



SEPTEMBER 2014

# Considerations for State Development of Performance Measure Sets

Prepared by Beth Waldman and Michael Bailit, Bailit Health Purchasing, LLC

## Introduction

As states play a more active role in health care delivery system and payment reform, Medicaid programs have joined other public and private sector purchasers in measuring performance as part of value-based purchasing initiatives. While essential to value-based purchasing, performance measurement can create a significant administrative burden for providers. This burden can grow significantly when individual payers (e.g., insurers, managed care plans, and third-party administrators) utilize different measures. There is a growing interest by Medicaid programs and other payers in developing common measure sets to reduce administrative burden on providers and send a common message to them about performance accountability.

This guide provides an overview of the steps states should take in developing a performance measure set—either on their own or in partnership with others—identifies critical considerations, and offers guidance in selecting measures.

## Key Initial Steps in Developing a Performance Measure Set

A number of basic questions must be answered in order to appropriately shape a discussion of what measures should be included in a measure set. It is essential to define early on whose performance is to be measured, for what purpose, and by whom. It is also important to decide who will participate in measure set development and how decisions will be made within the participant group.

### 1. Whose Performance is Being Measured?

States may choose to measure health plans and/or providers. Most current state measure set development work is focused on provider organizations, including one or more of the following: patient-centered medical homes (PCMHs), health homes, hospitals, and accountable care organizations (ACOs). In some cases, states are also developing general, procedure-specific, and condition-specific measures to support episode-based payment programs. There are also efforts to measure the performance of behavioral health and long-term services and supports providers. Measurement of ambulatory health care, however, is most common.

While this guide focuses on developing measure sets for providers, the processes described here are also applicable to health plan measure set development.

### 2. What is the Purpose of Measuring Performance?

There are a number of reasons why a state chooses to measure provider performance. Historically, states have measured provider performance as a component of a quality monitoring system, and have used performance results to inform selection of quality improvement initiatives. More recently, states have begun using performance measures to provide consumers with information about the performance of a provider and to inform discussions with contracted provider groups about their performance. State

### ABOUT BAILIT HEALTH PURCHASING

Bailit Health Purchasing is a health care consulting firm dedicated to ensuring insurer and provider performance accountability on behalf of public agencies and private purchasers.

### ABOUT STATE HEALTH AND VALUE STRATEGIES

State Health and Value Strategies, a program funded by the Robert Wood Johnson Foundation, provides technical assistance to support state efforts to enhance the value of health care by improving population health and reforming the delivery of health care services. The program is directed by Heather Howard at the Woodrow Wilson School of Public and International Affairs at Princeton University.

### ABOUT THE ROBERT WOOD JOHNSON FOUNDATION

For more than 40 years the Robert Wood Johnson Foundation has worked to improve the health and health care of all Americans. We are striving to build a national Culture of Health that will enable all Americans to live longer, healthier lives now and for generations to come. For more information visit [www.rwjf.org](http://www.rwjf.org). Follow the Foundation on Twitter at [www.rwjf.org/twitter](http://www.rwjf.org/twitter) or on Facebook at [www.rwjf.org/facebook](http://www.rwjf.org/facebook).

purchasers and their contracted health plans are also introducing new payment models that tie reimbursement to performance. Some state employer purchasers also use performance measures to tier a provider network or to identify a center of excellence.

It is not uncommon for states to use measures for more than one of the above purposes, or to use some measures for some purposes, and other measures for different purposes. For example, the state of Vermont organized a multistakeholder process to establish a performance measure set for a large ACO pilot. Some measures were selected for ACO reporting only, some for reporting and for influencing payment, and still others for measurement at the health plan level due to high baseline performance.

### 3. Is Measurement Specific to a State Program or Part of a Multipayer Initiative?

It is important to determine whether state programs will measure performance on their own or as part of a larger, multipayer initiative. For example, it is increasingly common for state Medicaid programs, state-operated insurance exchanges, and agencies charged with purchasing state employees' health coverage to use the same measure sets commercial payers use. When deciding if a single or multipayer measure set is desirable, state staff must determine the following:

- Is there a shared set of providers from whom services are being purchased?
- Are there common areas of measurement interest?
- Is there a shared purpose or intended use for the measures?

If the answer to each of these questions is “yes,” then it may make sense for a state agency to embark on a multipayer measure set initiative. Where feasible, there are advantages to both payers and providers. First, it offers a way to consistently assess performance across the entire health system within a state or geographic region. Second, depending upon the approach utilized, it can increase the measure denominator, resulting in greater ability to measure with statistical certainty. Third, it reduces the burden on providers of supplying data and attempting to improve across a large number of measures.\* Fourth, it gives providers a clear message on what aspects of care are most important to purchasers and payers, and encourages them to focus on those areas.

### 4. How Often Will Measurement Occur?

As part of the initial planning process, states and/or multipayer initiatives should consider whether measurement will be one-time or ongoing, and if ongoing, how often. In most cases, measurement occurs on an annual basis as many quality measures use 12-month measurement periods. There may be a desire to measure more frequently to track progress toward an established

goal (something the Oregon Health Authority does two to three times a year), or for certain types of measures, such as utilization (something which the Vermont Green Mountain Care Board does when it tracks ACO member service utilization on a year-to-date basis).

### 5. Who Participates in the Process and How Are Decisions Made?

When developing a measurement set internal to a state agency, it is important to include the right staff from across the organization to ensure appropriate consideration is given to the entirety of the agency's measurement goals, and that the appropriate decision-makers are in the room. There will be some difficult decisions about how to prioritize measures and whether the agency has sufficient resources to implement a particular measure or set of measures. At a minimum, an agency's quality, informatics, medical management, and finance departments should be represented, and there should be a clear decision prior to the start of the project as to who will own the project and serve as the ultimate decision-maker.

In addition, the participation of external stakeholders, such as affected providers, health plans, and consumer advocates can not only increase the likelihood of obtaining buy-in from key constituents, but also contribute to a better-reasoned and effective measure set.

If state agencies are participating in a multistakeholder effort to develop a measure set, it is important to have the right staff from all participating organizations actively engaged. Participants must be able to make decisions and commit their organization to an approach. Individuals who are neither technically informed (e.g., an insurer's regional sales manager) nor empowered will be unable to contribute to the process or ensure that the resulting measure set will be adopted by their organizations.

Multistakeholder initiatives must clearly delineate up front how decisions will be made within the group and how measures will be prioritized when there are differing goals or disagreement on how to move forward. At the start of the process, participants should lay out how decisions will be made and how disagreements will be addressed.

### Measure Selection

The first step in selecting measures is to set out selection criteria that allow for a consistent review of potential measures that is informed by the overall goals and desired outcomes for the measurement program.

Selection criteria typically address:

1. clinical and technical merits of the measure;
2. the relation of the measure to goals and improvement opportunities;

\* Cambridge Health Alliance (MA) reported having 546 payer-defined measures. (Somava Stout, personal communication, May 14, 2014).

3. operational considerations for generating the measure; and
4. the relation of the measure to other pre-existing measure sets of interest.

### Selection Criteria

There are a number of important questions to consider when selecting measures. States should leverage the *Buying Value Measure Selection Tool*, which provides both technical and programmatic criteria for each measure, and a set of criteria for the overall measure set. Further, the tool also provides examples of measure set criteria and can help states track whether the measures under consideration meet measure selection criteria adopted by the state. Such criteria can and should be applied both to individual measures and the entire set, the latter to ensure that the entirety of the measure set is balanced and complete. Examples of criteria commonly adopted include whether measures:

- are collectively consistent with the overall goals of those involved in measure set development;
- are valid and reliable;
- represent opportunities for performance improvement;
- measure the provider's performance in an area within the targeted providers' control;
- have been endorsed by a national body, such as the National Quality Forum (NQF) or the National Committee for Quality Assurance (NCQA);
- have sufficient denominators to produce reliable measurement, be they intended for assessment of statewide, multi-provider, practice site or individual practitioner performance;
- have relevant benchmarks;
- are focused on outcomes;
- are feasible to implement, and are not overly burdensome to generate, report, and if applicable, aggregate;
- are aligned with existing state measure sets and initiatives;
- are aligned with measures currently in use by health plans; and
- are aligned with national and federal measurement initiatives.

One potential criterion is the size of the set. It is often difficult to set a limit on size before knowing the types of measures to be adopted and their intended use. For example, a measure set that includes both physician and hospital measures, as well as access, quality, patient experience, and efficiency measures,

should be expected to be larger than one including only physician ambulatory care quality measures. Should the state desire to adopt a measure set size criterion, however, the number should not be set in stone, but should be used to help filter and prioritize potential measures.

### Use of New and Innovative Measures

As states look to develop measure sets, they often begin with a desire to look at outcome measures rather than process measures, and to focus on areas that may currently be under-measured, such as care integration, social determinants of health, and social supports. Such measures can pose implementation challenges. This is not to say that a state should not strive to innovate, or adopt “transformational measures,” but in so doing the state should ensure that implementation is feasible, recognizing that it will require significant time and resources to develop and/or implement such measures. The state may want to consider staging the implementation of innovative measures, piloting and testing them before using them for transparency or payment purposes.

### Designating Measures for Specific Uses and Specific Populations

As indicated above, measures may be selected for one or more uses. The Maine Health Management Coalition organized a multistakeholder measure selection process on behalf of the state with the specific purpose that the measures would be employed in both the state's and commercial insurers' contracts with ACOs. Other states, however, have designated different measures for distinct purposes, including performance monitoring, value-based payment, public reporting, and measure testing.

In addition, measures may be selected for use across populations or for a specific population. For example, Medicaid and commercial payers may agree that common measures of diabetes care are a priority for both of their populations. They may differ in opinion, however, when considering measures specific to persons with serious and persistent mental illness due to the greater prevalence of the condition in the Medicaid population. In such circumstances, the parties may agree to adopt a measure set that is common to commercial and Medicaid populations, but also allows for a limited number of Medicaid-only measures.

This measure designation process can occur during measure set development, or following initial development of the measure set.

### Identifying Populations, Performance Domains and Services for Measurement

To develop a comprehensive measure set, the state should include measures that comprehensively address patient populations, performance domains, and services. Table 1 provides a description of potential populations, domains, and clinical service areas. Not all of the categories are mutually exclusive.

**Table 1: Potential Measurement Categories**

Populations	Performance Domains	Service Areas
<ul style="list-style-type: none"> <li>▪ children</li> <li>▪ adolescents</li> <li>▪ non-disabled adults</li> <li>▪ adults with disabilities</li> <li>▪ pregnant women</li> <li>▪ seniors</li> </ul>	<ul style="list-style-type: none"> <li>▪ provider infrastructure</li> <li>▪ access</li> <li>▪ clinical process</li> <li>▪ clinical outcomes</li> <li>▪ health status</li> <li>▪ function</li> <li>▪ consumer experience</li> <li>▪ patient engagement</li> <li>▪ patient safety</li> <li>▪ cost</li> <li>▪ efficiency</li> <li>▪ utilization</li> <li>▪ overuse and misuse</li> </ul>	<ul style="list-style-type: none"> <li>▪ prevention</li> <li>▪ acute care</li> <li>▪ chronic illness care</li> <li>▪ dental care</li> <li>▪ behavioral health care</li> <li>▪ inpatient care</li> <li>▪ ambulatory care</li> <li>▪ long-term services and supports</li> <li>▪ care management</li> </ul>

States sometimes identify sub-populations, performance domains, and service areas of special interest to them. It is quite common for states to identify specific diseases that are prevalent within a population or program being measured. Where diabetes and asthma are common across populations, Medicaid programs might want to target care for behavioral health conditions, such as serious and persistent mental illness and substance use disorders. The specific conditions and/or procedures to be measured depend on the goals of the measurement program, the participants in the measure selection process, and the criteria that they adopt at the outset of their work.

**Resources for Locating Measures**

There are many sources that may be used to identify potential measures. In addition to the 700 NQF-endorsed measures, measure set developers should consider the following resources:

- Federal measure sets (partial list)
  - Consumer Assessment of Healthcare Providers and Systems (CAHPS®) surveys
  - Children’s Health Insurance Program Reauthorization Act (CHIPRA) core set
  - Center for Medicare and Medicaid Innovation (CMMI) core measures
  - Hospital Compare
  - Meaningful Use Clinical Quality measures\*
  - Medicare Advantage Stars Program measures

- Medicare Shared Savings Program measures
- Medicare-Medicaid Financial Alignment Model measures
- Medicaid adult core set
- Nursing Home Compare
- Pre-existing state measure sets (partial list, not applicable to all states)
  - Measure sets currently in place in state health plan and third-party administrators contracts
  - Measures sets currently in place in state ACO, PCMH, and health home contracts
  - Measure sets defined through state-facilitated processes for multipayer and provider use. For example, Massachusetts’ [Standard Quality Measure Set](#) and California’s [CalQualityCare.org](#).
- Pre-existing multistakeholder coalition measure sets, such as those developed by the Wisconsin Collaborative for Health Care Quality (WCHQ), Better Health *Greater* Cleveland, Minnesota Community Measurement, and the New Mexico Coalition for Healthcare Quality.
- Agency for Healthcare Research and Quality’s prevention quality indicators
- NCQA’s Healthcare Effectiveness Data and Information Set (HEDIS)
- Long-term services and supports scorecard

\* These measures are a subset of the larger Physicians Quality Reporting System and Physician Value Based Payment Modifier Program measure set.

Web links to several of the measure sets cited above are in the Buying Value Measure Selection Tool. The tool also includes a list of the most frequently used measures by domain. In addition, the tool provides a scoring template states can use to organize the measures in use or under consideration and apply their selection criteria. Through an automated crosswalk, the template lets states identify whether a measure is included within a federal measure set.

### Measure Set Fidelity

Most current measure set activity involving multiple payers is aimed at achieving true alignment, where each payer agrees to adopt the common measure set in full, with the exception of a few population-specific measures.

An alternative approach, however, involves the adoption of a common measure set from which each participating payer (or payer and provider) chooses which measures to use. While this “menu” approach reduces variation across payer measure sets, it leaves the door open to non-alignment.

## Producing the Measure Set

The process of collecting data and producing measures can be resource-intensive. It is important to understand the data that are needed to produce a particular measure, and to consider the ability of the state and/or its health plans to access, collect, and analyze such data prior to selecting a measure for use.

### Data Sources

A variety of data sources can be used to generate measures. For the most part, measures that use claims or encounter data are the easiest to produce, because they are readily available to the state and/or its health plans. Measures that require a consumer survey are also relatively easy to produce, particularly if the survey process is already in place.

More difficult to produce are measures that require a review of clinical records. If performed manually, reviewing clinical records is time-consuming and expensive for providers and states and/or health plans. If performed using electronic data sources, there are additional challenges, including:

- limitation in the numbers of providers able to capture and report the designated measures;
- inconsistent reporting across electronic health records (EHRs), creating problems in the reliability of reported data; and
- the inability of many health information exchanges to facilitate electronic measure reporting.

Despite the current difficulties associated with generating measures using clinical data sources, there is little question that current trends toward expanded EHR adoption and health information exchange development will result in increasing use

of clinical data-based measures over time. States should anticipate this trend and make provision for testing or including some clinical data-based measures in their measure set.

### Identifying Benchmarks

In addition to identifying data sources for measure generation, it is also important to identify benchmarks to which a provider’s performance will be compared. This is particularly true if the state anticipates using the measure set for quality improvement, public reporting, or adjusting payment. In all three applications, it is often necessary to assess performance relative to a benchmark to identify opportunities for improvement.

Unfortunately, there are limitations in the number of measures for which national benchmarks are available. Many states select NCQA’s HEDIS measures for their measure sets, because NCQA annually publishes Medicaid, Medicare, and commercially-insured population benchmarks for most of the HEDIS measures.<sup>1</sup> Yet, use of the HEDIS health plan measure benchmarks for provider performance can be troublesome. As reported by WCHQ at the *Buying Value* meeting in March 2014, differences in specifications necessary to make a health plan measure applicable to a provider entity can significantly impact the comparability of the two rates.

Other sources for national benchmarks exist, but these too have their limitations, as noted below:

- Health Resources Service Administration (HRSA): HRSA collects and reports on a number of clinical data-based measures. The rates are reported from the EHRs operated by federally-qualified health centers (FQHCs) and reflect FQHC performance only.
- Centers for Disease Control and Prevention (CDC): The CDC publishes the results of the Behavioral Risk Factor Surveillance System, the world’s largest, ongoing telephone health survey system. While research has shown the reliability of patient-reported measures to be good, states cannot be certain of the comparability of each measure relative to measures generated from other data sources.
- Medicare Hospital Compare: The CDC publishes benchmarks for hospitals using Medicare performance data, as well as for nursing homes (Nursing Home Compare).

States and state and regional quality improvement organizations have often created their own internal state benchmarks; these can also be a resource.

## Reviewing and Modifying the Measure Set

It will be important to develop a process for both ad hoc and regular periodic review of current measures to determine whether they should be retained or modified, or if new measures should be included based on changing circumstances or priorities.

Ad hoc measure review is necessary because changes in national clinical guidelines have direct impact on commonly used, nationally endorsed measures. For example, the new American College of Cardiology/American Heart Association guidelines issued in late 2013 on cholesterol management had significant impact on the LDL-C control measure employed in many measure sets.<sup>2</sup> As a result, many state and multi-payer/multistakeholder organization measure sets had to be modified based on the new guidelines.

Periodic measure set review should occur well in advance of the implementation of any measure set changes so that affected provider organizations will have adequate time to react. For example, the Oregon Health Authority created a calendar of planned measure review activities to inform affected provider organizations 60 days prior to their effective date. As with initial measure set development, a set of explicit criteria should be used to inform decision-making.

## Pitfalls in Performance Measurement

While there are important opportunities in performance measurement, it is also important to be mindful of the potential pitfalls. While performance measurement can serve to align goals and incentives, it has the potential to narrowly focus providers and health plans on aspects of care that are being measured, and especially so when the measure is tied a reward or penalty. This narrow focus could lead to unintended consequences, such as paying too little attention to other important health care components that are not being measured. One way to reduce this potential pitfall is to include both monitoring and incentive measures within a performance measurement set. Monitoring measures can be promoted to incentive measure status if performance slides.

As mentioned previously, the development of homegrown measures can be problematic for a number of reasons, including validity, reliability, and the inability to access a performance benchmark. As states try to measure social determinants of health as part of measurement initiatives, it is important to consider whether it is appropriate to hold health care providers accountable for things over which they have little or no control, such as education, environment, and poverty.

## Conclusion

In developing a performance measurement initiative, the state should consider how measurement can evolve over time. While there may be short-term limitations to the depth and breadth of measures that can be implemented, the consideration of a broader array of measures gives states a pathway for expanding their measurement set and increasing their options for incentives.

In addition to developing a measure set as part of a multipayer initiative—the state and its payer partners if in a multipayer initiative—should engage the participating providers to help them achieve success on these measures. While quality-based incentives offer providers extrinsic motivation to improve the quality of care and the health status of Medicaid beneficiaries, they are not sufficient. Providers must not only want to change, they must also know what and how to change in order to improve care. States and other payers will need to continue their efforts to actively manage health plans and providers, including setting strategic direction and providing ongoing performance review and support for quality improvement activities. They must also consider how to provide technical and data support to providers to ensure that measurement and other activity yield desired results.

## Endnotes

1. Bazinsky, Kate, and Michael Bailit. 2013. “The Significant Lack of Alignment Across State and Regional Health Measure Sets.” *Buying Value*. [www.buyingvalue.org/wp-content/uploads/2014/02/buying-value-common-measures-Bailit-State-Measure-Set-Brief-9-10-13-FINAL-FINAL.docx](http://www.buyingvalue.org/wp-content/uploads/2014/02/buying-value-common-measures-Bailit-State-Measure-Set-Brief-9-10-13-FINAL-FINAL.docx).
2. Stone, Neil, et al. 2013. “2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidance.” American Heart Association. doi: 10.1161/01.cir.0000437738.63853.7a.