PCPI Outcomes Ad Hoc Committee: A Toolkit for Outcome Measure Development

Approved by the PCPI, October 2015

Introduction

Background
The quality measurement enterprise is evolving with an increased demand for measures that are broad-based, meaningful, and patient-centered. There is a need for measures that can assess performance in all six priority domains of the National Quality Strategy: clinical care, patient experience and engagement, population and community health, safety, care coordination, and cost and efficiency. At the core of this advancement is a call for measures that assess the outcomes of health care, for they represent the fundamental purpose for seeking health care from the patient’s perspective and providing health care from a provider’s perspective. Outcome measures are considered by many to more closely reflect the quality of care and to potentially have the greatest impact in improving care. Despite their appeal, outcome measures remain the most underutilized and underdeveloped type of performance measure in the US. This can be evidenced by the number and types of measures in use in national governmental and nongovernmental programs as well as in the measures endorsed by the National Quality Forum (NQF).

The AMA-convened PCPI is a driving force in evidence-based clinical performance measurement, having been the sole developer or a collaborating party for a portfolio that includes more than 300 measures spanning 46 clinical conditions or topic areas. The PCPI measure portfolio is also reflective of a pattern seen across the quality measurement community, with a vast majority of the more than 300 measures being classified as process measures. The PCPI remains committed to advancing performance measurement to support improvement in health outcomes and, as a result, is seeking to expand the portfolio of available clinical and patient-reported outcome measures developed by the PCPI, its members, and other stakeholders in the measurement community. Toward that end, in October 2013, the PCPI convened an Outcomes Ad Hoc Committee comprised of experts with methodological expertise in the development of outcomes measures and representing diverse areas of medicine (see Appendix A for a list of committee members). The Outcomes Ad Hoc Committee was charged with building upon the framework developed in 2011 by the Measures Development, Methodology, and Oversight (MDMO) Advisory Committee of the PCPI, titled “Recommendations to PCPI Work Groups on Outcome Measures.”


© 2015 American Medical Association. All Rights Reserved.
Objectives, Intended Use, and Intended Audience

The Outcomes Ad Hoc Committee was charged with creating a toolkit to assist users with the development of outcomes measures relevant to their specialty or topic area. The toolkit includes a discussion of several key issues that are particularly relevant to outcomes measurement and recommended ways to address them. These topics include guiding conceptual framework – selection of outcomes to measure, level of attribution, feasibility and availability of data, risk adjustment, and testing. The toolkit is intended to provide guidance for the most significant obstacles to outcome measure development and for those areas in which outcome measurement is distinct from process measurement. The toolkit is designed to have a wide reach and broad applicability. The audience for the toolkit is measure developers such as medical specialty societies, health plans, or health systems that are relatively new to the development of outcome measures and who could therefore benefit from the recommendations and examples provided herein. Additionally, in putting forward this toolkit, we wanted to build upon the work of others as much as possible and refer to existing resources that could serve as excellent starting points for measure developers (see Appendix B).

A Few Foundational Principles for the Toolkit

There are various ways to define health outcomes. Consistent with definitions from the NQF, Agency for Healthcare Research and Quality’s (AHRQ) National Quality Measures clearinghouse (NQMC), and Avedis Donabedian, the committee members agreed that a health outcome can be defined as “The health state of a patient (or change in health status) resulting from healthcare—desirable or adverse.” An outcome may address a vast range of health states including mortality, physiologic measures, patient-reported health states, functional status, and patient satisfaction. Outcome measures can then be described as: assessing the results of healthcare that are experienced by patients.

However, we acknowledge that certain diagnostic specialties, such as pathology and radiology, may have difficulty conforming to the above definition of outcomes. Diagnosis is an important intermediate step in achieving positive patient outcomes and a correct diagnosis is a necessary part of successful treatment. Since pathologists and other diagnosticians play key roles in coordinating care with other medical specialties and have positive impacts on patient care, we hope to better address diagnostic outcomes in the future. The ultimate goal of measurement is to obtain meaningful data to improve care. Strategies to achieve this ultimate goal may include using measures in quality improvement, benchmarking, accountability, and/or payment programs. Some measures may be used for accountability and public reporting while others may solely be intended for quality improvement purposes. As described in several recent PCPI approved measure sets, measures that can be used for accountability provide information that can assist stakeholders in making choices about

---


providers. Accountability measures, such as those used in public reporting to facilitate comparisons across clinicians or facilities or possibly pay for performance programs, must have certain attributes that would permit such a designation. In particular, the PCPI methodology requires that these measures be developed through a rigorous measure development process including a public comment and peer review period. These measures must also be based on guideline recommendations that were prioritized due to several important variables [eg, for process measures, the clinical importance of the intervention and link to desired outcomes, the strength of evidence and strength of recommendation, gaps in care, validity, and feasibility]. Accountability measures are also intended to promote quality improvement. On the other hand, measures only intended for quality improvement use provide information that should be of value to stakeholders responsible for improving the quality of care and should not be used for accountability purposes. Measures designated only for quality improvement lack at least one of the aforementioned desired attributes and require further validation and testing. The majority of the recommendations contained in this toolkit are particularly applicable to measures that are intended to be used for accountability purposes.

Developers need to identify health outcomes and corresponding outcome measures that reflect the goals of care from the patient’s perspective. In most circumstances, the patient’s perspective and the provider’s perspective become aligned through a process of discussion and shared decision-making. This alignment of goals may not always be achievable, but it should represent a guiding principle for measure development and implementation efforts. The relative utility of any outcome measure is based upon its importance for the patient and the patient’s caregivers, given the trajectory of the patient’s health status and disease.

---

9 Ibid.
Section 1: Guiding Conceptual Framework—Selecting Outcomes to Measure

Introduction
Given the breadth of outcomes resulting from healthcare including mortality, physiologic processes, patient-reported health states, functional status, and patient satisfaction, outcome measures can be similarly broad and far-reaching to address any number of health states.

In selecting which outcomes to measure, measure developers should begin their work by considering measures across a wide spectrum of outcomes. Several models exist to guide in the identification of outcomes that can subsequently serve as the subject of performance measures designed to assess healthcare quality.

Frameworks for Identifying Outcomes
Avedis Donabedian, a thought leader in the evaluation of healthcare quality and developer of the structure/process/outcome paradigm for quality measures, first described two elements in the performance of practitioners—technical and interpersonal. These can simply be described as the science and art of medical care, respectively. Although he acknowledged the challenge in attempting to consider the management of the interpersonal process when assessing the quality of care, he emphasized the importance of it as “the vehicle by which technical care is implemented and on which its success depends.” As a result, Donabedian suggested that these distinct yet interdependent components must both be addressed in assessments of the quality of care. Outcomes can therefore focus on the degree of physical or physiological function, as a result of the technical process of care, or more broadly on the degree of psychological and social function, as a result of the technical process of care.\(^{10}\)

More recently, management expert Michael Porter\(^{11}\) described an outcome measures hierarchy to define an appropriate set of outcome dimensions, specific metrics, and associated risk factors. This hierarchy includes three tiers:

1) Patient health status achieved or retained, as may be appropriate with some degenerative conditions, including the concepts of survival and the degree of health or recovery achieved or retained

2) Process of recovery, including the concepts of time required to achieve recovery and disutility of the care process (e.g., missed diagnosis, failed treatment, anxiety)

3) Sustainability of health including the concepts of recurrences of the original disease or associated longer-term complications and new health problems created as a consequence of the treatment itself, or care-induced illnesses

Porter suggests that the hierarchy can guide the identification of the full set of outcomes for any medical condition.

Another useful construct to facilitate identification of outcomes and subsequent outcome measures is “the five Ds”—death, disease, disability, discomfort, and dissatisfaction—first

---

\(^{10}\) Op. cit., Donabedian.

described by Elinson\textsuperscript{12} in the late 80s. This concept was reconceived in the positive by Lohr\textsuperscript{13}, who described survival, states of physiologic, physical, and emotional health, and satisfaction. Building on this concept, the committee has similarly recommended an updated lexicon with a modern focus for increased relevance in today’s environment. To this categorization, one may add a new category, as shown below, focusing on patients’ acquisition of the knowledge and skills necessary to manage their health better, through improved self-efficacy.

<table>
<thead>
<tr>
<th>Outcome Measure Categories (note: no one category is more important than any other)</th>
<th>Original</th>
<th>Updated Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>Survival</td>
<td></td>
</tr>
<tr>
<td>Disease</td>
<td>Disease management</td>
<td></td>
</tr>
<tr>
<td>Disability</td>
<td>Functioning</td>
<td></td>
</tr>
<tr>
<td>Discomfort</td>
<td>Quality of life, symptom management</td>
<td></td>
</tr>
<tr>
<td>Dissatisfaction</td>
<td>Patient experience and engagement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Patient knowledge and skills</td>
<td></td>
</tr>
</tbody>
</table>

Examples of outcome measures within some of these categories have previously been described by the PCPI in the MDMO guidance document “Recommendations to PCPI Work Groups on Outcome Measures”.\textsuperscript{14} Some of these outcomes may represent improved or preserved health states, functional capacities, or symptoms, but others may represent intermediate steps toward those longer-term outcomes. Such intermediate steps may include physiologic metrics such as blood pressure, laboratory values such as glycosylated hemoglobin, stages or intentionality of behavior change (e.g., “contemplation” versus “precontemplation” in Prochaska’s Transtheoretical Model\textsuperscript{15}), and the knowledge, beliefs, or skills necessary to achieve better long-term outcomes.

\textsuperscript{14} Op cit., American Medical Association-convened Physician Consortium for Performance Improvement. Measures Development, Methodology, and Oversight Advisory Committee.
<table>
<thead>
<tr>
<th>Types of Outcome Measures</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Outcomes</td>
<td></td>
</tr>
<tr>
<td>• Mortality</td>
<td>Infant death rates</td>
</tr>
<tr>
<td>• Morbidity</td>
<td>Reduced rates of preeclampsia; reduced rates of gestational diabetes</td>
</tr>
<tr>
<td>• Intermediate clinical outcomes</td>
<td>Blood pressure levels, blood glucose levels, patient/family goal setting</td>
</tr>
<tr>
<td>• Symptoms</td>
<td>Reduction in chest pain; reduced depression episodes</td>
</tr>
<tr>
<td>• Clinical events</td>
<td>Stroke, AMI; cancer; adverse events during childbirth</td>
</tr>
<tr>
<td>Patient-Reported Outcomes</td>
<td></td>
</tr>
<tr>
<td>• Health status</td>
<td>Quality of life-health status as perceived by the individual; Functional measures- SF-36, PROMIS</td>
</tr>
<tr>
<td></td>
<td>Pre- and post-treatment physical function</td>
</tr>
<tr>
<td></td>
<td>Pre- and post-treatment mental health (e.g., depression severity)</td>
</tr>
<tr>
<td></td>
<td>Pre- and post-treatment social/role function (e.g., ADLs, return to work or school)</td>
</tr>
<tr>
<td></td>
<td>Other measures of health status such as pain, vitality, perceived well-being, health risk status, etc.</td>
</tr>
<tr>
<td>• Patient experiences with care (patient satisfaction; patient engagement/ patient preference; patient education)</td>
<td>Consumer assessment (CAHPS); quality of life, adherence to treatment regimen; provider retention; shared decision making; engagement of family and friends in process of care; patient knowledge and understanding</td>
</tr>
<tr>
<td>Economic Outcomes</td>
<td></td>
</tr>
<tr>
<td>• Resource Use/Costs</td>
<td>Healthcare service utilization; cost per episode of care; cost per quality-adjusted life year</td>
</tr>
</tbody>
</table>
A useful illustration of this broad-based approach to the identification of outcomes can be seen in the work of the International Consortium for Health Outcomes Measurement (ICHOM) in conceiving their standard sets for various conditions. The following represents the outcomes that ICHOM working group members agreed would matter most to patients with low back pain and which they recommend all providers track.\footnote{International Consortium for Health Outcomes Measurement. ICHOM Standard Set for Low Back Pain. 2013. Available at: \url{http://www.ichom.org/project/low-back-pain/}. Accessed 12/5/2014.}

Patient-reported outcomes (PROs) are health outcomes that are directly reported by a patient or caregiver without modification or interpretation by the healthcare provider or anyone else. These outcomes represent an emerging area of interest for performance measurement, as increasing emphasis has been placed on the importance of patient-centered care and they merit special
discussion. In their report on the measurement of PROs, the National Quality Forum distinguishes between patient-reported outcome measures (PROMs), which are the tools used to assess patient-reported health status, and patient-reported outcome-based performance measure (PRO-PM), which is a performance measure based on aggregated PRO data. For example, with regards to measuring pain as a PRO, an example of a PROM would be a 10-point numeric pain scale while an example of a PRO-PM might be the percentage of patients with a pain score ≤4 following surgery. The PROM is a survey, questionnaire, or single-item measure that is completed by the patient or caregiver. The response provided by the patient to the PROM is then compiled into a PRO-PM to allow comparison between patients or providers, or to allow for monitoring for changes in response from the same patient over time. This type of measure is often used for capturing patient-focused areas of care such as symptom burden, health-related quality of life, health behaviors, and patient satisfaction. Because of the person-centered nature of PRO-PMs, psychometric soundness of the PROM, or tool used to assess the PRO, is an important consideration for this type of measure in addition to reliability and validity of the PRO-PM. For example, language used in PROMs needs to be clear, concise, unambiguous, and ideally accessible to patients of all literacy levels, cognitive abilities, and cultural and linguistic backgrounds to ensure comparability of results. Some patients will require assistance or even proxies (eg, family members, caretakers) to answer questions. Only a few PROMs have been evaluated for use with proxies. Additionally, the burden of implementation of the PROM needs to be assessed in the development process. In particular, PROMs require varying degrees of time and effort from the clinician and staff administering the PROM as well as the patient or proxy providing the information.

The American College of Surgeons (ACS) developed a measure that serves as a useful example of a PRO-PM. The ACS measure, Patient Experience with Surgical Care Based on the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Surgical Care Survey, was designed to assess aspects of surgical quality that are important to consumers and for which consumers are the best source of information. The PRO-PM incorporates use of the Surgical Care survey created by the ACS in partnership with other surgical and anesthesia organizations and the Agency for Healthcare Research and Quality’s (AHRQ) Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Consortium. The survey assesses surgical patients’ experiences before, during, and after surgical procedures as a means of identifying opportunities to improve quality of care, surgical outcomes, and patient experiences of care, as well as for public reporting. The CAHPS Surgical Care survey is a standardized patient survey that produces clear and usable comparative information for consumers and health care providers. The CAHPS Surgical Care Survey has been available to the public since 2010 and the ACS measures is the only NQF-endorsed measure designed to assess surgical quality from the patient’s perspective.18

The National Quality Forum’s 2012 report on the incorporation of PROs in performance measurement is a good resource for further guidance on the development and use of PRO-PMs.


© 2015 American Medical Association. All Rights Reserved.
Key Principles for Developing Outcome Measures

After identifying the broad set of outcomes for the medical condition or topic area under study, measure developers need to determine which outcomes can and should be measured. Several key principles described by others and supported by this committee should be considered.

In his seminal work, Donabedian described several key considerations for selecting outcomes as indicators of quality:

- The outcome selected should be relevant to the objective of care and achievable by the application of scientific medicine.
- The outcome must be attributable to medical care and not to factors other than medical care that may have intervened.
- The duration of the outcome should be taken into account, in some cases requiring follow-up over long periods of time which may affect the utility of the appraisal and challenges comparability.
- The type and magnitude of the outcome needs to be considered and is subject to the values, judgments, and preferences of the patient. For example, the prolongation of life may have to be weighed against increasing disability and loss of function.
- The outcome cannot stand alone. The means used to achieve the outcome also have to be considered in order to shed light on the deficiencies or strengths to which the outcome might be attributed.\(^{19}\)

Porter also suggested several guiding principles to define the relevant outcomes to measures:

- Outcomes should be patient-centered and involve the health circumstances most relevant to patients
- Outcomes should address short term and long term patient health to allow for evaluation of the ultimate results of care
- Outcomes should cover the full range of services that determine the patient’s results, outcome measurement needs to consider risk factors and adjust as appropriate.\(^{20}\)

The aforementioned work of the MDMO Advisory Committee described several guiding principles, the following of which remain valid today:

1. While assessing full measures of outcome, such as mortality, is desirable, when assessment is impractical, PCPI should begin with implementation and testing of intermediate outcomes measures
2. Initial clinical outcome measures should be developed around episodes of care (eg, Heart Failure). More complex outcome measures that cut across multiple measurement areas will be considered later.
3. Patient-reported outcome measures should be based on a validated tool or instrument and should be relevant to the objective of care (the ultimate aim of the practitioner) and measurable through testing.

4. Outcome measures must be clinically relevant, meaningful, measureable, and actionable by the clinician.
5. Given the complexity inherent in risk adjustment, data collection should be made available through the proper denominator construction, where applicable.
6. Outcome measures should be actionable and focus on linking the outcome to the process component of the measure.
7. Information on the relevant outcome must be available; obtaining this information may require follow-up over long periods of time.
8. Outcome measures should be empirically tested for validity and reliability (see Section 5).
9. The PCPI acknowledges that early outcomes measures may be imperfect and that they will evolve over time; testing and evaluation will provide greater guidance on the feasibility and reliability of the measure.\(^{21}\)

The Outcomes Ad Hoc Committee elected to omit one guiding principle that was included in the MDMO framework, in light of the evolution of the measurement field since the framework was first developed and increasing emphasis on and understanding of risk adjustment and its value in outcomes measurement. It may be appropriate for a developer’s first foray in outcome measure development to avoid complex risk adjustment methodologies, as recommended by the MDMO, and opt for a simpler approach that can advance incrementally. However, this committee recognizes the important role of risk adjustment in producing outcome measures that can accurately reflect the quality of care provided and didn’t want to appear to discourage risk adjustment even for novel developers. Please refer to section 4 of the document for more on risk adjustment.

Additionally, the Outcomes Ad Hoc Committee has incorporated a truncated version of the sixth principle described above which had previously included a statement suggesting that an outcome measure cannot stand on its own without an associated process measure. Process measures play a very important role in improving care as they provide insight into the process changes that need to occur to support improvement in outcomes. However, it may not be possible to link all outcomes and outcome measures with a specific process of care that could then form the basis of a process measure.

Another principle to consider that has yet to be expanded upon by the aforementioned authors comes from the Hippocratic Oath – “primum non nocere” or “first, do no harm.” The principle of *primum non nocere* applies to the use of outcome measures; potential measures that are likely to cause harm in practice should, of course, be avoided. When harm is theoretically possible, such as by creating incentives to avoid high-risk patients who would actually benefit from physician intervention, monitoring programs should be put into place.

Finally, it’s important for any measure developer to understand that the development of outcome measures is iterative in nature. Outcome measures, even good ones, may need to be modified as medical care changes and patient populations change. For example, an outcome as seemingly

important and permanent as mortality following surgery is something for which the assessment may change depending upon advances in surgical care and the risk of the population being referred for surgery. Therefore, commitment to outcomes measure development is a commitment to continuous ongoing assessment of the parameters by which the outcomes are evaluated and the manner in which they are used.
Section 2: Level of Attribution

Introduction
When developing a performance measure, especially an outcome measure, it is important to consider whose performance is actually being measured. Is the measure intended to capture the performance of an individual provider? A care team? A facility? The level of performance at which measurement is aimed helps shape the structure of the measure itself, as well as its potential use. Measures that are developed to capture the performance of individual providers are typically used for accountability and payment purposes, while measures intended to capture performance at a higher level such as a healthcare facility or a health plan can be used for internal quality improvement, benchmarking, and accreditation purposes, as well as for accountability and payment. When determining the appropriate level of attribution for your measure, there are several important factors to consider.

Actionability and Variability
When determining the appropriate level of attribution for an outcome measure, it is important to consider whether the outcome is actionable by individual providers of care. In other words, could changes in an individual provider’s behavior impact the outcome of interest or would action be required at a higher level, such as a facility-wide policy change? In order to attribute an outcome to an individual provider, that provider needs to have a certain level of control over that outcome. For this reason, outcome measures that are specified at the individual-provider level are generally associated with procedures or processes that are clearly under the control of an individual provider, such as foreign bodies left in a patient after surgery. Outcomes that are impacted by far-reaching factors such as staffing or resource allocation are better attributed to higher levels (e.g., provider team, facility, health system). Some outcomes are so dependent on patient factors (health literacy, compliance, socioeconomic status, family support, etc.) that provider attribution is not ultimately possible.

When considering the degree of control an individual provider may have over the selected outcome, developers should also consider the potential role of external factors such as patient adherence that may fall outside of the scope of care for a provider. Patient adherence or non-adherence to medical advice can be influenced by factors both related (e.g., literacy level of educational materials, degree of patient follow-up) and unrelated (e.g., family support, financial concerns) to the quality of care provided. Careful consideration needs to be given to how external environmental and socioeconomic factors could impact the selected outcome and to the expected relationship between those factors and the provider.

Once a measure developer has ascertained the level of attribution for the selected outcome, it is then important to determine whether there is variability in that outcome. Variability in outcome demonstrates inconsistency in the provision of care and room for improvement. Variability in care can occur on multiple levels. Individual provider preferences can result in differences in the way individual clinicians provide care. Alternatively, policy, staffing, and resource issues can result in inconsistencies in the way care is provided between facilities or even geographic areas. Identifying the level or levels at which inconsistencies in care arise can help determine the most appropriate level of attribution for an outcome measure. For example, differences in surgical techniques that result in differential outcomes represent variability and an opportunity for improvement at an individual provider level. Differences in availability of equipment that impact
care choices and patient outcomes represent an opportunity for improvement at a facility, system, or geographic level.

**Sample Size and Event Rate**

In order for an outcome measure to be stable and representative, an adequate sample size and event rate is necessary. Sample size refers to the number of patients eligible to undergo the chosen outcome, while event rate refers to the frequency of the chosen outcome in the target population. For example, in an outcome measure looking at the postoperative infection rate for patients undergoing open heart surgery, the sample size refers to the number of patients who undergo open heart surgery, while the event rate refers to the frequency of infection among those people. In order to have a statistically reliable measure, both numbers need to exceed a minimum threshold at the specified level over the measurement period. To achieve an adequate sample size and event rate for an individual provider-level measure, it is necessary to measure a relatively common outcome that affects a large number of people. This ensures that the provider sees enough eligible patients throughout the course of the measurement period (usually a year) to achieve an accurate measurement of the chosen outcome. Outcomes, such as mortality, that are rare or only affect a small number of people are best measured at a higher level (eg, facility, health plan, health system) in order to achieve a large enough sample to ensure statistically reliable measurement. Other outcomes that are not rare but are complex, such as functional status or patient engagement in the process of care, might best be measured at a team or multiple provider level. In addition to changing the level of measurement, data collection over a longer period of time, adjustment of measure inclusion criteria, or a composite measure structure (ie, combining two or more measures into a single measure with a single score) can also be used to achieve an appropriate sample size for the outcome of interest. Each approach has strengths and weaknesses. For example, using longer data collection period hinders your ability to use the data to provide real-time feedback and drive timely improvements. Similarly, expanding the inclusion criteria for an outcome measure could increase the risk of false positive results. When choosing an approach to ensure an adequate sample size, it is important to consider these trade-offs to determine which approach makes the most sense for your outcome measure of choice.

**Potential Unintended Consequences**

Because outcomes measurement is intended to guide clinical behavior, careful consideration must be given to all of the potential ways a measure may influence behavior. Unintended consequences are outcomes that are not the ones intended by a purposeful action (in this case, performance measurement). In the area of outcomes measurement, unintended consequences often occur as a tradeoff between different potential outcomes. For example, a hospital may discharge patients early in order to reduce the average patient length of stay. However, shorter lengths of stay in the hospital could result in increased patient readmissions to the hospital if they are being discharged before they are fully ready. Unintended consequences are not always harmful and, in some cases, they may be unavoidable. However, it is important to anticipate and prevent these potential consequences as early as possible.

One particular type of unintended consequence is commonly described as “gaming”. Gaming occurs when the measured entity (eg, provider, facility) attempts to influence the measured outcome by deliberately altering variables other than clinical quality. For example, in order to achieve a high outcome measure score, a provider may choose to avoid high-risk patients. Anticipated gaming and other unintended consequences can be handled through a variety of
methods including proper risk adjustment methodology. These effects can also be minimized by attributing the outcome measure to the appropriate level. Outcome measures that may be subject to gaming on an individual provider level may work more appropriately as facility or health plan measures.

Section 3: Feasibility and Availability of Data

Introduction
The ability to measure meaningful outcomes largely depends on the availability of quality data. The current information technology infrastructure in our health care system presents several challenges to the effective measurement of outcomes. Often, outcome data with the most meaning to providers and patients are also the most complex to collect, requiring a combination of patient reports and provider assessments. It is important to consider whether the necessary data elements are available as well as where and how they would be captured if they are not already being collected. Another important consideration is the time and effort required to capture the necessary data elements. Some data elements are routinely captured within the electronic health record (EHR) and can be easily queried, while other data elements require manual chart abstraction, which can be burdensome and resource-intensive. The administrative burden and associated costs present a significant challenge for outcome measurement. Early consideration of these questions in the measure development process will prevent confusion and mitigate feasibility challenges later. There are several data sources available for use in outcome measurement; each with its own strengths and weaknesses.

Administrative Claims Data
Administrative claims data have been widely used in performance measurement due to their ready availability. Because claims data are standardized and universally-available, they are often used as a data source in the development of performance measures. However, the structure of administrative claims data was developed to facilitate billing and payment, not performance measurement. As a result, administrative claims data often lack the level of clinical detail necessary to construct a meaningful outcome measure. Additionally, because the primary purpose of administrative claims data is billing, there is usually a time-lag associated with the coding and collection of the data which hinders real-time performance measurement. Due to these significant limitations, investment in infrastructure to capture clinical data is recommended whenever possible. In cases where this is not a feasible option, utilizing administrative claims data may be a good first step for outcomes measurement.

Electronic Health Records
In order to overcome the challenges associated with the use of administrative claims data, many current initiatives emphasize the use of clinical data contained within EHRs. EHR data provide additional flexibility and clinical detail compared to that provided by administrative claims data. Additionally, EHR data are often collected as part of the clinical workflow at the point of care. This allows for more real-time application of the data. While EHR data allow for more precision and flexibility when capturing outcomes, they are not currently universally available in a useable form. Many EHR data elements are captured in a free text format, or in a non-standardized structured format. Data in these formats can be challenging to extract from the EHR for
performance measurement purposes. However, there are some tools under development, such as natural language processing software, that could potentially enable this data to be used in performance measures in the future. Workflow and infrastructure changes are often required to ensure that data are captured in the EHR in a usable way, which can be time and resource intensive and burdensome for clinicians. While EHR technology is constantly improving and expanding, infrastructure, resource limitations, and lack of interoperability still represent a significant barrier to widespread adoption and use. To better facilitate its use in performance measurement, data in the EHR need to be stored in a standardized and structured format. This is necessary to ensure data elements and performance measures are captured reliably across providers, sites, and EHR systems.

**Clinical Data Registries**

A clinical data registry captures information about the health status of patients and the health care they receive over varying periods of time. Similar to EHR data, data contained within clinical registries is also based on information generated within the course of clinical care. Clinical data registries typically focus on patients who share a common reason for needing health care. There are many types of clinical data registries including those that focus on a disease or condition (e.g., cystic fibrosis), a procedure (e.g., coronary artery bypass grafting surgery) or to track the performance of a device (e.g. artificial joint).

Patients often receive care from different organizations over time. Each time a patient participating in a registry sees their health care professional or is admitted to a hospital, detailed data are recorded about their health status and the care received. Health care professionals then send encrypted data about the patients to the clinical data registry through a secure web portal or from their EHR. As data enters the clinical data registry, quality checks are performed to ensure the correctness and completeness of the data. Data elements are highly structured and standardized within each registry, as compared with EHR data which has variances across health care professionals and settings. Additionally, some organizations with clinical registries contribute to the development of national clinical quality measures in their respective specialties, linked to their registry data. All of these characteristics of clinical data registries make them an increasingly important source of data for quality measurement.

One significant difference between EHR and registry data is the focus of data collection. EHR data enable a focus on the care of the patient as a whole, whereas clinical registries often have a condition-specific or procedure-specific focus. As a result, EHR data may be able to provide data for outcomes related to the entirety of care for the patient, while registry data may be more appropriate for condition-specific outcomes. Because registries can be costly and resource intensive to create and maintain, they are not available for all conditions. However, there are several national initiatives to promote the use of clinical data registries which could lead to their playing a more important role in outcomes measurement in the future.
Section 4: Risk Adjustment

Risk Adjustment Defined
The National Quality Forum defines risk adjustment as “the statistical methods to control or account for patient-related factors when computing performance measure scores.”22 Theoretically, risk adjustment would allow an outcome measure and its subsequent results to be isolated to factors that relate to health care and the entity being measured while controlling for significant confounding factors that might affect the outcome.

The Need for Risk Adjustment
In order for an outcome measure to produce valid results, the measure needs to account for those factors that may impact the outcome being measured and those that are outside the control of the entity being measured. Risk factors such as patient demographic characteristics, comorbid conditions, and patient preferences which are present at the time of health care delivery may contribute to different outcomes regardless of the quality of care provided. For example, the mortality of patients with stroke is associated with the initial severity of the stroke, age, and comorbidities - none of which can be modified.23 Risk adjustment enables outcome measure results to more accurately reflect the quality of care provided and facilitates making comparisons across physician or facilities.

Although rare, there may be some outcome measures for which risk adjustment is not needed. For example, measures that are focused on so-called “never events” that refer to adverse events that are unambiguous, serious, and usually preventable such as wrong site surgery may not need risk adjustment. These outcomes are not likely to be impacted by patient risk factors and are more singularly under the purview of the entity being measured. Another category of outcome measures that might not require risk adjustment is measures that look at the change over time in individual patients, provided the population under study can be narrowly defined and the time period is relatively short to ensure attribution to the entity being assessed is appropriate. These types of measures are using the patient’s own baseline information to assess performance on the outcome over time. As a result, although patient risk factors would potentially impact the outcome, they would not impact the results of the outcome measure since the patient’s own data is used as the comparison value.

Key Considerations in Identifying Risk Adjustment Strategies for Outcomes Measures
1. The intended use of the outcome measure may impact the nature and type of risk adjustment. Given the greater stakes for accountability measures, any risk adjustment models need to be well vetted and carefully constructed. They must be based on variables that are fairly objective and not subjective. Risk adjustment models must have

demonstrated adequate discrimination (extent to which the model predicts higher probabilities of the outcome for patients who experienced the outcome than for those who did not) and calibration (the match between predicted and actual outcome rates within subgroups of the data such as risk deciles).\textsuperscript{24}

2. Risk adjustment variables will differ depending on the outcome of interest and condition or procedure under study. For example, with regards to the outcome of interest, a prime risk factor affecting mortality would be comorbidities, while disease severity would impact outcomes related to disease management and caregiver support would have an effect on functional outcomes. Risk adjustment for special populations (such as children or the elderly) will need to account for the unique, compelling differences for that patient population such as the substantial physiologic changes that are associated with each age group in childhood and adolescence. At the same time, risk adjustment variables should make clinical sense for the outcome of interest and should include clinical variables known to be associated with the outcome of interest. For example, a recent study demonstrated that the health-related quality of life outcomes for patients following an acute myocardial infarction were negatively impacted by a number of factors including younger age, depressive symptoms, and female gender.\textsuperscript{25} This concept is aligned with guidance from the National Quality Measures Clearinghouse stating that, “risk adjustment is best developed when it is purposefully designed for the condition under study.”\textsuperscript{26} The inclusion or exclusion of socioeconomic status (SES) and other factors such as race and ethnicity as variables in a risk adjustment strategy has been carefully considered by an expert panel convened by NQF. As a result of the panel’s recommendations, NQF has instituted a two-year trial period for risk adjustment for sociodemographic factors during which measure developers will be required to document whether or not SDS adjustment is appropriate for any individual outcome measure under consideration for endorsement. For more information on the role of SES in risk adjustment please refer to the NQF report included in the Directory of Resources at the end of this document.

3. Risk adjustment models are often times dependent on the data that is available. Although research may indicate that a particular variable has an impact on a particular outcome, incorporating this variable in a risk adjustment model may be hampered by the ability to capture it in a consistent and standardized way. Going back to the stroke example, a group of experts recently recommended that pre-stroke physical function and stroke severity be included as one of several variables in any risk-adjustment model examining functional status 30 days after ischemic stroke.\textsuperscript{27} However, they acknowledged that these variables are not collected in a uniform fashion and that additional research would be needed to support quality assessment.\textsuperscript{28} In identifying an appropriate risk adjustment model, consideration will need to be given to the information that can be captured. In so doing, measure developers have to acknowledge that risk-adjustment models which fail

\textsuperscript{26} Op. cit., National Quality Measures Clearinghouse™
\textsuperscript{27} Op. cit., Katzan IL.
\textsuperscript{28} Ibid
to control for potentially significant variables due to feasibility concerns may result in performance misclassifications.

4. Risk adjustment models can take on several different forms. Some simpler approaches may define and focus the measure on the subgroup of patients for whom an expected outcome is not likely to be impacted by factors beyond the health care received. Most sophisticated risk adjustment models require significant statistical expertise and judgment to develop and maintain. A number of statistical models are available to support risk adjustment and include multivariable modeling, indirect standardization or direct standardization.29 As a result, a statistical expert should be involved in the design of any risk adjustment model. Additionally, risk stratification may be considered as an alternative to risk adjustment. Regardless of the risk adjustment model selected, a clear description to users regarding the goal and intended impact of the risk adjustment strategy is needed.

Several outcome measurement pioneers have previously articulated their recommendations or preferences for risk-adjustment models. These recommendations are provided here to serve as guiding principles for measure developers newly embarking on outcome measure development and risk adjustment.

American Heart Association: Preferred Attributes of Models Used for Publicly Reported Outcomes

- Clear and explicit definition of an appropriate patient sample
- Clinical coherence of model variables
- Sufficiently high-quality and timely data
- Designation of an appropriate reference time before which covariates are derived and after which outcomes are measured
- Use of an appropriate outcome and a standardized period of outcome assessment
- Application of an analytical approach that takes into account the multi-level organization of data
- Disclosure of the methods used to compare outcomes, including disclosure of performance of risk-adjustment methodology in derivation and validation samples

Society of Thoracic Surgeons: Essential factors to insuring the accuracy and usefulness of risk adjustment models31

- Selection of an appropriate clinical database
- Inclusion of critical core variables
- Precise definitions for predictor variables and endpoints
- Proper model development, validation, and audit

The Society for Vascular Surgery proposed a novel approach to outcome measurement with the ultimate goal of creating clinically and statistically valid procedural outcome measures without adding the cost of burdensome data collection for risk adjustment. Their initial outcome measures are focused on in-hospital post-procedure stroke (based on ICD-9 claim) or death for carotid endarterectomy (CEA) or carotid stenting (CAS) in asymptomatic patients and in-hospital death for open abdominal aortic aneurysm (AAA) repair or endovascular aneurysm repair (EVAR) in non-ruptured small or moderate aneurysms. The main principle of the measures is that by focusing on the population where the patient risk with non-intervention is low (asymptomatic carotid stenosis and small/moderate non-ruptured AAA) the risk assessment should be part of the surgical decision making and therefore no additional risk adjustment is necessary for assessment of the outcomes. In fact it would be inappropriate to “credit” a physician through risk adjustment for performing an intervention on a high medical risk patient. While the measures are currently in use in a registry, the SVS is exploring the opportunity to collect them using existing administrative (claims) data combined with the use of CPT-II codes and no further risk-adjustment. The CPT-II coding would be used to stratify the population in the case of carotid disease by symptom status (asymptomatic vs. symptomatic) and in the case of AAA by aneurysm size neither of which are validly captured in existing administrative data.32

Finally, in appendix C, we have included several diverse examples of outcomes measures that were well-constructed and that employ different but thoughtful approaches to risk adjustment and that can serve as models for future outcome measure development.

Section 5: Testing of Outcome Measures

The necessity of empirically establishing the reliability and validity of performance measures has long been acknowledged. The methodology for this type of testing of outcome measures is, for the most part, no different than that designed for other types of performance measures.

Reliability testing demonstrates that the results are repeatable, producing the same results a significant proportion of the time when assessed in the same population in the same time period. Examples of reliability testing include but are not limited to inter-rater/abstractor or intrarater/abstractor studies, internal consistency for multi-item scales, signal-to-noise analyses, and test-retest for survey items. Reliability testing may address the data elements or computed measure score.

Validity testing establishes that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. Examples of validity testing include but are not limited to a systematic assessment of face validity, correlation of measure scores with another valid indicator of quality for the specific topic, and content validity for multi-item scales/tests.

In 2006, the PCPI’s Measures Implementation and Evaluation (MIE) Advisory Committee developed a Measure Testing Protocol which outlined a standardized methodology to test performance measures. This document, later updated in 2010, includes additional detail regarding the various testing approaches mentioned above and should serve as a useful resource in guiding outcome measure testing activities.

The consistency in testing methodologies across different types of measures (ie, process and outcome) can also be seen in NQF’s measure endorsement evaluation criteria, with a few exceptions:

- For outcome measures, measure developers need to specify an evidence-based risk-adjustment strategy, based on patient factors that influence the measured outcome and are present at the start of care, and has demonstrated adequate discrimination and calibration. Alternatively, measure developers need to provide a rationale or data that supports a lack of risk adjustment or stratification.

- For PRO-PMs, reliability and validity should be demonstrated for the computed performance score. Additionally, an analysis should be conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders). Information should be provided to specify how the handling of missing data minimizes bias.

---

**Threats to Validity**

With outcome measurement, in particular, measure developers need to understand and reduce the source of a number of potential biases to assure that potential threats to validity are minimized or eliminated.

- **Confounding bias** - Threat that differences in measure scores are due to differences in severity of conditions of patients served rather than differences in quality\(^{34}\) For example, results from a measure assessing remission from depression would differ based on the patient’s baseline severity of depressive symptoms. A well-crafted risk adjustment approach (see previous section) could mitigate the possible threats to validity from confounding bias.

- **Information bias** - Threat that differences in measure score are due to differences in data type and/or differences in data collection practices\(^{35}\) For example, a measure relying on patient or provider recall may be inherently subject to differing results since the recall of information depends entirely on memory which can often be flawed and therefore unreliable. There may be limited ways to address recall as a type of information bias. In the measure development process for outcome measures, it may be best to identify alternative ways to collect the information needed.

- **Selection/Attrition bias** - Threat that differences in measure score are due to missing or “incorrect” data or exclusions\(^{36}\) For example, a measure assessing physical functioning at the first postoperative visit following a total knee replacement would need to carefully consider the impact of missing data since patients lost to follow up may in fact have the best results and therefore be less likely to return for their post-operative visit. In this instance, missing data would need to be assessed to understand the possible impact on the outcome of interest.

- **Ascertainment bias** – Threat that differences in measure score are due to the methods of ascertainment of outcomes For example, if a measure was assessing the incidence of peri-procedural myocardial infarction, different methods for identifying an MI would yield different results. The advent of new, sensitive diagnostic methods allows for detection of very small amounts of myocardial injury\(^{37}\) and these techniques may not be universally utilized. So, any outcome measures addressing this patient population would need to have very clearly defined specifications and data element definitions to ensure consistent ascertainment of the outcome of interest.


Conclusion
As the demand for measures assessing the outcomes of health care continues to increase, the PCPI hopes that this toolkit can serve as a useful resource for measure developers newly seeking to meet that demand. Measure development is a significant commitment for which there is no clear endpoint, for measures once developed require ongoing maintenance to ensure they maintain their clinical relevance and value for quality improvement.

As more outcome measures are developed and the science of outcome measurement evolves, a next iteration of this toolkit may need to be developed – one that can address the experienced outcome measure developer and such concepts as financial toxicity.
Appendix A

PCPI Outcomes Ad Hoc Committee Members

Patrick Romano, MD, MPH (Chair)
Amy Aronsky, DO
Dickson Cheung, MD, MBA, MPH
Mark Jarrett, MD, MBA
Arif Kamal, MD
Jay Lieberman, MD
Timothy Kresowik, MD, MS
Paul Kurlansky, MD
Frederick Masoudi, MD, MSPH
R. Scott McKenzie, MD
Matthew Nielsen, MD, MS
M. Elizabeth Sandel, MD
Junko Takeshita, MD, PhD

PCPI Staff to the Outcomes Ad Hoc Committee

Mark Antman, DDS, MBA
Toni Kaye, MPH
Samantha Tierney, MPH
## Appendix B

### Directory of Resources - Publications

<table>
<thead>
<tr>
<th>Title</th>
<th>Publication Date</th>
<th>Website</th>
</tr>
</thead>
</table>
Directory of Resources – Organizational Websites

International Consortium for Health Outcomes Measurement:  http://www.ichom.org/
Instititue for Healthcare Improvement:  http://www.ihi.org/
## Appendix C

### Existing Outcome Measure Examples

<table>
<thead>
<tr>
<th>Measure title</th>
<th>In-Hospital Risk Adjusted Rate of Mortality for Patients Undergoing PCI</th>
</tr>
</thead>
</table>
| Measure steward and copyright information | American College of Cardiology Foundation  
© 2010 American College of Cardiology Foundation All Rights Reserved |
| Measure description | Risk adjusted rate of mortality for all patients age 18 and over undergoing PCI |
| Numerator statement | Patients 18 years of age and older with a PCI procedure performed during admission who expired |
| Denominator statement | Patients 18 years of age and older with a PCI procedure performed during admission |
| Exclusions | 1. NCDR Registry patients who did not have a PCI (Patient admissions with a diagnostic cath only during that admission);  
2. Patient admissions with PCI who transferred to another facility on discharge |
| Risk adjustment model | Data from 1,208,137 PCI procedures performed between July 2009 and June 2011 at 1,252 CathPCI Registry sites were used to develop both a “full” and pre-catheterization PCI in-hospital mortality risk model using logistic regression.  
Weights were assigned to risk factors or variables reflecting the strength of their association to PCI in-hospital mortality. Each patient in a facilities submission is given a risk score to predict risk of in hospital mortality and accurately report risk adjusted mortality rates during hospitalization.  
The most noteworthy risk factors or variables in the model include:  
1. ST-segment elevation MI defined as a patient who had a STEMI on admission, with an onset within 24 hours, or the procedure indication was primary, rescue or facilitated PCI.  
2. Discharge status (alive or expired). The interaction between this variable with other variables were key in the analysis.  
3. The glomerular filtration rate (GFR) variable is calculated using abbreviated MDRD formula \( \text{GFR} = 186 \times (\text{last creatinine}) - 1.154 \times (\text{age}) - 0.203 \times (\text{gender factor}) \times (\text{race factor}) \text{ where (gender factor)} = 1 \text{ for male and 0.742 for female, (race factor)} = 1.21 \text{ for black and 1 for others }.  
4. The body mass index (BMI) (kg/m2) is calculated from height (cm) and weight (kg): \( \text{BMI} = \text{weight} \times 10000 / (\text{height})^2 \). |
| All Risk Adjustment Variables | STEMI patients  
Age  
BMI  
Cerebrovascular disease  
PAD  
Chronic lung disease  
Prior PCI  
Diabetes  
GFR  
Renal Failure |
<table>
<thead>
<tr>
<th>Left Ventricular Ejection Fraction</th>
<th>Cardiogenic shock and PCI status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Failure NYHA within 2 weeks</td>
<td>Cardiac arrest within 24 hours</td>
</tr>
<tr>
<td>At least 1 previously treated lesion within 1 month with in-stent thrombosis</td>
<td>Highest risk lesion: segment category</td>
</tr>
<tr>
<td>Number of diseased vessels: 2,3, vs 0,1</td>
<td>Chronic total occlusion</td>
</tr>
<tr>
<td><strong>Current/planned use</strong></td>
<td>Public Reporting, Quality Improvement with Benchmarking (external benchmarking to multiple organizations)</td>
</tr>
<tr>
<td><strong>Care setting</strong></td>
<td>Hospital/Acute Care Facility</td>
</tr>
<tr>
<td><strong>Data source</strong></td>
<td>Electronic Clinical Data: Registry</td>
</tr>
<tr>
<td><strong>Level of Analysis</strong></td>
<td>Facility</td>
</tr>
<tr>
<td><strong>NQF number</strong></td>
<td>0133</td>
</tr>
<tr>
<td><strong>Website URL</strong></td>
<td><a href="https://www.ncdr.com/webncdr/cathpci/home/datacollection">https://www.ncdr.com/webncdr/cathpci/home/datacollection</a></td>
</tr>
<tr>
<td>Measure title</td>
<td>Risk-Adjusted Operative Mortality for CABG</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Measure steward and copyright information</td>
<td>The Society of Thoracic Surgeons</td>
</tr>
<tr>
<td>Measure description</td>
<td>Percent of patients aged 18 years and older undergoing isolated CABG who die, including both 1) all deaths occurring during the hospitalization in which the CABG was performed, even if after 30 days, and 2) those deaths occurring after discharge from the hospital, but within 30 days of the procedure</td>
</tr>
<tr>
<td>Numerator statement</td>
<td>Number of patients undergoing isolated CABG who die, including both 1) all deaths occurring during the hospitalization in which the operation was performed, even if after 30 days, and 2) those deaths occurring after discharge from the hospital, but within 30 days of the procedure</td>
</tr>
<tr>
<td>Denominator statement</td>
<td>All patients undergoing isolated CABG</td>
</tr>
<tr>
<td>Exclusions</td>
<td>N/A</td>
</tr>
<tr>
<td>Risk adjustment model</td>
<td>The details of the risk adjustment model development were published in 2009. The list of candidate risk predictors were selected by a surgeon panel based on prior research and clinical expertise. Initial models were selected using a backwards approach with a significance criterion of 0.001 for removal. Three variables were preselected and forced into the models. These included all of the continuous variables (age, BSA, date of surgery [in 6-month intervals], creatinine, ejection fraction), plus sex and dialysis. In addition, atrial fibrillation was included a priori in the model for permanent stroke. Shahian DM, O’Brien SM, Filardo G, Ferraris VA, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1—coronary artery bypass grafting surgery. Ann Thorac Surg. 2009 Jul;88(1 Suppl):S2-22. The definitions of all the variables in the final 2008 CABG model are provided below. (Note not all were included in the final model for this measure.) Variable Definition</td>
</tr>
<tr>
<td></td>
<td>Intercept = 1 for all patients</td>
</tr>
<tr>
<td></td>
<td>Atrial fibrillation = 1 if patient has history of preoperative atrial fibrillation, = 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>Age = Patient age in years</td>
</tr>
<tr>
<td></td>
<td>Age function 1 = max (age – 50, 0)</td>
</tr>
<tr>
<td></td>
<td>Age function 2 = max (age – 60, 0)</td>
</tr>
<tr>
<td></td>
<td>Age by reop function = Age function 1 if surgery is a reoperation, = 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>Age by status function = Age function 1 if status is emergent or salvage, = 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>BSA function 1 = max (1.4, min [2.6, BSA]) − 1.8</td>
</tr>
<tr>
<td></td>
<td>BSA function 2 = (BSA function 1)^2</td>
</tr>
<tr>
<td></td>
<td>CHF but not NYHA IV = 1 if patient has CHF and is not NYHA class IV, = 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>CHF and NYHA IV = 1 if patient has CHF and is NYHA class IV, = 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>CLD mild= 1 if patient has mild chronic lung disease, = 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>CLD moderate = 1 if patient has moderate chronic lung disease, = 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>CLD severe = 1 if patient has severe chronic lung disease, = 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>Creatinine function 1 = max (0.5, min [creatinine, 5.0]) if patient is not on dialysis, = 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>Creatinine function 2 = max ([creatinine function 1] – 1.0, 0)</td>
</tr>
<tr>
<td></td>
<td>Creatinine function 3 = max ([creatinine function 1] – 1.5, 0)</td>
</tr>
<tr>
<td></td>
<td>CVD without prior CVA = 1 if patient has history of CVD and no prior CVA, = 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>CVD and prior CVA = 1 if patient has history of CVD and a prior CVA, = 0</td>
</tr>
</tbody>
</table>
Diabetes, noninsulin = 1 if patient has diabetes not treated with insulin, = 0 otherwise
Diabetes, insulin = 1 if patient has diabetes treated with insulin, = 0 otherwise
Ejection fraction function = max (50 – ejection fraction, 0)
Female = 1 if patient is female, = 0 otherwise
Female by BSA function 1 = BSA function 1 if female, = 0 otherwise
Female by BSA function 2 = BSA function 2 if female, = 0 otherwise
Hypertension = 1 if patient has hypertension, = 0 otherwise
IABP or inotropes= 1 if patient requires IABP or inotropes preoperatively, = 0 otherwise
Immunosuppressive treatment = 1 if patient given immunosuppressive therapy within 30 days, = 0 otherwise
Insufficiency, aortic = 1 if patient has at least moderate aortic insufficiency, = 0 otherwise
Insufficiency, mitral = 1 if patient has at least moderate mitral insufficiency, = 0 otherwise
Insufficiency, tricuspid = 1 if patient has at least moderate tricuspid insufficiency, = 0 otherwise
Left main disease = 1 if patient has left main disease, = 0 otherwise
MI 1 to 21 days = 1 if history of MI 1 to 21 days prior to surgery, = 0 otherwise
MI > 6 and < 24 hours = 1 if history of MI >6 and <24 hours prior to surgery, = 0 otherwise
MI 6 hours = 1 if history of MI 6 hours prior to surgery, = 0 otherwise
No. diseased vessel function = 2 if triple-vessel disease, = 1 if double-vessel disease, = 0 otherwise
PCI 6 hours = 1 if patient had PCI 6 hours prior to surgery, = 0 otherwise
Peripheral vascular disease = 1 if patient has peripheral vascular disease, = 0 otherwise
Race black = 1 if patient is black, = 0 otherwise
Race Hispanic = 1 if patient is nonblack Hispanic, = 0 otherwise
Race Asian = 1 if patient is nonblack, non-Hispanic, and is Asian, = 0 otherwise
Reop, 1 previous operation = 1 if patient has had exactly 1 previous CV surgery, = 0 otherwise
Reop, 2 previous operations = 1 if patient has had 2 or more previous CV surgeries, = 0 otherwise
Shock = 1 if patient was in shock at time of procedure, = 0 otherwise
Status urgent = 1 if status is urgent, = 0 otherwise
Status emergent = 1 if status is emergent (but not resuscitation), = 0 otherwise
Status salvage = 1 if status is salvage (or emergent plus resuscitation), = 0 otherwise
Stenosis aortic = 1 if patient has aortic stenosis, = 0 otherwise
Unstable angina = 1 if patient has unstable angina, no MI within 7 days of surgery, = 0 otherwise

Current/planned use
Public Reporting. Quality Improvement (Internal to the specific organization), Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

Care setting
Hospital/Acute Care Facility

Data source
Electronic Clinical Data: Registry

Level of Analysis
Clinician: Group/Practice, Facility

NQF number
0119

Website URL
http://www.sts.org/sites/default/files/documents/STSAadultCVDataCollectionForm2_73_Annnotated.pdf
<table>
<thead>
<tr>
<th>Measure title</th>
<th>Rate of Postoperative Stroke or Death in Asymptomatic Patients Undergoing Carotid Endarterectomy (CEA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure steward and copyright information</td>
<td>Society for Vascular Surgery</td>
</tr>
<tr>
<td>Measure description</td>
<td>Percent of asymptomatic patients undergoing CEA who experience stroke or death following surgery while in the hospital</td>
</tr>
<tr>
<td>Numerator statement</td>
<td>Patients who experience stroke or death in the hospital following CEA</td>
</tr>
<tr>
<td>Denominator statement</td>
<td>Patients aged 18 and older who are asymptomatic undergoing CEA</td>
</tr>
<tr>
<td>Exclusions</td>
<td>Symptomatic carotid stenosis: Ipsilateral carotid territory TIA or stroke less than 120 days prior to procedure OR Other carotid stenosis: Ipsilateral TIA or stroke 120 days or greater prior to procedure or any prior contralateral carotid territory or vertebrobasilar TIA or stroke</td>
</tr>
<tr>
<td>Risk adjustment model</td>
<td>No risk adjustment necessary</td>
</tr>
<tr>
<td>Current/planned use</td>
<td>Payment Program</td>
</tr>
<tr>
<td>Care setting</td>
<td>Hospital/Acute Care Facility</td>
</tr>
<tr>
<td>Data source</td>
<td>Electronic Clinical Data: Registry</td>
</tr>
<tr>
<td>Level of Analysis</td>
<td>Clinician: Group/Practice, Clinician: Individual, Facility</td>
</tr>
<tr>
<td>NQF number</td>
<td>N/A</td>
</tr>
<tr>
<td>Website URL</td>
<td>N/A</td>
</tr>
<tr>
<td>Measure title</td>
<td>Depression Response at Six Months</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Measure steward and copyright information</td>
<td>Minnesota Community Measurement  © MN Community Measurement, 2014. All rights reserved.</td>
</tr>
<tr>
<td>Measure description</td>
<td>Adult patients age 18 and older with major depression or dysthymia and an initial PHQ-9 score &gt;9 who demonstrate remission at six months defined as a PHQ-9 score less than 5. This measure applies to both patients with newly diagnosed and existing depression whose current PHQ-9 score indicates a need for treatment. This measure additionally promotes ongoing contact between the patient and provider as patients who do not have a follow-up PHQ-9 score at six months (+/- 30 days) are also included in the denominator.</td>
</tr>
<tr>
<td>Numerator statement</td>
<td>Adults age 18 and older with a diagnosis of major depression or dysthymia and an initial PHQ-9 score greater than nine who achieve remission at six months as demonstrated by a six month (+/- 30 days) PHQ-9 score of less than five.</td>
</tr>
<tr>
<td>Denominator statement</td>
<td>Adults age 18 and older with a diagnosis of major depression or dysthymia and an initial (index) PHQ-9 score greater than nine.</td>
</tr>
<tr>
<td>Exclusions</td>
<td>Patients who die, are a permanent resident of a nursing home or are enrolled in hospice are excluded from this measure. Additionally, patients who have a diagnosis (in any position) of bipolar or personality disorder are excluded.</td>
</tr>
<tr>
<td>Risk adjustment model</td>
<td>This measure is risk adjusted based on severity band of the PHQ-9 which is based on the initial PHQ-9 score. Severity bands are defined as 10 to 14- moderate depression, 15 to 19- moderately severe depression and 20 to 27- severe depression. The measure is also risk adjusted for insurance product type (commercial, Medicare, and MN government programs/ self-insured) and age bands (18-25, 26-50, 51-65 and 66+).</td>
</tr>
<tr>
<td>Current/planned use</td>
<td>Payment Program, Public Reporting, Quality Improvement (Internal to the specific organization), Quality Improvement with Benchmarking (external benchmarking to multiple organizations)</td>
</tr>
<tr>
<td>Care setting</td>
<td>Ambulatory Care: Clinician Office/Clinic, Behavioral Health/Psychiatric: Outpatient</td>
</tr>
<tr>
<td>Data source</td>
<td>Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records</td>
</tr>
<tr>
<td>Level of Analysis</td>
<td>Clinician : Group/Practice, Facility</td>
</tr>
<tr>
<td>NQF number</td>
<td>0711</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure title</th>
<th>Patient Experience with Surgical Care Based on the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Surgical Care Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure steward and copyright information</td>
<td>American College of Surgeons, Division of Advocacy and Health Policy</td>
</tr>
<tr>
<td>Measure description</td>
<td>The following 6 composites and 1 single-item measure are generated from the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Surgical Care Survey. Each measure is used to assess a particular domain of surgical care quality from the patient’s perspective. Measure 1: Information to help you prepare for surgery (2 items) Measure 2: How well surgeon communicates with patients before surgery (4 items)</td>
</tr>
</tbody>
</table>

© 2015 American Medical Association. All Rights Reserved.
| Measure 3: Surgeon’s attentiveness on day of surgery (2 items) |
| Measure 4: Information to help you recover from surgery (4 items) |
| Measure 5: How well surgeon communicates with patients after surgery (4 items) |
| Measure 6: Helpful, courteous, and respectful staff at surgeon’s office (2 items) |
| Measure 7: Rating of surgeon (1 item) |

The Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Surgical Care Survey is administered to adult patients (age 18 and over) having had a major surgery as defined by CPT codes (90 day globals) within 3 to 6 months prior to the start of the survey.

**Numerator statement**

We recommend that CAHPS Surgical Survey composites be calculated using a top-box scoring method. The top box score refers to the percentage of patients whose responses indicated excellent performance for a given measure. This approach is a kind of categorical scoring because the emphasis is on the score for a specific category of responses.

The composite measures do not have a typical numerator. This section is used to describe the composite score. The composite score is the average proportion of respondents who answered the most positive response category across the questions in the composite. The top box numerators for items within Composite measures 1, 2, 4, 5, and 6 is the number of respondents who answered “Yes, definitely” across the items in each composite. The top box composite score is the average proportion of respondents who answered “Yes, definitely” across the items in the composite.

The top box numerator for items within Composite measure 3 is the number of respondents who answered “Yes” across the items in this composite. The top box composite score is the average proportion of respondents who answered “Yes” across the items in this composite.

The top box numerator for the Measure 7, the Global Rating Item, is the number of respondents who answered 9 or 10 to the Global Rating Item.

Note that for users who want to case-mix adjust their scores, case-mix adjustment can be done using the CAHPS macro and the adjustment is made prior to the calculation of the total score. For more, see section 2e.2.

See also Attachment H: Reporting Measures for the CAHPS Surgical Care Survey.

**Denominator statement**

The composite does not have a typical denominator statement. This section describes the target population.

The major criteria for selecting patients were having had a major surgery as defined by CPT codes (90 day globals) within 3 to 6 months prior to the start of the survey. [For the full list of CPT codes, see Attachment J].

**Exclusions**

The following patients would be excluded from all composites:
- Surgical patients whose procedure was greater than 6 months or less than 3 months prior to the start of the survey.
- Surgical patients younger than 18 years old.
- Surgical patients who are institutionalized (put in the care of a specialized institution) or deceased.
- Surgery performed had to be scheduled and not an emergency procedure since emergency procedures are unlikely to have visits with the surgeon before the surgery.
- Multiple surgery patients within the same household can be included in the sampling frame. However, once one patient in the household is sampled, any
other patients in the same household would be excluded from being sampled in order to minimize survey burden to the household.

Risk adjustment model

Case-mix risk adjustment.

Variable selection

We chose an initial set of adjusters to evaluate based on both historical use – some variables, such as age and education, have been subjected to extensive case-mix analysis in other CAHPS studies, and there is ample evidence that they are important case-mix adjusters – and the conceptual appropriateness of the variable as an adjuster. The next step was to select a subset of these potential case-mix adjusters for further analysis. The strength of the relationship of each potential adjuster to overall ratings of provider care quality was evaluated using step-wise regression in which each potential adjuster was regressed onto three global ratings of care.

Our initial pool of potential adjusters consisted of the following:
- Age (q53)
- Education (q56)
- Gender (q54)
- Hispanic (q57)
- Mode (paper vs. web)
- Form (A vs. C)
- Number of Prior Surgeries (q55)

Three global measures of care served as the outcome variables in the models:
- The global rating of the surgeon (q50),
- The global rating of the anesthesiologist (q30)
- The question asking if the results of the surgery were as good or better than expected (q51)

Three separate stepwise regression models were estimated in which the potential adjusters listed above were regressed on each of the outcome variables. Variables selected in any of these models formed a core set eligible for final selection. For each outcome, we present the parameter estimate for each potential adjuster and indicate which adjusters stayed in the model (i.e., each variable for which the parameter estimate met the retention criterion of being <.05). Of the potential adjusters evaluated, age, education, and the number of prior surgeries were each significantly related to at least one of the ratings. The set of variables retained for further assessment included:
- Age (q53)
- Education (q56)
- Number of Prior Surgeries (q55)

<table>
<thead>
<tr>
<th>Current/planned use</th>
<th>Public Reporting, Quality Improvement (Internal to the specific organization)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Care setting</td>
<td>Ambulatory Surgery Center, Ambulatory Clinician Office/Clinic, Hospital/Acute Care Facility</td>
</tr>
<tr>
<td>Data source</td>
<td>Patient Reported Data/Survey</td>
</tr>
<tr>
<td>Level of Analysis</td>
<td>Clinician: Group/Practice, Clinician: Individual</td>
</tr>
<tr>
<td>NQF number</td>
<td>1741</td>
</tr>
<tr>
<td>Website URL</td>
<td>N/A</td>
</tr>
</tbody>
</table>