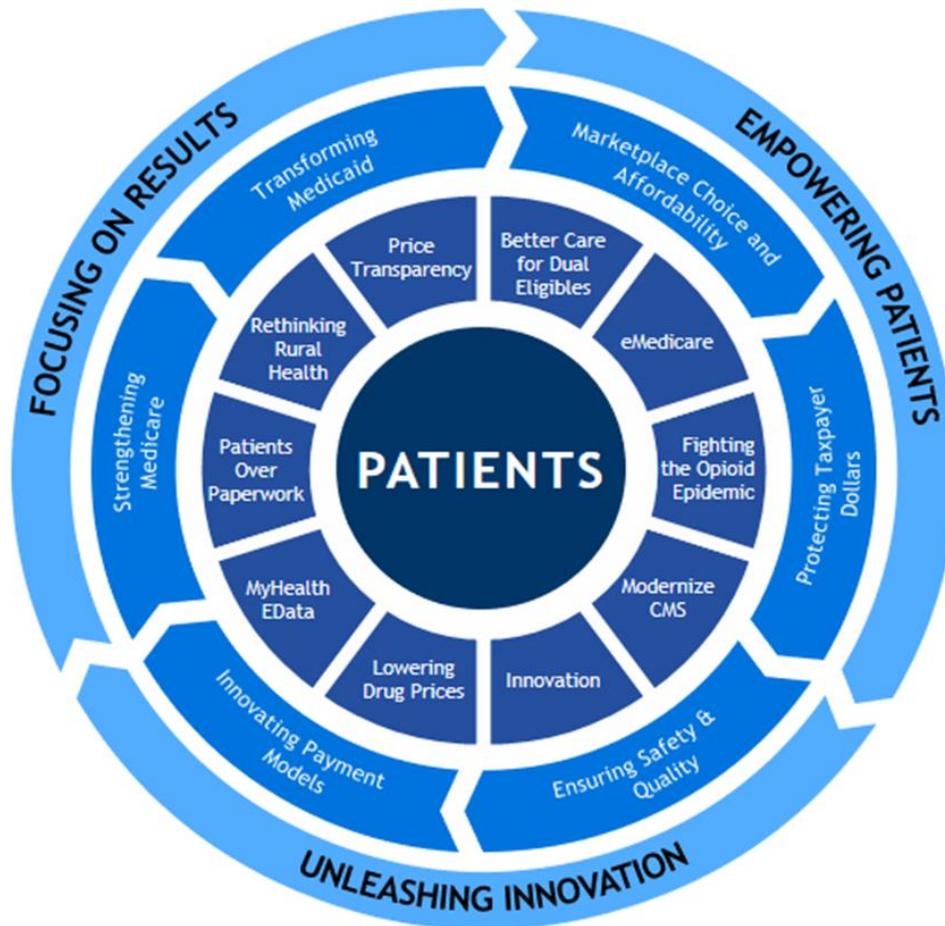


Figure 2. CMS Strategic Priorities

Much of the quality measurement work across CMS aligns with the strategic initiative, [Ensuring Safety & Quality](#). With a focus on better patient health outcomes, CMS holds providers accountable for providing safe and effective care, while minimizing administrative burden to ensure clinicians can spend more time with patients. The focus is on empowering beneficiaries to make decisions about their healthcare based on quality and cost information by moving our quality programs to measure value and to provide consumers access to information in an understandable and actionable way.

2 MEASURE PRIORITIZATION AND PLANNING

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

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

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

2.1 CMS MEASURE PLANNING INPUTS

2.1.1 Meaningful Measures Initiative

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

2.1.2 Patients, Public, and Other Stakeholders

CMS conducts its measurement activities in a transparent manner. The information gathered through various methods described in [Chapter 4. Measure Conceptualization](#), informs the Department of Health and Human Services (HHS) and CMS about future measurement needs. Additionally, Section 101(f) of MACRA requires that CMS solicit, accept, and respond to input from stakeholders, including physician specialty societies, applicable practitioner organizations, and other stakeholders for episode groups (i.e.,

care episode groups and patient-condition groups). Care episode groups include those patients whose care included similar treatments and procedures, taking into consideration patients' clinical diagnoses and problems during the care episode, care setting, and level of acuity, and principal procedures or services furnished. Patient-condition groups include those patients with similar conditions, taking into consideration patients' medical and surgical histories, comorbid conditions, overall health status, and eligibility or dual-eligibility status. Patients and families are extremely important stakeholders in the quality measurement enterprise, and CMS is committed to gathering their input during priorities planning. More detail about ways in which to hear the patient's voice is found in [4.3.2, Person and Family Engagement](#) and the [Person and Family Engagement in Quality Measurement](#)² supplemental material.

2.1.3 Legislative Mandates

Most CMS quality reporting and incentive programs are born out of legislation, which in turn amend the Social Security Act (SSA). MACRA, ACA, and the [American Recovery and Reinvestment Act of 2009 \(ARRA\)](#)², including the [Health Information Technology for Economic and Clinical Health \(HITECH\) Act](#)¹, have the largest influence on CMS's quality measurement priorities, which have led to the broad payment reform and quality-based payment models. A list of the CMS quality programs and initiatives and the initiating legislation and SSA location is in the [Legislative Mandates for Quality Measurement and Reporting](#)² supplemental material.

2.1.4 Quality Measure Development Plan (MDP)

On May 2, 2016, CMS finalized the MDP, mandated under the MACRA, to support the new Medicare Incentive Payment System (MIPS) and advanced APMs. MACRA supports a transition to value-based payment incentives for physicians and other clinicians to be based on quality, rather than quantity, of care.

The MDP is an essential resource in this transition, as it provides the foundation and a strategic framework for building and implementing a measure portfolio to support the quality payment programs under MACRA. The MDP highlights known clinical and specialty measurement and performance gaps and recommends prioritized approaches to close these gaps through the development, adoption, and refinement of [quality measures](#)¹.

Through the application of the principles included in the MDP and the quality measure development funded by MACRA, CMS is committed to increased transparency and partnerships with persons and families, clinicians, and professional societies to develop measures that are meaningful, applicable, and useful across payors and healthcare settings. These quality measures are essential to address critical performance gaps, facilitate [alignment](#)¹ across settings and payors, and promote efficient data collection. CMS intends for the MDP and related quality measures to be key levers of delivery system reform, promoting movement toward paying for value rather than volume and improved national healthcare delivery.

For a copy of the MDP, MDP Annual Reports, and MDP [Environmental Scan](#)¹ and Gap Analysis Reports and more information, view the [CMS Quality MDP and Annual Report](#)² website.

2.1.5 Impact Assessment and Other Reports

Once a measure is in use, it requires ongoing monitoring and [maintenance](#)¹ in addition to formal periodic reevaluations to determine whether it remains appropriate for continued use. The measure

developer conducts [measure](#) ① trend analyses, evaluates barriers, and identifies unintended consequences associated with specific measures in their purview.

[Measure maintenance](#) ① reports yield information that CMS leadership may find valuable for setting priorities. This information may include barriers to implementation of measures, unintended consequences, lessons learned, measure impact on providers, [care disparities](#) ①, and gaps in care. Measure maintenance includes assessment of the performance of the measure, including trend analyses, and comparison to the initial projected performance, found in the [Business Case](#) ①. CMS uses this input to decide whether to remove, retire, modify, suspend, or retain measures in use.

In addition to measure maintenance, CMS conducts various evaluations and assessments of its measures and programs to determine the effectiveness of its programs. Many of these programs use [quality measures](#) ①, and these analyses evaluate the usefulness of the measures as used in the programs.

The triennial [National Impact Assessment of the CMS Quality Measures Reports](#) ① required by section 1890A(a)(6) of the Social Security Act aims to contribute to the overall, cross-cutting evaluation of CMS quality measures. The intent of the analyses in these reports is not to replace or duplicate program-specific assessments nor to replace the analyses individual measures must undergo as part of ongoing measure maintenance. Rather, the intent is to help the federal government and the public understand the overall impact of the government's investments in quality measurement and reflect on future needs.

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

- [Medicare Payment Advisory Committee \(MedPAC\)](#) ① and [Medicaid and CHIP Payment and Access Commission \(MACPAC\)](#) ① quality reports
- Agency for Healthcare Research and Quality (AHRQ) [National Healthcare Quality and Disparities Reports](#) ①
- [CMS Office of Minority Health](#) ①
- CMS Office of Enterprise Data and Analytics – [Chronic Conditions among Medicare Beneficiaries](#) ①
- universities, researchers, and healthcare facilities, including their journal articles and conference presentations

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

2.2 ROLE OF THE MEASURE DEVELOPER IN PRIORITIES PLANNING

The measure developer plays a key role in supporting CMS's priorities planning. It is important for measure developers to be knowledgeable about how CMS plans its measure development and maintenance activities so appropriate measures are based on CMS-established priorities.

Measure developers also play an important part in [measure harmonization](#) ① and [alignment](#) ①. During measure development, it is important that measure developers conduct a thorough [environmental scan](#) ① and are knowledgeable about measures that may be like those they are seeking to develop. To the extent possible, measure developers are to avoid developing [competing measures](#) ①—those that essentially address the same concepts for the target process, condition, event, or outcome, and the same [target patient population](#) ①. Competing measures are conceptually similar, but their technical

specifications^① may differ. For more information on harmonization^①, see the [Harmonization, Respecification, and Adoption](#)[↗] supplemental material.

Measure developers should consider the HHS and CMS goals and priorities, e.g., measure^① priorities found on the [CMS Pre-Rulemaking](#)[↗] webpage, when identifying a list of potential measures for pre-rulemaking, rulemaking, and eventual program adoption.

3 QUALITY MEASURE DEVELOPMENT

3.1 STRUCTURE-PROCESS-OUTCOME

At a high level, measures ① used to assess and compare the quality of healthcare organizations have the designation as structure ①, process ①, or outcome measures ①. This triad is also known as the Donabedian model named after the physician who developed it. Considered the father of modern healthcare quality management, Avedis Donabedian was a prolific writer penning 11 books and more than 100 articles addressing the topic of healthcare quality ([Best & Neuhauser, 2004](#)^[2]). His 1966 article, Evaluating the Quality of Medical Care, laid the groundwork for the Framework for Health Care Quality. Donabedian's Framework outlines the structure, process, and outcome model, which is the basis for most current quality measures ① ([Donabedian, 1966/2005](#)^[2]).

Donabedian defined process of care as “a set of activities that go on within and between practitioners and patients” (Donabedian, 1980, p. 79) and process is the primary object of assessment. Process is the normative care and is dependent on the state of the science of healthcare, values, and ethics. Donabedian defined outcome as “a change in a patient's current and future health status that can be attributed to antecedent health care” (Donabedian, 1980, pp. 82-83). He noted assessing outcome is an indirect method of measuring quality of care. He argued that it is indirect because one cannot judge changes in health status as quality of care until you have eliminated other causes of the change. With respect to process, once you have established a process as clearly associated with good results, there is acceptance that the presence or absence of the process is evidence of good or bad quality. With outcome, you must also establish that there are no other possible factors to explain the change in status. The third piece, structure, is also an indirect method of measuring quality of care. Structure includes the physical setting, organizational policies, financial resources, the tools and resources available to providers of care, and much more. Donabedian noted that structure “is relevant to quality in that it increases or decreases the probability of good performance” (Donabedian, 1980, p. 82).

The triad of structure, process, and outcome have a relationship in that the structural aspects have an influence on the processes of care which, in turn, influence the effect of care on health status. Donabedian noted other aspects of quality of care can fit into these three categories. He acknowledged the triad designation is somewhat arbitrary because the differentiation of reality into these three parts is not clear and we should treat the triad as a guide, not as a straitjacket. Thus, quality measures have categorizations that extend beyond the triad. [Chapter 3.6](#) discusses the categories of measures with definitions for the different measure types.

3.2 OVERVIEW

There are five stages in the Measure Lifecycle: measure conceptualization; measure specification ①; measure testing ①; measure implementation; and measure use, continuing evaluation, and maintenance ① ([Figure 3](#)). The stages are not necessarily sequential, but are iterative, and can occur concurrently.

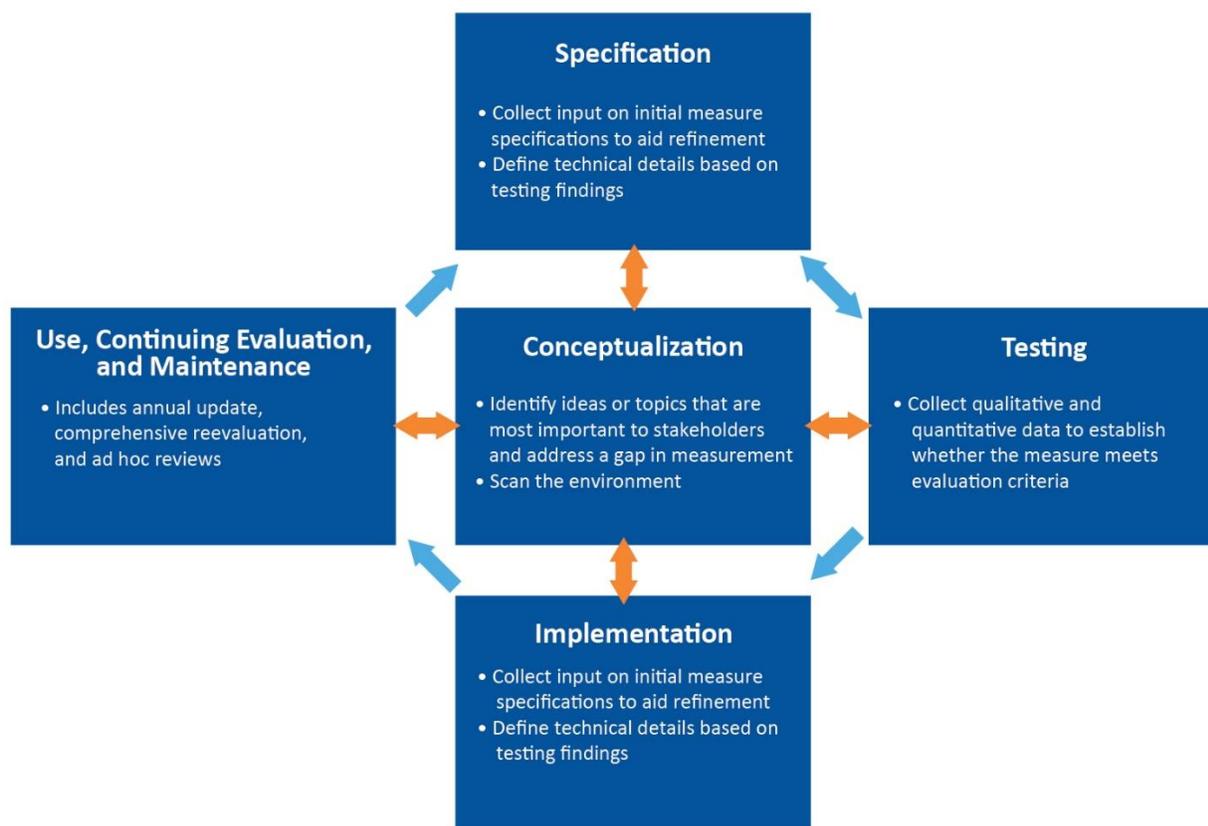


Figure 3. The Measure Lifecycle

The [measure](#) ^① conceptualization stage initiates information gathering, [business case](#) ^① development, and stakeholder outreach. The [measure specification](#) stage involves establishing the basic elements of the measure, including the [numerator](#) ^①, [calculation algorithm](#) ^①, and [data source](#) ^① identification. The [measure testing](#) ^① stage examines the [specifications](#) ^①, usually with a limited number of real settings, to make sure the measure is [scientifically acceptable](#) ^① and feasible. Measure specification and measure testing are iterative. The [measure implementation](#) stage begins with measure selection through various processes and then measure rollout. The [measure use, continuing evaluation, and maintenance](#) stage involves continued monitoring of the measure's use, performance, importance, accuracy, and impact on patients.

The end product of measure development is a precisely specified, valid, and reliable measure that is meaningful to clinicians, patients, and caregivers. Find more details about the Measure Lifecycle stages in [Chapters 4, 5, 6, 7, and 8](#).

[Figure 4](#) depicts a high-level view of the major tasks and timeline involved in developing measures, from initial measure conceptualization through measure implementation and maintenance. Although the graphic depicts the five stages of the Measure Lifecycle in a linear, sequential fashion, measure developers can adjust the sequence or carry out steps concurrently and iteratively. Measure developers conduct [feasibility](#) ^① evaluation, information gathering, and stakeholder engagement on an ongoing basis throughout the Measure Lifecycle. CMS and other stakeholders are working to shorten the measure timeline for more rapid development and implementation of new measures.

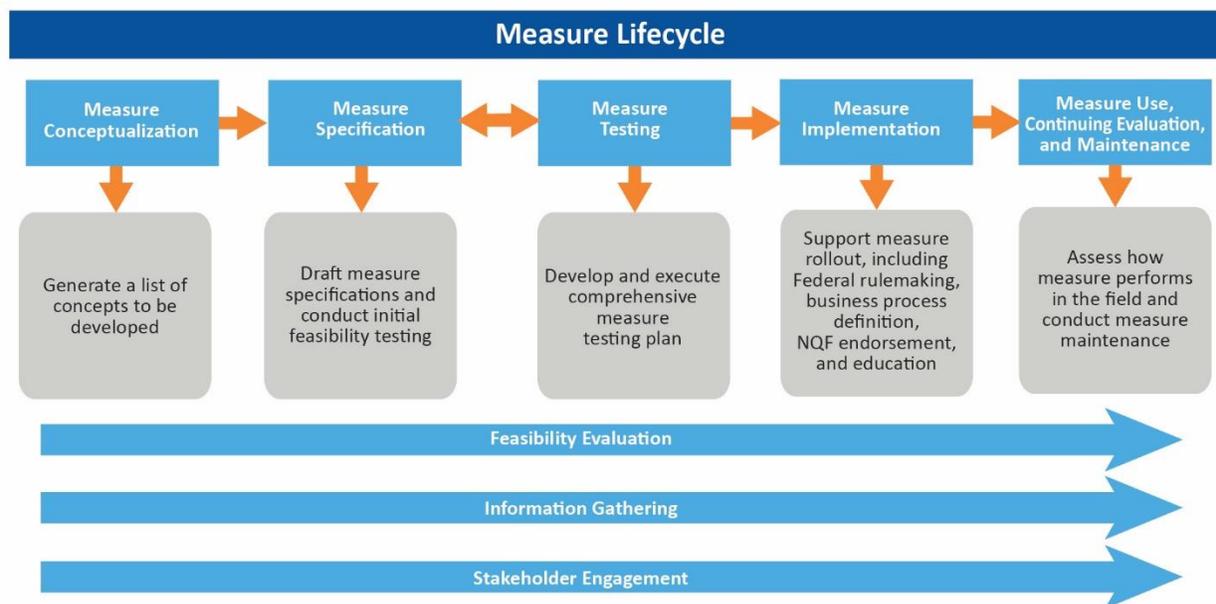


Figure 4. High-Level View of Major Measure Lifecycle Tasks

3.3 GENERAL PRINCIPLES FOR MEASURE DEVELOPMENT

General principles for measure development serve as overarching guidelines for measure development that meet the standards and rigor expected of a meaningful, valid, and useful measure. Development of measures should:

- Focus on what is best for patients and most meaningful to patients, caregivers, and providers.
- Explicitly align with Meaningful Measures and its goals and objectives.
- Align across payors, including Medicare, Medicaid, the Exchanges, other federal partners, and private payors, to the extent feasible based on data availability for each payor type, differences in populations served, and level of analysis.
- Address a performance gap where there is known variation in performance, not just a measure gap.
- Use resources efficiently in a rapid-cycle fashion, including using process improvement techniques, such as Lean and human-centered design, and considering respecification instead of de novo measure development.
- Encourage collaboration among measure developers and share best practices/new learnings freely.
- Reorient and align around patient-centered outcomes that span across clinical settings, which may require different “versions” of the same measure (i.e., different cohorts, but same numerator). Test each of these setting-specific versions for reliability and validity.
- Promote value-based care that produces quality outcomes.
- Focus on outcomes (including patient-reported outcomes [PROs]), safety, patient experience, care coordination, appropriate use/efficiency, and cost.
- Identify disparities and promote health equity in the delivery of care.

- Guard against negative unintended consequences of measure implementation, including overuse and underuse of care.
- Engage stakeholders early and often during the measure development process.
- Strive to reduce clinician burden in reporting measures.⁴

3.4 TECHNICAL PRINCIPLES FOR MEASURE DEVELOPMENT

Measure developers should apply technical principles when developing measures for consideration for quality reporting and value-based purchasing programs.

- Develop a rigorous business case for an evidence-based measure concept—a critical first step in the development process.
- Prioritize digital data sources (e.g., electronic health records [EHRs], registries, electronic administrative claims), where appropriate, and reduce dependency on data from chart abstraction whenever possible. Maintain a focus on iterative testing using both real and synthetic data.
- Consider approaches to aggregate multiple data sources (e.g., hybrid measures) to achieve the most accurate assessment of quality until interoperability is universal.
- Define outcomes, risk factors, cohorts, and inclusion/exclusion criteria based on clinical and empirical evidence.
- Judiciously select exclusions to capture as broad a patient population as possible and appropriate; consider developing a paired measure to capture and measure the care excluded patients received if a significant number of patients are excluded from the measure calculation (e.g., for all patients seen in the emergency department, if those patients who were transferred directly to another acute care facility for tertiary treatment are excluded, a paired measure would address those patients who were transferred out of the original facility).
- Develop risk adjustment models to distinguish performance among providers rather than predict patient outcomes.
- Include measure stratification and risk adjustment approaches to show differences in quality or outcomes among demographic groups and allow for quality comparisons between providers after considering differences in patient characteristics that would not influence the care received. Harmonize measure methodologies, data elements, and specifications, when applicable and feasible.
- Develop each measure with sufficient statistical power to detect and report statistically significant differences in provider performance.
- Consider strategies to enable clinicians that have smaller practices and low-volume facilities to report a measure reliably.
- Strive to develop measures that can progress to multi-payor applicability using all-payor databases where available.⁵
- Consider the clinical workflow needed in the electronic record for electronic clinical quality measures (eCQMs).

⁴ Adapted from the Measure Development Plan (MDP).

⁵ Adapted from the MDP.

3.5 MEASURE EVALUATION IN THE MEASURE LIFECYCLE

Measure developers should apply standardized evaluation criteria to their measure throughout the Measure Lifecycle. The more effectively measure properties meet evaluation criteria, the more likely the measure will be robust and meaningful. Measure developers should strive to identify weaknesses in the justification for their measure ^①—through applying evaluation criteria—and revise and strengthen the measure during development. The [Measure Evaluation Report Template and Instructions](#) [↗] and documents on the [NQF Submitting Standards page](#) [↗] are available to assist the measure developer in documenting measure evaluation. The measure evaluation criteria are

- importance to measure and report ^①, including evidence and performance gaps, and priority (i.e., impact)
- scientific acceptability of measure properties ^①, including reliability ^① and validity ^①
- feasibility ^①
- usability ^① and use
- comparison to related ^① or competing measures ^①—harmonization ^①

3.5.1 Applying Measure Evaluation Criteria

The Blueprint addresses the evaluation criteria in several chapters and supplemental materials. In the [Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement](#) [↗], NQF provides measure developers with guidance on applying the criteria. Guidance also facilitates a systematic approach for applying measure evaluation criteria, rating strength of the measure, and tracking results. Results help the measure developer identify how to refine and strengthen the measure as it moves through the development and evaluation process. The aforementioned documents function as a grading rubric, enabling measure developers to anticipate results of the measure evaluation when reviewed by the TEP and the public, and potentially, when submitted to NQF. Throughout measure development, measure developers should evaluate the measure to determine the degree to which the measure is consistent with standardized evaluation criteria. Measure developers use resulting evaluation information to determine whether and how to modify the measure to increase its importance, scientific acceptability, feasibility, and usability and use.

[Figure 5](#) shows the process of applying the measure evaluation criteria. Please note, use of the [Measure Information Form \(MIF\)](#) [↗], [Measure Justification Form \(MJF\)](#) [↗], and [Measure Evaluation Report](#) [↗] templates is voluntary for non-CMS-contracted measure developers.

Measure developers should apply measure evaluation criteria during

- information gathering to guide the search for appropriate measures and measure concepts
- TEP meetings to inform TEP members and contribute to meaningful deliberation
- testing and refinement of specifications ^① to strengthen the measure
- development of a testing plan
- evaluation of test results

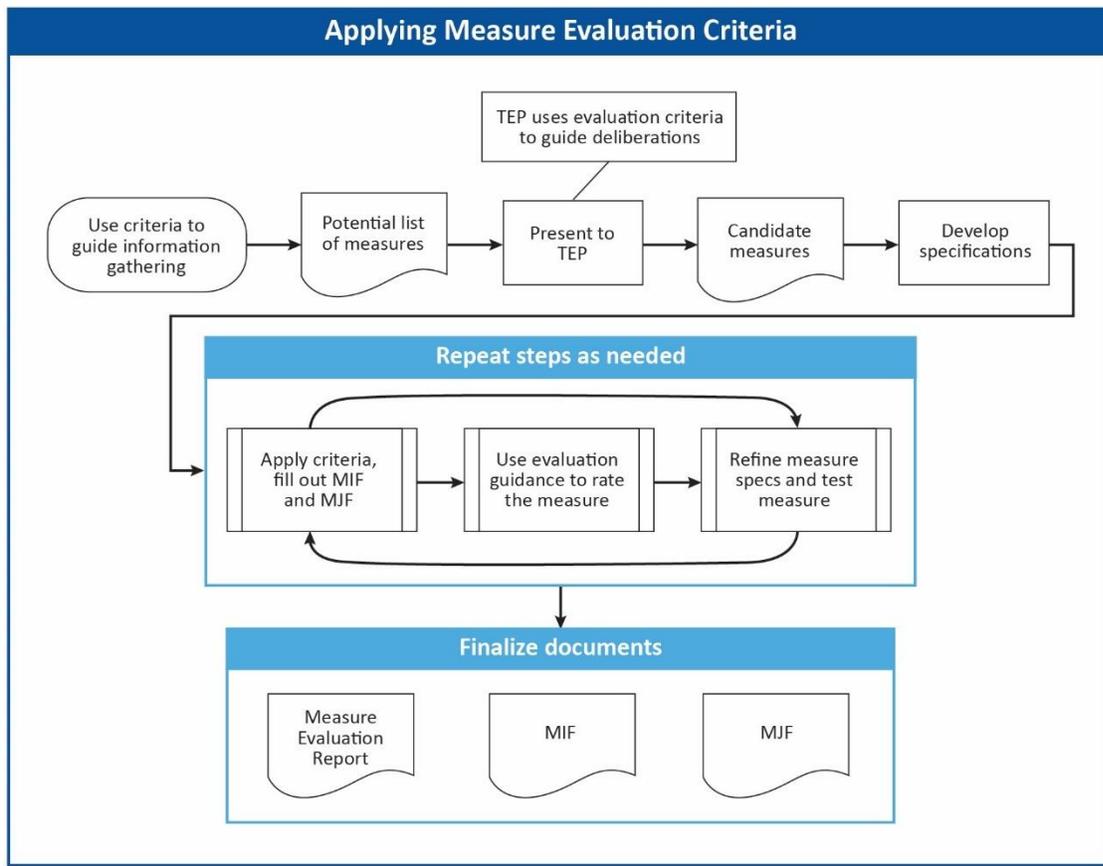


Figure 5. Applying Measure Evaluation Criteria

3.5.2 Timing of Measure Evaluation

Evaluation may be useful during all stages of the Measure Lifecycle. An updated evaluation enables the measure developer to make corrections and strengthen any weaknesses at each point rather than waiting for the end of the lifecycle. This section provides guidance for the measure developer at each stage of the Measure Lifecycle.

Measure [ⓘ] Conceptualization

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

Measure Specification [ⓘ]

- Report how the measure’s proposed technical specifications function.
- Evaluate how the risk model works for outcome measures [ⓘ].

Measure Testing [ⓘ]

- Apply evaluation criteria when analyzing test results.
- Review updated measure specifications and justification according to evaluation criteria.

Measure Implementation

- During endorsement consideration, respond to questions or suggestions made by the NQF Standing Committee by updating the report.
- Support CMS by providing requested information on the [business case](#) during [MAP](#) deliberations.

[Measure Use](#), [Continuing Evaluation](#), and [Maintenance](#)

- Apply evaluation criteria during comprehensive reevaluation to review performance.
- Update [measure specifications](#) and justification based on the evaluation.

The measure developer must evaluate the measure as objectively as possible, for example, to help anticipate any issues that may arise if submitted to NQF for endorsement.

3.6 MEASURE CATEGORIZATION

Measures may be categorized according to a variety of schemes, including measurement domain,⁶ Meaningful Measurement area, or [NQF submission types](#). Legislation, consensus, or other methodology may dictate measure categorization types and names and the types and names can and do change over time. As such, complete alignment in types and names would be difficult, if not impossible, to attain. [Table 2](#) provides the primary types of measures with Blueprint definitions.

Table 2. Measure Definitions

Measure Type	Definition	Source
Composite Measure	A composite measure is a measure that contains two or more individual measures, resulting in a single measure and a single score.	Adapted from CMIT
Cost/Resource Use Measure	A cost/resource use measure is a measure of health services counts (in terms of units or dollars) applied to a population or event (including diagnoses, procedures, or encounters). A resource use measure counts the frequency of use of defined health system resources. Some may further apply a dollar amount (e.g., allowable charges, paid amounts, or standardized prices) to each unit of resource use.	Adapted from CMIT and NQF definition of Cost/Resource
Efficiency Measure	An efficiency measure is the cost of care (inputs to the health system in the form of expenditures and other resources) associated with a specified level of health outcome.	Adapted from CMIT, 2020 MUC , List, and Cylus, Papanicolas, & Smith (2016)
Intermediate Outcome Measure	An intermediate outcome measure is a measure that assesses the change produced by a healthcare intervention that leads to a long-term outcome.	Adapted from MUC User Guide and NQF
Outcome Measure	An outcome measure is a measure that focuses on the health status of a patient (or change in health status) resulting from healthcare — desirable or adverse.	NQF
Patient-Reported Outcome-Based	A patient-reported outcome-based performance measure (PRO-PM) is a performance measure that is based on	Adapted from the MUC User Guide

⁶ Note that CMS and other HHS agencies define and use the term “domain” and classify measures differently from one another; within this Blueprint, the term “domain” is defined differently in different contexts, depending on the relevant agency within the discussion and different measure classification types.

Measure Type	Definition	Source
Performance Measure (PRO-PM) ❶	patient-reported outcome measure (PROM) ❶ data aggregated for an accountable healthcare entity. The data are collected directly from the patient using the PROM tool, which can be an instrument, scale, or single-item measure.	
Population Health Quality Measure ❶	A population health quality measure is a <u>measure ❶</u> of the health outcomes of a group of individuals, including the distribution of such outcomes within the group.	Adapted from Kindig & Stoddard (2003)
Process Measure ❶	A process measure is 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.	CMIT
Structure Measure ❶	A structure measure, also known as a structural measure, is a measure that assesses features of a healthcare organization or clinician relevant to its capacity to provide good healthcare.	CMIT and NQF

3.7 ROLES IN MEASURE DEVELOPMENT

Many entities are involved in measure development. Throughout the Blueprint there is an emphasis on stakeholder engagement. These stakeholders include TEPs, persons and families, clinicians (measured entities), and the public.

3.7.1 Measure Stewards

The NQF [Phrasebook](#) ❷ defines measure steward ❶ as an individual or organization that owns a measure and is responsible for maintaining the measure. Measure stewards may also be measure developers. Measure stewards are the ongoing point of contact for others interested in a given measure. CMS is the steward for most measures developed under contract for CMS. Stewards have permission to approve, reject, and publish measures that their assigned measure developer creates and submits. Stewards provide overall coordination and management of the measures created by measure developers under a specific program or for a specific purpose. Stewards are responsible for approving measure content. Stewards may withdraw measures from approval.

3.7.2 Measure Developers

Measure developers, as directed by the measure steward, are responsible for the development, implementation, and maintenance of measures ❶. Measure developers create, edit, and submit measures to a designated steward. CMS contracts with organizations to develop measures, but other organizations also develop measures, for example specialty societies and the National Committee for Quality Assurance (NCQA). Measure developers submit measures to their assigned stewards for approval. It is also the responsibility of the measure developer to circulate their measure content for feedback and to collaborate on potential measure changes suggested by other authors or other entities.

3.7.3 The Measures Management System (MMS)

The [MMS](#) ❷ is a standardized system for developing and maintaining the quality measures ❶ used in CMS's various quality initiatives and programs. The primary goal of the MMS is to provide guidance to measure developers to help them produce high-caliber healthcare quality measures. CMS-funded

measure developers (or contractors) should follow this Blueprint, which documents the core set of business processes and decisions criteria when developing, implementing, and maintaining measures.

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

3.7.4 Stakeholders

CMS conducts its measurement activities in a transparent manner. Section 101(f) of [MACRA](#) requires that CMS solicit, accept, and respond to input from stakeholders, TEPs, persons and families, the public, physician specialty societies, applicable practitioner organizations, clinicians, stakeholders for episode groups (e.g., care episode groups, patient-condition groups), and others.

When the Blueprint discusses stakeholders, the reference is to these external stakeholders. Throughout the Blueprint there is an emphasis on stakeholder engagement. Stakeholders need to be involved early and often throughout the Measure Lifecycle. These varying perspectives allow for more balance and transparency in measure development and maintenance processes. Patients and families are extremely important stakeholders in the quality measurement enterprise and CMS is committed to gathering their input during priorities planning and throughout the Measure Lifecycle. [Figure 6](#) depicts actions, goals, and sample activities of stakeholders in the stages of the Measure Lifecycle.

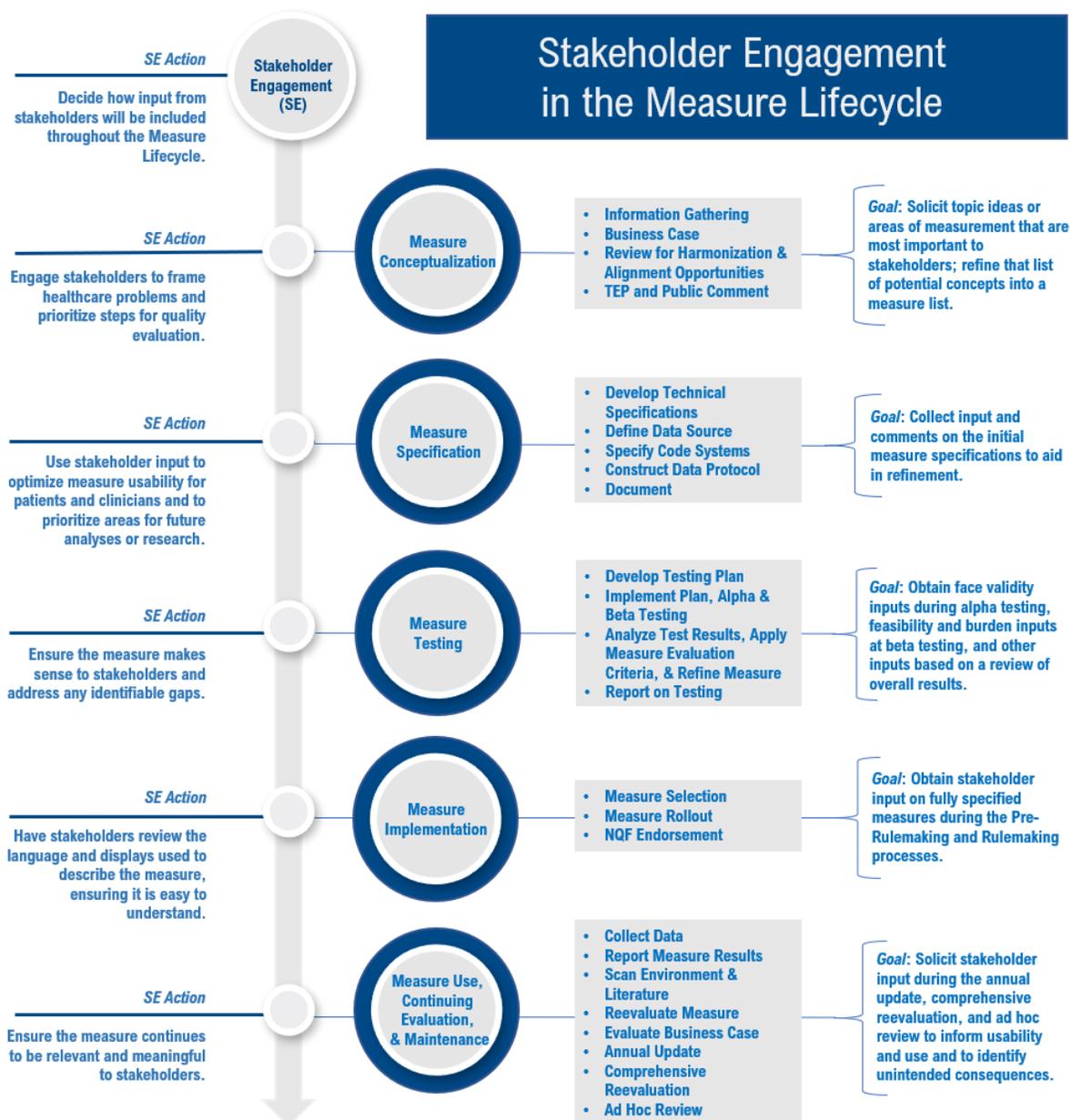


Figure 6. Stakeholder Engagement in the Measure Lifecycle

More detail about ways measure developers can elicit the patient’s perspective is included in [Chapter 4.3.2, Person and Family Engagement \(PFE\)](#), and in the [Person and Family Engagement in Quality Measurement](#) supplemental material.

3.7.4.1 Measured Entities

Measured entities are the front-line clinicians and their organizations, including [health information technology](#), collecting quality measurement data. Measured entities are the implementers of [quality measures](#). The effect of quality measure data collection on clinician workflow can be negative. There may be effects on their payments, positive and negative, with respect to reporting and actual

performance on [quality measures](#)^①. Because of these potential effects, measured entities should be involved in all aspects of the Measure Lifecycle.

3.7.4.2 *Persons and Families*

Strengthening persons and families as partners in care is important to CMS. Involving persons and family representatives in the [measure](#)^① development process is one of the ways that CMS is striving to achieve this goal. Patients and families are extremely important stakeholders in the quality measurement enterprise and their input is particularly valuable to CMS. Engaging persons and family representatives benefits consumers by helping to identify issues that are important and meaningful from their perspective. It also supports identification of information that consumers need in order to make informed healthcare decisions. PFE helps measure developers and CMS produce easily understood, high-quality measures, relevant and useful to consumers. The involvement of persons and families helps CMS develop messaging that resonates with and reflects healthcare quality issues that are important to the public.

Per the CMS [Person & Family Engagement Strategy](#)[↗], “the term ‘person’ is used to reflect an individual’s identity as more than a patient, to recognize his or her participation in prevention and wellness.” Also, *family* “is used broadly to include participants in a person’s healthcare, including informal caregivers, along with the primary caregivers of persons who are in need of the support of their caregivers to make informed healthcare decisions.” Advocates and advocacy groups can also be involved to provide the person and family perspective.

There are multiple ways to obtain information from patients early in the process, including having informal conversations with patients, conducting focus groups, and adding patients or their caregivers to a TEP.

Find more information on PFE in [Chapter 4.3.2, Person and Family Engagement \(PFE\)](#) and in the [Person and Family Engagement in Quality Measurement](#)[↗] supplemental material.

3.7.4.3 *The Public*

The measure developer should obtain comments from the public at several points during the Measure Lifecycle. Defined public comment periods are consistent with Lean principles because they enable early identification of potential issues. Addressing issues raised in public comments can prevent errors and rework later. Once resolution of such issues occurs, the measures will perform better when proposed for use in specific programs. There is some flexibility to determine the best time to obtain comments during measure development, depending on the needs of the measure developers, relative to specific measures and programs.

For more information on the public comment processes, see [Chapter 4.3.3, Public Comment](#).

3.7.4.4 *Technical Expert Panel*

One of the most commonly used stakeholder engagement methods is the TEP: a group of experts and other stakeholders who contribute guidance and thoughtful input to the measure developer in every stage of the measure development process, from conceptualization through maintenance. TEPs are composed of representatives from multiple stakeholder groups for the purpose of obtaining balanced input that represents varied perspectives. Since one main purpose of quality measures is to provide information to patients and caregivers, measure developers should include them as members of the TEP. Inclusion on the TEP affords patients and caregivers the opportunity to provide their perspective on

what is important and useful to measure and evaluate. Find more information on TEPs in [Chapter 4.3.1, Technical Expert Panel](#) and the [Technical Expert Panels](#)^[2] supplemental material.

3.7.4.5 *Other Stakeholders*

Other stakeholders can be engaged in the Measure Lifecycle. For example, measure developers may interview subject matter experts (SMEs), conduct focus groups, and provide other opportunities for people to weigh in during the Measure Lifecycle outside of the more formal TEP and public comment periods. For eCQMs^①, the [Measure Collaboration \(MC\) Workspace](#)^[2] is a place for other stakeholders to provide comments on different stages of the eCQM Measure Lifecycle. There are numerous help desks available to assist stakeholders. Questions addressed to help desks can guide education and outreach and help refine [specifications](#)^① and data collection guidance.

4 MEASURE CONCEPTUALIZATION

In the first stage of the Measure Lifecycle, [measure](#) conceptualization, measure developers should

- Identify an opportunity for healthcare quality improvement.
- Begin to quantify in the [business case](#) the potential impact on patients by compiling evidence for the resulting concept or quality construct and the basic elements of the measure (e.g., [numerator](#), [denominator](#)).
- Demonstrate a need for the measure by identifying the healthcare priority area, defining a [conceptual framework](#), performing an [environmental scan](#) and gap analysis, and gathering stakeholder input.

Measure conceptualization initiates the Measure Lifecycle. It includes identification of measure concept(s) by researching and scanning the environment. This exploration encompasses researching a variety of sources, analyzing measure gaps and conducting other types of analyses, developing a business case, and engaging multiple stakeholders. The end goal of measure conceptualization is a meaningful, well-researched measure concept with well-defined initial components (e.g., [initial population](#), [denominator](#), [numerator](#)).

This chapter discusses the main components of measure conceptualization, each of which plays a critical role throughout the Measure Lifecycle:

- information gathering
- business case development
- stakeholder input
 - TEP
 - PFE
 - public comment

Before beginning measure development, measure developers should consider whether to adopt or respecify an existing [related](#) or [competing measure](#) to fit the desired purpose. Examples of resources measure developers may use to identify existing measures are [CMIT](#), the [NQF Quality Positioning System \(QPS\)](#), and [Qualified Clinical Data Registry \(QCDR\)](#) lists of measures. To review QCDR measures, go to the [Quality Payment Program \(QPP\) Quality Measures Requirements](#) website. If the measure developer does not identify a measure that matches the desired purpose, they should work with a TEP to develop a new measure. The TEP will consider concepts and measures throughout the information gathering process, including application of the measure evaluation criteria.

[Figure 7](#) depicts measure conceptualization in the context of the entire Measure Lifecycle. The steps are iterative and not necessarily sequential.

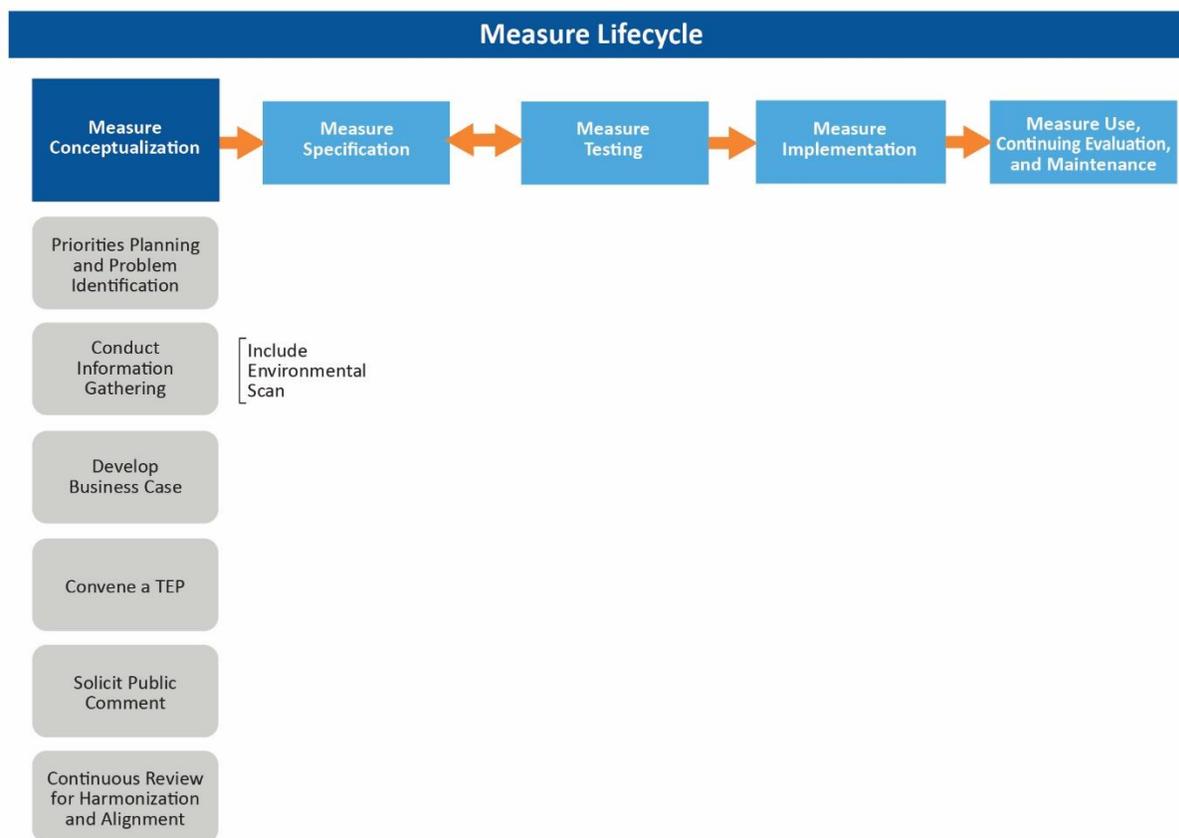


Figure 7. Flow of the Measure Lifecycle—Measure Conceptualization

4.1 INFORMATION GATHERING

Information gathering activities include an environmental scan (e.g., review of literature, search for clinical practice guidelines and existing measures, input from experts, other related activities) and empirical data analysis. These activities yield information that will guide prioritization of topics or conditions, gap analysis, the business case, and the compilation of existing related and competing measures. This section describes the various sources of information as well as instructions for documentation and analysis.

Comprehensive information gathering should result in a significant knowledge base that includes the quality goals, relative strength or weakness of scientific evidence pertinent to the topics or conditions of interest, and information with which to build a business case for the measure. It will also produce evidence of either general agreement or diverse and conflicting views on the quality issues pertinent to the topics or conditions of interest. Measure developers should explore underlying costs associated with the condition, procedure, or healthcare issue early in information gathering to contribute to an eventual return on investment calculation as a part of the business case.

At a minimum, five measure evaluation criteria—importance, scientific acceptability, feasibility, usability and use, and related and competing measures—serve as a guide for gathering information and for identifying priority topics/conditions and measurement areas. The fifth criterion—consideration of related and competing measures—refers to measure harmonization;

measure developers should consider this criterion from the very beginning of the measure development process.

Early in measure conceptualization, it is important for measure developers to consider what other measure developers have learned regarding [feasibility](#)^① or scientific evidence from existing measures, previous measure concepts, or [measure data elements](#)^①. Examination of prior work allows the measure developer to use data elements previously identified as feasible in the measure they are developing and/or determine ways to improve feasibility. Awareness of prior work also helps the measure developer identify possible unintended negative consequences and avoid or mitigate them to avoid rework later in the process.

Information gathering consists of eight steps, which may not occur sequentially

- identify the healthcare quality issue to be addressed and determine its priority area
- conduct an [environmental scan](#)^① (refer to the [Environmental Scanning for Quality Measurement](#)[↗] supplemental material)
- analyze empirical data, when available
- evaluate information collected during the environmental scan and empirical data analysis
- conduct a measurement gap analysis to identify areas for new measure development
- justify creation of a new measure
- apply measure evaluation criteria
- prepare an initial list of measures or measure topics

While listed under measure conceptualization, measure developers conduct information gathering in other stages and steps, e.g., during respecification and reevaluation.

Additional details about these steps are provided in [Chapter 4.1.1 - 4.1.8](#) and the [Environmental Scanning for Quality Measurement](#)[↗] supplemental material. Several templates are available to guide documentation of information gathering, measure information, and measure justification.

4.1.1 Identify the Healthcare Quality Issues to be Addressed and Determine Their Priority Area

The measure developer should clearly define the nature of the measure's focus and scope of the measure's construct and its relation to healthcare needs and quality improvement. The measure developer should consider quality priorities as well as Medicare, Medicaid, and other payor top volume and top cost conditions, as appropriate.

4.1.2 Conduct an Environmental Scan

The environmental scan is essential in building the case for a [quality measure](#)^①. It serves as the foundation for the measurement plan. Developing a broad-based environmental scan that includes a strong review of the literature, regulatory environment, economic environment, and stakeholder needs and capabilities will guide thinking and decision-making. A strong, comprehensive environmental scan will improve the likelihood of project success.

Among the many important areas to scan, measure developers must consider the six domains of quality, outlined in [Crossing the Quality Chasm \(2001\)](#)[↗], which include safety, timeliness, efficiency, effectiveness, equitability, and patient centeredness. Also refer to the [Meaningful Measures Initiative](#)[↗], which identifies the priorities for quality measurement and improvement. Measure developers must explore various dimensions of quality to develop informative quality measures. The resulting report of the environmental scan will reflect the [data sources](#)^① depicted in [Figure 8:](#)

- identification of related ① or competing measures ①, including opportunities for consolidation, harmonization ①, and alignment ①
- listing of clinical guidelines pertinent to the clinical domain or topic
- review of studies that document the success of measures ① in the same or similar healthcare setting or domain
- discussion of scientific evidence supporting clinical solutions that might serve as a basis for the measure

The environmental scan ① includes a review and evaluation of both peer-reviewed and grey literature ①, clinical practice guidelines ①, legislation and regulations and their implications on measurement, the study of clinical decision support artifacts, existing related and competing measures, empirical data, expert input (including input from the TEP and other experts), and stakeholder input—inclusive of all relevant stakeholders, including patients and caregivers.

The [Environmental Scanning Support Tool \(ESST\)](#) ②, a tool within CMIT, is available to aid measure developers in reviewing articles pertinent to existing measures. The ESST reduces the time needed to scan literature from months to hours, saving substantial resources. Instructions for how to use the ESST are available on the [ESST site](#) ②. In the case of a new measure, the measure developer must identify any measures in current use that might be appropriate for the specific healthcare need or project. In addition to CMIT, QPS, and the QCDR ① list of measures, these measures are identified through analysis of resources, including employer plans, commercial plans, managed care plans, Core Quality Measure Collaborative (CQMC), NQF, Medicare Payment Advisory Commission (MedPAC), National Academy of Medicine (NAM), Institute for Healthcare Improvement (IHI), Veterans Health Administration (VHA), and the Defense Health Agency (DHA) (e.g., TRICARE). The measure developer may also conduct interviews or post a Call for Measures to identify measures currently in use or under development. Other calls may go out during measure implementation, such as a call for fully developed measures for consideration for implementation in CMS programs.

For more information about conducting an environmental scan, see the [Environmental Scans for Quality Measurement](#) ② supplemental material.

4.1.3 Analyze Empirical Data, as Appropriate

If empirical data are available, measure developers should analyze the data statistically to support the importance of the measure, identify gaps or variations in care, and provide incidence/prevalence information and other data, e.g., return on investment (ROI), necessary for development of the business case ①. Empirical data analysis may also provide quantitative evidence for inclusion or exclusion of a set of populations ① or geographic regions or other considerations for development of the measure.

Measure developers can also analyze empirical data to test the feasibility ① of data elements ① required for a measure, such as data availability (including standardization) and accuracy of data information. They may use empirical data to help identify feasibility concerns early in development of the measure. Measure developers may need to replace or revise data elements, consider an alternative

Environmental Scan Data Sources

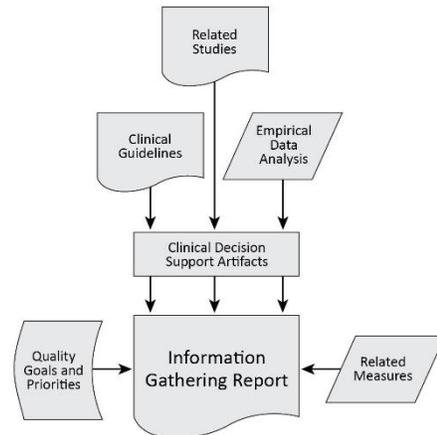


Figure 8. Environmental Scan Data Sources

measure type, assess implementation burden versus value of the measure①, or recommend halting further development of the measure concept.

If developing a risk-adjusted measure, measure developers should assess feasibility of the risk variables early on.

4.1.4 Evaluate Information Collected During Environmental Scan and Empirical Data Analysis

If the environmental scan① discovers related measures①, measure developers should evaluate the measures to assess whether they meet the needs of the project. If a related measure is found with a measure focus appropriate to the needs of the project, but the measure is specified for a different population①, setting, or data source①, the measure developer may be able to respecify the measure for the new use and test it for reliability① and validity① specific to the new population.

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

The purpose of a gap analysis is to identify measure types or concepts that may be missing for the measure topic or focus. The measure developer uses information collected from the environmental scan, measure gap analysis, and other information gathering activities to identify existing competing or related measures before deciding to develop new measures. If no related or competing measures① can be respecified or adopted, then it is appropriate to develop a new measure. Measure developers should establish a framework to organize any existing measures.

4.1.6 Justify the Creation of New Measures

If no existing measures are suitable for adoption or respecification, then the measure developer may develop a new measure. They must justify the new measure by gathering supporting information, which will vary by type of measure, and which will contribute to the business case①.

The goal is to develop a measure most proximal to the outcome desired. Measure developers should avoid selecting or constructing a measure that can be met primarily through documentation—often satisfied with a checkbox, date, or code—for example, completing an assessment, care plan, or delivered instructions—without evaluating quality of the activity. Measure developers should consider these guidelines in their justification of a new measure.

- For an outcome measure①, there should be a rationale supporting the relationship of the health outcome to processes or structures of care. Specifically, there must be at least one healthcare-related structure, process, service, or intervention that can improve performance on the outcome.
- For an intermediate outcome① measure, there should be a body of evidence that the measured intermediate outcome leads to a desired health outcome.
- For a process measure①, there should be a body of evidence that links the measured process to a desired health outcome.
- For a structure measure①, provide evidence that there is a link from specific structural elements to improved care and improved health outcomes.
- For a cost and resource use measure①, link with measures of quality care for the same topic.
- For all types of measures, ensure the unit of analysis aligns with an appropriate accountable entity (e.g., payor, hospital, or clinician). Consider the extent to which processes are under control of the measured entity. Attribute the measure topic to an appropriate provider or setting. In some cases, there is “shared accountability.” For example, for measures of

functional outcomes and care coordination, no single provider controls performance results so the unit of analysis may be at the ACO- or payor-level rather than at the clinician-level.

- For a composite measure ①, provide the conceptual rationale.

For more information about developing a business case ①, see [Chapter 4.2, Business Case Development](#).

4.1.7 Apply Measure Evaluation Criteria

If the measure developer identifies many measures ① or concepts, they should narrow the list of potential measures by applying measure evaluation criteria—especially the importance ① and feasibility ① criteria—to determine which measures should move forward. At a minimum, they should consider the measure’s relevance to the population ①, effects on healthcare costs, gaps in care, availability of well-established, evidence-based clinical guidelines, and/or translate supporting empirical evidence ① into meaningful quality measures ①. The measure developer should explore possible data sources ① while considering feasibility (e.g., understanding the data captured in EHRs ①) and include other criteria depending on the specific circumstances of the measure or measure set ①.

4.1.8 Prepare an Initial List of Measures or Measure Topics

The measure developer should create an initial list of measures based on results of the previous steps. The measure developer then provides recommendations based on results of the environmental scan ①, measure gap analysis, initial feasibility assessment, and other information collected during the information gathering process. This list may contain adopted measures ①, respecified measures ①, new measures, or measure concepts.

Before proposing a new measure, the measure developer should evaluate the literature for quality, quantity, and consistency. The measure developer then reviews the appropriateness of any clinical guidelines used as the basis for the measure to make sure the measure is based on a key leverage point and explores possible data sources for the new measure.

4.2 BUSINESS CASE DEVELOPMENT

Once the measure developer has narrowed the list of candidate measures, they should develop a business case for each remaining concept. The business case documents anticipated impacts of a quality measure, including financial outcomes, and resources required for measure development and implementation. Despite what the name suggests, there is no limitation to a description of economic benefits and costs. Impacts and outcomes resulting from a measure may include positive clinical outcomes such as preservation of healthy lifestyles for patients, lives saved, costs reduced, complications prevented, clinical practice improved, and patient experience enhanced.

Anticipated benefits made explicit in the [business case](#) ① should outweigh the costs and burden of measure development, data collection, and implementation for the quality measure. The measure developer should evaluate and report on all potential positive and negative impacts ([see the call out box](#)). For example, a [measure](#) ① intended to reduce long-term [mortality](#) ① through early detection and treatment may cause increased short-term costs and potential complications from screening tests. The business case should demonstrate

- the need for the measure
- how the measure will further the aims and objectives of the project
- the value of the measure
- why this measure is the best balance of cost, benefits, and risks
- whether the measure is sensitive to changes in behavior or policy such that improvements in measure performance reflect improvements in care delivery
- realistic and affordable costs
- sufficient capacity within the system to implement the measure

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

- better care through reduction of harm and positive influence on patients' perception of their care
- better health through reduction in mortality and [morbidity](#) ① and improvements in quality of life
- more affordable care through cost savings

By documenting the potential improvement anticipated from implementing a specific [measure](#) ①, the measure developer can make a strong case by explaining why the organization should invest resources in development (or continued use) of the specific measure in its quality initiatives. At a minimum, the business case for a measure should state explicitly, in economic and societal terms, the expected costs and benefits of the measure.

The business case supports the measure importance evaluation [criterion](#) ①, in part, by creating a model that predicts measure performance and the impact it will have on health and financial outcomes. As such, the measure developer should refer to the [importance criterion](#) ① in the Measure Evaluation Report and in the [Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement](#) ② before developing a business case. Development of the business case starts early during measure

Potential Costs, Benefits, and Savings

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

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

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

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

Materials: product purchase, maintenance, storage, and disposal

conceptualization. The business case ❶ evolves throughout the Measure Lifecycle and the measure developer uses it to compare anticipated to actual results during measure ❶ reevaluation and maintenance ❶ (Table 3). While NQF does not require a formal business case, the measure developer may use many of the elements outlined in the business case in the [NQF Measure Submission Form](#) ❷.

Table 3. Refinement of the business case is an ongoing process that occurs throughout the Measure Lifecycle.

Measure Conceptualization	Measure Specification	Measure Testing	Measure Implementation	Measure Use, Continuing Evaluation, and Maintenance
Gather data to assess pros and cons of measure implementation and to inform continued measure development Begin drafting business case	Initial business case	Business case update(s)	Further business case update(s) based on implementation	Assess measure performance pertaining to the business case Determination whether the business case adequately captures benefits, outcomes, and costs

4.2.1 Business Case Best Practices

A well-constructed business case consists of five key elements.

- precise statement of need
- business impact
- proposed solution/alternatives
- benefits estimation
- cost estimation

The business case executive summary should focus on what information is available and provide a concise, high-level overview (maximum 500 words).

[Figure 9](#) shows the overall flow of inputs to business case development, and [Figure 10](#) diagrams business case inputs and the impact of the business case throughout the Measure Lifecycle.

The measure developer should evaluate the business case periodically during measure development and maintenance. Evaluating the strength of the business case is ongoing during measure development because the business case helps to justify continued development of the measure.

While other models may be used, a cost savings model is the most prevalent for evaluating the potential quality measure's ❶ business case (i.e., the aggregate effect of cash inflows and outflows accruing to an organization as a result of implementing a specific process or treatment). This model presents a more easily interpreted result, given its quantitative method, and the measure developer can make reliable comparisons to ranking multiple events. If there is no expectation of savings until future years, adjust the savings to a net present value. This model also applies to many outcome measures ❶. For example,

if the requirement is for increased physician follow-up visits to reduce hospital readmissions, the savings equals the cost saved by not being readmitted minus the cost of the additional physician visits.

Flow of Inputs into Business Case Development

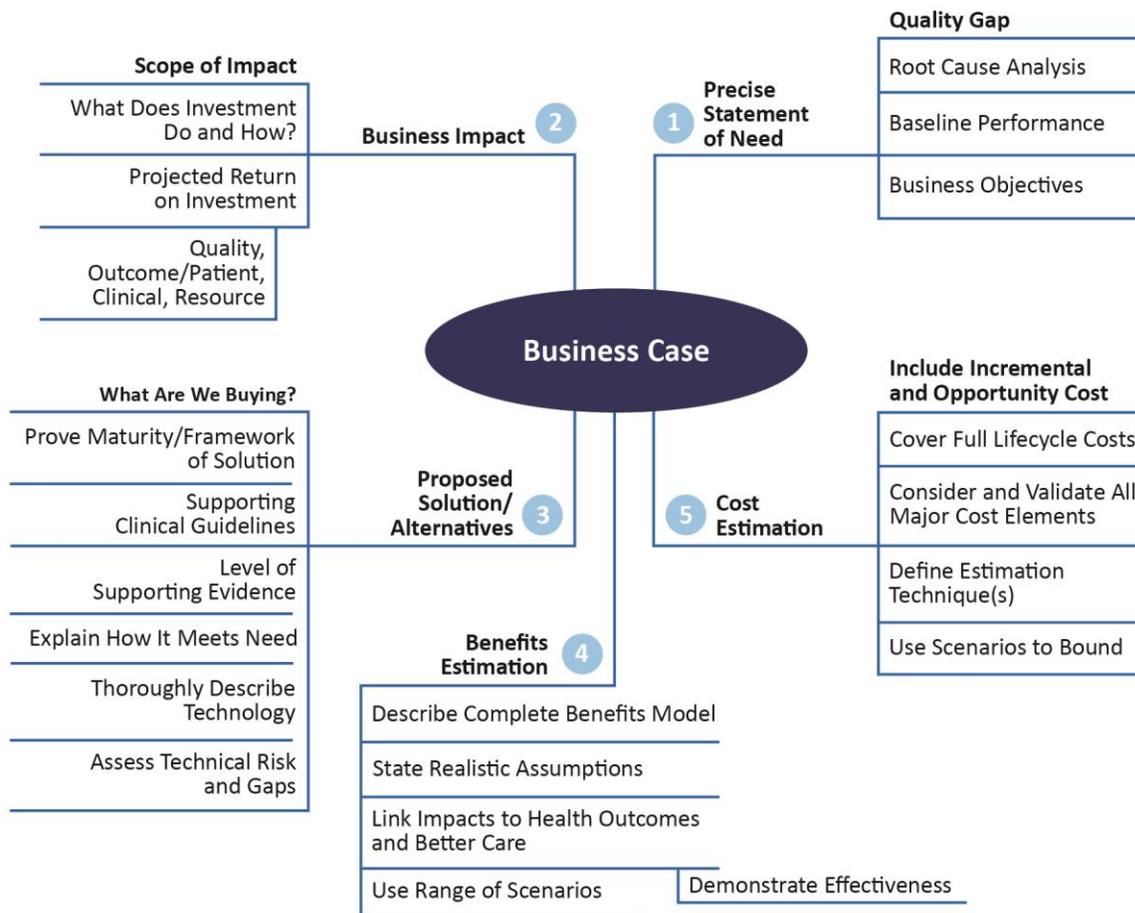


Figure 9. Flow of Inputs into Business Case Development

Measure developers should measure better health and better care with quantifiable anticipated benefits assigned to a model for testing. The measure developer should support these assertions with high-quality, consistent evidence. Using the example of increased physician follow-up visits, improved care coordination may not only reduce expenses associated with unnecessary readmissions, but could also reduce mortality in selected populations and improve patient satisfaction.

Regardless of the model used, the business case should include an explicitly stated hypothesis for use in later testing and, at a minimum, predict how the measure will have an effect over time (the trajectory). This enables the measure developer to make comparisons during measure use, continuing evaluation, and maintenance. When possible, the measure developer should include historical and baseline data—that is, data collected from the measure (if completing for maintenance) or similar measures (identified during the environmental scan).

Inputs and Uses for the Business Case

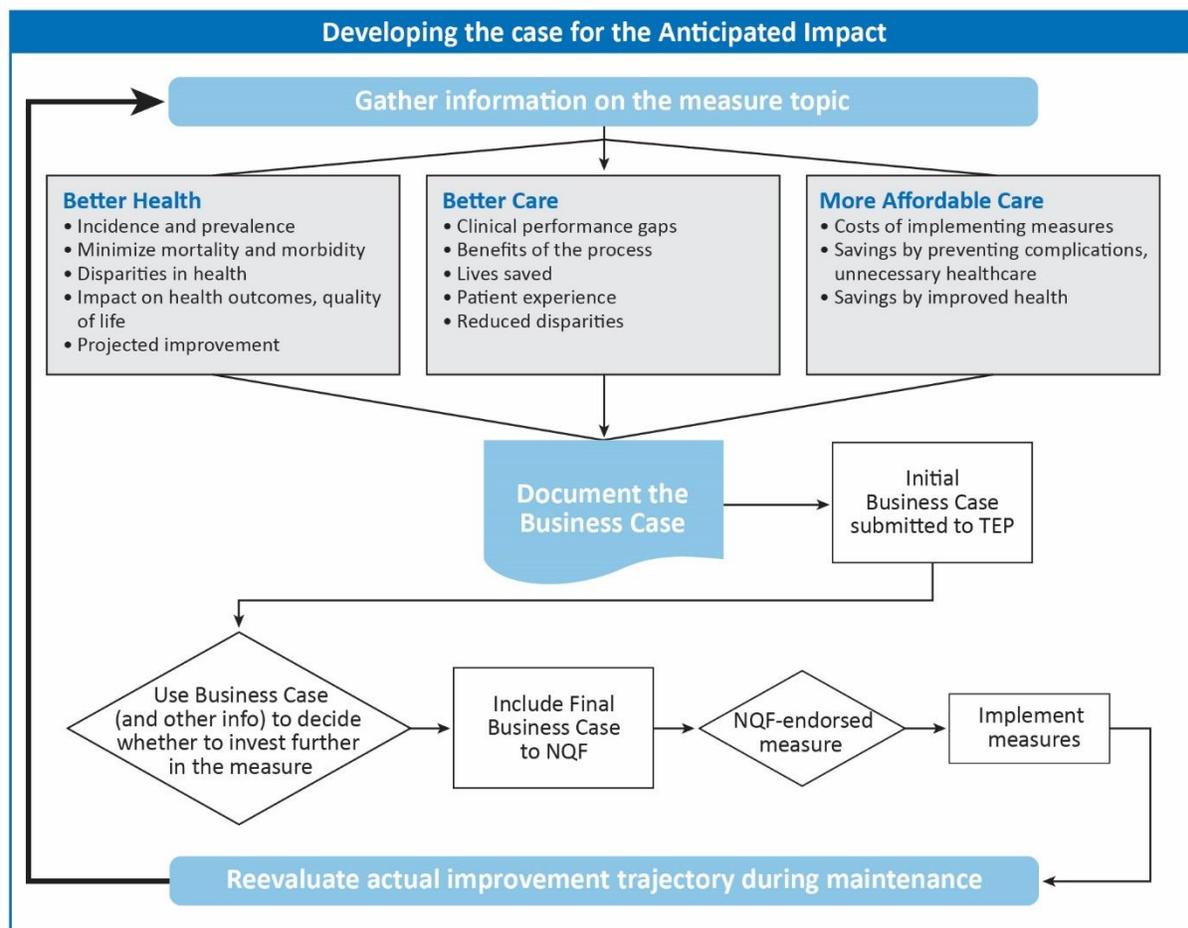


Figure 10. Inputs and Uses for the Business Case

Acronym Legend	
TEP:	Technical Expert Panel
NQF:	National Quality Forum

After implementation of the measure ①, the measure developer should compare the business case ① and predictions about measure performance, which helped inform decision-making during measure development and selection for use, against actual performance. If anticipated improvements in health, provider care performance, and cost savings occur as predicted, then the measure is succeeding in terms of the business case. If anticipated improvements are absent, then the measure developer should reexamine the data, reevaluate the justification for the measure, and analyze why improvements are not occurring. The measure developer then should adjust the business case for any changes in the environment and revisit the initial assumptions as needed. For annual updates of measures in use and for continuing evaluation, simply reporting performance relative to predictions may be sufficient. For the comprehensive reevaluation, the measure developer should conduct a full analysis, and then report recommendations for improvement.

For more information, refer to the [Business Case Form and Instructions](#)²⁴ to assist in the documentation of the business case.

4.3 STAKEHOLDER ENGAGEMENT

During the [measure](#) conceptualization stage, different types of stakeholders have the opportunity to suggest topics or areas of measurement important to them. It is important for stakeholders to be able to present ideas to those who develop measures very early in the process to influence the direction of measure development. The measure developer should initiate TEPs during measure conceptualization and solicit public comments during measure conceptualization and information gathering. This section discusses the role of three types of stakeholders: TEPs, persons and families, and the public. [Chapter 4.3.4](#) discusses other stakeholder engagement.

4.3.1 Technical Expert Panel

[Chapter 3.7.4.4](#) introduced TEPs, noting that TEPs are one of the most commonly used stakeholder engagement methods. TEPs are composed of representatives from multiple stakeholder groups for the purpose of obtaining balanced input that represents varied perspectives. As noted in [Chapter 3.7.4](#), measure developers should include persons (patients) and family member representatives on their TEP because these individuals provide a vital perspective on what is important and useful to measure and evaluate. For more information about PFE, see the [Person and Family Engagement](#) supplemental material.

Refer to the [Technical Expert Panels](#) supplemental material for step-by-step instructions for how to establish and conduct TEPs. Several templates are available to assist with the management and documentation of TEPs.

4.3.2 Person and Family Engagement

Strengthening persons and families as partners in care is important to CMS. In this context, a *person* is a non-healthcare professional representing those who receive healthcare. A person may be a patient or a patient representative. Involving persons and family representatives in the measure development process is one of the ways that CMS is striving to achieve this goal. Engaging persons and family representatives benefits consumers by helping to identify issues that are important and meaningful from their perspective. It also supports identification of information that consumers need in order to make informed healthcare decisions. PFE helps measure developers and CMS produce easily understood, high-quality measures, relevant and useful to consumers. The involvement of persons and families helps CMS develop messaging that resonates with and reflects healthcare quality issues that are important to the public.

There are multiple ways to obtain information from patients early in the process, including having informal conversations with patients, conducting focus groups, and adding patients or their caregivers to the TEP.

The measure developer must keep the person's/caregiver's point of view central throughout measure development and provide opportunities for person input during the information gathering process. Involving persons and family representatives in the measure development process (e.g., on TEPs, in focus groups, during testing) is among the many ways that measure developers can accomplish the goal of strengthening PFE. PFE in the measure development process is the process of involving persons and/or family representatives in a meaningful way throughout the Measure Lifecycle.

Prior to measure conceptualization, measure developers should compile a comprehensive plan outlining how they will incorporate person and/or family representative input at each stage of the Measure Lifecycle. Regardless of the engagement methods used, the measure developer provides all individuals

involved with [measure](#) development efforts with clear expectations about what their participation will entail. Measure developers may also consider the principles in the [Patient-Centered Outcomes Research Institute \(PCORI\) Engagement Rubric](#) when engaging consumers (see call out box) and observe best practices for conducting qualitative research, survey and interview construction, and testing, as applicable.

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

- **Reciprocal Relationships:** Define roles and decision-making authority of all involved collaboratively and clearly.
- **Co-Learning:** It is important to ensure that participants understand the measure development process, PFE, and person-centeredness.
- **Partnership:** Value the time and contributions of person partners. Time commitment and attendance requests for persons need to be thoughtful and reasonable. The research team is committed to diversity and demonstrates cultural competency, including disability accommodations, as appropriate.
- **Trust, Transparency, Honesty:** Encourage measure developers to express commitment to open and honest communication with person stakeholders, in a meaningful and usable way, and ensure to make major decisions inclusively.

A discussion of the best practices for engaging persons and family members in measure development activities, including obtaining input from persons and family member stakeholders, is in the [Person and Family Engagement in Quality Measurement](#) supplemental material.

4.3.3 Public Comment

The public comment process is an essential way that measure developers ensure their measure development is using a transparent process with balanced input from relevant stakeholders and other interested parties. The public comment period provides an opportunity for the widest array of interested parties to provide input on the measures under development and to provide critical suggestions not previously considered by the measure developer or the TEP. Defined public comment periods are consistent with process improvement principles, such as Lean, because they enable early identification of potential issues. The *Blueprint* recommends a public comment period for each stage of the Measure Lifecycle because addressing issues raised in public comments can prevent errors and early input can reduce rework in later stages. Once resolution of such issues occurs, the measures will perform better when 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 the measure developers, relative to specific measures and programs. At a minimum, before submitting the fully specified measure for [MAP](#) and NQF endorsement consideration, the measure developer should obtain comments on the fully specified measure. Public comments obtained during measure development (and maintenance) are separate from—and complementary to—the public comments obtained during the NQF endorsement process.

Several templates (e.g., [Public Comment Call Web Posting](#)) are available to assist in creating the Call for Public Comment and documenting results.

4.3.4 Other Stakeholder Engagement

As noted in [Chapter 3.7.4.5](#), other stakeholders can be engaged in the Measure Lifecycle. In the conceptualization stage, measure developers may interview subject matter experts (SMEs), convene focus groups, and present other opportunities for people to weigh in on [measure](#) concepts outside of the more formal TEP and public comment periods. One example is the [Measure Collaboration Workspace](#) (MC Workspace) which has an [eCQM Concepts](#) module that permits anyone with an [eCQI Resource Center](#) account to suggest a concept for an [eCQM](#). Stakeholders have the option to submit an eCQM concept for feedback from other stakeholders and to ultimately submit the eCQM concept to CMS for review. While the *MC Workspace* is likely to attract healthcare and health information technology (IT) professionals, anyone with an interest in eCQI can get an [account](#).

5 MEASURE SPECIFICATION

Development of measure ① specifications ① is an iterative process throughout the Measure Lifecycle. Measure specification consists of both technical specification and harmonization ①, along with stakeholder engagement through public comment and TEPs. Final technical specifications provide comprehensive details that allow collection of measure data and implementation of the measure to be consistent, reliable, and effective.

This chapter provides guidance for measure developers to ensure measures have complete, detailed, clear, rigorous, and precise technical specifications. Measure specification follows the conceptualization stage, when the measure developer identified the initial intent of the measure based on clinical practice guidelines ①, and evidence identified in the environmental scan ①. Special types of measures (e.g., eQMs ①) require additional steps. Multiple supplemental materials document these additional steps for special measure types.

This chapter first addresses measure specification by measure category, followed by a discussion of the processes of defining data sources ①, developing specifications and definitions, specifying the code and/or code systems ①, constructing the data protocol, documenting measures, and harmonization.

Figure 11 depicts the measure specification portion of the Measure Lifecycle.

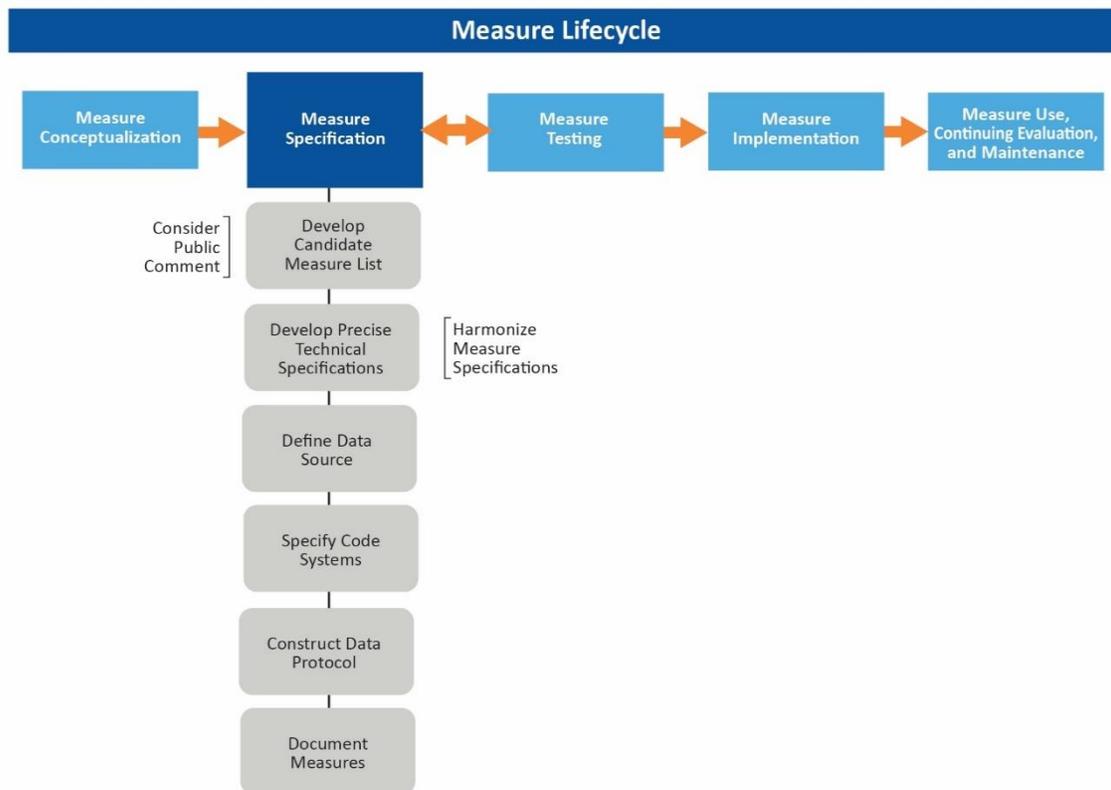


Figure 11. Flow of the Measure Lifecycle—Measure Specification

Measure technical specifications¹ are instructions for how to build and calculate a measure¹. The intent of measure specification is that each measure should reach its appropriate target population¹, but not over-reach or under-reach. Errors in specifying the target population not only waste resources, but also may generate misleading conclusions about care quality.

Developing technical specifications is an iterative process. Prior to drafting initial specifications, the measure developer should consider the data elements¹ necessary for the proposed measure and conduct preliminary feasibility¹ assessments. There could be additional benefit in obtaining preliminary input during the measure specification phase from standards SMEs regarding data model, terminology, data elements and content, [Clinical Quality Language \(CQL\)](#)¹ expression, and impact on clinician workflow. The measure developer then drafts initial specifications, which the TEP and possibly other stakeholders, such as work groups, SMEs, and other developers, will review and may suggest changing.

At this stage, technical specifications are likely to include high-level numerator¹ and denominator¹ statements and initial information on potential exclusions¹, if applicable. The measure developer should continue to detail these specifications and refine them throughout the development process as they obtain more information.

For special types of measures such as cost and resource use measures¹ and composite measures¹, technical specifications may differ slightly in their execution. For more information about cost and resource use, composites, and other special measures refer to the supplemental materials.

The building blocks of a measure in the technical specifications may include, but not limited to

- measure name/title
- measure description
- initial population¹
- numerator statement and definitions
- denominator statement and definitions
- denominator exclusions¹
- numerator exclusions¹
- denominator exceptions¹
- target population
- time interval¹
- stratification¹ scheme, or how to split results to show differences across groups
- risk adjustment¹ methodology
- calculation algorithm¹, or how to calculate results
- sampling methodology
- data source(s)¹
- key terms, data elements, codes, and code systems¹
- level of analysis¹
- attribution¹ model, or how to attribute data to measured entities
- care setting

Different information sources influence development of technical specifications for a measure.

- literature review
- clinical practice guidelines¹
- clinical decision support artifacts

- existing measures
- TEP, SME, and stakeholder input
- public comment
- alpha testing
- beta testing

These inputs will improve the precision of technical specifications¹ and increase validity¹ and reliability¹ of the measure¹. Measure developers should specify measures with sufficient details to be distinguishable from other measures and to support consistent implementation. The *NQF Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement*² notes measures should be specified with the broadest applicability (e.g., target population¹, setting, level of measurement/analysis¹) as supported by the evidence.

Technical specifications inform and are informed by later stages of the development process, particularly measure testing¹. During the development process, alpha (formative) testing of the measure occurs. For measures based on electronic, administrative, or claims-based data, the measure developer may provide draft technical specifications to the programming staff responsible for data retrieval and for developing programming logic necessary to produce the measure. Programmers will assess feasibility of the technical specifications as written and may provide feedback. For measures based on chart abstraction, the measure developer develops and tests data collection tools. Beta (field) testing occurs when there are more fully developed specifications (refer to Chapter 6, Measure Testing). As a result of testing, technical specifications will continue to evolve, becoming more detailed and precise.

5.1 MEASURE SPECIFICATIONS BY MEASURE CATEGORY

Almost all measure technical specifications depend, at least in part, on the category of the specified measure. Measure categories, for the purposes of this chapter, are ratio¹, proportion¹, and continuous variable (CV)¹ measures. A calculation algorithm¹ provides a score¹, and the derivation of that score determines a measure's category.

A ratio is a score that is derived by dividing a count of one type of data by a count of another type of data (e.g., number of patients with central lines who develop infection divided by the number of central line days). The key to the definition of a ratio is that the numerator¹ and the denominator¹ represent the count of different kinds of people, things, events, or objects.

A proportion is a score derived by dividing the number of cases that meet a criterion¹ for quality (the numerator) by the number of eligible cases within a given time frame (the denominator) where the numerator cases are a subset of the denominator cases (e.g., percentage of eligible women with a mammogram performed in the last year).

A CV is a score in which the individual value for the measure can occur along a continuous scale and be aggregated using a variety of methods such as the calculation of a mean or median (e.g., mean number of minutes between the time when a patient presents with chest pain and the time when thrombolytic medications are administered).

Table 4 defines the measure components relevant to the three measure categories.

Table 4. Measure Specifications by Measure Category

	Ratio	Proportion	CV
Initial Population ^①	R	R	R
Denominator ^②	R	R	R*
Denominator Exclusion ^③	O	O	O*
Denominator Exception ^④	NP	O	NP
Numerator ^⑤	R	R	NP
Numerator Exclusion ^⑥	O	NP	NP

R = Required
 O = Optional
 NP = Not Permitted
 *CV measures use *measure population* instead of denominator and *measure population exclusion* instead of denominator exclusion.

The steps outlined and described in detail in [Chapters 5.2 through 5.6](#) are performed to develop the full measure technical specifications^①. With completion of each step, the measure developer updates the [measure^①](#) documentation accordingly. The measure developer should

- define the [data source^①](#)
- develop specifications and definitions
- specify codes and/or [code systems^①](#)
- construct the data protocol
- document measures (ongoing across all steps)

5.2 DEFINE THE DATA SOURCE(S)

Measure specifications should include data sources and methods of data collection that are acceptable. The data source used to calculate a measure will determine the [reliability^①](#), [validity^①](#), [feasibility^①](#), and [usability^①](#) of the measure. If calculated from more than one data source (e.g., registry and [eCQM^①](#)), the measure developer should generate detailed specifications for each data source. Collect evidence that results calculated from the different data sources are comparable.

Examples of data sources include

- administrative data
- claims data
- patient [medical records^①](#) (paper and electronic)
- electronic clinical data (e.g., device data)
- registries
- standardized patient assessments
- patient-reported data and surveys

When contemplating the source of data, the measure developer must consider the feasibility and methods of collecting data from that source. Included in [Table 5](#) are examples of the strengths and limitations of using the different data sources. For more information on different data sources, refer to the [Data Sources for Quality Measurement^①](#) supplemental material.

Table 5. Strengths and Limitations of Different Data Sources

Data Source	Strengths	Limitations
Administrative data	Can provide information not usually found in a clinical database	These data are not collected primarily for the purpose of quality measurement, they are collected for other purposes, e.g., admissions, discharges, and transfers.

Data Source	Strengths	Limitations
	Less burdensome than manual abstraction to providers for data collection	
Claims data	Professionally coded; drawn from large populations ^① (i.e., more representative of the populations of interest) Standardized data Less burdensome than manual abstraction to providers for data collection	These data are not collected primarily for the purpose of quality measurement, they are collected for other purposes, e.g., billing; therefore, they can have varying degrees of clinical detail and are often limited in content, completeness, timeliness, and accuracy.
Paper patient medical record ^①	Detailed clinical data with a rich description of care	Identifying test sites that can serve as data sources can be difficult; abstraction is time-intensive; requires expert staff (cost and time) to interpret each record and input data findings into a format suitable for analysis; abstraction can be open to subjectivity and interpretation or lack of consistency in how data are abstracted.
Electronic patient medical record	Reduced cost of accessing clinical information from the patient medical record; detailed clinical data with a rich description of care	Identifying test sites that can serve as data sources can be difficult; inconsistent adoption of EHR ^① systems, especially across settings; extracting the data requires expertise, time, and money; hurdles related to continuing use of paper notes for point-of-care documentation; use of drop downs and structured fields can reduce the richness of the clinical data and descriptions of care; structured data not always using or mapped to standard terminologies; and potential negative impact on clinical workflow.
Electronic clinical data	Reduced cost of accessing clinical information from the patient medical record; or personal health device (e.g., home blood glucose monitor)	Identifying test sites that can serve as data sources can be difficult; extracting the data requires expertise, time, and money; hurdles related to continuing use of paper notes for point-of-care documentation; device data may be external to the patient medical record; and still only partially implemented in most settings.
Registry	Data from multiple sources and across care settings; often available as an electronic upload ⁷	It is unknown how registry requirements impact workflow; feasibility of data collection is determined by the data requirements imposed by the registry. Registries may impose fees so there may not be representation from all relevant providers and some selection bias for those who choose to participate.
Standardized patient assessment	Well validated and tested	There is a potential for bias as some have mixed use for determining reimbursement, meeting conditions of participation, and assessing quality; may be proprietary, therefore no available non-proprietary reliable or valid tool.
Patient-reported data and survey	Unique source of data available only from the patient or patient’s family/significant other; direct way to collect patient experience	Validated/reliable assessment tools are needed (these may be proprietary); some patient-reported data are not often used in the delivery of care, so likely small number of responses (<i>n</i>); not always reliably or consistently collected; may be costly and time-intensive for data collection.

5.3 DEVELOP SPECIFICATIONS AND DEFINITIONS

The measure developer begins construction of measure^① specifications^① by outlining the initial population^①, numerator^①, denominator^①, denominator exclusions^①, numerator exclusions^①,

⁷ Data are submitted electronically directly from the registry.

exceptions①, and measure logic①. Then, the measure developer gives the measure concept increasing amounts of detail, including precisely defined data elements① and the appropriate values or value sets①. Every part of the measure specification① requires explicitly defined elements with accompanying analysis to identify constraints and criteria of the specification. Additional considerations for both numerator① and denominator① include alignment with other measures① conceptually and technically.

5.3.1 Define the Initial Population

The initial population① refers to the cohort from which to select the denominator population①. Some measures (e.g., ratio① measures) require multiple initial populations, one for the numerator and one for the denominator.

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

If the measure is part of a measure set①, the broadest group of population for inclusion in the set of measures is the initial population. The measure developer must specify codes or other data necessary to identify this cohort, as well as any sequencing of steps needed to identify cases for inclusion.

The measure developer should consider the attribution① model early in development. The attribution model “can affect which patients are included in the population addressed by a value-based purchasing program or included in the denominator of a performance measure①” ([NQF, 2016, p. 2](#) [↗]). NQF offers these considerations for attribution approaches ([NQF, 2016](#) [↗]):

- Is the attribution model for the new measure evidence-based?
- To what degree can the new accountable unit influence the outcomes?
- Are there multiple units to which the attribution model will be applied?
- What are the potential consequences?
- What are the qualifying events for attribution, and do those qualifying events accurately assign care to the right accountable unit?
- What are the details of the algorithm used to assign responsibility?
Did the measure developer consider multiple methodologies for reliability①?

5.3.2 Define the Denominator

The denominator statement describes the population evaluated by the individual measure. The target population① defined by the denominator can be the same as the initial population, or it can be a subset of the initial population to further constrain the population for the measure. The measure developer must describe the denominator statement sufficiently so the reader understands the eligible population or composition of the denominator, and should not use codes in lieu of words to express concepts in written descriptions. The measure developer should define the denominator precisely and include parameters such as

- age ranges
- setting
- diagnosis
- procedures
- time interval①
- other qualifying events, e.g., look back period

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

Examples

- All patients aged 18 and older with a diagnosis of Chronic Obstructive Pulmonary Disease (COPD) (NQF 0091) ([CMIT Reference Number 326](#))
- All patients at least 18 years old as of the first day of the reporting month who are determined to be maintenance hemodialysis (HD) patients (in-center and home HD) for the complete reporting month at the same facility (NQF 2978)([CMIT Reference Number 5642](#))

5.3.3 Determine if there is a need for a Denominator Exclusion

Denominator exclusions ① refer to criteria that result in removal of patients or cases from the denominator ① before calculating the numerator. An exclusion ① means that the numerator ① event is not applicable to those covered by the exclusion; an example of an exclusion is to “exclude women who have had bilateral mastectomy from the denominator for a measure ① of screening mammography.”

The goal of denominator exclusion criteria is to have a population ① or sample ①, all of whom share a similar profile in terms of their likelihood of meeting the numerator criteria.

Format: denominator-eligible patients who [have some additional characteristic, condition, procedure]

The measure developer must not specify systematically missing data as an exclusion. The NQF 2011 Consensus Standards Approval Committee ([CSAC Guidance on Quality Performance Measure Construction](#)) ② notes systematic missing data (e.g., when poor performance is selectively not reported) reduces validity ① of conclusions that can be made about quality.

The measure developer should support an allowable exclusion with

- evidence that the exclusion condition occurs with such frequency that will distort the measure results without the exclusion, and
- evidence that the exclusion significantly improves measure validity ①, and/or
- evidence of both empiric and face validity ①.

Also consider

- Conditions that are present on admission (POA) should not count as an adverse event.
- An adverse event can be very difficult to prevent in a population of interest, and therefore not an indication of substandard care.
- Some inclusion criteria identify populations who are at very low risk for the adverse event, but then the measure developer incorrectly made an exclusion to prevent dilution of the quality improvement denominator.
- Some inclusion criteria are for the purpose of enhancing face validity with clinicians.
- Some inclusion criteria are an inherent part of the quality improvement definition.
- The inclusion criteria may conflict with the patient’s goals of care (e.g., advanced illness, terminally ill).

5.3.4 Define the Numerator

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

- event or events that will satisfy the numerator inclusion criteria

- performance period or time interval ① in which the numerator event must occur, if it is different from that used for identifying the denominator

Format: patients who received/had [measure focus] {during [time frame] if different than for target population ①}

Examples

- Patients with documented spirometry results in the medical record ① (FEV1 and FEV1/FVC) (NQF 0091) ([CMIT Reference Number 326](#) ^{L3})
- The number of adult patient-months in the denominator who were on maintenance hemodialysis using a catheter continuously for three months or longer as of the last hemodialysis session of the reporting month (NQF 2978) ([CMIT Reference Number 5642](#) ^{L3})

5.3.5 Determine if there is a need for a Denominator Exception

An exception ① allows the provider to get credit when the provider performs the quality action, but not penalized if not done for an appropriate reason. It allows the exercise of clinical judgment and implies that the provider at least considered treatment, or offered to, each potentially eligible patient in the denominator. Exceptions are most appropriate when contraindications to drugs or procedures being measured are relative ([Spertus et al., 2010](#) ^{L3}). The measure developer should only use a denominator exception ① in proportion ① measures. It is not appropriate for ratio ① or CV ① measures.

An example of an exception allowing for clinical judgment in the case of a patient with two chronic conditions

- Asthma is an allowable denominator exception for the performance measure ① of the use of beta blockers for patients with heart failure. Thus, physician judgment may determine there is greater benefit for the patient to receive beta blockers for heart failure than the risk of a problem occurring due to the patient's coexisting condition of asthma. If the provider gives the medication, the measure implementer does not search for exceptions, and the patient remains in the denominator. If the provider did not give the medication, the implementer looks for relevant exceptions and removes the patient—in this example, a patient with asthma—from the denominator. If the provider did not give the medication and the patient does not have any exceptions, the patient remains in the denominator and the provider fails the measure.

A measure developer should specifically define an exception when capturing the information in a structured manner fits the clinical workflow. Allowable reasons for exceptions fall into three general categories: medical, patient, and system.

- Medical reasons should be precisely defined and evidence-based. The events excepted should occur often enough to distort measure results if 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. For example, medication specified in the numerator ① is shown to cause harm to fetuses, and the patient's pregnancy is documented as the reason for not prescribing an indicated medication. If in the course of a measure's use, the measure developer finds that medical reasons resulting in an exception to occur in a high enough volume and are of universal applicability, then the measure developer can consider the exception for redefinition as an exclusion.

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

The measure developer must capture the exception ① with explicitly defined data elements ① that allow analysis of the exception across providers to identify patterns of inappropriate exception and gaming ① and to detect potential healthcare disparity issues. Analysis of rates without attention to exception information has the potential to mask disparities in healthcare ① and differences in provider performance.

Examples

- Inappropriate exception: a notation in the medical record ① indicates a reason for not performing the specified care, and scientific evidence does not support the reason.
- Gaming: patient refusal may be an exception; however, it has the potential for overuse. For example, a provider does not actively encourage the service, explain its advantages, or attempt to persuade the patient, and then uses patient refusal as the reason for nonperformance.
- Disparity issues: the use of a patient reason for exception for mammograms are noted to be high for a minority population, which may indicate a need for more targeted, culturally appropriate patient education or closer examination of patient access issues such as lack of transportation or lack of childcare.

To ensure transparency, the measure developer should capture an allowable exception in a way that the provider can report it separately, in addition to the overall measure rate. The measure developer should support an allowable exception with evidence

- of sufficient frequency of occurrence such that distortion of the measure results occurs without the exception
- that the exception is clinically appropriate to the eligible population for the measure ①

Although no single agreed-upon approach to exceptions exists, there seems to be consensus that exceptions provide valuable information for clinical decision-making. Measure developers who build exceptions into measure logic ① should be cautioned that—once implemented—exception rates may be subject to reporting, auditing ①, endorsement/maintenance review, and validation of appropriateness. The measure developer should account for these factors in measure design and development.

5.3.6 Determine if there is a need for a Numerator Exclusion

The measure developer should use numerator exclusions ① only in ratio ① measures to define elements that should not be included in the numerator ① data.

Example

- If the number of central line bloodstream infections per 1,000 catheter days were to exclude infections with a specific bacterium, the measure developer would list that bacterium as a numerator exclusion.

5.3.7 Define Stratification Scheme

Measure developers may define a stratification ① scheme in lieu of risk adjustment ① by stratifying the population based on their risk for an outcome or procedure. They may also stratify according to a reporting scheme (e.g., if reported data are in strata by age groups). For more information, refer to the supplemental material, [Risk Adjustment in Quality Measurement](#) ②, Table 1. Framework for Risk Adjustment Strategies.

Measure developers should always consider stratifying by sociodemographic characteristics. Stratification may effectively detect potential disparities in care ①/outcomes among populations ① related to the measure ① focus. If the measure developer stratifies results by population characteristics, then the measure developer must describe population stratification variables.

When the measure definition includes stratification, the measure developer reports each population in the measure definition both without stratification and by each stratification criteria. For measures with multiple numerators ① and/or strata, they should consider scoring ① each patient/episode for inclusion/exclusion ① to every population. For example, if a measure has two numerators, and the patient is included in the first numerator, the patient should also be scored for inclusion/exclusion from the populations related to the other numerators (e.g., Antidepressant Medication Management [NQF 0105] [[CMIT Reference Number 3044](#) ②]).

Measure developers should stratify measures by organizational characteristics. This is known as peer grouping stratification, and it is appropriate in any circumstance when there is unmeasured systematic and persistent patient heterogeneity and characteristics of the organizational setting are related to that unmeasured patient heterogeneity (e.g., location). There should be an explicit hypothesis or rationale as to why the characteristic is related to the unmeasured patient heterogeneity, but not the quality construct.

If there is a reporting requirement of multiple rates or stratifications, the measure developer should state this in the specifications ①. If the measure developer includes the allowable exclusion in the numerator, they should specify that the measure reports the overall rate as well as the rate of each exclusion. If stratification of results is by population characteristics, they should describe the variables used.

Examples of possible stratification schemes:

- Vaccination measure numerator that includes (1) healthcare worker who received the vaccine; (2) healthcare worker who was offered the vaccine and declined; or (3) healthcare worker who has an allergy, a condition, or another medical contraindication to the vaccine (e.g., Influenza Vaccination Coverage Among Healthcare Personnel [NQF 0431]) [[CMIT Reference Number 854](#) ②])
- Measure is to be stratified by a population type (i.e., race, ethnicity, age, social risk factors, income, region, gender, primary language, disability) (e.g., Chlamydia Screening for Women [NQF 0033] [[CMIT Reference Number 2513](#) ②])

 For eCQMs ①, include a Reporting Stratification section in the human-readable document. If a measure does not have reporting strata defined, the default display is “None”. If a measure contains reporting stratification, list each of the reporting strata separately under the Population Criteria section.

5.3.8 Use Positive Evidence

Inquiries for measure specifications ^① should be based on the principle of positive evidence, defined as data used to confirm a given criterion was met. The principle is particularly relevant when there are no data or there are conflicting data. Where, for instance, a numerator criterion ^① is “low density lipoprotein (LDL) cholesterol is less than 100” and there is no LDL cholesterol result in the patient record, then there is no positive evidence, and the criterion is not met. When, for instance, a denominator ^① 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 same patient’s record, then because there is positive evidence of an ejection fraction less than 40%, the criterion is met.⁸

5.4 SPECIFY THE CODE AND/OR CODE SYSTEMS

Most measures ^① rely at least in part on the use of various standardized codes or code systems ^① for classifying healthcare provided in the United States. The measure developer should list all required codes (plus the code system and the version that the codes came from) for the measure and explicitly state the source of the codes and instructions pertaining to their use. Measure developers must remember that versions of code systems should align with the timeframe of testing data, which may span multiple versions of a code system. Specifications may require that certain codes accompany other codes, occur in specific locations in the record, or occur on claims from specific provider types. Some code sets such as Current Procedural Terminology (CPT) codes may require copyright statements to accompany their use.

Some claims-based measures use Quality Data Codes (QDCs) to report quality measure ^① data. QDCs are CPT Category II or Level II G-codes (Healthcare Common Procedure Code System [HCPCS]). When appropriate, QDCs are added to the [CMS form 1500 version 02/12](#) [↗]. Measure specifications require use of codes, so measure developers must identify the QDCs when developing measure specifications as they would other codes (e.g., International Classification of Diseases, 10th Revision, Clinical Modification [ICD-10-CM] codes) for determining numerators ^① and denominators ^①.

For relevant guidance, see the [Codes, Code Systems, and Value Sets](#) [↗] supplemental material.

5.5 CONSTRUCT DATA PROTOCOL

The measure developer must explicitly identify types of data and how to aggregate or link these data so that calculation of the measure is reliable and valid. The measure developer should proceed carefully when merging data from different sources or systems to prevent errors in assumptions. 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 ^① (e.g., different age groups, different time frames)
- different quality controls used by data sources

It may be necessary to clean the merged data. If the measure developer finds inaccurate, incomplete, or unreasonable data, they should correct data errors or omissions.

⁸ Many measures will be more specific with respect to which observation to use when comparing against a criterion, such as “MOST RECENT ejection fraction is less than 40%.”



For eCQMs^①, conduct preliminary feasibility^① assessments to confirm availability of the information within an EHR, ideally in a structured format. The feasibility assessments should include both the data model and how various EHR systems map and store the data elements^①. Also review the specifications on either the draft external document (prior to Measure Authoring Tool [MAT] entry) or on the MAT-exported files, and other documentation for criteria such as

- The eCQM header includes appropriate information in the data fields or contains preferred terms.
- Correct mapping of the measure^① data elements^① to the correct category and datatype in the [Quality Data Model \(QDM^①\)](#). See the [Codes, Code Systems, and Value Sets](#)[↗] supplemental material.
- Each QDM data element is subsequently linked to an appropriate value set(s)^① or direct reference code^①.
- Value sets and direct reference codes used represent the most current [Interoperability Standards Advisory](#)[↗] recommendations. See the [Codes, Code Systems, and Value Sets](#)[↗] supplemental material.
- The addition of CMS additional supplemental data elements.
- Testing of the eCQM logic^① in [Bonnie](#).[↗]

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

Measure developers must precisely define terms used in the numerator^① or denominator^① statement, or in allowable exclusions^① and exceptions^①. Measure developers construct some measures by using precisely defined components or discrete pieces of data, often called data elements. Technical specifications^① include the “how” and “where” to collect the required data elements. The measure developer should fully specify measures, including all applicable definitions and codes. Precise specifications are essential for successful implementation.

Example

- Up-to-date vaccination status requires a clear definition of which type of vaccinations need assessment along with the definition of “up-to-date.”



Patient medical record data from EHRs^① and other clinical systems (for eCQMs or measures specified for use in an EHR and other clinical systems) consist of patient-level information coded in such a way for extraction in a format that can be used in a measure. Information entered in an EHR and other electronic clinical systems, but not coded in a structured field, may require special processing by measure implementers.



Patient medical record^① data from paper charts, EHRs, and other electronic clinical systems (if not specified for an EHR) will require instructions for abstraction. It is important to specify which value when there will be multiple values in a record, e.g., average of all lab values in a year, only the most recent lab value, or the first lab value during a hospitalization. The level of detail may require specifying allowable terms, allowable places in the record, and the allowable values. The measure developer should assess inter-rater reliability^① to ensure the specifications are clear and unambiguous, see [Chapter 6.2.2.1.1, Types of Reliability^①](#).

Examples

- allowable terms used from the record: hypertension; HTN; high blood pressure; ↑BP

- allowable places within the record: problem list, history and physical, and progress notes
- allowable values: systolic blood pressure <130, urine dipstick result +1 or greater.

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

Example

- The Acute Myocardial Infarction (AMI) mortality ① 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 ICD-10-CM code I21.xx, excluding those with I22.xx (AMI, subsequent episode of care).

The measure developer should include enough detailed information in the denominator ①, numerator ①, exclusions ①, and exceptions ① so that each person collecting data for the measure ① will interpret specifications ① in the same way. If the measure developer is allowing multiple data collection methods, they should produce detailed specifications for each method.

5.5.2 Describe the Level of Measurement/Analysis

The unit of measurement/analysis is the primary entity upon which to apply the measure. The measure developer should clearly state and justify the procedure for attributing the measure. The measure developer should specify measures with the broadest applicability (e.g., target population ①, setting, level of measurement/analysis ①) as supported by the evidence. However, a measure developed for one level may not be valid for a different level. Therefore, the measure should be respecified and retested for reliability ① and validity ① in each setting/population ①.

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 developing a claims-based measure for Medicare use, and the literature and guidelines support the measure for all adults, the measure developer should consider expanding the data sources ① beyond “Medicare Parts A and B claims.”

 Given the circumstance that multiple entities are using a shared EHR ①, measure developers should consider all relevant testing needed to minimize the possibility of quality actions performed by one entity inappropriately attributed to another entity.

5.5.3 Describe Sampling

If allowing sampling, the measure developer should describe the sample ① size or provide guidance in determining the appropriate sample size and describe any prescribed sampling methodologies explicitly.

 Sampling is not applicable to eQMs ①.

5.5.4 Determine Risk Adjustment

In certain cases, risk adjustment ① of the measure is also a component of the specification process, specifically for outcome measures ①.

Risk adjustment is the modeling of health outcomes or costs as a function of various risk factors. Risk adjustment is important because health outcomes and costs are often a result of the interplay of

demographic, clinical, and socioeconomic factors. In the context of quality measurement, the purpose of risk adjustment is to enhance meaningful comparison across healthcare providers, health plans, or individual clinicians. It facilitates a fairer comparison of providers with different patient populations^① and aims at leveling the playing field (Iezzoni, 2013).

The measure developer should risk adjust performance measures^① to facilitate fair and accurate comparisons of outcomes across healthcare organizations, providers, or other groups. Risk factors that exist outside healthcare encounters^① may affect outcomes regardless of the quality of care received. Adjusting for these risk factors can avoid misleading comparisons. However, risk adjustment models for publicly reported quality measures^① should not obscure disparities in care^① that are associated with race, gender, socioeconomic status, or other demographics and social factors.

The measure developer should fully disclose all measure specifications^①, including the risk adjustment^① methodology, to ensure transparency and accountability. The measure developer must update and recalibrate the risk adjustment model as needed to adjust for changes in cohort/population change, coding changes (e.g., provider coding practice, change from ICD-9 to ICD-10 coding, new codes added), newly added clinically relevant data elements^① (e.g., National Institutes of Health [NIH] Stroke Scale for stroke measures), or social risk factors (e.g., patient dual status).

The [Risk Adjustment in Quality Measurement](#)[↗] supplemental material provides additional guidance for determining when risk adjustment is necessary and describes the steps involved.

5.5.5 Clearly Define Any Time Intervals

The measure developer must explicitly state in the measure specification when they use such intervals to determine cases for inclusion in the denominator^①, numerator^①, or exclusion^①. The measure developer must clearly indicate the index event used to determine the time intervals^①. Also, the specification must identify how often to report the numerator for each patient as well as how often to include a patient in the denominator. For example, if the count of the event or action in the numerator is performed during an episode of community-acquired pneumonia, how is that episode of community-acquired pneumonia captured correctly if a patient has three episodes of pneumonia during the measurement period?

Measure developers must

- Avoid using ambiguous semantics when specifying time intervals.
- State the exact interval units required to achieve sensitivity^① necessary for measurement.
- State the exact interval units required to achieve the level of granularity necessary to ensure validity^① and reliability^① of the measure calculation.
- Explicitly state any look back periods. For example, looking at patient's history for a previous diagnosis of cancer, which may be an exclusion.

International Organization for Standardization (ISO) 8601:2004 defines data elements and interchange formats for the representation of dates and times, including time intervals. The [Health Level Seven International® \(HL7\) Clinical Quality Language \(CQL\)](#)^① specification, [Appendix H](#)[↗] also provides conventions that are intended to standardize time calculation units for durations (e.g., difference between two date/time elements). The measure developer should use these standards in time interval calculations for any type of CQM^①, not just eCQMs^①.

Example

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

5.5.6 Describe How to Score and Report the Measure Results

Most [quality measures](#) (1) produce rates; however, there are other [scoring](#) (1) methods such as categorical value, [CV](#) (1), count, frequency distribution, non-weighted score/composite/scale, [ratio](#) (1), and weighted score/composite/scales. Measure information must include a description of the scoring type.

The measure developer should describe the type of scoring, accompanied by an explanation of how to interpret the score, such as

- higher score indicates better quality; improvement noted as increase in rate
- lower score indicates better quality; improvement noted as decrease in median value
- score within a defined interval (or a “passing score” over or under a certain threshold) indicates better quality

5.5.7 Develop the Calculation Algorithm

The [calculation algorithm](#) (1)—sometimes referred to as the performance calculation, [measure logic](#) (1), or [measure flow](#)—is an ordered sequence of [data element](#) (1) retrieval and aggregation through which the measure identifies the [numerator](#) (1) and [denominator](#) (1) events or CV values. At the [measure specification](#) (1) stage, the calculation algorithm is not necessarily an equation or a body of computer programming code, but instead is a depiction of the path from the raw data to the result. The measure developer must describe how to combine and use 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 a required item in the NQF Intent to Submit and measure submission. Revisions and updates continue to the calculation algorithm through to the measure implementation phase where it is known as the implementation algorithm (see [Chapter 7.3](#)).

Development of the calculation algorithm should be based on the written description of the [measure](#) (1). The measure description must contain enough information to develop the algorithm. The measure developer needs to check the calculation algorithm for consistency with measure text, as the calculation algorithm will serve as the basis for development of computer programming to produce measure results. The measure developer should account for each scenario and ensure there is a logical end point for each scenario. They should establish this through alpha testing and preliminary [feasibility](#) (1) assessments. These assessments will inform beta testing and also minimize implementation and reporting burden on the provider. [Chapter 6, Measure Testing](#) (1), has additional details on responsibilities during this process.

5.6 DOCUMENT THE MEASURES

The measure developer must complete the detailed technical specifications, including any additional documents required to evaluate and implement the measure as intended. The MIF, MJF, and several documents on the [NQF Submitting Standards page](#) (2) are available to assist in documentation of specifications.

5.6.1 Finalize the Measure Name and Description

The measure name (or measure title) should be a very brief description of the measure’s focus and target population①. If it is an NQF-endorsed measure, use the NQF-endorsed title.

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

Examples

- Diabetes: Medical Attention for Nephropathy (NQF 0062) ([CMIT Reference Number 4021](#)🔗)
- Percent of Residents Whose Need for Help with Activities of Daily Living has Increased Long Stay (NQF 0688) ([CMIT Reference Number 4066](#)🔗)
- COPD: Spirometry Evaluation (NQF 0091) ([CMIT Reference Number 326](#)🔗)

For measures based on appropriate use criteria① addressing overuse of certain services, there are three standardized title lead-ins.

- Appropriate Use of ...
- Appropriate Non-Use of ...
- Inappropriate Use of ... (for inverse measures①—the least desirable approach)

For the measure description, measure developers should briefly describe the type of score① (e.g., percentage, percentage rate, proportion①, number), target population①, and focus of measurement.

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

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

- “The percentage of”
- [gender qualifier] if applicable (e.g., “female”)
- “patients or individuals”
- “during visit or event”
- [environment qualifier] (e.g., admitted to a post-anesthesia care unit [PACU])
- [age qualifier] (e.g., aged 18 years and older)
- [denominator① definition] (e.g., who are under the care of an anesthesia practitioner)
- [numerator① 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 to word performance measures① positively when possible (i.e., to demonstrate which clinical activity to capture in the numerator).

Examples

- Percentage of patients 18–75 years of age with diabetes who had a nephropathy screening test or evidence of nephropathy during the measurement period (NQF 0062) ([CMIT Reference Number 4021](#)🔗).
- Percentage of healthcare personnel who receive the influenza vaccination (NQF 0431) ([CMIT Reference Number 854](#)🔗).
- Percentage of patients 18 years of age and older who were treated with antidepressant medication, had a diagnosis of major depression, and who remained on an antidepressant

medication treatment. Two rates are reported: a) percentage of patients who remained on an antidepressant medication for at least 84 days (12 weeks); b) percentage of patients who remained on an antidepressant medication for at least 180 days (6 months) (NQF 0105) ([CMIT Reference Number 2503](#)[↗]).

- Percentage of adult hemodialysis patient-months using a catheter continuously for three months or longer for vascular access (NQF 2978) ([CMIT Reference Number 5642](#)[↗]).

5.6.2 Making Changes

Information from [measure testing](#)^①, the public comment period, updated information gathering, or other stakeholder input may require the measure developer to make changes to [technical specifications](#)^①.

5.7 STAKEHOLDER ENGAGEMENT IN MEASURE SPECIFICATION

A variety of stakeholders need to be involved in the measure specification stage. In addition to the measure developer and TEP, patients, caregivers, and clinicians need to be involved to address the [feasibility](#)^① of data collection before a measure progresses too far in development. Measure developers can use focus groups, interviews, and other informal methods to engage other stakeholders.

The MC Workspace has an [eCQM Clinical Workflows](#)[↗] module that provides stakeholders—clinicians in particular—the opportunity to review and comment on measure flow and clinical context for [eCQMs](#)^① under development. Measure developers strive to develop [measures](#)^① that complement clinical workflow and this feedback helps achieve that objective. As with the eCQM Concepts module, users need an eCQI Resource Center [account](#)[↗] to make comments.

5.8 SPECIAL CONSIDERATIONS FOR MEDICAID-FOCUSED MEASURES

Measures developed for Medicaid programs may need special considerations:

- There is a tendency to aggregate Medicaid measures at a higher level (e.g., plan, state).
- Medicaid measures tend to be tools to advance policy, which often relates to containing costs and improving quality.
- NQF endorsement may not be a high-priority end goal for state-level stakeholders.
- Ensure the entities calculating measures (plans, states) have access to the data and are capable of using it for the intended purpose.
- States vary in the collection, analytics, and presentation of data.
- Federal Medicaid data reporting is voluntary, not mandatory.

5.9 HARMONIZATION

When specifying measures, measure developers should consider whether a similar measure exists for the same condition, process of care, outcome, or care setting. Measure developers should consider [harmonization](#)^① for every measure under development or maintenance throughout the Measure Lifecycle and harmonize measures unless there is a compelling reason for not doing so (e.g., significant risk variation by age, comorbidity, race) that would justify a separate measure. Harmonization standardizes similar measures when their differences do not make them scientifically stronger or more valuable. Harmonization should not result in inferior measures, but in measures that are scientifically strong, clinically valuable, evidence-based, and important to persons/families/caregivers. [Quality measures](#)^① should be based on the best way to capture and specify the measure based on the current

scientific information and guidelines. Do not assume that an NQF-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 a fully developed and specified measure. The earlier in the process that measure developers identify related or competing measures ⓘ, the sooner there is resolution to problematic issues.

Harmonization may include comparison and reconciliation of

- age ranges
- measurement period
- allowable values for medical conditions or procedures (e.g., codes, code systems ⓘ, code lists, descriptions)
- allowable conditions for inclusion in the denominator ⓘ (e.g., codes, code systems, code lists, descriptions)
- exclusion ⓘ categories, whether the exclusion is from the denominator or numerator ⓘ, and whether optional or required
- calculation algorithm ⓘ
- risk adjustment ⓘ methods

Examples:

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

If harmonization of the measure can occur with one or more existing measures, then the measure developer should use existing definitions for those attributes. Other resources (e.g., [CMIT](#)²³ and the [eCQI Resource Center](#)²⁴) contain specifications ⓘ to help identify opportunities for further harmonization. If the measure developer determines not to harmonize measures, they must document the reasons and include any literature used to support this decision. Some reasons not to harmonize include

- The science, such as clinical practice guidelines ⓘ, behind the new measure does not support using the same variable(s) found in the existing measure.
- The measures' intentions vary across programs/payors, which requires the measures to be distinct.
- The measures have differing denominator populations ⓘ at significantly different risk (i.e., the denominators are risk stratified).

Examples of measures for harmonization

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

For more detail, refer to the [Measure Harmonization, Respecification, and Adoption](#)  supplemental material.

6 MEASURE TESTING

Chapter 6 provides an overview of the types of testing needed to assess measure ① evaluation criteria and outlines the process for development, implementation, and reporting of test plans, results, and associated artifacts. Information in this chapter is not meant to be prescriptive or exhaustive. Measure developers can use other testing approaches that employ appropriate methods and rationales. Measure developers should always select testing that is appropriate for the measure under development and provide empirical evidence ① for the importance to measure and report ①, feasibility ①, scientific acceptability ①, and usability ① and use.

Measure testing ① is an iterative process conducted concurrently with measure specification ①. Iterative testing provides measure developers an opportunity to refine draft specifications before finalization; augment or reevaluate earlier judgments about the measure’s importance; and assess feasibility, usability, and scientific acceptability of the measure.

Measure testing enables a measure developer to assess suitability of the quality measure’s ① technical specifications and acquire empirical evidence to help assess strengths and weaknesses of a measure with respect to the NQF [Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement](#) ①. To evaluate a measure, the measure developer should use information gathered through measure testing in conjunction with expert judgment. For Blueprint purposes, measure testing refers to evaluating the draft specifications of quality measures, including components of the quality measures, such as the data elements ①, instruments, and performance score ①.

Figure 12 describes how testing fits into the flow of the Measure Lifecycle.

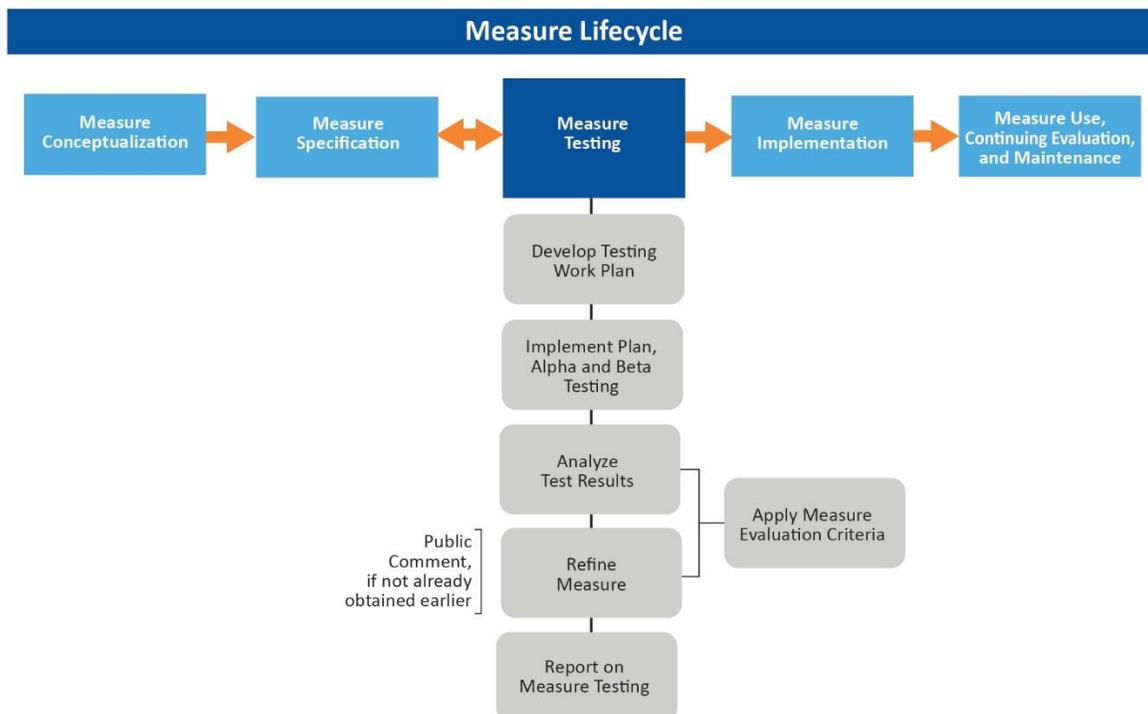


Figure 12. Flow of the Measure Lifecycle—Measure Testing

6.1 THE MEASURE TESTING PROCESS

Proper testing and analysis are critical to development of a feasible, reliable, and valid measure ①. Chapters 6.2-6.9 describe types of testing that may be conducted during measure development (alpha and beta testing), the procedure for planning and testing, and key considerations when analyzing and documenting results of testing and analysis, including incorporation of stakeholder inputs after testing is complete.

The measure developer should conduct initial testing during development (i.e., pilot testing) within the framework of alpha and beta tests. Although considered part of measure testing ①, alpha testing may occur as early as information gathering and repeated iteratively during development of measure specifications ①. Measure developers should test early and often.

Alpha testing (i.e., formative testing) is of limited scope since it usually occurs before full development of detailed specifications. Measure developers may conduct alpha testing, particularly regarding feasibility ① of the concept in the context of the data source ①, as part of information gathering empirical analysis and may occur concurrently with development of technical specifications as part of an iterative process. Alpha tests include methods to determine whether individual data elements ① are available and whether the form in which they exist is consistent with the intent of the measure. Types of testing used in an alpha test vary widely and often depend on the measure’s data source or uniqueness of the measure specifications. Measures that use data sources similar to existing measures may require minimal alpha testing. In contrast, measures that address areas with no development of specifications may require multiple iterations of alpha testing. For example, an alpha test may include a query to a large, integrated, delivery system database to determine how it captures specific data, where the query originates, and how to express the query. Results can impact decisions about measure specifications.

Beta testing (i.e., field testing) generally occurs after development of initial technical specifications 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 ① and usability ① of a measure. Measure developers can use beta tests to evaluate the measure’s suitability for risk adjustment ① or stratification ① 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 6 shows the attributes of alpha and beta testing. The measure developer should consider these attributes when developing a work plan for alpha and beta tests.

Table 6. Features of Alpha and Beta Testing

Feature	Alpha Testing	Beta Testing
Timing	<ul style="list-style-type: none"> Usually conducted prior to completion of technical specifications May conduct multiple times in quick succession 	<ul style="list-style-type: none"> Conducted after development of measure developer’s detailed and precise technical specifications
Scale	<ul style="list-style-type: none"> Typically, smaller scale <ul style="list-style-type: none"> Only enough records to ensure the data set contains all elements needed for the measure Only enough records to identify common occurrences or variation in the data 	<ul style="list-style-type: none"> <u>Samples</u> ① strive to achieve representative and adequate 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, <u>medical record</u> ①)

Feature	Alpha Testing	Beta Testing
Sampling	<ul style="list-style-type: none"> Convenience sampling 	<ul style="list-style-type: none"> Sufficient to allow adequate testing of the <u>measure's scientific acceptability</u> Representative of the <u>target population</u> Representative of the people, places, times, events, and conditions important to the measure If based on administrative or claims data, uses entire eligible <u>population</u> Randomized, if possible
<u>Specification Refinement</u>	<ul style="list-style-type: none"> Permits early detection of problems in technical specifications (e.g., identification of additional inclusion and exclusion criteria) 	<ul style="list-style-type: none"> Used to assess or revise complexity of computations required to calculate the measure
<u>Importance</u>	<ul style="list-style-type: none"> Designed to look at volume, frequency, or costs related to a measure topic (e.g., cost of treating the condition, costs related to procedures measured) Establishes, on a preliminary basis, that the measure can identify gaps in care Provides support for further development of the measure 	<ul style="list-style-type: none"> Allows for enhanced evaluation of a measure's importance, including evaluation of performance thresholds, disparities analysis, 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 <u>topped-out</u>, where most groups achieve similarly high performance levels approaching the measure's maximum possible value)
Scientific Acceptability	<ul style="list-style-type: none"> Limited in scope if conducted during the formative stage Usually occurs later in development 	<ul style="list-style-type: none"> Assesses measure <u>reliability</u> and <u>validity</u> Reports results of analysis of <u>exclusion</u> (if any used) Tests results of the <u>risk adjustment</u> model, quantifying relationships between and among factors
<u>Feasibility</u>	<ul style="list-style-type: none"> Provides initial information about feasibility of collecting required data and calculating measures using technical specifications Identifies barriers to implementation Offers initial estimate of costs or burden of data collection and analysis 	<ul style="list-style-type: none"> Provides enhanced information regarding feasibility, including greater determination of barriers and provider burden to implementation and costs associated with measurement Evaluates feasibility of <u>stratification</u> factors based on occurrences of target events in the <u>sample</u>
<u>Usability and Use</u>	<ul style="list-style-type: none"> No formal analytic testing at this stage; may use qualitative testing with patients and providers May use the TEP to assess potential usability of the measure 	<ul style="list-style-type: none"> Identifies unintended consequences, including susceptibility to inaccuracies and errors Reports strategies to ameliorate unintended consequences May consist of focus groups or similar means of assessing usefulness of the measure by consumers May not be in the scope of measure development contract Can use TEP to assess potential usability

A measure developer should develop specific reports when testing a measure (or set of measures). Although completion of reports usually occurs after beta testing, measure developers should consider the need to report the results of formative alpha testing, especially if the intent is for alpha testing to precede beta testing. The first few steps of measure testing address planning and execution of testing and are identical for alpha and beta testing; the last steps address reporting and follow up after the conclusion of testing. During measure testing the measure developer

- develops the testing work plan
- performs sampling

- implements the plan
- analyzes test results
- refines measure ①, including incorporation of stakeholder inputs
- retests the refined measure
- updates the measure documentation

6.2 TESTING AND MEASURE EVALUATION CRITERIA

The measure developer should use results of measure testing ① to demonstrate a measure's alignment with most measure evaluation criteria. Because testing is often an iterative process, both alpha and beta test findings may provide information that address measure evaluation criteria.

- Alpha testing often supplies information that demonstrates feasibility ① of the measure's implementation.
- Measure developers may use the findings from one or more beta tests to demonstrate scientific acceptability ① and usability ① and use, as well as to augment previous information on the importance ① and feasibility of the measure.

Find additional information on measure testing in several special measure supplemental materials, e.g., [Electronic Clinical Quality Measures \(eCQMs\) Specification, Testing, Standards, Tools, and Community](#) [↗].

Chapters 6.2.1 through 6.2.4 describe application of testing results to four measurement areas — importance, scientific acceptability, feasibility, and usability and use.

6.2.1 Importance

Information from testing often provides additional empirical evidence ① to support prior judgments of a measure's importance. 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 no measurement of rare or low-cost events
- identifying substantial variation among comparison groups or suboptimal performance for a large proportion of the groups
- demonstrating that methods for scoring ① and analysis of the measure allow for identification of statistically significant and practically/clinically meaningful differences in performance
- showing disparities in care ① related to race, ethnicity, gender, income, or other classifiers
- identifying evidence that a measure is associated with consistent delivery of effective processes or access that leads to improved outcomes

Reported data to support the importance of a measure may include

- Descriptive statistics (e.g., means, medians, standard deviations, confidence intervals for proportions, 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 R² ① or intraclass correlation.

6.2.2 Scientific Acceptability

Scientific acceptability ① 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 use of the measure can draw reasonable conclusions about care in a given domain. Because many measure scores ① are composed of patient-level data elements ① (e.g., blood pressure, laboratory values, medication, surgical procedures) that are aggregated at the comparison group level (e.g., hospital, nursing home, physician), evidence of reliability ① and validity ① is often needed for both the measure score ① and data elements, and the measure developer should ensure that both facets are addressed. Some examples of common measure testing ① and reporting errors, which can reduce scientific acceptability, include

- Reporting is limited to descriptive statistics. Descriptive statistics demonstrate that data are available for analysis, but do not provide evidence of reliability or validity.
- A lack of testing of a respecified measure ①. When respecifying a measure (e.g., using similar process criteria for a different population ① or denominator ①), the newly respecified measure still requires testing to obtain empirical evidence ① of reliability and validity.
- Inadequate evidence of scientific acceptability for commonly used data elements. Data elements (e.g., diagnosis codes, EHR ① 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 ① for justifying an exclusion ①. The measure developer should report analyses and/or clinical guidelines justifying an exclusion or demonstrating reliability for different methods of data collection.
- Not properly accounting for missing data.
- Lack of risk adjustment ① or stratification ①.

Since expression of reliability and validity is along a scale or continuum (i.e., they are not all-or-nothing properties), the measure developer may need to address many issues to supply adequate evidence of scientific acceptability. The complexity of different healthcare environments, data sources ①, and sampling constraints often preclude ideal testing conditions. As such, judgments about a measure's acceptability are often a matter of degree. The assumption is that a measure developer will contract or employ experienced methodologists, statisticians, and SMEs to select testing that is appropriate and feasible for the measure(s) under development and ensure demonstration of measure reliability and validity. The measure developer must also engage experts to review testing data and determine the measure's reliability and validity.

Although not intended to replace expert judgment of the measure development team, the next subsections describe general factors for a measure developer to consider when evaluating reliability and validity of both a measure score and its component elements. The descriptions should acquaint the measure developer with specialized terminology that testing, evaluation, and statistics experts may use in assessing scientific acceptability.

6.2.2.1 Reliability

Reliability testing ① demonstrates that the 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.

6.2.2.1.1 Types of Reliability

Depending on the complexity of the measure specifications, assessment of one or more types of [reliability](#) may occur. A description of several general types of reliability testing are in the next paragraphs.

[Inter-rater \(i.e., inter-abtractor\) 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 frequently employed to assess reliability of [data elements](#) used in exclusion specifications, as well as the calculation of [measure scores](#) when the [measure](#) requires review or abstraction. Quantitatively summarize the extent of inter-rater/abstractor reliability, concordance rates, and [Cohen's Kappa](#) with confidence intervals which are acceptable statistics to describe inter-rater/abstractor reliability.

 [eCQMs](#) implemented as direct queries to [EHR](#) databases may not use abstraction. Therefore, there may be no need for inter-rater reliability for eCQMs.

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, across automated data extraction from different [data sources](#), or testing agreement between the known values from a simulated data set and the elements obtained when the specifications are applied to the data set. Measure developers may quantify form equivalence reliability using a coefficient of equivalence when possible to calculate a correlation between the forms. As part of the analysis, the measure developer should investigate and document the reasons for discrepancies between methods (i.e., mode effects—for example, when results from a telephone survey are different from results of the same survey when mailed).

[Test-retest reliability](#), also known as temporal reliability, assesses the extent to which a measurement instrument elicits the same response from the same respondent across two measurement time periods. The measure developer should use the coefficient of stability to quantify the association for the two measurement occasions and when assessing information not expected to change over a short or medium interval of time. Test-retest reliability is not appropriate for repeated measurement of disease symptoms nor for measuring [intermediate outcomes](#) that follow an expected trajectory for improvement or deterioration. The measure developer assesses test-retest reliability when there is a rationale for expecting stability—rather than change—over the time period.

[Internal consistency reliability](#) is testing of a multiple-item test or survey that assesses the extent to which items designed to measure a given construct are inter-correlated. Cronbach's alpha has been used to evaluate internal consistency reliability for several decades ([Cronbach, 1951](#)). Use Cronbach's alpha when developing multiple survey items that assess a single construct.

With respect to other approaches to reliability across all types of reliability estimation, the shared objective is to ensure replication of measurements or decisions. In terms of comparisons of groups, the measure developer should extend reliability to assess stability of the relative positions of different groups or determination of significant differences between groups. These types of assessments address the proportion of variation in the measure attributable to the group. The measure developer describes the proportion as true differences (or “signal”) relative to variation in the measure due to other factors, including chance variation (or “noise”). Measure developers may consider measures with a relatively high proportion of signal variance reliable because of their power for discriminating among providers and the repeatability of group-level differences across samples. Provided that the number of observations within groups is sufficiently large, these questions can be partially addressed using

methods such as analysis of variance (ANOVA), calculation of intraclass correlation coefficients (ICC), estimation of variance components within a hierarchical mixed (i.e., random-effects) model, or [bootstrapping](#) ^① simulations. Changes in group ranking across multiple measurements may also add to an understanding of the stability of group-level measurement ([Adams, 2009](#) ^②).

6.2.2.1.2 Measure Data Elements versus Reliability Measure Score

Because many [measures](#) ^① are composed of multiple [data elements](#) ^①, [reliability testing](#) ^① ideally applies to both the data elements comprising the measure and the computed [measure score](#) ^①. However, for measures that rely on many data elements, measure developers may only occasionally test the individual data elements for critical elements that contribute most to the computed measure score, rather than all data elements. Similarly, measure developers may occasionally exclude from [reliability](#) ^① testing commonly used data elements, for which there is an assumption of reliability (e.g., gender, age, date of admission). NQF does not require data element reliability testing if demonstrating [data element validity](#) ^①.

Flexibility in the reliability testing of data elements contrasts with assessment of the measure score. The measure developer should always assess the measure score under development for reliability using data derived from testing.

6.2.2.2 Validity

In measure development, the term “[validity](#) ^①” has a specific application known as test validity, which refers to the degree to which evidence, clinical judgment, and theory support interpretations of a measure score. Stated more simply, test validity is an empirical demonstration of the ability of a measure to record or quantify what it purports to measure; validity represents the intersection of intent (i.e., what is being assessed) and process (i.e., how it is assessed).

6.2.2.2.1 Types of Validity

Measure developers may test validity of a measure score in many ways. Although some experts view all types of validity as special cases or subsets of [construct validity](#) ^①, researchers commonly reference the types of validity separately: [construct validity](#), [discriminant validity](#) ^①, [predictive validity](#) ^①, [convergent validity](#) ^①, [criterion](#) ^① validity, and [face validity](#) ^① ([Messick, 1994](#) ^②).

Construct validity refers to the extent to which the measure quantifies what the theory says it should. Construct validity evidence often involves empirical and theoretical support for the interpretation of the construct. Evidence may include statistical analyses such as confirmatory factor analysis of data elements to ensure they cohere and represent a single construct.

Discriminant/contrasted groups validity examines the degree to which a test of a concept is not highly correlated with other tests designed to measure theoretically different concepts. The measure developer demonstrates discriminant validity 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 refers to the ability of measure scores to predict scores of other [related measures](#) ^① in the future, particularly if the original measure scores predict a subsequent patient-level outcome of undisputed importance (e.g., death, permanent disability). Predictive validity also refers to scores on the same measure for other groups at the same point in time.

Convergent validity refers to the degree to which multiple measures/indicators of a single underlying concept are interrelated. Examples include measurement of correlations between a measure score and other indicators of processes related to the target outcome.

Reference strategy/criterion validity refers to verification of data elements against some reference criterion determined to be valid (i.e., the gold standard). Examples include verification of data elements obtained through automated search strategies of EHRs ① compared to manual review of the same medical records ① (i.e., the gold standard).

Face validity ① is the extent to which a measure ① appears to reflect what it is supposed to measure “at face value.” It is a subjective assessment by experts about whether the measure reflects its intended assessment. Face validity for a quality measure ① may be adequate if accomplished through a systematic and transparent process, by a panel of identified experts, when there is a recording of the formal rating of the validity ① and it is appropriately aggregated. The expert panel should explicitly address whether measure scores ① provide an accurate reflection of quality, and whether use of the scores can distinguish between good and poor quality. Because of the subjective nature of evaluating the face validity of a measure, measure developers should take special care to standardize and document the process used.

In [Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties](#) ①, NQF recommends that a formal consensus process, such as a modified Delphi approach, be used for the review of face validity. In the Delphi approach, participants systematically rate their agreement, and formal aggregating and follow consensus failure processes. Likewise, in [Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement](#) ①, NQF allows the use of face validity in lieu of empirical testing for new measures if a systematic assessment is performed and targeted to reflect accuracy of the targeted care measured. Since this is the weakest form of validity testing ①, experts involved in measure development should not be the same experts who perform face validity tests. Measure developers may use this type of formal process when addressing whether specifications of the measure are consistent with medical evidence. Maintenance review requires empirical validity testing. Justification is necessary if empirical validity testing is not possible.

6.2.2.2 Measure Data Elements Versus Performance Measure Score

Patient-level data elements ① are the building blocks for a performance measure and measure developers should assess them for reliability ① and validity. Although patient-level data elements are important, measure developers should use computed measure scores to draw conclusions about the targeted aspect of care. According to [Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement](#) ①, NQF will accept data element and/or measure score validity testing. Instrument-based and composite measures ① need performance score validity testing.

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 comparisons of

- claims data that have codes used to represent primary clinical data (e.g., ICD-10-CM/Procedure Coding System [PCS], CPT) to manual abstraction from a sample ① of patient medical records
- standardized patient assessment instrument information (e.g., minimum data set [MDS], Outcome and Assessment Information Set [OASIS], registry data) that is not abstracted, coded, or transcribed with “expert” assessor evaluation (conducted at approximately the same time) for a sample of patients

- EHR information extracted using automated processes based on measure technical specifications to manual abstraction of the entire EHR

6.2.2.3 Prior Evidence of Reliability and Validity for Measure Data Elements

According to NQF's [Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties](#), when prior evidence of reliability or validity of the data elements comprising the measure exists, it can sometimes be used in place of updated or additional testing of the measure's data elements. Although prior evidence can augment findings for a calculated measure score under development, prior evidence is not acceptable to demonstrate score-level reliability or validity. The measure developer always assesses commonly used data elements for reliability and validity within the context of the new measure specifications using data derived from the beta test. They should use data from prior validity or reliability testing of data elements from the same data source to calculate the measure score or computed measure score, since the two concepts are both mathematically and conceptually related. Prior evidence of reliability or validity testing may include published or unpublished testing results of same data elements, same data type, and/or a representative sample of sufficient size.

According to NQF [Guidance](#)

- Use of prior evidence of validity of data elements is acceptable if the measure under development uses the same data elements and data type and obtains a representative sample of sufficient size.
- There is no requirement for separate reliability testing of data elements if the measure developer conducted validity testing on the data elements. If using patient scores from an instrument/scale in the measure under development, measure developers can use testing and documentation of the reliability of the scale as evidence of data element reliability. If the measure developer did not conduct validity testing of the data elements, they should use prior evidence of reliability of data elements.

Refer to the NQF [Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement](#) for more information and guidance on validity testing.

6.2.2.4 Testing of Exclusions/Exceptions

The review of measure exclusions and exceptions should be based on the testing data and should include, at a minimum

- evidence of sufficient frequency of occurrence of exclusions/exceptions
- evidence that data elements (e.g., codes) used to identify exclusions/exceptions are valid

Review may also include evidence that there is distortion of measure results without the exclusions/exceptions. For example, evidence that an exclusion distorts a measure may include variability of an exclusion across comparison groups and sensitivity analyses of the measure score with and without the exclusion.

Additional review is required when patient preference or other individual clinical judgment based on unique patient conditions is allowed as an exception category. The measure developer should analyze whether the exception will make a major change to measure results. The measure developer must also consider whether patient preference represents a clinical exception to eligibility or whether provider intervention can influence patient preference. Measure developers should always evaluate these

measures with and without the exception and include the proportion of exceptions for any group-level tabulations.

6.2.2.5 Risk Adjustment and Stratification

For outcome measures ①, the measure developer should use beta testing to evaluate an evidence-based risk adjustment ① strategy. Typically, process measures ① do not need risk adjustment.

The measure developer must provide empirical evidence ① for adequacy of the risk adjustment ① or a rationale that risk adjustment is not necessary to ensure fair comparisons.

Measure developer-provided information should include the analytic methods used and evidence of meaningful differences. If using stratification ①, the measure developer should include stratification results. Additional information about risk adjustment, stratification, and social factors is available in the [Risk Adjustment for Quality Measurement](#) [↗] supplemental material.

6.2.3 Feasibility

Ideally, the measure developer should perform feasibility ① testing very early in the development process, preferably, right after establishment of the measure ①'s description, numerator ①, and denominator ①; after identification of the required data for measure calculation; and before finalization of specifications.

 Feasibility testing is especially important for eQMs ①. Given disparate EHRs ① and other electronic clinical systems, collection and storage of required data elements ① may be different. Therefore, the presence or absence of a data element ① in the EHR and other electronic clinical systems will inform whether the required data elements will be in the measure specifications ① or whether the measure developer will need to explore other similar data elements. eQMs have a separate [feasibility scorecard](#) [↗] assessing data availability, data accuracy, data standards, and workflow. eQMs developers can use the *MC Workspace* [eQMs Test Results](#) [↗] module to publish an eQMs data element feasibility template to gather feedback from stakeholders.

The measure developer should use testing to assess measure feasibility. One aspect of feasibility is the extent to which required data are available and retrievable without undue burden, and the extent to which measure developers can collect data and process for performance measurement. Some feasibility information may be obtained when assessing validity of the measure score ① or data elements (e.g., quantifying the frequency of absent diagnosis codes when a target condition is present). The measure developer should obtain other feasibility information using systematic surveys (e.g., survey of physician practices tasked with extracting the information). They may choose to gather more in-depth information by conducting focus groups composed of professionals who may be responsible for a measure's implementation.

Feasibility assessments should address

- availability of data (e.g., evidence of routine generation and use in care delivery of required data, including any exclusion criteria)
- extent of missing data, measure susceptibility to inaccuracies, and the ability to audit ① data to detect problems
- estimates of the costs or burden of data collection, data entry, and analysis including the impact on clinician workflow, diagnostic thought processes, and patient-physician interaction

- barriers encountered in implementing performance measure ❶ specifications, data abstraction, measure calculation, or performance reporting
- ability to collect information without violation of patient confidentiality, including circumstances where measures based on patient surveys or the small number of patients may compromise confidentiality
- identification of unintended consequences

6.2.4 Usability and Use

All measures ❶ may not need formal usability ❶ testing. In some cases, the TEP may review measure characteristics (e.g., descriptive statistics, dispersion of comparison groups) to determine usability of the measure for performance improvement and decision-making and whether negative unintended consequences are likely. When there is a requirement for more formal testing to assess the understandability and decision-making utility of the measure with respect to intended audiences (e.g., consumers, purchasers, providers, policy makers), a variety of methods are available:

- focus groups
- structured interviews
- surveys of potential users

These different methods often focus on the discriminatory ability of the measure and the meaning of the score as applied to evaluation of comparison groups or decision-making. For example, the measure developer should use a survey of potential users 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 ❶. Measure developers should consider developing a plan for implementation.

6.3 DEVELOP THE TESTING WORK PLAN

The [MIF](#) ❷, [MJF](#) ❷, [Measure Evaluation Report](#) ❷, and several testing documents on the [NQF Submitting Standards page](#) ❷ are available to assist with developing and documenting the testing work plan and updating measure specifications ❶.

Testing work plans vary depending on measure type and complexity. Measure developers can test a single measure or a set of measures. If testing targets a set of measures, the measure developer should construct a work plan that describes the full measure set ❶. The work plan for alpha testing is usually prepared early in the measure development process; therefore, the exact number of measures for testing may not be known, and many work plan areas listed may not be appropriate. In contrast, the work plan for a beta test should be prepared after measure specifications have been developed and should include sufficient information to help understand how sampling and planned analyses aim to meet scientific acceptability ❶, feasibility ❶, and usability and use criteria required for endorsement by NQF.

The testing plan should contain

- name(s) of measure(s)
- type of testing
- study objective(s)
- timeline for testing and report completion
- data collection methodology

- description of test population①, including number and distribution of test sites/data sets, when available
- description of data elements① for collection
- sampling methods, if applicable
- if using multiple sites or data sets, a description of strategy to recruit providers/obtain test data sets
- analysis methods planned and description of test statistics to support assessment. This will be less extensive for an alpha test. For a beta test, methods and analysis should address these evaluation criteria
 - importance①—including analysis of opportunities for improvement such as reducing variability in comparison groups or disparities in healthcare① related to race, ethnicity, age, or other classifications
 - scientific acceptability①—including analysis of reliability①, validity①, and exclusion① appropriateness
 - feasibility①—including evaluation of reported costs or perceived burden, frequency of missing data, and description of data availability
 - Usability① and Use—including planned analyses to demonstrate the measure is meaningful and useful to the target audience. The TEP may accomplish this by reviewing the measure results (e.g., means and detectable differences, dispersion of comparison 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)

6.4 PERFORM SAMPLING

The need for sampling often varies depending on the type of test (i.e., alpha or beta) and type of measure. When determining the appropriate sample① size during testing, the measure developer must evaluate the burden placed on providers and/or beneficiaries to collect the information. For example, measure developers may test measures that rely on administrative or claims data by examining data from the entire eligible population with limited drain on external resources, depending on the nature of the analysis. However, to test some measures, it is necessary to collect information from service providers or beneficiaries directly, which can become burdensome to measure developers, service providers, and beneficiaries. Outcome-dependent and covariate-dependent sampling are two approaches to reduce the burden of data collection while maintaining the ability to conduct meaningful testing ([Ding, Lu, Cai, & Zhou, 2017](#)^[2]). Outcome-dependent sampling may be an efficient, but statistically equivalent to simple random samples, method for developing a risk model. For example, assume a measure developer wanted 30 cases for each covariate to estimate the coefficients. For a relatively infrequent event, such as <10%, it would be more cost effective for them to use a higher sampling probability for Y=1 than Y=0.

As previously noted, alpha testing frequently uses a convenience sample; however, beta testing may involve measurement of a target population①, which requires careful construction of samples to support adequate testing of the measure's scientific acceptability. The analytic unit of the specific measure (e.g., physician, hospital, home health agency) determines the sampling strategy. In general, samples used for reliability① and validity testing① should

- Represent the full variety of measured entities (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 reliability① and validity① analyses using the planned statistical methods. When possible, observations should be randomly selected.
- Be of high-quality. Measure developers must ensure data used for risk adjustment① are of high-quality. Refer to the [Risk Adjustment in Quality Measurement](#)[†] supplemental material for specific considerations.
-  Test measure calculation against an appropriate data set that reflects multiple reporting entities (e.g., providers, provider groups, or hospitals) to evaluate the impact of measure calculations when there may be an attribution①-related concern for providers using shared EHRs①.

6.5 IMPLEMENT THE PLAN

The measure developer should execute the work plan iteratively and refine the work plan based on intermediate results.

6.6 ANALYZE THE TEST RESULTS

After completion of data gathering from the test sites, the measure developer conducts a series of analyses to characterize evaluation criteria of the measures①. The assumption is that a measure developer will contract or employ experienced methodologists, statisticians, and SMEs to select testing that is appropriate and feasible for the measure(s) under development.

6.7 REFINE THE MEASURE

The measure developer may need to modify measure specifications①, data collection instructions, and calculation of measure results based on analysis of testing results.

Examples

- Following alpha testing, the measure developer may undertake measure respecification or efforts to overcome implementation barriers.
- Following beta testing, changes in the definition of the population① or adjustments to the comparison group definition may occur.
- If making changes to the measure, the measure developer should consult with the TEP prior to retesting the measure.

6.8 RETEST THE REFINED MEASURE

Measure testing① is an iterative process. The measure developer should continue to refine specifications and retest measures as deemed necessary.

6.9 MEASURE TESTING SUMMARY

When reporting measure testing results, the measure developer's assessment of each of four measurement criteria is a matter of degree. For example, not all revisions will require extensive reassessment for all testing criteria, and not all previously endorsed measures will be strong—or equally

strong—among each set of criteria. Assessment is often a matter of judgment and expertise. Given the difficulty of assessment, the expectation is that measure developers contract or employ clinical experts in addition to experienced statisticians and methodologists to provide expert judgment when reporting [measure reliability](#) and [validity](#). The measure testing summary should reflect expert findings/consensus with respect to the measure, including [importance](#), [scientific acceptability](#), [feasibility](#), and [usability](#) and use.

The measure testing summary may include

- name of measure or [measure set](#)
- executive summary of tests and resulting recommendations
- type of testing conducted (i.e., alpha or beta), and overview of testing scope
- description of any deviation from the work plan along with rationale for deviation
- data collection and management method(s)
- description of test [population\(s\)](#) and description of test sites, if applicable
 - description of test [data elements](#), including type and source
 - [data source](#) description (and export/translation processes, if applicable)
 - sampling methodology, if applicable
 - description of [exclusions](#), if applicable
 - patient [medical record](#) review process, if applicable, including abstractor/reviewer qualifications and training, and process for adjudication of discrepancies between abstractors/reviewers
- detailed description of measure [specifications](#) and [measure score](#) calculations
- description of the analysis conducted, including
 - summary statistics (e.g., means, medians, [denominators](#), [numerators](#), descriptive statistics for exclusions)
 - importance — specific analyses demonstrating importance, such as suboptimal performance for a large proportion of comparison groups and analysis of differences between comparison groups
 - scientific acceptability
 - reliability — description of reliability statistics and assessment of adequacy in terms of norms for the tests, and rationale for analysis approach
 - validity — specific analyses and findings related to any changes observed relative to analyses reported during the prior assessment/endorsement process, or changes observed based on revisions to the [measure](#); these may include assessment of adequacy in terms of norms for the tests conducted, panel consensus findings, and rationale for analysis approach
 - [exclusions/exceptions](#) — discussion of the rationale, which may include listing citations justifying exclusions; 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 need for [risk adjustment](#) and [stratification](#) as described in the [Risk Adjustment in Quality Measurement](#) supplemental material
- feasibility—discussion of feasibility challenges and adjustments made to facilitate obtaining measure results, and description of estimated costs or burden of data collection
- [usability](#) and use—if materially changing the measure, the recommendation is to provide a summary of findings related to measure interpretability and methods used to provide a qualitative and quantitative usability assessment (e.g., TEP review of measure results)

- any recommended changes to the measure specifications and an assessment as to whether there is a need for further testing
- detailed discussion of testing results compared to NQF requirements, including whether testing results sufficiently met NQF requirements or whether there is a need for additional testing
- any limitations of the alpha or beta testing, such as
 - sample limited to two sites or three EHR applications
 - sample used registry data from only one state, and registry data are known to vary across states
 - testing was formative alpha test only and not intended to address validity and reliability
- recommend approval of a candidate measure for further development
- recommend approval of a fully tested and refined measure for implementation
- plan for comprehensive reevaluation

6.10 STAKEHOLDER ENGAGEMENT DURING MEASURE TESTING

It is often appropriate for the measure developer to obtain stakeholder input at several points during the testing process. This process may include obtaining face validity assessments at alpha and beta testing, feasibility and burden inputs at beta testing, and other inputs based on a review of overall results. These inputs can take many forms, including formal TEPs, consultation with SMEs, outreach to professional associations or patient advocacy groups, and public comments. Once collection of these inputs occurs, it is important to follow up on the communications by providing additional opportunities for stakeholders to comment on the results of their inputs at future stages. Follow up increases the likelihood that the measure developer operationalizes the inputs consistent with the stakeholders' needs. It also improves the likelihood that stakeholders will remain engaged for ongoing support on current or future measures. If a TEP reviews testing results and updated specifications, the measure developer may post those summaries for further public comment.

Clinicians and facilities can volunteer to test measures under development. Additionally, through the *MC Workspace* [eCQM Test Results](#) module, eCQM data elements ready for testing are posted in a Data Element Feasibility Testing template. Stakeholders may use the template to share feasibility based on their processes and workflows.

6.11 TESTING AND EVALUATION FOR SPECIAL MEASURES

Special types of measures may require the measure developer to assess different measure aspects. For example, while measure developers assess all measures for feasibility, data element feasibility is a major focus in testing and evaluating eCQMs. To assist in assessing an eCQM's feasibility, the measure developer needs to include the [eCQM Feasibility Scorecard](#), not only as part of the testing plan, but also during feasibility testing as changes happen. For the measure testing summary related to an eCQM, documentation must show

- evidence of testing with at least two different EHRs
- QDM data elements and the feasibility ratings of those elements

Because composite measures include component measures, there are additional considerations when testing and evaluating composite measures. For more information on testing and evaluating special types of measures, refer to the individual special measure supplemental materials.

7 MEASURE IMPLEMENTATION

This chapter discusses the implementation process, CMS pre- and rulemaking processes (measure selection), and measure rollout. Quality measure implementation includes all activities associated with progressing a measure from the development state into an active, in-use state, including consensus endorsement processes, measure selection processes, and measure rollout.

Measure implementation processes vary depending on the program in which the implementation occurs. There is a statutory requirement for some CMS programs to use the pre-rulemaking and rulemaking processes for measure selection. The process of implementing measures varies significantly from one measure set to another, depending on a number of factors.

- scope of measure implementation
- measured entity
- data collection processes
- ultimate use of the measure (e.g., quality improvement, public reporting, pay-for-reporting, value-based purchasing)
- program into which the measure is being added

The scope of measure implementation could entail a measure or measure set being

- implemented in a new program
- added as a new measure or measure set for an existing program

7.1 THE IMPLEMENTATION PROCESS

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

- [NQF endorsement](#), if applicable
- measure selection
- measure rollout

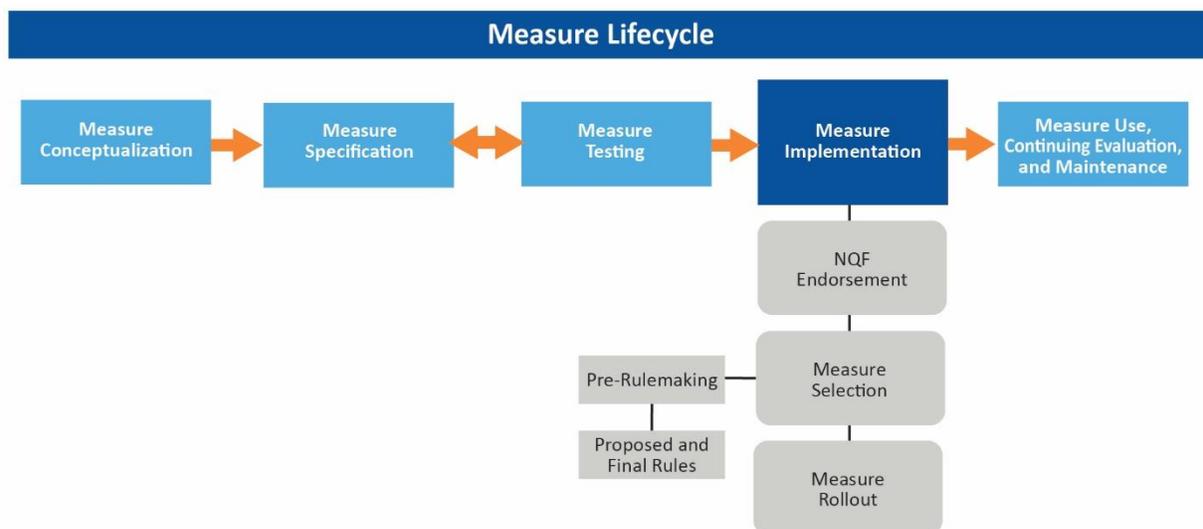


Figure 13. Flow of the Measure Lifecycle—Measure Implementation

More information about the NQF endorsement process is available in the [NQF Endorsement and Maintenance](#) supplemental material.

7.2 MEASURE SELECTION

The path a measure takes for selection and implementation depends on the program, as not all measures go through the pre-rulemaking and rulemaking process. The requirement is that only measures for Medicare quality programs subject to Section 3014 of the ACA go through the pre-rulemaking and rulemaking process. For a list of the programs, and information on their needs and priorities, see the [MUC List Program-Specific Measure Needs and Priorities](#). Other quality initiatives, like those for Medicaid and the Health Insurance Marketplace Quality Initiatives, do not go through pre-rulemaking and follow another path.

7.2.1 Pre-Rulemaking Process

Section 3014 of the [ACA](#) mandated the establishment of a federal pre-rulemaking process for selecting quality and efficiency measures for specific programs within HHS. The pre-rulemaking process requires HHS to consider multi-stakeholder input on quality and efficiency measure selection. To meet these requirements, CMS develops a MUC list. The NQF-convened MAP is currently the multi-stakeholder group described in Section 3014 and it provides input to HHS on the list of measures for use in a specified program. By statute, HHS and CMS must consider MAP input and publish the rationale for selecting any measure (i.e., in proposed or final rules) not NQF-endorsed.

7.2.1.1 Measures Under Consideration

Over the past few years, CMS has articulated a number of measure selection criteria in its final rules for various programs. The term “measure selection” typically applies to determining whether a measure should be included in a measure set for a specific program, while “measure evaluation” applies to assessing the merits of an individual measure, not in the context of a specific program. CMS has established a set of measure selection criteria so that HHS can develop the MUC list for qualifying programs and make it available publicly by December 1 each year. These selection criteria are operationalized by CMS program staff and leadership, who decide which measures to place on the MUC list for review by the MAP.

Figure 14 contains a sample timeline for the pre-rulemaking process.



Figure 14. Sample Pre-rulemaking Process Timeline

After opening an online data collection system with database and accompanying user interface intended to gather specifications (i) and supporting information on new candidate measures (i), CMS publishes guidance on the [CMS Pre-Rulemaking website](#) (i). CMS may host educational webinars to kick off the official MUC (i) season. More information on the CMS pre-rulemaking process is available on the [CMS Pre-Rulemaking website](#) (i).

7.2.1.2 CMS Measure Selection Criteria

CMS measure selection criteria help to ensure that each measure

- is responsive to specific program goals and statutory requirements
- addresses an important condition or topic with a performance gap and has a strong scientific evidence base to demonstrate that, when implemented, the measure can lead to the desired outcomes and more affordable care. This requirement corresponds to NQF's importance criterion (i).
- addresses one of the Meaningful Measure Areas
- promotes alignment (i) with CMS program attributes and across HHS programs
- is feasible to report, fully developed, and tested. At a minimum, measures must be tested for reliability (i) and validity (i).
- Offers results and performance that identify opportunities for improvement. CMS will not select measures when evidence already identifies high levels of performance with minimal opportunity for improvement (i) (i.e., topped out (i) measures).
- avoids negative unintended consequences (e.g., discharged too soon, overuse or inappropriate use of treatment, limiting access to care)
- does not duplicate another measure currently implemented in one or more programs
- if it is an eCQM (i), was created using the MAT and is expressed in Health Quality Measure Format (HQMF) (i) using QDM (i) and CQL (i).

Applying the [measure](#) selection criteria, CMS develops the [MUC list](#). CMS may ask measure developers to provide details on the measures to help CMS develop the [MUC](#) list. CMS then provides this list to the [MAP](#).

7.2.1.3 MAP Recommendations

The MAP input to HHS on the annual list of [quality](#) and [efficiency measures](#) that are under consideration by one or more Medicare programs is due by February 1 of each year as a recommendation report. Each annual report can be found on the [MAP pages](#) on the NQF website and on the [CMS Pre-Rulemaking website](#). Measure developers are strongly encouraged to attend the MAP meetings.

7.2.1.4 CMS Considers MAP Input for Final Selection

After CMS receives the MAP input, a deliberation process begins. CMS determines the inclusion of the measures in the federal rulemaking processes. The measure selection criteria used during development of the MUC list, and identified in [Chapter 7.2.1.2](#), are the same criteria used for 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—not endorsed by NQF.

7.2.2 CMS Rulemaking Processes

After CMS completes the pre-rulemaking process and selects measures for potential inclusion in rulemaking, the next steps in the cycle are

- Proposed rules: CMS writes the proposed rules and publishes them in the Federal Register. A proposed rule is generally available for public comment for 60 days.
- Final rules: CMS considers the received comments and publishes the final rules in the Federal Register.

CMS treats existing measures that undergo substantive changes as new measures. Some examples of substantive changes are a change to the

- intent of the measure
- [numerator](#)'s inclusion and/or [exclusion](#) criteria
- methodology previously published in a final rule
- [denominator](#), [denominator exclusions/exceptions](#), or [numerator exclusions/exceptions](#) cohort, both a significant increase or decrease
- science impacting the primary medication, dosage, or medical device
- removal or addition of a component measure to a [composite measure](#)

7.2.3 Stakeholder Engagement in Measure Selection

CMS requests public comments on MUC for implementation either through the formal federal pre-rulemaking and rulemaking process or through an ad hoc public comment process for measures that are not subject to rulemaking. The MAP publicly discusses measures on the MUC list and posts their recommendations for which measures to implement in the different CMS programs for public comment.

Measure developers convene stakeholder meetings regarding implementation of considered measures and resolution of stakeholder questions about the measures occurs iteratively as the measure remains under consideration. The measure implementation process is completely transparent and open to the public for comments and questions.

7.3 ROLLOUT

When measures ① are selected for use, the measure developer prepares plans for implementation, including the initial rollout, data management and production, audit ① and validation ①, provider education, dry runs ①, and appeals processes. Measure rollout occurs after measure approval for use in a program. The rollout process may include collection of data for a dry run from all relevant providers across the country and share calculated rates with the providers. CMS does not use dry run data for payment, but may use them as a baseline for future payment years. As with the overall implementation process, the process of rolling out measures varies significantly from one measure set ① to another depending on a number of factors, which may include

- the scope of measure implementation
- the measured entity
- data collection processes
- the ultimate use of the measure (for example, quality improvement, public reporting, pay-for-reporting, value-based purchasing)
- the program into which the measure is added

If a dry run occurs, the measure developer ensures the rollout plan includes support for the providers to

- improve the usability of the measure report to the providers in advance of implementation
- identify and respond to questions and concerns about the measures
- address issues with the report production process for process improvements prior to implementation

The measure developer documents the results from the dry run and assesses the measures' success in meeting the program's intentions for the dry run, such as

- adequacy (specificity ①, accuracy) of the measure specifications ①
- accuracy of the data collection methods
- accuracy of the measure results calculations
- identification of unintended consequences, gaming ①, or misrepresentation (if any)
- accuracy and adequacy of the provider reports (whether they are useful to the facilities/entities involved and whether the measure developer is to respond to questions and concerns)

When communicating and coordinating with all parties involved in the rollout, the measure developer must consider the timelines of other processes (for example, rulemaking, NQF projects, and quality alliances). The measure developer prepares and presents education for the end users on what is being measured and how to interpret the results.

The measure developer also documents the results of any educational activities and assesses whether the activities were adequate to meet the needs of the end users of the measures. For example, the measure developer should report on the number of events, including the attendance at each

- conference call and recordings of the calls
- web-based presentation and recordings of the presentations
- workshop at conferences or scientific society meetings
- train-the-trainer event



For eCQMs^①, measure developers should refer measure implementers to the [CMS Data Element Repository \(DERep\)](#)[↗] for information about the data elements^① associated with eCQMs used in CMS quality reporting programs as well as the definitions and clinical focus for each data element^①.

8 MEASURE USE, CONTINUING EVALUATION, AND MAINTENANCE

Chapter 8 explains the processes of measure production and monitoring; continuing evaluation; and three types of [measure maintenance reviews](#) ^①, including the process for annual updates, comprehensive reevaluation, and ad hoc review. Measure use, continuing evaluation, and maintenance are necessary to assure the ongoing accuracy and value of the [measure](#) ^①.

Once a measure is in use, there is a requirement for periodic reevaluation to determine whether its strengths and limitations related to the measure evaluation criteria have changed since the last formal evaluation. Every measure should undergo a high-level review at least annually, along with a vigorous and more comprehensive review every three years. Reevaluation also gives insight on whether the measure should continue to be in use. To help ensure the continued soundness of the measure, the measure developer must provide strong evidence that a measure currently in use continues to add value to quality reporting and incentive programs and that its construction continues to be sound throughout its lifecycle. This work also helps ensure that measures obtain or maintain NQF endorsement, if desired.

There is discussion of the details of the requirements for the different reviews in the [Measure Maintenance Reviews](#) [↗] supplemental material.

As depicted in [Figure 15](#), there are multiple steps to measure maintenance. Reporting of these steps is via three basic types of measure maintenance reviews: annual updates, comprehensive reevaluations, and ad hoc reviews, with stakeholder inputs being a critical component of this review process.

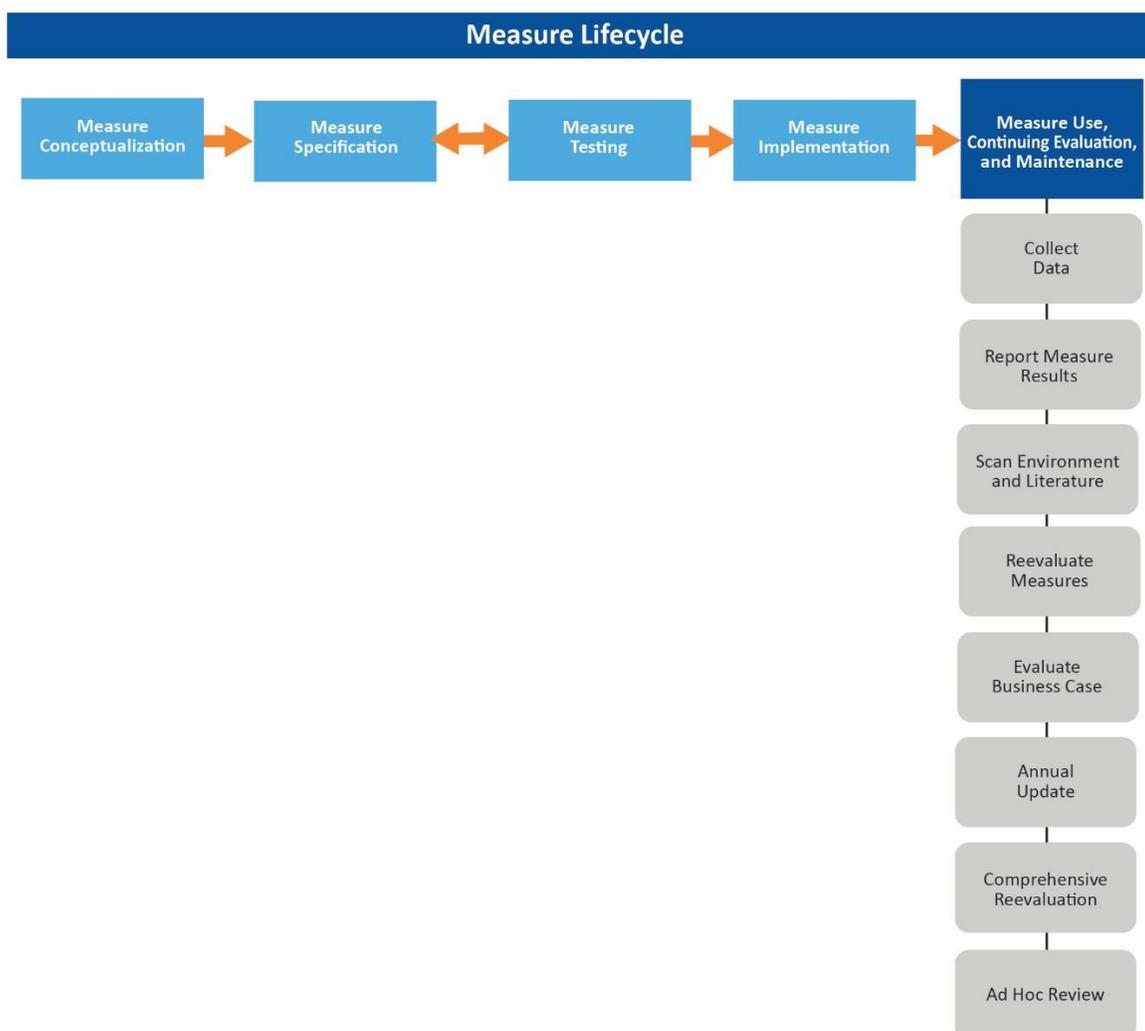


Figure 15. Flow of the Measure Lifecycle—Measure Use, Continuing Evaluation, and Maintenance

8.1 MEASURE PRODUCTION AND MONITORING

Measure production and monitoring include the ongoing tasks necessary to use the measure ^① over time. As with other aspects of implementation, the process of measure production and monitoring varies significantly from one measure set ^① to another depending on a number of factors, which may include

- scope of measure implementation
- measured entity
- data collection processes
- intended use of the measure (e.g., quality improvement, public reporting, pay-for-reporting, value-based purchasing)
- program measure use

The level of effort for each task may vary by different measure attributes.

Figure 16 is a diagram of the overall production and monitoring components of an implemented measure. There may be a requirement for measure developers to perform various tasks associated with ongoing implementation and production. A discussion of some examples of these steps are in the next subsections.

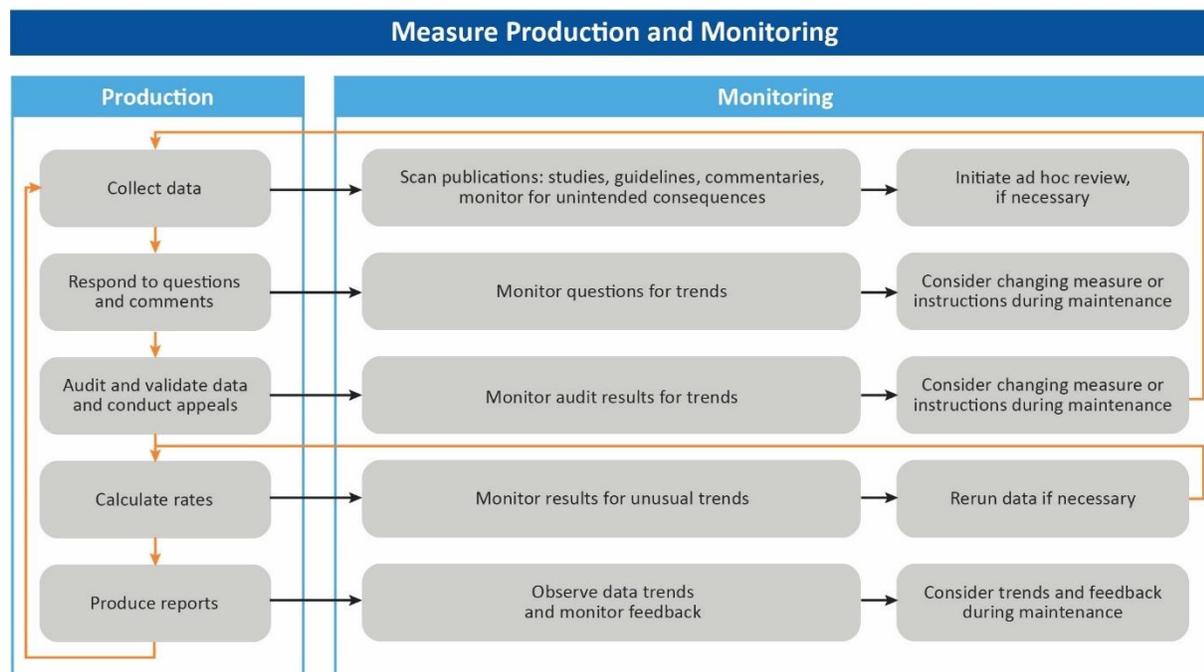


Figure 16. Overview of the Measure Production and Monitoring Process

8.1.1 Conduct Data Collection and Ongoing Surveillance

Once measure development is complete and dry run identified implementation issues resolved, the measure may go to full implementation (i.e., data are being collected, calculated). The measure developer’s ongoing surveillance should include scanning medical and scientific publications as well as the general media for articles and commentaries about the measure.

New information may trigger an ad hoc review if the concern needs immediate action. Ongoing information surveillance is similar to the information gathering stage of measure development, as covered in Chapter 4, Measure Conceptualization. The measure developer conducts similar analyses of the literature, with reports submitted or used for annual updates to the measure, as required by the project. Measure developers can use the [ESST](#) to assist with the review of the literature. Instructions on how to use the ESST are found on the [ESST website](#).

As use of the measure continues, the measure developer identifies publication of new studies that address the soundness of the measure. They must pay attention to any organizations that issue relevant clinical practice guidelines, especially for process measures. If the measure is based on a specific set of guidelines, the measure developer should monitor the guidelines’ publisher closely for any indication of plans to change their guidelines. If the measure is not based on guidelines, the measure developer can monitor the scientific and clinical literature for reports that would impact the scientific basis of the measure. These guideline changes or other statements may also trigger an ad hoc review.

After data collection begins, the measure developer monitors for unintended consequences the measure might have on clinical practice or outcomes. Articles in the literature may identify unusual

trends in data suggesting unintended consequences. If the measure developer identifies significant unintended consequences, especially if patient safety is the concern, they must not wait for a scheduled annual or comprehensive review. An ad hoc review may be necessary.

8.1.2 Respond to Questions About the Measure

The measure developer may be responsible for reviewing stakeholder feedback and responding promptly, which may include help desk questions. This stakeholder feedback may include direct questions or comments about the [measure](#) ① or the use of the measure in a specific program. Assuming the source of the feedback has provided contact information, the measure developer should reply immediately, alerting the submitter of receipt and review of the feedback. Questions may offer ideas for education and outreach.

8.1.3 Monitor and Analyze the Measure Rates and Audit Findings

The measure developer should monitor and analyze measure performance rates and [audit](#) ① findings periodically and at least once a year for

- overall performance trends
- variations in performance, gaps in care, and extent of improvement
- [disparities](#) ① in resulting rates by race, ethnicity, age, social risk factors, income, region, gender, primary language, disability, or other classifications
- frequency of use of [exclusions](#) ① or [exceptions](#) ① and how they influence rates
- discretionary exclusion, evaluate carefully for [gaming](#) ①, unintended consequences, and uneven application that could influence comparability
- patterns of errors in data collection or rate calculation
- changes in practice that may adversely affect rates
- impact of measurement activities on providers
- correlation of the data to either confirm the measure's efficacy or identify weaknesses in the measure

Ongoing monitoring should continually assess a measure's progression; any marked departures may be cause for concern. If the [business case](#) ① predicted performance targets, the measure developer should investigate any measure whose performance over time falls short of its target. This information is reported during reevaluation as described in [Chapter 8.3, Measure Maintenance](#) ① Reviews and the [Measure Maintenance Reviews](#) [↗] supplemental material.

8.1.4 Perform Measure Maintenance or Ad Hoc Review, When Appropriate

The measure developer should review each measure at least annually to ensure that the codes used to identify the [populations](#) ① (e.g., [denominator](#) ①, [numerator](#) ①, [exclusions](#)) are current, and to address other [minor changes](#) ① that may be needed. The standardized processes for annual update and the triennial full reevaluation are described in [Chapter 8.3, Measure Maintenance Reviews](#) and the [Measure Maintenance Reviews](#) [↗] supplemental material.

If NQF has endorsed the measure, the measure developer reports the results of the maintenance review to NQF to reevaluate its endorsement at the time of NQF maintenance review.

8.2 CONTINUING EVALUATION

The measure developer uses the continuing evaluation process to make any changes to the technical specifications ① to demonstrate that

- The aspects of care included in the specifications continue to be highly important to measure and report because the measurement results can supply meaningful information to consumers and healthcare providers.
- The measurement results continue to drive significant improvements in healthcare quality and health outcomes where there is variation in and/or overall, less-than-optimal performance.
- The data elements ①, codes, and parameters included in the specifications are the best ones to use to quantify the specific measure ① because they most accurately and clearly target the aspects of the measure that are important to collect and report, and they do not place undue burden on resources in order to collect the data.
- The calculation methods included in the specifications remain valid because they reflect a clear and accurate representation of the variation in the quality or efficiency of the care delivered or the variation in the health outcome of interest.
- The measure continues to be either unique for its topic or it is the “best in class” when compared to competing measures ①.
- The measure is comparable to other measures in its clinical significance or difficulty.

8.3 MEASURE MAINTENANCE

During measure maintenance ①, it is important that measure developers analyze measure performance trends, including feedback through help desks and trainings, to determine whether the measure undergoing reevaluation is still the best or most relevant measure, and whether there are unintended consequences that need to be addressed.

After implementation of measures, the measure developer monitors measures’ performance, responds to ongoing feedback, and continuously scans the environment regarding the measures. For example, for eCOMs ①, the ONC Project Tracking System ① (Jira) ① is one method for collecting and monitoring feedback on measure implementation.

In addition, two measure maintenance activities 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. Find a description of these reviews in 8.5 8.5, Measure Maintenance Reviews and the Measure Maintenance Reviews ① supplemental material. CMS uses one of five outcomes following maintenance review of CMS measures.

- Retain—keep the measure active with its current specifications and minor changes ① (refer to Figure 17).
- Revise—update the measure’s current specifications to reflect new information (refer to Figure 17).
- Retire—cease to collect or report the measure indefinitely. This applies to measures unowned or maintained by any measure steward ①. If it is necessary to retire a measure from a set, other replacement measures may be available to complement the remaining measures in the set (refer to Figure 18).

- Suspend—temporarily cease to report a measure ①. Data collection and submission may continue.
- Remove—a specific program set no longer includes a measure for one or more reasons. This does not imply that other payors/purchasers/programs should cease using the measure. If CMS is the measure steward ① and another CMS program continues to use the measure, CMS may continue to maintain the measure. If another entity is the steward, the other payors/purchasers/programs that may be using the measure are responsible for determining if the steward should continue to maintain the measure (refer to Figure 19).

Figure 17, Figure 18, and Figure 19 list the criteria CMS uses to make decisions regarding the various dispositions.⁹ Some programs may have other possible outcomes.

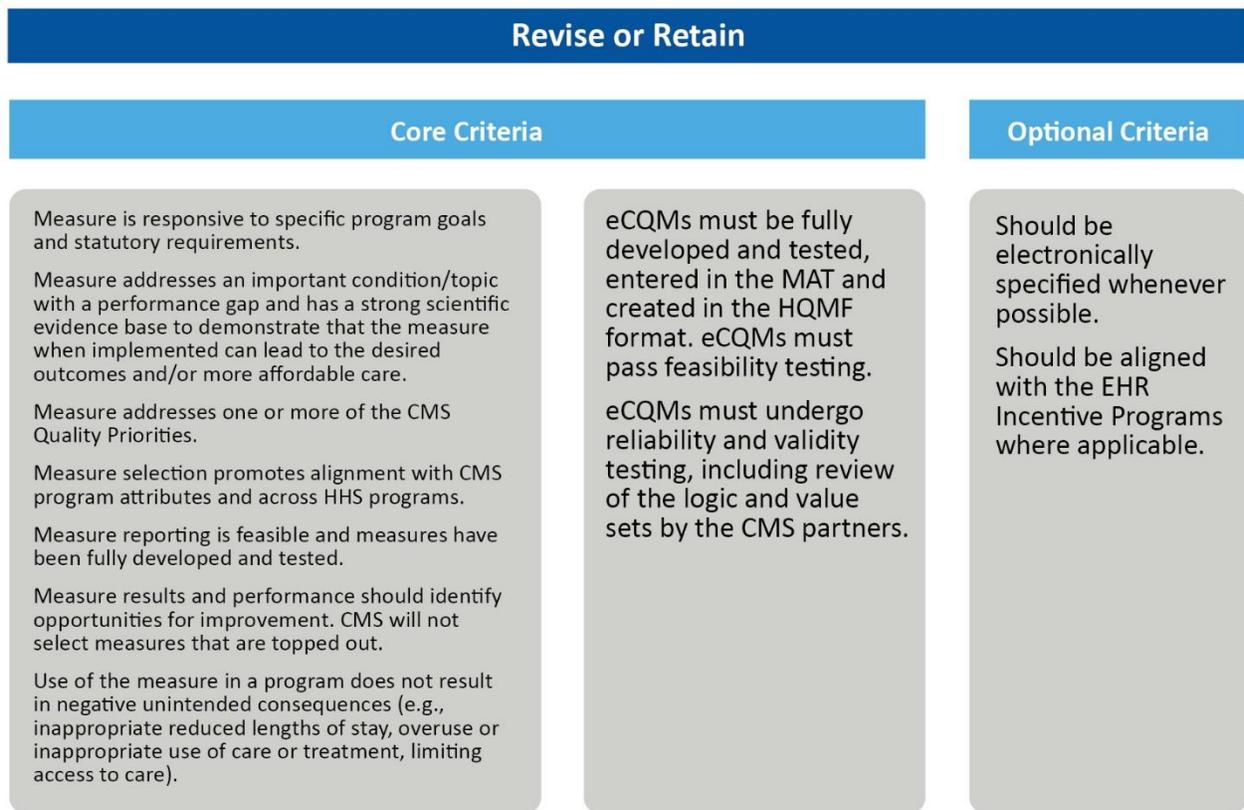


Figure 17. CMS Criteria for Measure Disposition: Revise or Retain

⁹ Adapted from 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.

Retire	
Core Criteria	Optional Criteria
<p>Measure is owned by CMS, and CMS will no longer maintain the measure.</p>	<p>No longer adds value commensurate with the cost of data collection and reporting.</p> <p>Performance or improvement on a measure does not result in better outcomes.</p> <p>Collection or public reporting of a measure leads to negative unintended consequences other than patient harm.</p> <p>Does not align with current clinical guidelines or practice.</p> <p>Measure performance is so high and unvarying that meaningful distinctions and improvements in performance can no longer be made.</p> <p>The availability of a better measure that is more (1) broadly applicable (across settings, populations, or conditions); (2) proximal in time to desired outcomes for the particular topic; (3) strongly associated with desired outcomes for the particular topic; or (4) aligned with other CMS/HHS programs.</p>

Figure 18. CMS Criteria for Measure Disposition: Retire

Remove	
Core Criteria	Optional Criteria
<p>Measure is no longer used in a CMS program. If the measure is owned by CMS, CMS may continue to maintain it even after removal.</p> <p>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.</p>	<p>Collection or public reporting of a measure leads to negative unintended consequences other than patient harm.</p> <p>The measure, as currently specified, cannot be reported.</p> <p>A measure does not align with current clinical measure performance or is so high and unvarying that meaningful distinctions and improvements in performance can no longer be made.</p> <p>The availability of a better measure that is more (1) broadly applicable (across settings, populations, or conditions); (2) proximal in time to desired outcomes for the particular topic; (3) strongly associated with desired outcomes for the particular topic; or (4) aligned with other CMS/HHS programs.</p>

Figure 19. CMS Criteria for Measure Disposition: Remove

8.4 EVALUATION DURING MEASURE MAINTENANCE

Full development of a measure ❶ does not end measure evaluation. The measure must also be continuously reevaluated during maintenance ❷, with measure documents updated regularly. Although evaluation details may differ for specific reevaluations, the general principles are the same.

As they did during measure development, the measure developers, TEP members, and other stakeholders involved in [measure maintenance](#) ① work toward ensuring use of sound [measures](#) ① to drive healthcare quality improvement and inform consumer choice. During measure maintenance, the measure developer must continue to evaluate measures and provide strong evidence that they conduct measure construction in a sound manner, and that they are continuing to add value to quality reporting programs.

8.4.1 Apply Measure Evaluation Criteria

Each measure should undergo an update at least annually, along with a rigorous, comprehensive reevaluation every three years to assess its continued value, based on the most current set of measure evaluation criteria.

Evaluation during maintenance should also document how the measure is performing compared to the projected trajectory in the [business case](#) ① during measure development. Throughout the measure evaluation process, measure developers update justification for the measure and any changes to the [technical specifications](#) ① to demonstrate

- Aspects of care included in the specifications continue to be highly important to measure and report, supply meaningful information to consumers and healthcare providers, and drive significant improvements in healthcare quality and health outcomes.
- [Data elements](#) ①, codes, and parameters included in the specifications are the best ones to use to quantify the specific measure, and data collection imposes only a modest, proportionate 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.

With respect to [usability](#) ①, the measure developer should demonstrate improvement with year-over-year performance data and assess for implementation challenges and obtain feedback from measure users.

8.4.2 Document Results of Evaluation

Measure evaluation criteria and subcriteria are detailed in the [Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement](#) [↗]. The measure developer must document how the measure is performing during the maintenance stage.

8.5 MEASURE MAINTENANCE REVIEWS

This section describes three types of maintenance reviews.

- annual update
- comprehensive reevaluation
- ad hoc review

Find detailed steps for each of these reviews in the [Measure Maintenance Reviews](#) [↗] supplemental material.

8.5.1 Annual Update

The annual update process ensures an update to measures as there are updates to the code sets on which the measures rely. Consider any comments and suggestions received after implementation during annual updates to determine whether there is a need for revision beyond updating the codes.

For measures① proposed for revision, suspension, removal, or retirement, the measure developer evaluates the impact of the decision on the program using the measure.

8.5.2 Comprehensive Reevaluation

Measure developers should conduct a formal review of the measure every three years. NQF requires comprehensive reevaluation for endorsed measures. For more information, see the [NQF Endorsement and Maintenance](#) [↗] supplemental material. In many ways, the comprehensive reevaluation process parallels the measure development process.

The comprehensive reevaluation process ensures that measures continue to be of the highest caliber possible. By periodically reviewing the measures against standard measure evaluation criteria, the measure developer helps maintain the best measures over time.

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.

For the updated documentation, the measure developer must ensure that any updates to the evidence, clinical guidelines, impact of the measure, experience of measure use in programs, gaps in care, and measure performance are documented and disseminated to users and other stakeholders in a timely manner. NQF may require additional measure testing① for maintenance review. At the time of initial submission, the measure developer must be aware of any testing requirements needed at maintenance. As the science of measurement advances, evaluation and endorsement processes also evolve. Therefore, the measure developer must also stay abreast with changes in measure testing standards for the purposes of measure maintenance①. Refer to [Chapter 6, Measure Testing](#), for additional details.

The measure developer should include specific elements in the updated Business Case①.

- comparison of the actual performance of the measure with the projected rates
- statement of the impact of the measure on the focus quality topic
- updated projections for the next evaluation period

The updated documentation must adequately address any areas of potential concern that the measure developer anticipates arising during public comment or NQF review.

8.5.3 Ad Hoc Review

An ad hoc review is a focused examination of the measure's methodology or data collection process based on new information from clinical evidence or feedback from the field. If there is new evidence that a measure may have significant, adverse effects on patients, the measure developer must undertake an ad hoc review. Measure developers should conduct the ad hoc review only when new evidence indicates a major overhaul of the methodology.

8.6 STAKEHOLDER ENGAGEMENT DURING MEASURE MAINTENANCE REVIEWS

The Blueprint describes the annual update, comprehensive reevaluation, and ad hoc review as distinct and separate activities; however, in practice, these activities sometimes overlap and can occur concurrently. All steps require solicitation of inputs via public comment, usually through the rulemaking process. The measure developer should publicly report results from, and progress on, each of these review processes. Stakeholders are engaged for comment and, in some cases, participate in a formal panel review. For more information on [measure maintenance](#) reviews, see [Chapter 8.5](#) and the [Measure Maintenance Reviews](#) supplemental material.

Ideally, there is alignment between the measure maintenance schedule for [measures](#) used in federal programs and the NQF endorsement maintenance cycle, which also includes requirements for public review and comment. However, in practice, these schedules may not align completely. NQF-endorsed measures are listed on the [NQF QPS](#) and have a mechanism enabled for comment.

The [ONC Project Tracking System \(Jira\)](#) provides the opportunity for public comments on [eCQMs](#) in use. The measure developer may choose to include the comments as part of the Change Review Process and contribute to changes in eCQMs during the annual update. For more information on the eCQM Change Review Process and Annual Update, refer to the [eCQMs Specification, Testing, Standards, Tools, and Community](#) supplemental material.

9 TOOLS AND RESOURCES FOR MEASURE DEVELOPERS

Numerous tools and resources are available to assist measure developers in the different stages of the Measure Lifecycle. New tools and resources are always under development. [Chapter 9](#) provides illustrative examples of currently available tools and resources.

9.1 *BONNIE*

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

9.2 *CLINICAL QUALITY LANGUAGE (CQL) STYLE GUIDE*

The [CQL Style Guide](#)  provides standardized expressions of measure concepts across eCQMs and defines a uniform “look and feel” to eCQM logic using [CQL](#) . The guide focuses on an implemented set of common best practices across CQL-based eCQMs in CMS quality reporting programs. The Style Guide also promotes the use of consistent language within the framework of CQL, including libraries, aliases, definitions, and functions, as well as guidance on other conventions, such as operator precedence. [Measure stewards](#)  or measure developers who are developing or specifying eCQMs for future inclusion in CMS programs should align with these best practices.

9.3 *CMS DATA ELEMENT LIBRARY (DEL)*

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

- serve as a centralized resource for assessment data elements (questions and response options)
- promote sharing of electronic assessment data sets and [health IT](#)  standards
- influence and support industry efforts to promote [EHR](#)  and other health IT interoperability

9.4 *CMS MEASURES INVENTORY TOOL (CMIT)*

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

9.5 *ENVIRONMENTAL SCANNING SUPPORT TOOL (ESST)*

The [ESST](#)  is a tool within the CMIT environment, intended to automate [environmental scans](#)  required in the information gathering process to develop and maintain [quality measures](#) . The ESST uses an automated natural language processing (NLP) approach that rapidly scans literature in [PubMed](#) , [PubMed Central](#) , and the *Cumulative Index of Nursing and Allied Health Literature (CINAHL)*  to

- identify relevant documents (abstracts and full-text articles)
- identify and extract specific knowledge within each relevant document that applies to the [measure](#)'s potential [opportunity for improvement](#).

9.6 *CQL-TO-EXPRESSION LOGICAL MODEL (ELM) TRANSLATOR REFERENCE IMPLEMENTATION*

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

9.7 *CYPRESS*

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

9.8 *ELECTRONIC CLINICAL QUALITY IMPROVEMENT RESOURCE CENTER (ECQI RESOURCE CENTER)*

The [eCQI Resource Center](#) is a website that provides eCQI resources and connections for the community of professionals dedicated to eCQI for better health. It serves as “the one-stop shop for the most current resources to support electronic clinical quality improvement.” It is the source of truth for specifications of eQMs in CMS programs, CMS [QRDA](#) Implementation Guides, and is the home to the [MC Workspace](#), which includes the [eCQM Data Element Repository](#), [eCQM Concepts](#), [eCQM Clinical Workflows](#), and [eCQM Test Results](#).

The eCQI Resource Center has an [eCQI Tools and Resources Library](#) providing information and links to tools and resources used in eCQI.

9.9 *ELECTRONIC CLINICAL QUALITY MEASURE (ECQM) LOGIC AND IMPLEMENTATION GUIDANCE*

The [eCQM Logic and Implementation Guidance](#) is a resource document which provides guidance for understanding, using, and/or implementing eQMs.

9.10 *GUIDE FOR READING ECQMs*

The [Guide for Reading eQMs](#) is a resource to assist stakeholders in interpreting and understanding eQMs. The Guide provides information on eQMs such as file naming conventions, understanding an eCQM human-readable rendition, [value sets](#), [QDM](#) data criteria, and more.

9.11 *MEASURE AUTHORIZING TOOL (MAT)*

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

9.12 *MEASURES MANAGER*

The Measures Manager is a MIDS contractor responsible for numerous tasks supporting measure development. The Measures Manager's role in supporting CMS is to research and consider a wide variety of measure-related information and materials and to help CMS prioritize and coordinate measure development activities, which may include the actions

- reviewing HHS and CMS strategic plans, goals, and initiatives
- monitoring the progress of CMS measure development and maintenance projects against quality priorities and identify Meaningful Measurement areas in need of measure development
- producing [harmonization](#) and [alignment](#) tools and reports
- developing white papers to help CMS formulate measurement policies
- researching legislative mandates, proposed and final rules, and priorities of key external stakeholders
- supporting various HHS, CMS, and interagency working groups that focus on coordination of measure development, measure alignment, and harmonization
- supporting CMS's collection of measures for and management of the [MUC](#) list for pre-rulemaking
- maintaining a CMS Inventory of measures for policy and program use and update it three times per year
- supporting MIDS measure developers

9.13 *MEASURES MANAGEMENT SYSTEM WEBSITE*

The [MMS website](#) provides information on the MMS, measures development, how to get involved in the Measure Lifecycle, tools and resources, links to education, and other sources of information. It is the home of the Blueprint.

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

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

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

The [QPS](#) is a web-based inventory tool developed by the NQF to help stakeholders select and use NQF-endorsed measures. It enables a user to search for NQF-endorsed measures in many ways (e.g., by type of measure), and then export search results. A QPS user can find NQF-endorsed measures on particular topics, track and receive reminders about measures that are important to them, provide feedback on measures, and discover which measures others are using.

9.16 VALUE SET AUTHORITY CENTER (VSAC)

The [VSAC](#) is a repository and authoring tool for public [value sets](#) created by measure developers. Value sets are lists of codes and corresponding terms from National Library of Medicine (NLM)-hosted standard clinical vocabularies (e.g., [SNOMED CT](#), [RxNorm](#), [Logical Observation Identifiers Names and Codes \[LOINC\]](#)) and billing terminologies (e.g., [ICD-10-CM](#)) that define clinical concepts to support effective and interoperable health information exchange. The VSAC does not create value set content.

APPENDIX A: SUMMARY OF CHANGES TO BLUEPRINT

[Appendix A](#) presents a high-level summary of the changes found in this version of the CMS MMS Blueprint. These changes are arranged by Chapter.

Section/Chapter	Changes
Throughout	Updated graphics
Throughout	Added references to the MC Workspace
Preface – Using the Blueprint	Revised to discuss new format for the Blueprint information
Preface – Organization of the Blueprint Materials	Revised to discuss the new organization of the Blueprint package
Chapter 3	Added introduction to Structure-Process-Outcome
Chapter 3	Added the section, Measure Evaluation in the Measure Lifecycle, from the Measure Evaluation chapter
Chapter 3	Expanded Roles in Measure Development section – added Measured Entities and Other Stakeholders
Chapter 3	Added new Stakeholder Engagement graphic
Chapter 3	Added Blueprint definitions for different measure types
Chapter 5	Added section on Special Considerations for Medicaid-focused Measures
Chapter 6	Incorporated most of the measure evaluation content
Chapter 9	Added several tools and resources

APPENDIX B: REFERENCES

- Adams, J. L. (2009). *The reliability of provider profiling. A tutorial*. <https://doi.org/10.7249/TR653>
- Agency for Healthcare Research and Quality. (n.d.). *National healthcare quality & disparities reports*. Retrieved June 25, 2020, from <https://www.ahrq.gov/research/findings/nhqrdr/index.html>
- American Psychological Association. (2019). *Publication manual of the American Psychological Association*. [VitalSource Bookshelf]. Retrieved from <https://bookshelf.vitalsource.com/#/books/9781433832185/>
- American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 13001, 123 Stat. 227 (2009). <https://www.govinfo.gov/content/pkg/PLAW-111publ5/pdf/PLAW-111publ5.pdf>
- Best, M., & Neuhauser, D. (2004). Avedis Donabedian: Father of quality assurance and poet. *BMJ Quality & Safety*, 13, 472-473. <http://dx.doi.org/10.1136/qshc.2004.012591>
- Centers for Disease Control and Prevention. (n.d.). *International classification of diseases, tenth revision, clinical modification (ICD-10-CM)*. Retrieved June 30, 2020, from <https://www.cdc.gov/nchs/icd/icd10cm.htm>
- Centers for Medicare & Medicaid Services. (2018b). *The 2018 national impact assessment of the Centers for Medicare & Medicaid Services (CMS) quality measures report*. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Downloads/2018-Impact-Assessment-Report.pdf>
- Centers for Medicare & Medicaid Services. (2020, March). *2020 measures under consideration list. Programs-specific measure needs and priorities*. Retrieved June 30, 2020, from <https://www.cms.gov/files/document/cms-measurement-priorities-and-needs.pdf>
- Centers for Medicare & Medicaid Services. (2020, May). *Clinical quality language (CQL) style guide version 4.0*. <https://ecqi.healthit.gov/sites/default/files/CQL-Style-Guide-v4.pdf>
- Centers for Medicare & Medicaid Services. (2020, May). *Electronic clinical quality measure logic and implementation guidance, version 4.0*. <https://ecqi.healthit.gov/sites/default/files/eCQM-Logic-and-Guidance-v4.pdf>
- Centers for Medicare & Medicaid Services. (2020, May). *Guide for reading electronic clinical quality measures (eCQMs) version 6.0*. <https://ecqi.healthit.gov/sites/default/files/Guide-for-Reading-eCQMs-v6.pdf>
- Centers for Medicare & Medicaid Services. (n.d.). *Antidepressant medication management (AMM-AD)*. CMS Measures Inventory Tool. Retrieved from https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=2503
- Centers for Medicare & Medicaid Services. (n.d.). *Chlamydia screening for women*. CMS Measures Inventory Tool. Retrieved June 29, 2020, from https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=2513

- Centers for Medicare & Medicaid Services. (n.d.). *Chronic conditions overview*. Retrieved June 25, 2020, from <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/>
- Centers for Medicare & Medicaid Services. (n.d.). *Chronic obstructive pulmonary disease (COPD): Spirometry evaluation*. CMS Measures Inventory Tool. Retrieved June 29, 2020, from https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=326
- Centers for Medicare & Medicaid Services. (n.d.). *CMIT environmental scan*. CMS Measures Inventory Tool. Retrieved June 29, 2020, from https://cmit.cms.gov/CMIT_public/EnvironmentalScanAbout
- Centers for Medicare & Medicaid Services. (n.d.). *CMS measures inventory tool*. Retrieved June 25, 2020, from https://cmit.cms.gov/CMIT_public/ListMeasures
- Centers for Medicare & Medicaid Services. (n.d.). *CMS office of minority health*. Retrieved June 25, 2020, from <https://www.cms.gov/About-CMS/Agency-Information/OMH/OMH-Mapping-Medicare-Disparities.html>
- Centers for Medicare & Medicaid Services. (n.d.). *CMS quality payment program measure development*. Retrieved June 25, 2020, from <https://www.cms.gov/Medicare/Quality-Payment-Program/Measure-Development/Measure-development.html>
- Centers for Medicare & Medicaid Services. (n.d.). *Data element library*. Retrieved June 30, 2020, from <https://del.cms.gov/DELWeb/pubHome>
- Centers for Medicare & Medicaid Services. (n.d.). *Diabetes care kidney disease monitoring*. CMS Measures Inventory Tool. Retrieved June 29, 2020, from https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=4021
- Centers for Medicare & Medicaid Services. (n.d.). *Email updates*. Retrieved August 14, 2020, from <https://public.govdelivery.com/accounts/USCMS/subscriber/new>
- Centers for Medicare & Medicaid Services. (n.d.). *Ensuring safety & quality*. Retrieved June 30, 2020, from <https://www.cms.gov/About-CMS/Story-Page/Ensuring-safety-and-quality-fact-sheet.pdf>
- Centers for Medicare & Medicaid Services. (n.d.). *Hemodialysis vascular access: Long-term catheter rate*. CMS Measures Inventory Tool. Retrieved June 29, 2020, from https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=5642
- Centers for Medicare & Medicaid Services. (n.d.). *How can I use CMIT?* CMS Measures Inventory Tool. Retrieved June 29, 2020, from https://cmit.cms.gov/CMIT_public/UseEnvironmentalScanSupportToolESST
- Centers for Medicare & Medicaid Services. (n.d.). *Influenza vaccination coverage among healthcare personnel*. CMS Measures Inventory Tool. Retrieved June 29, 2020, from https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=854

- Centers for Medicare & Medicaid Services. (n.d.). *Measure authoring tool*. Retrieved June 30, 2020, from <https://www.emasuretool.cms.gov/>
- Centers for Medicare & Medicaid Services. (n.d.). *Measures management system*. Retrieved June 29, 2020, from <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/MMS-Content-Page>
- Centers for Medicare & Medicaid Services. (n.d.). *Meaningful measures framework*. Retrieved July 14, 2020, from <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy>
- Centers for Medicare & Medicaid Services. (n.d.). *Meaningful measures hub*. Retrieved June 25, 2020, from <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/MMF/General-info-Sub-Page>
- Centers for Medicare & Medicaid Services. (n.d.). *National impact assessment of the Centers for Medicare & Medicaid Services (CMS) quality measures reports*. Retrieved June 25, 2020, from <https://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>
- Centers for Medicare & Medicaid Services. (n.d.). *Percentage of long-stay residents whose need for help with daily activities has increased*. CMS Measures Inventory Tool. Retrieved June 29, 2020, from https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=4066
- Centers for Medicare & Medicaid Services. (n.d.). *Person & family engagement strategy*. Retrieved June 29, 2020, from <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/Person-and-Family-Engagement-Strategy-Summary.pdf>
- Centers for Medicare & Medicaid Services. (n.d.). *Pre-rulemaking*. Retrieved June 25, 2020, from <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Pre-Rulemaking.html>
- Centers for Medicare & Medicaid Services. (n.d.). *Public comments*. Retrieved June 29, 2020, from <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Public-Comments.html>
- Centers for Medicare & Medicaid Services. (n.d.). *Quality improvement organizations*. Retrieved June 25, 2020, from <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityImprovementOrgs/index.html?redirect=/QualityImprovementOrgs>.
- Centers for Medicare & Medicaid Services. (n.d.). *Quality measures requirements*. Retrieved July 1, 2020, from <https://qpp.cms.gov/mips/quality-measures>
- Centers for Medicare & Medicaid Services. (n.d.). *Technical expert panels*. Retrieved June 29, 2020, from <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Technical-Expert-Panels>

- Centers for Medicare & Medicaid Services. (n.d.). *Using the environmental scan support tool (ESST)*. CMS Measures Inventory Tool. Retrieved June 30, 2020, from https://cmit.cms.gov/CMIT_public/UseEnvironmentalScanSupportToolESST
- Centers for Medicare & Medicaid Services. (n.d.). *Welcome to get involved*. Retrieved June 29, 2020, from <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Get-Involved-Landing-Page>
- Centers for Medicare & Medicaid Services, & Office of the National Coordinator for Health Information Technology. (n.d.). *Bonnie: A tool for testing electronic clinical quality measures*. Retrieved from <https://bonnie.healthit.gov/>
- Centers for Medicare & Medicaid Services, & Office of the National Coordinator for Health Information Technology. (n.d.). *Cypress*. Retrieved June 30, 2020, from <https://www.healthit.gov/cypress/>
- Conway, P. H., Mostashari, F., & Clancy, C. (2013). The future of quality measurement for improvement and accountability. *JAMA*, 309(21), 2215-2216. <https://doi.org/10.1001/jama.2013.4929>
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-33. <https://doi.org/10.1007/bf02310555>
- Cylus, J., Papanicolas, I., & Smith, P. C. (Eds.). (2016). *Health System Efficiency. How to make measurement matter for policy and management*. World Health Organization. https://www.euro.who.int/data/assets/pdf_file/0004/324283/Health-System-Efficiency-How-make-measurement-matter-policy-management.pdf?ua=1
- Ding, J., Lu, T-S, Cai, J., & Zhou, H. (2017). Recent progresses in outcome-dependent sampling with failure time data. *Lifetime Data Analysis*, 23, 57-82. <https://doi.org/10.1007/s10985-015-9355-7>
- Donabedian, A. (1966/2005). Evaluating the quality of medical care. *The Milbank Quarterly*, 83(4), 691-729. <http://dx.doi.org/10.1136/qshc.2004.012591> (Reprinted from "Evaluating the quality of medical care" 1966, *The Milbank Memorial Fund Quarterly*, 44[3, Pt. 2], pp. 166–203. <https://doi.org/10.2307/3348969>).
- Donabedian, A. (1980). *Explorations in quality assessment and monitoring: Vol. 1. The definition of quality and approaches to its assessment*. Health Administration Press.
- EBSCO Information Services. (n.d.) *CINAHL complete*. Retrieved June 30, 2020, from <https://www.ebscohost.com/nursing/products/cinahl-databases/cinahl-complete>
- Electronic Clinical Quality Improvement Resource Center. (n.d.). *CQL- Clinical quality language*. Retrieved August 14, 2020, from <https://ecqi.healthit.gov/cql>
- Electronic Clinical Quality Improvement Resource Center. (n.d.). *Create new account*. Retrieved June 29, 2020, from <https://ecqi.healthit.gov/user/register>
- Electronic Clinical Quality Improvement Resource Center. (n.d.). *eCQI tools & key resources*. Retrieved June 30, 2020, from https://ecqi.healthit.gov/ecqi-tools-key-resources#quicktabs-tabs_tools3

- Electronic Clinical Quality Improvement Resource Center. (n.d.). *Home*. Retrieved June 29, 2020, from <https://ecqi.healthit.gov/>
- Electronic Clinical Quality Improvement Resource Center. (n.d.). *Measure collaboration (MC) workspace – eCQM concepts*. Retrieved June 29, 2020, from <https://ecqi.healthit.gov/mc-workspace-2/ecqm-concepts>
- Electronic Clinical Quality Improvement Resource Center. (n.d.). *Measure collaboration (MC) workplace – new electronic clinical quality measure (eCQM) clinical workflow*. Retrieved from <https://ecqi.healthit.gov/mc-workspace-2/ecqm-clinical-workflows>
- Electronic Clinical Quality Improvement Resource Center. (n.d.). *Measure collaboration (MC) workplace – electronic clinical quality measure (eCQM) test results*. Retrieved June 30, 2020, from <https://ecqi.healthit.gov/mc-workspace-2/ecqm-test-results>
- Electronic Clinical Quality Improvement Resource Center. (n.d.). *Measure collaboration (MC) workspace – MC workspace*. Retrieved June 30, 2020, from <https://ecqi.healthit.gov/mc-workspace-2>
- Electronic Clinical Quality Improvement Resource Center. (n.d.). *QDM- Quality data model*. Retrieved August 14, 2020, from <https://ecqi.healthit.gov/qdm>
- Electronic Clinical Quality Improvement Resource Center. (n.d.). *QRDA - quality reporting document architecture*. Retrieved June 30, 2020, from <https://ecqi.healthit.gov/grda>
- GitHub. (n.d.). *CQL-to-ELM translator reference implementation*. Retrieved June 30, 2020, from https://github.com/cqframework/clinical_quality_language/blob/master/Src/java/cql-to-elm/OVERVIEW.md
- HL7 International. (n.d.). *Appendix H – Time interval calculation examples*. Retrieved July 14, 2020, from <https://cql.hl7.org/STU4/15-h-timeintervalcalculations.html>
- Iezzoni, L. (Ed.). (2013). *Risk adjustment for measuring health care outcomes* (4th edition). Foundation of the American College of Health Care Executives.
- Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. The National Academies Press. <https://doi.org/10.17226/10027>
- Institute of Medicine. (2015). *Vital signs: Core metrics for health and health care progress*. <https://doi.org/10.17226/19402>
- Kindig, D., & Stoddard, G. (2003). What is population health? *American Journal of Public Health*, 93, 380-383. <https://doi.org/10.2105/ajph.93.3.380>
- LOINC. (n.d.). *Home*. Retrieved June 30, 2020, from <https://loinc.org/>
- Medicaid and CHIP Payment and Access Commission (MACPAC). (n.d.). *Medicaid and CHIP payment and access commission*. Retrieved June 25, 2020, from <https://www.macpac.gov/>

- Medicare Access and CHIP Reauthorization Act of 2015, Pub. L. No. 114-10, 87 Stat. 129 (2015).
<https://www.congress.gov/114/plaws/publ10/PLAW-114publ10.pdf>
- Medicare Payment Advisory Committee (MedPAC). (n.d.). *Reports*. Retrieved June 25, 2020, from
<http://www.medpac.gov/-documents-/reports>
- Messick, S. (1994). *Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning* (ED 380 496).
<https://files.eric.ed.gov/fulltext/ED380496.pdf>
- National Quality Forum. (2011a). *CSAC guidance on quality performance measure construction*. Retrieved June 29, 2020, from
<http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdIdentifier=id&ItemID=68999%20>
- National Quality Forum. (2011b). *Guidance for measure testing and evaluating scientific acceptability of measure properties*. Retrieved June 30, 2020, from
<http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdIdentifier=id&ItemID=70943>
- National Quality Forum. (2016, December). *Attribution: Principles and approaches*. Retrieved July 29, 2020, from
<http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdIdentifier=id&ItemID=84236>
- National Quality Forum. (2017). *Measure submission form*. Retrieved June 29, 2020, from
<http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdIdentifier=id&ItemID=86103>
- National Quality Forum. (2019). *Measure evaluation criteria and guidance for evaluating measures for endorsement*. Retrieved June 29, 2020 from
<http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdIdentifier=id&ItemID=88439>
- National Quality Forum. (n.d.). *Consensus development process*. Retrieved June 30, 2020, from
http://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process.aspx
- National Quality Forum. (n.d.). *Measure applications partnership*. Retrieved June 25, 2020, from
<http://www.qualityforum.org/map/>
- National Quality Forum. (n.d.). *NQF Feasibility scorecard for electronic clinical quality measures*. Retrieved June 30, 2020, from
<https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdIdentifier=id&ItemID=89036>
- National Quality Forum. (n.d.). *NQF measure incubator®*. Retrieved June 30, 2020, from
https://www.qualityforum.org/NQF_Measure_Incubator.aspx
- National Quality Forum. (n.d.). *Phrasebook: a plain language guide to NQF jargon*. Retrieved June 29, 2020, from <http://public.qualityforum.org/NQFDocuments/Phrasebook.pdf>
- National Quality Forum. (n.d.). *Quality positioning system*. Retrieved June 29, 2020, from
<http://www.qualityforum.org/QPS/>

- National Quality Forum. (n.d.). *Submitting eCQMs for NQF endorsement*. Retrieved June 30, 2020, from http://www.qualityforum.org/Electronic_Quality_Measures.aspx
- National Quality Forum. (n.d.). *Submitting standards*. Retrieved June 29, 2020 from http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx
- National Uniform Claim Committee. (2012, February). *Health insurance claim form*. <https://www.cms.gov/Medicare/CMS-Forms/CMS-Forms/Downloads/CMS1500.pdf>
- Office of the National Coordinator for Health Information Technology. (n.d.). *Interoperability standards advisory*. Retrieved June 29, 2020, from <https://www.healthit.gov/isa/>
- Patient Protection and Affordable Care Act of 2010, Pub. L. No. 111-148, 124 Stat. 119 (2010). <https://www.govinfo.gov/content/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf>
- Patient-Centered Outcomes Research Institute. (2016). *Engagement rubric for applicants*. <https://www.pcori.org/sites/default/files/Engagement-Rubric.pdf> (Original work published 2014)
- Rajkumar, R., Conway, P.H., & Tavenner, M. (2014). *CMS—engaging multiple payers in payment reform*. *JAMA*, 311(19), 1967–1968. <https://doi.org/10.1001/jama.2014.3703>
- SNOMED International. (n.d.). *Home*. Retrieved June 30, 2020, from <https://www.snomed.org>
- Spertus, J.A., Bonow, R.O., Chan, P., Diamond, G.A., Drozda, J.P., Kaul, S., ... Rumsfeld, J.S. (2010). ACCF/AHA new insights into the methodology of performance measurement. *Circulation*, 122, 2091-2106. <https://doi.org/10.1161/CIR.0b013e3181f7d78c>
- U.S. General Services Administration. (n.d.). *Federal advisory committee act (FACA) management overview*. Retrieved June 29, 2020, from www.gsa.gov/faca
- U.S. General Services Administration. (n.d.). *Section508.gov*. Retrieved June 30, 2020, from <https://www.section508.gov/>
- U.S. National Library of Medicine. (n.d.). *RxNorm overview*. Retrieved June 30, 2020, from <https://www.nlm.nih.gov/research/umls/rxnorm/overview.html>
- U.S. National Library of Medicine. (n.d.). *Value set authority center*. Retrieved June 30, 2020, from <https://vsac.nlm.nih.gov/>
- U.S. National Library of Medicine, National Center for Biotechnology Information. (n.d.). *PubMed.gov*. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed>
- U.S. National Library of Medicine, National Center for Biotechnology Information. (n.d.). *PubMed central (PMC)*. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/>