ACCELERATING AND ALIGNING
POPULATION-BASED PAYMENT MODELS:

PERFORMANCE MEASUREMENT

Draft White Paper

Written by:

The Population-Based Payment (PBP) Work Group

Version Date: 4/22/2016

Table of Contents

Executive Summary 3
Health Care Payment Learning & Action Network (LAN) 3
Overview 3
Work Group Charge and Scope 3
Foundational Principles on Performance Measurement 7
Recommendations 10
Next Steps 23
Conclusion 23
Appendix A: Roster 24
PBP Work Group Co-Chairs 24
PBP Work Group Lead on Performance Measurement 24
PBP Work Group Members 24
CMS Alliance to Modernize Healthcare (CAMH) Staff 25
Appendix B: Compendium of Core Measure Sets 26
Appendix C: References 27

Executive Summary

[Placeholder: Formal abstract will be developed after incorporating feedback from the affiliate community.]

Overview

The Health Care Payment Learning & Action Network (LAN) established its Guiding Committee in May 2015 as the collaborative body charged with advancing alignment of payment approaches across and within the private and public sectors. This alignment aims to accelerate the adoption of alternative payment models (APMs) that reward quality and value in health care. The Centers for Medicare & Medicaid Services (CMS) Alliance to Modernize Health Care (CAMH), the federally funded research and development center operated by the MITRE Corporation, was asked to convene this large national initiative.

In keeping with the goals of the U.S. Department of Health and Human Services (HHS), the LAN aims to have 30% of U.S. health care payments in alternative payment models by 2016, and 50% by 2018. One possible form of qualifying APM is population-based payment, in which providers accept accountability for the cost, care quality, and health outcomes for a patient population across the full care continuum. This is a particularly promising approach for creating and sustaining a delivery system that values quality, cost effectiveness, and patient engagement.

Work Group Charge and Scope

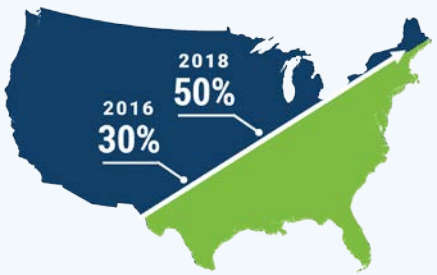
In October 2015, the LAN Guiding Committee convened the Population-Based Payment (PBP) Work Group (the PBP Work Group) and charged it with prioritizing methodologies and exploring alignment issues in support of the development, adoption, and success of population-based payment models under which providers accept accountability for a patient population across the full continuum of care.

APMs that hold provider organizations accountable for the total cost of care for a population across the care continuum are a particularly promising approach to creating and sustaining delivery systems that value quality, cost effectiveness, and patient engagement.

Health Care Payment Learning & Action Network (LAN)

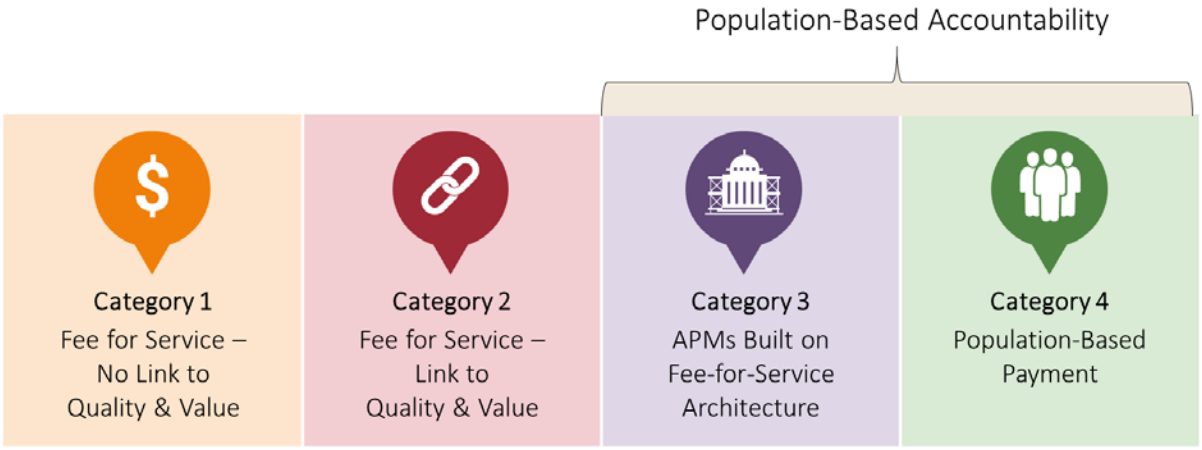
To achieve the goal of better care, smarter spending, and healthier people, the U.S. health care system must substantially reform its payment structure to incentivize quality, health outcomes, and value over volume. Such alignment requires a fundamental change in how health care is organized and delivered, and requires the participation of the entire health care ecosystem. The Health Care Payment Learning & Action Network (LAN) was established as a collaborative network of public and private stakeholders, including health plans, providers, patients, employers, consumers, states, federal agencies, and other partners within the health care ecosystem. By making a commitment to changing payment models, establishing a common framework, aligning approaches to payment innovation, sharing information about successful models, and encouraging use of best practices, the LAN can help reduce barriers and accelerate the adoption of APMs.

U.S. Health Care Payments in APMs



Many believe that these types of payment models have significant potential because they give providers more flexibility to coordinate and manage care for individuals and populations. Recommendations throughout this paper refer to APMs in which providers accept accountability for the full continuum of care can be classified in Categories 3 and 4 of the APM Framework, depending on how they handle financial risk for provider organizations. The principles and recommendations presented in this paper are directed toward that subset of APMs which, for the sake of convenience, are referred to as "PBP models" (Figure 1).

Figure 1: APM Framework (At-a-Glance)



Source: Alternative Payment Model (APM) Framework and Progress Tracking Work Group

Regardless of payment structure, all PBP models share one key element: They involve provider accountability for a patient population across the full continuum of care, including preventive care to end-of-life care and everything in between—with the goal of achieving better quality and outcomes and lower total cost for the population involved. The PBP Work Group determined that four priority issues are foundational for the success of population-based payment models. These include:

- Patient attribution;
- Financial benchmarking;
- Data sharing; and
- Performance measurement.

This paper focuses on performance measurement, which encompasses the development and implementation of metrics that assess the clinical quality, outcomes, patient care experience and cost of care provided to patients. Performance measurement can be used both for accountability and improvement purposes. Performance measurement makes it possible to monitor and quantify how well population-based payment models achieve and reward the Triple Aim of better care, better health, and lower costs. In conjunction with other elements in PBP models (e.g., financial benchmarking and patient attribution), performance measurement enables continuous quality improvement by highlighting those aspects of care delivery in need of optimization and by establishing clear performance goals. Performance measurement also provides a framework that patients, payers, purchasers, providers, and

other key stakeholders can use to productively collaborate to achieve their collective goals for patients, populations, and the health care system as a whole. For these reasons, the success of PBP models hinges on the proper implementation of measurement systems.

Although performance measurement systems hold great promise for improving the overall value of health care in the U.S., significant challenges currently prevent them from achieving their full potential. Available measures are typically not suitable for assessing how well a health system delivers care across the continuum; fragmentation in data reporting and quality measures is creating confusion and undue administrative burden; the data needed to evaluate outcomes that matter to patients are not systematically collected; and the national measure-development apparatus is not ideally configured to rapidly develop measures that will be needed in future PBP models. The challenges associated with performance measurement may be considerable, but the PBP Work Group and the LAN as whole believe strongly that alignment around common principles and approaches, across and within the public and private sectors, will accelerate progress toward unleashing the full potential of performance measurement to transform the health care system. To further this aim, the PBP Work Group deliberated at length and reached a set of baseline, consensual positions that recognize and defer to the roles and positions maintained by other multi-stakeholder groups.

This paper is grounded in the notion that payers, providers, purchasers, and patients should be collectively accountable for ensuring that the health care system delivers the highest possible value for patients and consumers. Moreover, the ideas articulated in the paper are predicated on the notion that the measures required for PBP models are fundamentally different from the measures used in traditional fee-for-service payment models. The existing measure sets are appropriate for use now and can contribute to significantly advancing quality, outcomes and cost while the more outcomes-oriented measure sets envisioned here are developed. This paper proposes sets of principles and recommendations on how to design measurement systems that can support the long-term success and sustainability of PBP models. It also addresses steps that must be taken to put into place the governance structure needed to prioritize and accelerate the development of measures that are suitable for PBP models.

Definitions

These are the key performance measurement terms used throughout this paper:

Population-Based Payment (PBP) Model: A payment model in which a provider organization is given a population-based global budget or payment and accepts accountability for managing the total cost of care, quality, and outcomes for a defined patient population across the full continuum of care. PBP models discussed in this paper correspond to payment models in Categories 3 and 4 of the LAN's APM Framework (refer to Figure 1).

Total Cost of Care (TCOC): A broad indicator of spending for a given population (i.e., payments from payer to provider organizations). In the context of PBP models, in which provider accountability spans the full continuum of care, TCOC includes all spending associated with caring for a defined population, including provider and facility fees, inpatient and ambulatory care, pharmacy, behavioral health, laboratory, imaging, and other ancillary services.

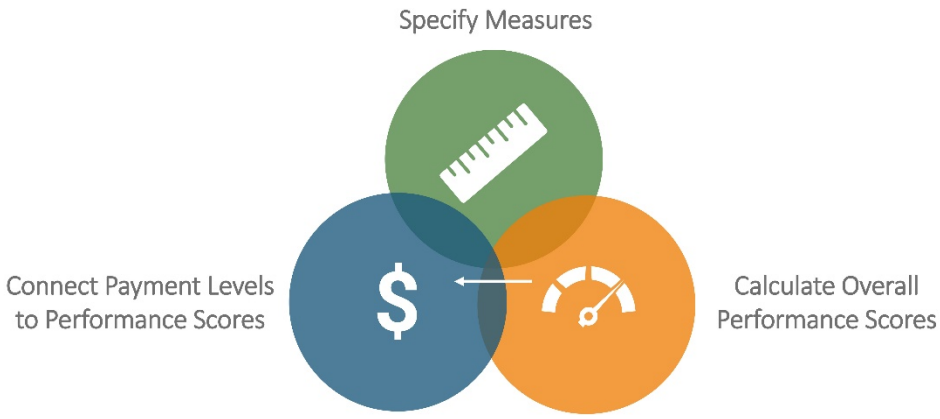
Full continuum of care: All aspects of care delivery, spanning preventive to end-of-life services in all settings.

Measure sets: A collection of measures that are mutually reinforcing to produce positive outcomes for a defined patient population.

The PBP Work Group conceives of measurement systems as being comprised of three components, each of which is needed to reward providers who deliver high-value health care: 1) measure sets; 2) methods for evaluating performance on measures (e.g., performance scoring); and 3) methods for using performance assessments to adjust payments (Figure 2). In other words, in order to reward providers who deliver high-value health care, measurement systems must necessarily specify measures, employ some method for calculating overall performance scores, and adopt some approach for adjusting payments in light of those performance scores. In order for measurement systems to function well, these three components must be designed deliberately to ensure they work properly together and are appropriate for a particular payment model.

By approaching the topic from the perspective of measurement systems and illuminating the idea that measures required for PBP models differ from those that have been useful under fee-for-service payment, the PBP Work Group hopes to chart new and fertile ground and drive alignments around broad-scale design elements in public and private sector PBP models—as confirmed by PBP-specific measurement systems.

Figure 2: Components of Measurement Systems



This paper does not seek to enumerate the measures that payers should use to evaluate accountable provider organizations participating in PBP models. For reasons discussed more fully later in the paper, measurement systems for PBP models should draw from existing core measure sets, but the paper intentionally does not advance recommendations on specific measures to be used.[1] Rather, the purpose of this paper is to recommend ways to design and implement measurement systems in PBP models, including important advance needed in the types of measures and the data and reporting infrastructure required for the long term success and sustainability of PBP models.

Understanding the incentives that motivate patients, providers, payers, and purchasers to participate in population-based payment models is paramount. Identifying the interests of different stakeholder groups (including states) is therefore critical for properly designing incentive structures and establishing their overall objectives. Keeping in mind that a single measurement system may not be able to meet all the interests of all health care stakeholders, the following list identifies some of the most critical

[1] A compendium of core measure sets available today can be found in Appendix B, and the Work Group encourages readers to consult those materials.

interests for each stakeholder in performance measurement.

- Patients have an interest in ensuring that the economic incentives in PBP models do not lead to underutilization of care or harm from over-treatment. Performance measurement is a vital tool for ensuring that patients receive the care they need through PBP models, and can assist in the selection of provider organizations and individual providers. As discussed at greater length later, using the proper types of measures and measure sets can help providers to focus on achieving outcomes that matter most to patients.
- Providers have a unique and particular interest performance measurement for multiple reasons. First, performance measurement is a vital tool through which providers can assess the quality and effectiveness of care, identify and address unexplained variations in practice patterns and outcomes, and continuously improve care provided to patients and populations. Providers also have an interest in performance measurement systems increasingly form the basis for accountability, both in payment and in public reporting of results.
- Payers have an important interest in performance measurement because it is foundational to new payment models, such as PBP models, that enable a shift away from incentives that reward volume rather than value in health care. To accomplish the goals of improved quality, outcomes, patient experiences, and cost, payers require measurement systems that address each of these. These measurement systems enable public and private payers to hold providers accountable for performance on these domains, and in turn, make payers accountable to purchasers and patients.
- Purchasers like patients, employers and public programs are interested in using performance measurement as one mechanism to ensure that the population is receiving needed care while not being over-treated with unwarranted services, and as a means of selecting plans and providers that they will make available for employees.

The PBP Work Group brings together public and private stakeholders, representing diverse constituencies. A roster of the PBP Work Group is provided in Appendix A. The PBP Work Group is aware that CMS is in the process of developing proposed regulations of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). MACRA is, explicitly and intentionally, not part of the Work Group's charge and therefore will not be addressed in this paper.

Foundational Principles on Performance Measurement

In order to ensure that performance measurement systems properly incentivize the delivery of high-value care in population-based payment models, the PBP Work Group established several principles that serve as a foundation for the PBP Work Group's thinking and approach to performance measurement. Although the principles focus on how to use performance measurement for provider accountability for providing the continuum of care for a population, they also consider the use of performance measurement in other contexts.

Four principles of performance measurement for PBP models:

- Foundational: Performance measurement is at the heart of PBP models' potential to advance the Triple Aim of better care, better health, and lower costs.
- Continuum of care: Measures for PBP models must cover the full continuum of care across time, providers and settings.
- Different from fee-for-service: Measures for PBP models must be more outcome-based than prevailing fee-for-service measures, which have focused largely on evaluation of specific care processes for individual conditions or care settings.
- Incentivizing care: PBP models must create meaningful incentives to improve health care quality, outcomes, and cost.

***Principle 1**: Performance measurement is foundational to the success of population-based payment models to advance better outcomes for patients and populations.*

Performance measurement makes it possible to monitor, in a transparent and quantifiable manner, how well PBP models achieve the Triple Aim of better care, better health, and lower costs—and also make it possible to reward performance through the incentive structures that are central to PBP models. Performance measurement also enables transparency regarding the overall results of PBP models, thereby enabling ongoing assessment and improvement of payment model designs. It also provides information on the results achieved by participating payers and providers, which can inform individual purchaser and consumer choices in the market and motivate ongoing improvement. Performance measurement enables continuous quality improvement by highlighting aspects of care delivery in need of optimization and by establishing common performance targets or goals for specific dimensions of care. Finally, performance measurement provides a framework that patients, payers, purchasers, providers, and other key stakeholders can use to collaborate productively toward the achievement of their collective goals for patients and the health care system as a whole.

The PBP Work Group used the framework from the Institute of Medicine's report *Vital Signs: Core Metrics for Health and Health Care Progress* (National Academies Press, 2015) in its discussions of performance measures and a measurement system for population-based payment. The domains, which include healthy people, quality care, affordable care, and engaged people, map well to other frameworks such as the CMS National Quality Strategy.

***Principle 2**: Because population-based payment models address the full continuum of care, measure sets have to span the full continuum across time, across providers, and across settings.*

The measure sets that are in widespread use today have arisen from, and have been largely rooted in, fee-for-service payment models. As a result, these measures have focused primarily on evaluating care within a particular setting or specialty, rather than measuring how a system as a whole performs across

the health care continuum. By contrast, many PBP models require system accountability (i.e., accountability that is broader than a single provider or specialty) and hold providers accountable for the full continuum of care. Therefore, measure sets used in these PBP models must cover the full continuum of care, meaning both specialty and primary care, as well as hospital and post-acute care. Measures for PBP models must also cover prevention and well-being as well as therapeutic services, and these services must span from the beginning to the end of life.

The PBP Work Group recognizes that despite the vast array of performance measures that exist today, there remain many significant gaps covering important areas of care for which there are no accepted or widely used measures today. The PBP Work Group also recognizes that the development of numerous measures to address these gaps could compound the challenges of "measure cacophony" that exist today—i.e., the fragmentation and burden associated with an abundance of overlapping measures that are difficult to reconcile and interpret. As discussed in the next principle, one way to minimize measure cacophony is to utilize measures that are different than those traditionally used in fee-for-service models. Furthermore, as discussed in Recommendations 2 and 3, additional and directed innovation in measurement science and development should aim to address critical gaps while also advancing the goals of a parsimonious, person-centered, outcomes-focused measurement system for use in PBP models.

***Principle 3**: The measures required for the long-term success and sustainability of population-based payment models are fundamentally different from the measures used in traditional fee-for-service payment models.*

Unlike the measures used in fee-for-service models, measures for PBP models can be more "macro" in their orientation and more outcomes based, rather than evaluating performance on narrow and specific care processes, conditions, or care settings. The PBP Work Group holds that existing core measures sets can support a strong transition to population-based payment models, but that during this transition, the development, testing, and adoption of more outcomes-oriented measure sets for the long-term sustainability and success of PBP models must be accelerated.

The development of core measure sets arose in response to an ever-increasing volume of performance measures, which in turn arose in response to an environment in which fee-for-service payment predominated. In that context, measurement that addresses specific care processes for each specialty, condition, and care setting was appropriate. By contrast, with providers and payers in PBP models accepting accountability for producing better care, better health, and lower cost across the full continuum of care, the measures required to assess and reward performance must be more outcomes oriented, and measured at the system level. This paper considers how to accelerate progress toward this form of measurement, and how doing so may alleviate some of the problems of measure cacophony that have arisen when a more atomistic measurement approach was required for performance-based payment.

***Principle 4**: To promote better results for patients and populations, the use of performance measurement for payment in PBP models must create meaningful incentives for improvement.*

The PBP Work Group believes that an important function of performance measurement in PBP models is to ensure that the population is receiving needed care while not being over-treated with unwarranted services. For this reason, it is especially important for purchasers and consumers to know that measurement systems are in place to monitor and hold providers accountable for providing quality care and health outcomes. Along similar lines, performance measurement systems must not incentivize providers to "play to the test" and neglect unassessed areas of care. A second and related function of performance measurement systems in PBP models is to motivate providers to continuously improve the quality and outcomes of care. Performance measurement systems will not achieve their goal of supporting progress toward the Triple Aim of better care, better health, and lower cost if providers do not see them as tools to improve care. In addition, the measures that PBP models use to reward performance must also create transparency about how well the models are performing in delivering the promised results of better care, better health, and lower cost, and allow purchasers and consumers to evaluate the performance of individual payers and provider systems on these dimensions.

Recommendations

In preparing this paper, the PBP Work Group included public and private stakeholders from many different perspectives related to measurement. Documents consulted during the development of this paper are referenced in Appendix C. The PBP Work Group's recommendations are intended to help measurement systems properly incentivize and reward accountable provider organizations that participate in PBP models.

***Recommendation 1:** To support the long-term success and sustainability of population-based payment models, future-state measures must be based as much as possible on results that matter to patients (e.g., functional status) or the best available intermediate outcomes known to produce these results.*

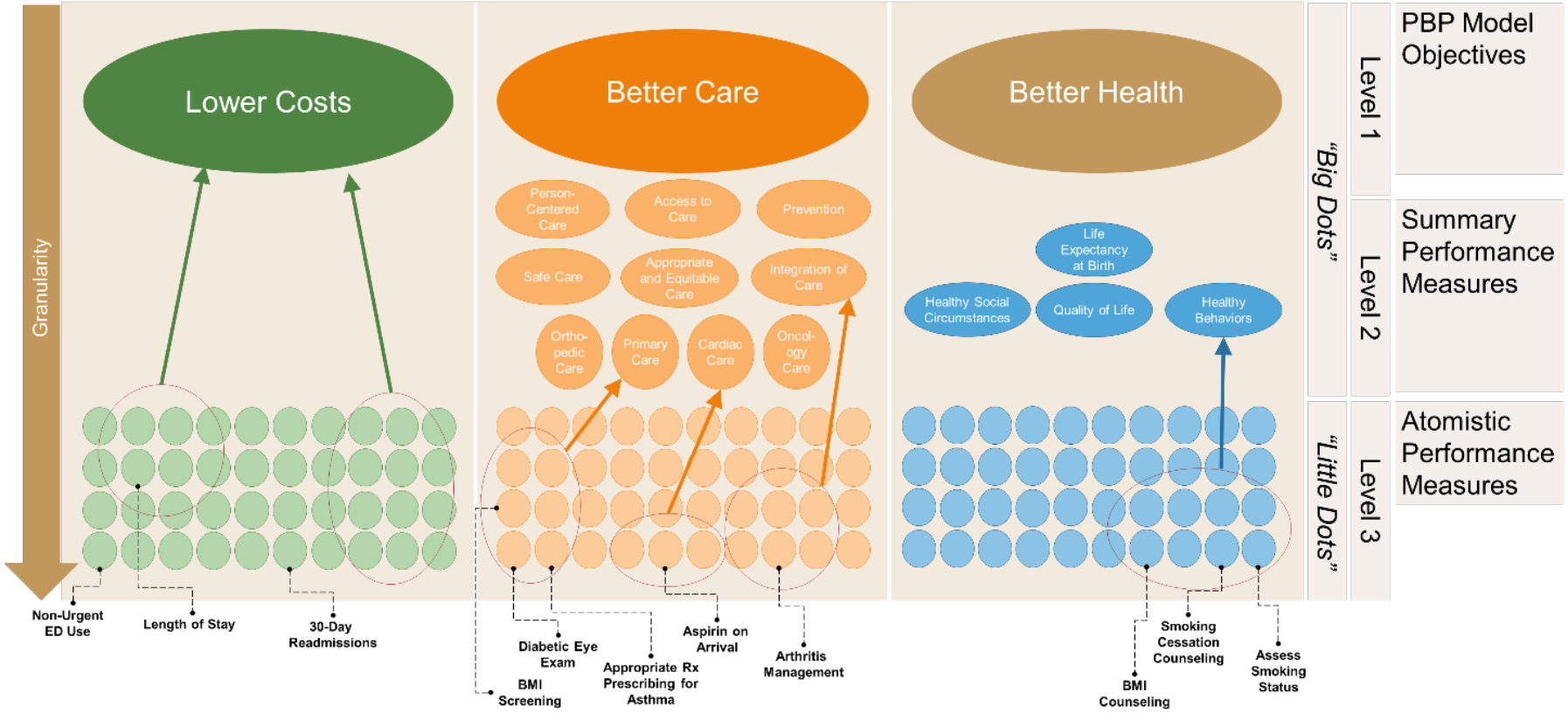
Given that the accountabilities that providers assume under PBP models differ fundamentally from those assumed under traditional fee-for-service payment models, the measures appropriate for use in PBP models are different from those that have been useful under fee-for-service payment. Specifically, under fee-for-service payment, accountability rarely, if ever, spans care settings or specialties. As a result, measurement was focused on individual care settings and/or specialties, and largely evaluated processes of care occurring in each.

By contrast, PBP models establish provider accountability for care across the continuum—from preventive care to end-of-life care and everything in between-and the unit of accountability is the provider organization or system. Thus, the measures suited to these models must address overall system performance and be oriented increasingly toward assessment of outcomes, not the processes

used to produce them. Indeed, given PBP provider accountability for care across the continuum, if the measurement systems that assess provider performance aimed to capture and incentivize all of the relevant care processes, the problems of measure cacophony that plague the system today could grow much worse. By contrast, moving toward a measurement system that is focused predominantly on the results being produced would require a more limited number of measures, although more granular process of care measures would continue to be useful for providers engaged in quality improvement activities. This move away from indicators of care processes in individual care settings and specialties toward system-level outcomes is referred to as a move toward "big dot" measures.

For the purposes of this paper, big dot measures assess the overall system performance based on the outcomes produced, not the processes used to produce them. Perhaps the best illustration of a big dot measure is total cost of care (TCOC). As described in the draft Financial Benchmarking White Paper, TCOC is a fundamentally important measure in PBP models, as it forms the basis for provider accountability on the cost side of the model. Although providers in PBP models may need to have information about many "little dot" cost measures to appropriately manage TCOC, the accountability measure in the cost domain is truly a big dot measure. Figure 3 illustrates this concept, with TCOC represented as a **Level 1** big dot measure, and the more atomistic cost measures of that providers may use to monitor and drive improvement on TCOC situated at **Level 3**.

Figure 3: Measures by Purpose Area



By contrast, there do not exist today **Level 1** measures of quality care or health that are analogous to TCOC. In fact, current measures of quality care and health are largely confined to the atomistic Level 3 types of measures that have arisen under fee-for-service payment models. Although the PBP Work Group acknowledges that defining a Level 1 global measure of quality care or health—analogous to TCOC on the cost side—is challenging, the PBP Work Group believes that significant progress can and should be made to develop Level 2 measures in the quality care and health domains and that these should form the basis for performance incentives in PBP models.

These **Level 2** big dot measures of quality care and health can take two different forms: (1) condition- or specialty-specific measures; or (2) cross-cutting and focused measures that address goals not connected to a particular condition or type of practice. The first category aligns with the types of measure sets recently release by the CMS/America's Health Insurance Plans (AHIP) Core Measure Collaborative, which included core sets on cardiovascular care, medical oncology, and other condition- and specialty-specific topics. The second category is consistent with the approach taken in the Institute of Medicine *Vital Signs* report, which outlines potential care measures on prevention, person-centered care, safety, and other cross-cutting topics.

"Big dot" which include **Level 1 & 2** measures, assess the overall system performance based largely on the outcomes produced, rather than on the processes used to produce them.

The PBP Work Group believes that, whenever possible, Level 2 measures used in PBP models should be outcomes-based. For example, the Level 2 measure of cardiac care would ideally comprise outcome measures reflecting the results achieved in this domain (e.g., 30-day mortality, health-related quality of life or well-being), not the individual care processes used (e.g., aspirin on arrival). Although the PBP Work Group believes that existing measures are sufficient to support a transition to PBP models, the PBP Work Group also recognizes that, at present and for the reasons outlined earlier, available measures and measure sets consist primarily of process measures, not outcomes. Thus, in the near-term, the PBP Work Group envisions that Level 2 measures will primarily be aggregates of the more atomistic **Level 3** measures, which may be largely process-based. Over time, however, and through the approaches outlined later in the paper, the PBP Work Group believes that PBP models can and should come to employ Level 2 measures that are primarily outcomes-based. This would represent a significant departure from historic performance incentive models, which have rewarded excellence in care processes. By incentivizing and rewarding the achievement of favorable health outcomes, PBP models will incentivize care that does, indeed, translate to better health.

It is important to note that purchasers and consumers will likely want performance information that is more granular than the system-level information required for evaluating and rewarding providers in PBP models. For example, while a system's outcomes in oncology care represent a desirable and appropriate Level 2 measure for financial incentives in a PBP model, the data that individual patients and consumers require to inform their care choices will be more granular—requiring information about the system's performance on specific types of cancers (e.g., breast, lung, prostate) and likely also requiring performance information for practice groups or even individual clinicians within the system. Further discussion about the development and use of Level 2 measures later in this paper seeks to illuminate how these dual purposes—measurement for payment and measurement for consumer choice—can be addressed.

Finally, and importantly, the PBP Work Group recognizes the fact that the more granular Level 3 measures in the cost domain (e.g., 30-day readmissions, ambulatory care sensitive admissions), while

vital information for provider systems to improve TCOC performance, should not play a prominent role in PBP quality incentive measure sets. Doing so would represent double-paying for the cost savings associated with these measures. That is, in PBP models structured to reward savings on TCOC (i.e., shared savings), success on Level 3 cost measures like readmissions are already incentivized and rewarded. Including such measures in PBP quality incentive measure sets would represent paying twice for the same results.

***Recommendation 2**: Because fragmentation across population-based payment models can undercut success, reliance on core measure sets is valuable. Continued innovation and refinement are needed to ensure measure sets are comprehensive, parsimonious, and outcome oriented.*

At a high level, core measure sets establish overarching domains as well as the measurement priorities within those domains. Core sets serve a critical function in measurement systems by driving alignment in the measures used by different payers and providers. This reduces providers' burden of reporting different measures to different payers, while simultaneously strengthening meaningful comparisons among different providers' performances.

In the near-term, available core measure sets should be used in PBP models. Despite the goal of having more outcomes-oriented measure sets form the basis for PBP models, the PBP Work Group strongly maintains that the use of existing measure sets will enable important and meaningful progress to improve quality. While largely composed of condition- or specialty-specific clinical process measures, available core sets for priority specialties and conditions can and should initially be used to form the basis for establishing Level 2 measures of these important clinical quality issues. Using the CMS/AHIP core sets would be one example of how PBP models could implement aggregated Level 2 measures. Many PBP models in place today use similar approaches and measure sets to establish overarching composite scores that are used for payment. However, in order to realize the vision of Level 2 outcome measures that evaluate more the extent to which clinical care results in better health outcomes, significant innovation will be required. The PBP Work Group maintains that models for accomplishing this innovation in outcomes measurement exist and should be leveraged to realize this vision. Three such models are highlighted.

International Consortium for Health Outcomes Measurement. First, the International Consortium for Health Outcomes Measurement (ICHOM) represents one worthy example to consider. ICHOM's work is predicated on the value of alignment in performance measurement, and it focuses exclusively on the measurement of outcomes that matter to patients, such as readmissions, pain, and health-related quality of life. For each health condition examined, ICHOM considers three broad categories of outcomes: (i) acute complications of treatment, (ii) patient-reported health and functional outcomes (PROM), and (iii) disease control and survival. The resulting measure sets employ a combination of clinically-rich data from health records or registries, administrative data, and patient-reported data to establish a holistic view of the results achieved for patients with a given condition. Importantly, when a given dimension of functional status or wellbeing is relevant to more than one condition, ICHOM's proposed tool for measurement of that dimension is the same across conditions. For example, for all conditions in which pain is a relevant dimension to be measured, the same patient-reported outcome tool for assessing pain is recommended; and same is true for physical functioning, emotional wellbeing,

and other domains of patient functional health status. This is important in that a condition-specific approach to measurement could otherwise exacerbate measure cacophony and disrupt the goal of a holistic view of the patient and a whole-person approach to care. ICHOM's condition-specific framework and process for developing outcome measure sets is worthy of consideration even though the PBP Work Group recommendation that PBP models ultimately need to include a combination of condition-specific and cross-cutting measures (e.g., access, person-centered care, safety, mortality, equity).

National Clinical Registry Network. Second, the work of the National Clinical Registry Network (NCRN) illustrates the potential for both the systematic development of data infrastructure needed to standardize measurement of clinical outcomes for specific conditions and procedures, and the use of these measures to improve patient care. The Society of Thoracic Surgeons' (STS) National Database is an exemplar of the value that such clinical registries can provide. The STS registry was established in 1989 as an initiative of cardiothoracic surgeons seeking to improve the safety and outcomes of care. The registry affords cardiothoracic surgeons nationwide a standardized format for collecting a set of data elements required to systematically measure and compare surgical outcomes. The system employs robust risk adjustment and benchmarks that enable comparison across providers and over time, and that form the basis for sharing best practices and motivating continuous quality improvement. Moreover, since 2010, the STS has facilitated the public reporting of results of surgical quality and outcomes, including for procedures such as coronary artery bypass graft (CABG) surgery and aortic valve replacement (AVR). The work of the STS and others within the NQRN could contribute importantly to the potential for incorporating clinically rich outcome measures for priority conditions and procedures into PBP models.

Yale Center for Outcomes Research & Evaluation. Third, given the increasing interest in and acceptance of PROMs and their potential value in PBP models, recent work by Yale Center for Outcomes Research & Evaluation (CORE) team is also illustrative. CORE's work to develop performance measures related to total hip and knee replacement provides a rich example of the steps required to develop a performance measure that employs data from a well-validated patient-reported outcome tool. There are many such tools today, often based on significant literature from clinical trials and health services research demonstrating their usefulness in evaluating important dimensions of patient health status and functioning. However, the use of these instruments for accountability purposes, such as in a PBP model, requires a number of additional methodological steps that have generally not been accomplished and cannot be overlooked. These include:

(i) Establishing the appropriate time intervals for administering the survey in order to standardize the measurement of change in patient status (e.g., number of days/weeks before and after a procedure);
(ii) Defining how much a measurement must change to be considered a clinically meaningful outcome (i.e., improvement or decline);
(iii) Identifying the appropriate risk adjustment model; and
(iv) Establishing the sample sizes required for reliable measurement at the clinician, group, or system level.

Summary of Needed Innovative Models

It is important to note that although these exemplar innovation models are condition-specific, the PBP Work Group believes that outcomes accountability will foster significant delivery system transformation, necessitating more whole-person, patient-centered approaches to care. To be successful in managing

and improving clinical and functional outcomes requires a different type of engagement with patients and communities and will challenge providers to transition away from traditional fee-for-service delivery models, which are focused largely on what happens in clinical offices, hospitals, and other delivery system settings. With accountability for outcomes, providers must think more holistically about patients, including their life circumstances and the social determinants that will support or impede the achievement of care goals. For example, in order to help patients keep chronic conditions (e.g., diabetes and hemoglobin A1C) under good control or achieve favorable functional health outcomes (e.g., reduced pain, and improved physical and mental status), providers must think beyond the practice setting. PBP models in place today, with accountability for outcomes, have demonstrated that to function effectively within a PBP environment, providers must transform their approaches to care—using new staffing models, new approaches to integrating care and prevention within and across settings, new uses of health information technology, new ways to engage patients and community resources, and new collaborations with community stakeholders. The PBP Work Group therefore sees accountability for the types of condition-specific and whole-person outcomes proposed for PBP models as an important step toward a delivery system that reaches beyond the confines of the clinical setting to significantly improve the health of the population.

The aforementioned models are useful illustrations of innovation and measure development processes that can help meet the goal of establishing outcomes-oriented PBP measure sets. The PBP Work Group believes that, in order to make meaningful progress toward this goal in the next 12 to 24 months, it will be necessary to agree upon a set of priority topics for outcomes measure development, and to define a process by which to accelerate the development, testing and adoption of these measures. While a number of processes currently exist through which priority measure gaps are identified, none of these is specifically aimed at addressing the needs of PBP models, and importantly, none is designed with the explicit goal of seeking alignment among payers, providers, purchasers and patients as a starting principle and priority. A model through which this might occur is presented later in this paper.

***Recommendation 3**: A governance process is needed to oversee and accelerate the development, testing, and use of new, high priority measures for population-based payment models.*

Performance measures and measure sets in use today have arisen from a vast and varied set of measure developers, each seeking to ensure that priority gaps are filled in the national portfolio of quality measures. Yet, as emphasized throughout this paper and elsewhere, stakeholders today agree that the field of performance measurement has both delivered too many measures and yet left significant gaps. Among the challenges is the need for a process that can be used to gain consensus among public and private sectors on the highest priority gaps to be filled, and a streamlined connection to another process that systematically develops, tests, and enables adoption of measures that meet these needs. The existing attenuated and uncoordinated processes for measure development will need to be dramatically accelerated to produce the types of measures needed for the long-term success of PBP models.

Setting Priorities for Measure Development

The establishment of clear and well-founded priorities for measure development is a necessary first step. The PBP Work Group agrees with the Institute of Medicine Vital Signs report that the Secretary of

Health and Human Services should lead the effort to establish measurement priorities. The PBP Work Group also believes that public-private partnerships, such as the LAN and the AHIP/CMS Core Quality Measures Collaborative, are a useful platform for establishing public-private sector consensus and alignment on priorities for measure development.

In establishing priority measures for development, testing, and implementation of PBP measures over the next 12 to 24 months, the PBP Work Group urges the Secretary and these multi-stakeholder partnerships to identify a set of high-prevalence and high-impact conditions for which systematic outcomes measurement is expected to meaningfully improve results for patients—for example, potentially accelerating and expanding upon current outcome measure development efforts for high prevalence musculoskeletal conditions, cardiovascular disease, and common cancers. A roadmap for systematically developing outcome measures corresponding to the national burden of illness by a certain time, similar to ICHOM's goal of establishing measure sets that cover 50% of the global disease burden by 2017, would represent a significant step. For each condition, considering a comprehensive outcome measure framework that would address acute treatment complications, patient-reported functional outcomes, and disease control and survival is recommended. At the same time, the PBP Work Group urges the Secretary and public-private partnerships to prioritize the development of cross-cutting measures that will not be addressed through a condition-specific approach, and that are critical to the success of PBP models. These include measures addressing access to care, integration and coordination of care within and across settings, and reduction in health disparities. Additional priorities to be considered for accelerated measure development in the immediate term are measures that evaluate how well care is matched with patient goals and measures evaluating the quality of advanced illness care (e.g., palliative care). These cross-cutting areas of measurement would enhance the assessment of specific priority conditions noted earlier and also contribute meaningfully to improving end-of-life care, which is recognized as a pressing national priority.

Accelerating Measure Development and Use

In addition to the establishment of priorities for measure development, the vision outlined herein calls for an infrastructure that will enable an accelerated and coordinated process for the development, testing, and adoption of these new measures. In light of its mandate under MACRA to develop measures, and its authorized budget of $15 million dollars per year from 2015-2019, CMS is the mostly likely source of funding for measurement innovation. The PBP Work Group therefore recommends that the Department of Health and Human Services, in close collaboration with private sector stakeholders, oversee the coordination of measure development and testing.

One manner in which this could proceed would be to establish a national network of qualified measure developers with proven credentials who would be eligible to bid on priority measure development projects. Such a system would require measure developers to demonstrate not only that they have the methodological expertise to develop, validate, and refine high priority outcome measures, but also that they are able to engage a broadly representative set of providers, patients, and other stakeholders to participate in the conceptualization and testing of new measures. As outlined later in this paper, the broad testing of a new measure in varied settings and populations is critical both to affording the necessary breadth of data for adequate psychometric testing and validation, and to gaining provider acceptance of a measure through its socialization and use. Indeed, APMs themselves may offer an important arena for testing new measures or measure sets. Therefore, the PBP Work Group envisions that measure developers would be evaluated on the extent to which their measures are widely adopted once they have been deemed ready for use. The establishment and oversight of a network of qualified measure developers could also help ensure coordination and alignment among these measure

development efforts and the resulting measure specifications. Further coordination would be achieved through the bidding process itself, which would use exclusive contracts to avoid duplications in effort.

Although the approach outlined here would go a long way toward fast-tracking measure development for high-priority topics, the PBP Work Group believes that this arrangement should not quash innovation that is taking place in other areas. For example, critical measure development work takes place via payer-provider partnerships and regional multi-stakeholder collaboratives. Examples include efforts to significantly advance PROMs occurring through the Pacific Business Group on Health's partnership with providers reporting data to the California Joint Replacement Registry, as well as Minnesota Community Measurement, which is measuring statewide implementation of PROM for effective treatment of depression. Further, the PBP Work Group expects that national organizations that have historically played a central role in measure development—including the National Committee for Quality Assurance (NCQA), the Joint Commission, and medical specialty societies—will continue to be important players in the development and testing of priority measures for PBP models. The PBP Work Group also envisions that the National Quality Forum (NQF), which plays a unique role in the process of certifying measures as ready and appropriate for accountability purposes, such as payment and public reporting, will continue to do so. In addition, NQF has recently launched a Measure Incubator program, designed to assist in the goal of accelerating measure development, which may be leveraged in the processes outlined earlier.

***Recommendation 4**: In service of a future state that employs measures that are outcomes-oriented, the infrastructure nationally must be sufficient to systematically collect, use, and report clinically rich and patient-reported data.*

The U.S. does not presently have the national infrastructure required to capture the types of clinically rich and patient-reported data needed for a comprehensive portfolio of outcome measures envisioned for PBP models. Setting aside the challenging task of how to increase the interoperability of health information technology, electronic health records in use today often do not collect data in structured or standardized ways that can afford truly meaningful use for broad and comprehensive clinical or health outcomes measurement. And while the availability of clinical data registries is growing, the current state cannot meaningfully advance the ability to rely on these systems for use in payment under PBP models. These aspects of the clinical and health data infrastructure significantly limit the development and implementation of more comprehensive sets of standardized clinical and health outcome measures as well as the potential for accessing the data elements that will be required for appropriate risk adjustment of these measures. The PBP Work Group believes that mitigating these issues must be part of the process of funding and accelerating measurement innovation.

Issues associated with data capture and use are compounded for PROMs of functional status, pain, emotional well-being, sleep, and other topics related to health-related quality of life. As noted previously, clinical trials and health services research have validated and relied on these tools for decades. And through the Patient Reported Outcomes Measurement Information System (PROMIS) initiative, the National Institutes of Health possesses a robust and scientifically sound collection of PROMs. However, care-delivery processes are rarely designed to systematically elicit this type of information from patients, and electronic health records are almost never configured to collect the resultant patient-reported data that are needed to calculate these measures. In addition, the cultural

barriers (e.g., clinical skepticism about the value of PROMs), professional/evidence barriers (e.g., the lack of understanding of the measures and the lack of an evidence base for improving results), and implementation barriers (e.g., how to use these tools in front-line provider offices) must be addressed. Part of the value of the collaborative models of developing PROMs-based performance measures—such as those employed by Yale's CORE, ICHOM and multi-stakeholder collaboratives described previously—is that they engage a broad and diverse set of provider organizations in measure conceptualization, implementation, testing, and refinement as integral to the process of measure development. Through this process, they accomplish both the important task of broad data collection needed for psychometric testing and analytics, and the task of broadly socializing the measure among providers and gaining feedback to inform refinements. In this way, at the conclusion of the process, the measure has both the empirical basis and professional acceptance to proceed to broad adoption and use.

Because patients and consumers care deeply about the outcomes captured by PROMs, the PBP Work Group maintains that these measures should be well represented in measure sets used to evaluate provider organizations in PBP models. Accordingly, the PBP Work Group believes that the practical and technological barriers to systematically capturing and interpreting patient-reported outcome data are in need of urgent remediation and should be an integral part of the funding and innovation model outlined earlier.

The Comprehensive Primary Care Initiative, sponsored by the CMS Innovation Center, requires practices to directly report quality measures from their EHRs. And the new Comprehensive Primary Care Plus initiative will also use patient-reported outcome measures and engage EHR vendors with the goal of incorporating the required data elements to report the PROM directly from the EHR. These programs may help stimulate the development and adoption of the types of data and reporting infrastructure outlined earlier in this paper.

***Recommendation 5**: Providers in population-based payment models should have meaningful incentives to deliver high-quality care, achieve favorable outcomes, and manage the total cost of care.*

Measurement systems used to measure performance should complement patient and provider goals for care delivery. Financial rewards for high performance can be structured in many different ways, but they should ensure that providers are incentivized to meet the Triple Aim of better care, better health, and lower costs for their patients and populations. For example, achieving a certain level of performance may be a prerequisite for participating in shared savings. It is also possible to establish an aggressive financial benchmark and then award incremental bonuses above that benchmark for different levels of performance measure attainment. The PBP Work Group does not believe that there is sufficient evidence to recommend one approach to rewarding high performance over another. Nevertheless, the PBP Work Group maintains that these rewards must be structured such that they enable, and do not impede, providers' capacity to provide better care and better health at lower costs.

***Recommendation 6**: Measurement systems should define performance targets in a way that motivates ongoing improvement across the performance continuum, promotes*

best practice sharing, avoids a forced curve that mandates winners and losers, and enables long-term planning and commitment to improvement.

Measure targets (i.e., the performance rates that accountable provider organizations need to achieve to receive different levels of rewards and penalties) can be set in a number of different fashions, and the manner in which performance targets are set carries significant implications for the activities they motivate.

The PBP Work Group does not believe there is sufficient evidence to recommend a detailed technical approach; however, the PBP Work Group maintains that measure targets should appropriately incentivize all provider organizations to improve, such that low performers receive strong incentives to improve rapidly and high performers receive strong incentives to maintain and/or raise high standards of quality. The following two sub-recommendations offer a blueprint for how measure targets can accomplish these goals.

***Recommendation 6a**: Whenever possible, measure targets should be set in absolute (not relative) terms, established prior to the measurement period and fixed for a minimum of one year, although ideally for the full contract term.*

Setting measure targets in "absolute" rather than "relative" terms is advantageous in performance incentive models for a number of reasons. Absolute measure targets specify a predetermined measure score (or range of scores) that an accountable provider organization needs to achieve in order to receive rewards or penalties in a PBP model. The absolute performance targets are based on criteria that demand that the target score represents high performance in real terms, not just relative to others' performance.

By contrast, relative performance targets are based on the performance of other providers, such as the 50th or 75th percentile of a regional or national distribution. The PBP Work Group believes that, where permitted under federal statute, absolute measure targets are preferable to relative measure targets for several reasons. First, because relative measure targets place provider organizations on a forced curve, they can create an environment that discourages collaboration and the sharing of best practices. Second, the use of absolute measure targets facilitates planning for quality improvement initiatives, because a target defined in absolute terms does not change from year to year. In other words, absolute performance scores by definition ought to represent high performance in real terms, and thus should not need to be re-set annually based on a changing distribution of scores.

Additionally, the PBP Work Group believes it is critical to establish measure targets as far ahead of the performance period as possible, because this facilitates planning and establishes certainty about provider expectations. Ideally, absolute targets should remain constant throughout the entire contract cycle; at minimum, targets should be fixed for an entire year, and under no conditions should the target change during the course of the performance period.

***Recommendation 6b**: Measure targets should include a range of scores on each measure to enable the incentive system to reward both performance and improvement.*

Once measure rates are evaluated against performance targets, the manner in which the resultant performance assessments impact payments can be structured in a variety of ways. For example, a single performance score can be used to establish a performance target or "gate" that must be cleared before an accountable provider organization can receive rewards. Alternatively, it is possible to establish a range of performance targets, and a corresponding series of incremental payment adjustments. The PBP Work Group believes that, where permitted under federal statute, the latter approach is preferable to the former because a single performance target does not accord with good behavioral economic principles. With everything riding on a single number, those whose performance misses the mark by even a fraction of a point get nothing, and those who surpass the number have no incentive to work toward further performance improvement.

Another consideration is the connection between performance scores and cost savings. PBP models may reward quality and cost savings separately, or they may link them. When treated separately, there is a discrete quality bonus, irrespective of the provider's performance relative to their financial benchmark. Some view this approach as undesirable because it can undercut the goal of focusing attention on managing spending, because quality bonuses can be achieved regardless of cost performance. Models that link quality and savings can do it in various ways. In most models where the two are linked, the quality score determines the magnitude of shared savings or deficit. Generally, the models that link quality and savings are structured such that a higher quality score is always advantageous, yielding a larger share of savings if a provider meets their financial benchmark, and yielding a smaller share of deficit if a provider overspends their target.

In some models, the quality score has an additional function, which is to determine the amount of a discrete quality bonus. The PBP Work Group believes that the public and private markets are evolving toward a model with a quality score that drives risk share without a separate quality payment. However, one might propose that additional payments for quality could be made if credible evidence supports that cost savings would occur in a reasonable timeframe. The Medicare accountable care organization (ACO) programs are examples of PBP models that use quality as one of many factors to determine the amount of shared savings. And the Blue Cross Blue Shield of Massachusetts' Alternative Quality Contract (AQC) is an example of a PBP model that uses the quality score for two purposes: to calibrate the provider's amount of shared savings or deficit and to determine the discrete quality bonus amount. These developments highlight the importance of monitoring emerging evidence on the results achieved under these varying incentive designs in order to begin to identify best practices.

***Recommendation 7**: Adherence to good measurement science and implementation (e.g., sample size requirements, demonstrated reliability and validity, national acceptability, clinical importance, and the opportunity for a provider to improve before being held accountable under the new model) is critical to achieving the desired results from performance measurements in population-based payment models.*

The PBP Work Group believes that the measures used to hold provider organizations accountable for the care they deliver must be based on rigorous measurement science and reliably differentiate provider performance. The PBP Work Group also believes that empirical evidence regarding the performance of measures themselves (and their intended and unintended consequences) increases most rapidly when they are put into widespread use. For example, a newly developed measure may have strong evidence of sound psychometric properties, but the measure development process generally involves measurement of a narrow sample of providers. Only once the measure is used more broadly can a more thorough assessment of the measure be conducted. It is analogous to what is known about a medicine after a clinical trial versus what is known after the medicine has been in widespread use.

Accordingly, and in light of the urgency of accelerating progress toward a more outcomes-based measure portfolio for use in PBP models, the PBP Work Group recommends the consolidation of what is normally a protracted and phased approach to measure development. Specifically, measure development might typically proceed along the following three phases:

- Phase 1: The initial development of the measure, typically based on a limited set of providers.
- Phase 2: The broader implementation of the measure for a much larger set of providers, which affords the data needed to gain deeper insights about the measures' properties, to gain feedback from providers that can inform refinements to the measure, and to allow providers the opportunity to improve on the measure before it is used for payment purposes.
- Phase 3: The stage at which the measure is ready for use in a payment model or other accountability purposes.

As previously described, the PBP Work Group envisions a process whereby these steps are accelerated—most likely, either by drawing upon measures that have already completed Phase 1 and accelerating their purposeful application in Phase 2 testing, or by consolidating the first two phases into a single process. It is also possible that, in some cases it may be valuable to incorporate a new measure into PBP programs on a "reporting-only" basis as a means to gather the broad types of data and experience with the measure as is ascribed to Phase 2. In each case, the result will be measures for which, over a period of 12 to 24 months, the PBP Work Group is able to establish both the strong empirical evidence and professional acceptance required for to move to Phase 3 adoption.

Given the intention that PBP models will comprise outcomes-oriented measures, proper risk adjustment will be essential. Development and testing of adequate risk adjustment models will be critical components of Phases 1 and 2, and must be assured before measures move to Phase 3.

Once in place, the measure would then need to be further monitored for unintended consequences and additional opportunities for refinement. The benefits of this approach are fourfold:

1) Additional data could be captured at each stage of the process, which measure developers can use to account for unanticipated scenarios in the measure calculations, and to further refine risk adjustment models.

2) Individual providers would have more opportunities to provide feedback on the measures and recommendations for improvements.

3) Accountable provider organizations would have enough time to become accustomed to the measure's results and reporting requirements before they are held accountable for their performance.

4) The process of ensuring that a measure's financial implications are matched by its demonstrated reliability and validity would be simplified.

Next Steps

- [Placeholder: The following section will be expanded when the PBP Work Group integrates feedback submitted during the public comment period. All stakeholders are invited to comment on the proposed next steps.]
- To accelerate progress toward the vision outlined in this White Paper, the PBP Work Group views the following as priority steps that could be taken in the next 12-24 months:
- Payers should proceed with PBP models, using existing core measure sets for population-based accountability. (Recommendation 2)
- HHS should work in collaboration with public-private partnerships to establish immediate priorities for measure development for PBP models—focusing on outcomes measurement for a set of high prevalence, high impact conditions, as well as important cross-cutting measures. (Recommendation 3)
- HHS should work in collaboration with public-private partnerships to establish a national network of qualified measure developers that can develop measures appropriate for PBP models. (Recommendation 3)
- Once established, the network of qualified measure developers should work with networks of providers and other interested stakeholders on the rapid cycle development, testing and use of measure prioritized for PBP models, using a consolidated or condensed version of the three-phase approach to measure development, validation, and acceptance. (Recommendation 7)
- HHS should work in collaboration with public-private partnerships to evaluate how best to develop the infrastructure required for the capture and use of the clinically rich and patient-reported data that will form the basis of outcomes-oriented measure sets for PBP models. (Recommendation 4)
- Wherever possible, public and private payers should begin to incorporate the use of absolute (vs. relative) performance targets in PBP models, and use a range of performance targets for each measures (vs. a single target) to motivate ongoing improvement across the performance continuum. (Recommendation 6a and 6b)
- Providers and payers should use existing and emerging performance measures in PBP models to develop engagement approaches with patients and communities that go beyond traditional clinical settings to advance improved health outcomes. (Recommendation 2)

Conclusion

[Placeholder: This section will summarize the PBP Work Group's main findings and recommendations following a process to gather feedback from the affiliated community.]

Appendix A: Roster

PBP Work Group Co-Chairs

Dana Gelb Safran, ScD
Chief Performance Measurement & Improvement Officer and Senior Vice President, Enterprise Analytics, Blue Cross Blue Shield of Massachusetts

Glenn Steele, Jr., MD, PhD
Chair, xG Health Solutions, Inc.

PBP Work Group Lead on Performance Measurement

Sanne Magnan, PhD, MD
Former President and Chief Executive Officer, Institute for Clinical Systems Improvement

PBP Work Group Members

Andy Baskin, MD
National Medical Director, Aetna

Mike Chernew, PhD
Leonard D. Schaeffer Professor of Health Care Policy, Harvard Medical School

Steve Hamman
Senior Vice President, Enterprise Network Solutions and Provider Partnerships, Health Care Service Corporation

Gretchen Hammer, MPH
Director, Medicaid Colorado

Amy Nguyen Howell, MD, MBA
Chief Medical Officer
California Association of Physician Groups

Kate Kinslow, EdD
President and Chief Executive Officer, Aria Health System

Elizabeth Mitchell
President and Chief Executive Officer, Network for Regional Healthcare Improvement

David Muhlestein, PhD, JD
Senior Director of Research and Development, Leavitt Partners, LLC

Hoangmai Pham, MD, MPH
Director, Seamless Care Models Group, Center for Medicare & Medicaid Innovation

Tom Raskauskas, MD
Formerly with St. Vincent's Health Partners

Andrew Sperling, JD
Director of Federal Legislative Advocacy, National Alliance on Mental Illness

Jeff White
Director, Health Care Strategy and Policy, the Boeing Company

CMS Alliance to Modernize Healthcare (CAMH) Staff

CAMH, sponsored by CMS, is an FFRDC operated by the MITRE Corporation. MITRE is chartered to work in the public interest.

Sarah Callahan, MHSA
Lead, LAN PBP Work Group

Grischa Metlay, PhD, MA
Lead Health Policy Analyst and Technical SME

Anne Gauthier, MS
LAN Project Leader

Amy Aukema, MPP
LAN Deputy Project Leader

Leina Slater
LAN PBP Work Group

Appendix B: Compendium of Core Measure Sets

[Placeholder]

Appendix C: References

[Placeholder]