



# Real-World Considerations for Implementing Pediatric Quality Measures: Insights From Key Stakeholders

*Claudia Schur, PhD; Margaret Johnson, MBA; Julia Doherty, MHSA; Kamila B. Mistry, PhD, MPH*

From the L&M Policy Research, LLC (C Schur, M Johnson, J Doherty), Washington, DC; and Child Health and Quality Improvement, Agency for Healthcare Research and Quality (KB Mistry), Rockville, Md.

The authors have no conflicts of interest to disclose.

Address correspondence to Margaret Johnson, MBA, L&M Policy Research, LLC, 1743 Connecticut Ave NW, Suite 200, Washington, DC 20009 (e-mail: [mjohnson@LMpolicyresearch.com](mailto:mjohnson@LMpolicyresearch.com)).

Received for publication January 22, 2021; accepted April 10, 2021.

## ABSTRACT

**OBJECTIVE:** Since its inception, the Pediatric Quality Measures Program has focused on the development and implementation of new and innovative pediatric quality measures (PQM) for both public and private use. Building the evidence base related to measure usability and feasibility is central to increasing measure uptake and, thereby, to increased performance monitoring and quality improvement (QI) for children in Medicaid or the Children's Health Insurance Program. This paper describes key stakeholder insights focused on measure implementation and increasing the uptake of PQM.

**METHODS:** The PQMP Learning Collaborative conducted semistructured interviews with 9 key informants (KIs) representing states, health plans, and other potential end users of the measures. The interviews focused on gaining KIs' perspectives on 6 research questions focused on assessing the feasibility and usability of PQM and strengthening the connection between measurement and improvement.

**RESULTS:** Our synthesis identified insights that highlight facilitators and barriers from the KIs' experience and the

strategies they employ when using measures to drive improvement "on-the-ground." Importantly, while the KIs agreed on how essential the research questions are to measure implementation and uptake, they uniformly acknowledged the complexity of the issues raised and pinpointed multiple unresolved issues.

**DISCUSSION:** The views expressed by these stakeholders point to several key issues – including incorporation of socio-economic status into quality measures and performance comparisons, use of benchmark data, and criteria for QI versus accountability – for developing a real-world research agenda to guide the future direction of quality measurement and implementation to improve children's health care.

**KEYWORDS:** measure implementation; pediatrics; quality measures

**ACADEMIC PEDIATRICS** 2022;22:S76–S80

## WHAT'S NEW

We sought insights from key expert stakeholders on issues related to the implementation and uptake of pediatric quality measures. Their views provide important insights to guide the future direction of pediatric quality measurement.

THE PEDIATRIC QUALITY Measures Program (PQMP) is focused on increasing the number of evidence-based and consensus-built pediatric quality measures (PQM) available for both public and private use. Several of the measures developed during the first phase of PQMP have been incorporated into the Centers for Medicare and Medicaid Services (CMS) Child Core Set and into other measure sets. The second phase of PQMP funding

emphasized research to better understand approaches to the dissemination and implementation of PQM in real-world settings and to build the evidence for increased performance monitoring and quality improvement (QI) for children in Medicaid or the Children's Health Insurance Program (CHIP). The grant funding identified the following research questions (Research Foci, RF) to be addressed:

- How can the same measure be used across multiple levels (ie, state, health plan, provider) to drive improvement across levels?
- How might relative performance be compared across different types of providers and organizations?
- What measurement challenges and successes exist at different levels of measurement?
- What methods can be used to benchmark progress?

- How might intermediate progress be measured to predict overall improvement?
- What are the different standards and criteria for measures depending on their intended use (ie, accountability versus quality improvement)?

Building the evidence base related to measure usability and feasibility with particular attention to these RF was considered central to increasing uptake of PQM and was a major focus of the grantees' work. While the questions were seen by CMS and AHRQ as ambitious, the underlying rationale for including these questions as part of PQMP grants was to ensure focus on key areas that have served as barriers to progress in pediatric (and adult) performance measurement and improvement efforts. To augment those findings, the PQMP Learning Collaborative conducted semistructured interviews with key informants (KIs) representing states, health plans, and other potential end users of the measures. This paper describes insights that emerged from the interviews related to the 6 RF with the intention of guiding future measure implementation work and increasing the uptake of quality measures and improve children's health.

## METHODS

The KI interviews were designed to gather information addressing the RF from individuals with on-the-ground experience implementing PQM. While the interviews are not representative of all views and are by their nature subjective, they offer a breadth and level of insight and real world examples not available from other sources.

In order to provide a broad range of end user perspectives, we developed a list of over twenty candidates representing states, health systems, and plans, with expertise in measure development and use, implementation science, and QI at the national, state and local level. Based on requirements to limit the data collection burden, AHRQ

selected 9 KIs, including senior executives who work with Medicaid/CHIP as medical directors and state program administrators, executives and physician leaders in health systems, health plans, or associations, and pediatric providers. The final list of KIs and their affiliations are listed in [Table 1](#).

We based the interview guide on the 6 RF that served as the basis of the second phase of PQMP and shared them with the KIs prior to the interview. Probes emphasized eliciting examples of implementation successes and challenges from the KIs' experience. Two senior PQMP-LC staff conducted the telephone interviews. The 60-minute interviews were conducted between June and September 2019 and recorded with the KIs' permission. The interview team debriefed after each interview to add probing questions to ensure the RF was fully addressed.

We transcribed interview recordings and developed a database organized by research question to enable analysis and identification of illustrative examples of measure implementation that exemplified the most frequently cited experiences and approaches. Three researchers conducted the analysis; any differences were discussed and settled through consensus.

## RESULTS

Our synthesis identified insights that highlight facilitators and barriers from the KIs' experience and the strategies they employ when using measures to drive improvement "on-the-ground." Importantly, while the KIs agreed on how essential these RF are to measure implementation and uptake, they uniformly acknowledged the complexity of the issues raised and pinpointed multiple unresolved issues. We report findings here, organized by the main themes identified, with facilitators and barriers to measurement and examples from the KIs interspersed throughout. Although many of the examples are specific to pediatrics, we included a few adult-focused examples because of their broad

**Table.** Key Informants

Name	Title and Organization at Time of Interview
Mary Applegate, MD	Medical Director, Ohio Department of Medicaid
Lindsay Cogan, PhD	Director of the Division of Quality Measurement, New York State Department of Health
Judy Dolins, MPH <sup>1</sup>	Chief Implementation Officer, Senior Vice President Community & Chapter Affairs and Quality Improvement, American Academy of Pediatrics (AAP)
Anne Edwards, MD	Senior Vice President for Primary Care and Subspecialty Pediatrics, AAP
Stephanie Greer	Director of Federal Advocacy, AAP
David Kelley, MD, MPA	Chief Medical Officer for the Pennsylvania Department of Human Services' Office of Medical Assistance Programs
Stephen Lawless, MD, MBA <sup>2</sup>	Senior Vice President and Chief Clinical Officer of Nemours Children's Health System; Professor of Pediatrics at Thomas Jefferson University
Elizabeth McGlynn, PhD	Vice President for Kaiser Permanente Research and Executive Director of the Kaiser Permanente Center for Effectiveness & Safety Research (CESR)
Susan Mills, PhD, RN	Senior Vice President, Population Health & Clinical Operations at Superior HealthPlan of Texas
Jeffrey Schiff, MD, MBA	Former Medical Director for Minnesota Health Care Programs, Minnesota Department of Human Services
Ryan Van Ramshorst, MD, MPH	Chief Medical Director for Medicaid/CHIP Services, Texas Health and Human Services Commission

<sup>1</sup>Ms Dolins was our original contact; she asked two colleagues to join the discussion. We have referred to them throughout as the "AAP colleagues." Ms Dolins has retired from/is no longer with AAP.

<sup>2</sup>Dr Lawless is currently Emeritus Professor at Nemours/Sidney Kimmel Medical College.

relevance to measure implementation and the paucity of pediatric examples. Although the KIs represent states, health plans, and providers, our findings are also limited by the small number of interviews conducted.

**TO USE THE SAME MEASURE ACROSS MULTIPLE LEVELS TO DRIVE IMPROVEMENT AT THE STATE LEVEL, DENOMINATORS (TARGET POPULATIONS) MAY NEED TO BE MODIFIED, APPROPRIATE LEVERS IDENTIFIED AT EACH LEVEL, AND GOALS ALIGNED ACROSS LEVELS**

When considering use of a measure developed for one level at another level, the KIs stressed the importance of ensuring that denominators are both adequate in size and consistent across levels. They also emphasized the importance of considering whether entities at each level are able to influence or impact the outcomes. To drive improvement at a higher level, these experts called for greater alignment of measures, incentives, and goals.

Attributing or assigning patients across levels was noted as a universal challenge, particularly when patients see multiple providers across different sites and in overlapping networks, use a specialist as a primary care provider, or switch health plans. Dr Van Ramshorst noted that the Medicaid population presents particular challenges for having consistent denominators across levels because of how frequently families shift on and off of Medicaid or between providers and plans; this results in difficulties determining correct target populations (denominators) at each level and potentially excludes frequent changers from being part of performance measurement. Dr Cogan noted that, since New York requires health plans to share quality data at the aggregate and patient levels, the state can look at quality measures across levels—for the practice, value-based contractor, health plan and state. In contrast, Dr Kelley noted that auto-assignment in the HealthChoices managed care program in Pennsylvania allows the state to attribute patients at the primary care physician (PCP) level. This allows Medicaid health plans to measure quality and offer incentives at the provider level. Dr Mills also described that the health plans in Texas are able to attribute patients to a PCP when a member has not made a selection.

A few of the KIs described using patient registries to more accurately target populations for measurement. When working on blood pressure control, Dr McGlynn shared that Kaiser Permanente could not reliably identify all of their hypertensive patients using claims data because of the myriad ways the codes are used and, at the time, electronic health record (EHR) data were not easily accessible. In order to more accurately identify their population, they built a patient registry which improved hypertension measurement and informed QI efforts.

Although measures may vary at different levels, Dr Van Ramshorst also noted that alignment can be achieved if the measures are focused on meeting similar goals. For example, the state may have adequate data and inputs to focus on an outcome measure, such as reducing pediatric

asthma mortality. At the practice level, providers can report a process measure — documentation of asthma severity — that ultimately supports reducing mortality. These measures are aligned in supporting a similar goal even though the reporting varies.

Similarly, Dr Applegate shared examples of how each level may have different levers available, all driving toward improvement in the same measure. When measuring emergency department (ED) visits for asthma at the level of the hospital or health system, providers at a lower level can impact use of the ED by expanding office hours or by providing 24/7 care, while health plans can reduce ED use by offering access to transportation and ensuring adequate pharmacy benefits.

**TO COMPARE RELATIVE PERFORMANCE ACROSS PROVIDERS, SIMILAR PROVIDER TYPES MAY BE GROUPED, AND PATIENT CASE MIX MAY BE ADJUSTED FOR CLINICAL SEVERITY OR SOCIO-ECONOMIC STATUS. BUT MAINTAINING CLINICAL STANDARDS OF CARE IS PARAMOUNT**

When making performance comparisons at each level, KIs discussed whether and when it is appropriate to group providers prior to comparison or to risk adjust measures, either for clinical severity or for social determinants of health (SDoH). Importantly, the AAP colleagues expressed discomfort with adult risk stratification and tools that have been imposed, citing a gap between the current and optimal state of stratification among pediatric patients.

Dr Applegate noted that Ohio Medicaid compares performance for all organization types in aggregate to identify patterns (eg, based on provider type, rural versus urban service area). The state applies the same threshold state-wide for management simplicity as well as to acknowledge that not all facilities at a given level—such as federally qualified health centers (FQHCs) or provider practices—are alike. The next, more granular examination is conducted at each level, such as independent PCPs, physicians who are part of a large health system, or an FQHC. A rationale for making comparisons within similar types of organizations is that opportunities and challenges faced by the organizations may differ in systematic ways; for example, FQHCs are more likely than independent practices to have colocated behavioral health personnel. Because of this difference, patients at independent practices would be more likely to require transportation and more telehealth facilitation.

Several KIs emphasized the importance of upholding standards of care, regardless of the patient population served. Ensuring providers are accountable for similar standards of care across practice types, Dr Van Ramshorst noted, *"Of course we want to increase our flu vaccination rate. That doesn't need to be different across the practices."* Dr Lawless also rejected the notion that expectations about care should be adjusted based on organization type (eg, FQHC providers; PCPs with a hospital affiliation). In contrast, Dr Applegate noted Ohio Medicaid makes some risk adjustment for SDoH, which helps to ensure the comparison is fair, given the potential for cherry picking

within Medicaid. Dr McGlynn discussed the importance of considering the intended use of measures before applying any risk adjustment. She noted that it may be useful to observe variation in performance not associated with differences in patient characteristics or other contextual factors; otherwise, providers who deliver unsatisfactory care will not be held accountable. While recognizing the importance of SDoH, Dr Mills stated that the health plan where she works does not currently adjust for SDoH, instead expecting that providers and health plans are responsible for identifying strategies to meet standards of care for the populations they serve.

#### **BENCHMARKS ARE COMMONLY USED BUT THERE ARE FEW GUIDELINES AS TO SELECTING AND IMPLEMENTING APPROACHES**

All of the KIs reported using benchmarks to compare performance across entities, whether to focus on variation in performance (often on low performers) or spur across-the-board improvement. Several KIs noted that benchmark data are also often used to set expectations for the pace of improvement. A variety of data sources for benchmarking were mentioned, including national data (such as HEDIS or Medicare star ratings), state data, patient registries and internal organizational data.

Dr Kelley explained that Pennsylvania publicly reports benchmarks based on state data, viewing this reporting as a useful tool to hold health plans accountable. The published data report, containing a weight-adjusted average of performance measures for the state's 9 Medicaid plans, draws attention to underperforming plans and is used as the basis for requiring implementation of QI interventions.

Three of the KIs described their state or system approach to setting targets for the pace of improvement. Dr McGlynn described how Kaiser Permanente establishes region- or medical center-specific benchmarks or targets, which differ depending on the circumstances and goals. To encourage forward movement, Kaiser Permanente uses dashboards that display both improvement metrics and absolute achievement scores so that low-performing providers or provider groups are not hidden, but instead are viewed in a broader context. Within an individual medical center, Kaiser Permanente sometimes convenes providers and staff to discuss achievement goals and negotiate targets collaboratively.

To set performance targets, Dr Applegate reported that Ohio views the 25th, median, and 75th performance percentiles of the distribution and examines variation across the state and by plan. The state sets benchmark levels accordingly, accounting for context and related priorities to determine how much improvement is achievable over several years.

Dr Lawless described how Nemours works with employees within the health system to set reasonable expectations around the pace of improvement. He noted that the system expects to see greater improvement earlier in an initiative compared to when the initiative is more

mature. He and other KIs also spoke of the importance of setting incremental improvement targets (eg, 20 percent improvement in the first year, 10 percent thereafter) and recognizing the realities of what can be realized after the initial gains are achieved.

#### **PROCESS AND ACCESS MEASURES USED TO ASSESS INTERMEDIATE PROGRESS TOWARD LONG TERM GOALS ARE LACKING**

The KIs uniformly agreed that identifying measures for intermediate progress when improvement is not likely to be seen in an annual (or longer cycle) is both necessary and challenging. In general, the KIs distinguished between long-term *outcome* measures and shorter-term, intermediate *process* measures. While the KIs provided several examples, their experiences often focused on adult rather than pediatric measures because of the dearth of measures and evidence about their predictive strength.

Dr Applegate emphasized the importance of understanding the pathophysiology of certain conditions as well as the context of care. For example, when measuring prescribing of antipsychotics to younger children, there may be short-term situations where children need antipsychotics while a longer-term strategy to address their behavioral needs is determined; thus, improvements in use of psychosocial services may be observed before the longer-term measure of safer antipsychotic use improves.

Dr Schiff identified the long-term goal of reducing the rate of low birthweight babies, asserting that the low birthweight measure is probably the closest to a population health measure in the Core Set. He suggested that, because improved birth spacing is a significant factor in decreasing low birthweight, rates of long-acting reversible contraception insertion immediately postpartum and at 6 weeks may be an appropriate intermediate measure that would indicate progress toward that goal. Dr Applegate described a related example of measuring intermediate progress through her department's Medicaid-level Progesterone Performance Improvement Project (PIP) with managed care plans. While the long-term goal was to drive improvement in rates of preterm birth and infant mortality, the plans and practices were brought together to focus on improvement in the initiation of progesterone between 16 and 24 weeks gestational age for the high-risk maternity Medicaid population.

#### **THERE IS AGREEMENT REGARDING STANDARDS FOR MEASURES USED FOR PAYMENT, BUT A LACK OF CONSENSUS REGARDING STANDARDS FOR MEASURES USED FOR QI**

Several KIs offered feedback specific to the standards and criteria suitable for measures used for payment, including that measures should be standardized and endorsed (ie, by NQF) to ensure the measure has been well-validated, tested, and is feasible and reliable; evidence-based and risk-adjusted; stable, valid, and linked to a national standard; and not easily manipulable.

The KIs expressed differing opinions about the standards to apply when using a measure for QI. Two KIs

asserted that these same standards and criteria should be applied when measures are used for QI. One KI noted that, when a measure is used solely for QI and is focused on a microsystem (unit, clinic), it may be reasonable to relax the statistical principles that are applied to the measure. However, this KI also noted that not being rigorous about the measures used for QI can create issues in interpreting the results, making it difficult to understand when trends are real versus “just noise.”

Several KIs commented that measures used solely for QI do not need to meet the same standards as those used for payment. Because QI measures need to be reported on a faster timeline, measures should be tailored to the entity conducting the QI, making it easier to use the information to take action. Another KI said that looking at reasonably timely quality measures (often derived from EHR measures) rather than claims-based measures, which have long lag times, is preferable for active QI efforts.

## DISCUSSION

These KIs’ perspectives on the RF that framed our inquiry revealed a number of commonalities but also differences in practices, priorities, and considerations. These experts emphasized that the RF posed, while raising critical issues related to measure implementation, are complex, multifaceted and many are not easily addressed.

We found several aspects of the RF where the issues and challenges are commonly known, and acceptable approaches or practices exist. While additional research, including empirical study, is needed to better establish guidelines or best practices, the need for work is apparent and there is little controversy. This applies to several dimensions of using the same measure across multiple levels, including attributing patients to providers across levels to ensure consistent target populations or measure denominators and identifying appropriate and actionable levers at each level to align goals and drive improvement. The use of benchmarks is also widely accepted—whether to show variation in performance, spur improvement, or set targets—yet multiple methods are used, and few criteria exist to guide selection of a particular approach.

In other areas, while issues and challenges are also known and methods exist, there is a more pronounced variation in perspective. Viewpoints differed regarding the standards and criteria to apply when using measures for accountability versus quality improvement. There is also considerable debate about how to compare performance across providers—whether to group ‘like’ providers for comparison and under what circumstances and for which measures to adjust for clinical complexity. One of the least settled issues is whether to adjust provider performance for patients’ socioeconomic characteristics. In the presence of large disparities in health and healthcare (eg, racial, ethnic, geographic, income), how we measure and incentivize quality has profound implications for improving the quality of pediatric care and achieving health equity. The design of current value-based payment models, however, does not always prioritize equity considerations.<sup>1–3</sup>

Finally, there were at least two aspects of the RF where there was little disagreement but also a paucity of evidence and approaches. These included the ability to use measures at lower levels where smaller numbers of patients may preclude reliable measurement and the identification of intermediate measures to predict progress and drive improvement for longer-term outcomes. In both areas, evidence exists but the KIs agreed that it is insufficient.

While the KIs were from a variety of levels of the health care system, findings are limited by the small number of experts and the range of perspectives within each level. Future research should include greater diversity and representation from different types of organizations within the levels.

We were unable to identify any recent formal efforts to gather broad-based perspectives on the challenges and facilitators facing stakeholders in the implementation of PQM. A study from over a decade ago<sup>4</sup> signals the progress made in many areas (eg, the development of new pediatric measures and related research capacity) but also demonstrates the persistence of obstacles related to insufficient resources and infrastructure for implementation. That many of the same challenges remain reinforces the need for developing a real-world research agenda to guide the future direction of quality measurement and implementation to improve children’s health care. Such research would benefit from increased collaboration and co-ordination across the different levels that have an impact on the quality of care delivered to children — state, health plan, hospital, and provider. Rigorous empirical research must also be coupled with understanding and experiences from end users relying on quality information to inform decision-making.

## ACKNOWLEDGMENTS

*Funding:* This study was conducted under contract to Agency for Healthcare Research and Quality (contract number HHS290201400003/75Q80119F32002).

The views expressed in this article are those of the authors, and no official endorsement by the Agency for Healthcare Research and Quality (AHRQ), the Centers for Medicare and Medicaid Services (CMS), or the Department of Health and Human Services (DHHS) is intended or should be inferred.

This article is published as part of a supplement sponsored by the US Department of Health and Human Services, the Centers for Medicare and Medicaid Services, and the Agency for Healthcare Research and Quality.

## REFERENCES

1. Chaiyachati KH, Bhatt J, Zhu JM. Time for value-based payment models to adopt a disparities-sensitive frame shift. *Annals of Internal Medicine* Ideas and Opinions. 2018;168:509–510. April 3.
2. DeMeester RH, Xu LJ, Nocon RS, et al. Solving disparities through payment and delivery system reform: a program to achieve health equity. *Health Affairs*. 2017;36:1133–1139. Jun 1.
3. Sandhu S, Saunders RS, McClellan MB, et al. Health equity should be a key value in value-based payment and delivery reform. *Health Affairs Blog*. 2020. <https://doi.org/10.1377/hblog20201119.836369>. November 25.
4. Shaller D. Implementing and using quality measures for children’s health care: perspectives on the state of the practice. *Pediatrics*. 2004;113:217–227. January.