

# Key Steps in Creating and Using Quality Measures



This resource provides an overview of the life cycle of quality measures and opportunities for consumer engagement. The image below displays the six stages: setting priorities, creating measure concepts, specifying measures, testing and endorsing measures, using measures, and maintaining measures. This information is excerpted from the report: [Engaging Consumers in the Quality Measurement Enterprise](#).

## **Setting priorities: What general things should we measure?**

To guide the overall direction of quality measure development and use, government and private organizations can make efforts to achieve consensus on which areas to prioritize for quality improvement. At this stage, initial decisions are made regarding which aspects of care will be prioritized and for which populations. Decisions emerging from these priority-setting efforts can influence subsequent investment in new measures and help determine how new and existing measures are used. General priorities include such things as the decisions to measure diabetes care, dementia care in long-term care settings, or congestive heart failure orders at discharge from inpatient hospitalizations. Participation in priority-setting efforts does not require clinical or methodological expertise.

Advocacy that truly reflects the priorities of consumers and communities is highly valuable at this stage. Specific opportunities for engagement include but are not limited to the National Priorities Partnership of the National Quality Forum and Measurement Advisory Panels of the National Committee for Quality Assurance.



## **Creating measure concepts within a priority area: What specific things should we measure?**

Once general priorities for measurement are identified, the next step is to create measure concepts that can serve as the basis for developing specific measures. When priorities involve clinical outcomes, measure concepts are frequently based on scientific evidence and clinical guidelines. Historically, there have been few measure concepts generated directly by patients. For priorities involving patient-centered care, measure concepts might include provider capabilities (e.g., having electronic health records that integrate patient generated data), as well as patients' experiences of care. Consumer advocates can create measure concepts on their own or in partnership with measure developers. Creating measure concepts requires knowledge of specific opportunities to improve care within a given priority area and benefits from reasonably specific ideas for how these might be improved. Consumer advocates who closely follow the scientific literature on treatment of their highest-priority conditions are likely to be well positioned to create measure concepts. Advocates' roles in this activity are especially important when they represent communities with health interests that are not widely prevalent in the general population. Without careful attention by advocates, these interests may be missed by measure developers who focus on the most prevalent health conditions.

### **Specifying measures: How exactly will we measure these concepts?**

Before a measure concept can be tested and used to improve care, it must be translated into detailed specifications. Measure specifications are rules that describe the data required and methods for calculating a quality measure. These methods include definitions of denominators (i.e., which patients and which health events to include or exclude), numerators (i.e., which health events count as better or worse care), time frames for measurement, and methods for adjusting measures by clinical severity of the population. Consumer advocates can make valuable contributions to measure specifications, even if they lack such methodological expertise in-house. This is because value judgments (i.e., decisions for which there is not a single, correct scientific answer) are nearly inevitable when translating measure concepts into detailed specifications. Such value judgments include determining which specific care scenarios do and do not meet the intent of the measure concept. By working closely with measure developers, consumer advocates can help ensure that measure specifications match their measure concept as closely as possible. Participation on technical expert and advisory panels for measure development are possible ways to do this.

### **Testing and endorsing measures: Are they accurate enough to use?**

Testing a measure means assessing its validity and reliability. Validity means that the measure assesses what it is intended to measure. Reliability is the degree to which *apparent* performance differences between providers are *true* differences, rather than being due to chance—i.e., some providers just being luckier or unluckier than others. Because measure specifications are complex and some degree of measurement error is unavoidable, assessing a measure's validity and reliability is highly advisable before using it in a high-stakes manner. To test validity, performance of the measure can be compared with a "gold standard" or with performance on other measures that are strongly related to the measure concept. Measure testing also includes assessment of reliability. Measures that are valid but unreliable might be true assessments of provider performance *on average* but still be inappropriate for high-stakes use because provider-to-provider differences in calculated performance mostly occur at random. An intuitive sign of an unreliable measure occurs when measured performance for a given provider fluctuates wildly from year to year, even if the provider is not doing anything differently. When a measure has acceptable validity and reliability, its developers can seek endorsement from a measure-vetting body, such as the National Quality Forum. Such endorsement can help potential users of the measure have confidence in its accuracy.

While testing a performance measure requires expertise in performance measurement and statistics, endorsement also requires input from stakeholders who have other kinds of expertise, including the experience of being a consumer. This is because statistics and measurement science alone cannot determine how much accuracy is good enough for a given use of the measure. For example, there is no mathematically "correct" answer for how much measurement inaccuracy is too much (i.e., the point at which an inaccurate report becomes worse than no report at all). Consumer advocates, therefore, are well positioned to provide this critical information by communicating to measure developers how much accuracy community members need from a quality measure.

### **Using measures to improve the care available to community members**

Following a conceptual model developed by Dana Safran in 2008, quality measures can be used in high-stakes and lower-stakes ways to improve the quality of care. High-stakes uses include financial incentives such as pay-for-performance (P4P), public performance reporting, and provider tiering, in which patients pay higher out-of-pocket prices to see lower-performing providers. In all of these high-stakes uses, measures are linked directly to external incentives targeting providers, patients, or both. Because inaccurate measures could misdirect patients to truly lower-quality providers and penalize truly higher-quality providers, most stakeholders have agreed that high-stakes uses require performance measures that are reasonably valid and reliable.

Lower-stakes uses of performance measures include confidential feedback to providers to help guide their quality improvement efforts and community-level public reporting to raise general awareness of quality issues and help set priorities for improvement. While accuracy is still important for these lower-stakes uses, they do not directly influence patients' choice of providers, nor do they change how providers are paid.

Consumer groups are frequently invited to help choose measures from a menu when new reporting and performance incentives programs are being designed. Deciding which measures to use for a given program is sometimes described as a prioritization activity. However, choosing from among a menu of existing measures should not be confused with the “setting priorities” phase of the life cycle for quality measures—which can identify priorities for which no quality measures exist.

### ***Maintaining measures***

Measure maintenance involves periodically determining whether the concepts underlying an existing measure need to be modified (e.g., to incorporate new research findings or revised guidelines) and updating measure specifications to accommodate changes in the measure concept or in the underlying data. Updating measure concepts or specifications can interfere with efforts to track year-to-year changes in performance. Therefore, modifying a measure can involve difficult tradeoffs. Measure maintenance might also include checking for unintended negative consequences of ways in which the measure is used, such as a worsening of disparities if P4P programs steer resources away from providers serving sociodemographically vulnerable communities. Some uses of measures might be challenging for consumers and providers alike. Such unintended consequences can be addressed by changing measure specifications, changing the way a measure is used, or removing a measure from use altogether. In addition, measures can be retired and no longer maintained when they have fulfilled their purpose. Measure stewards have primary responsibility for measure maintenance, and they are frequently the original developers of the measures. Consumer advocates with concerns or advice about maintaining a measure can always publish their concerns in a public forum, but in many cases simply contacting the measure steward directly might be sufficient.

### **Comment**

Not every quality measure follows this sequence. Sometimes a stage is skipped. For example, a new measure can be used before it is tested, even though such skipping might create bad incentives or misinform consumers—for example, if a measure turns out not to assess what it is supposed to. Moreover, the lengths of time within and between stages can be unpredictable. Measure developers and measure users can have differing agendas and time lines. Further, activities within one stage may be repeated based on lessons learned in other stages. Limitations of a new measure might not become apparent until its usage, when ideas for improved or substitute measures come into focus. In such cases, revisiting the earliest stages of measure development can guide its refinement. In principle, all stakeholders, including consumers, can participate in each stage of the quality measure life cycle. In practice, the ability of consumers to engage effectively in certain stages—most notably specifying, testing, and maintaining measures—depends in part on the willingness and experience of technical experts in quality measurement methods.

Consumer involvement in technical work is also dependent on their access to the resources that are necessary to support full participation, including funding, allocated time, training, and support. The magnitude of investment required to participate throughout the whole quality measurement life cycle can be considerable, and the payoff might be uncertain, especially in the short term. In our experience, meaningful consumer engagement in all stages depends on both the ability of technical experts to work collaboratively with consumers and on ongoing support for their engagement.