

# Lessons from the Field: Using Pediatric Quality Measures across Multiple Levels

---

Prepared for the Agency for Healthcare Research and Quality by L&M Policy Research, LLC with guidance from the Pediatric Quality Measure Program (PQMP) Grantees

## Table of Contents

<b>List of Acronyms .....</b>	<b>1</b>
<b>Lessons from the Field: Using Pediatric Quality Measures at Multiple Levels.....</b>	<b>2</b>
<i>Introduction.....</i>	<i>2</i>
<i>Grantee Challenges to Implementation .....</i>	<i>6</i>
<i>Grantee Approaches.....</i>	<i>7</i>
<i>Grantee Key Findings and Implications .....</i>	<i>8</i>

## List of Acronyms

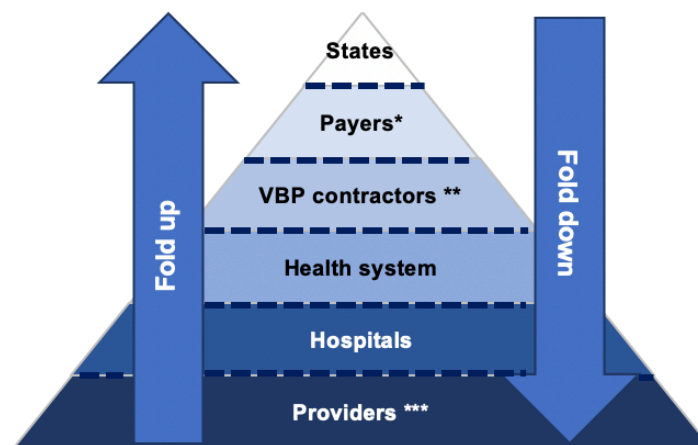
AHRQ	Agency for Healthcare Research and Quality
CEPQM	Children's Hospital Boston/Center of Excellence for Pediatric Quality Measurement
CheQ	University of Florida/Child Health Quality Partnership
DMOs	Dental Managed Care Organizations
EMR	Electronic Medical Record
HCAHPS	Hospital Consumer Assessment of Healthcare Providers and Systems
KI	Key Informant
KII	Key Information Interviews
NCINQ II	National Committee for Quality Assurance/National Collaborative for Innovation in Quality Measurement: Implementing and Improving
PQMP	Pediatric Quality Measure Program
PQMP-LC	Pediatric Quality Measure Program Learning Collaborative
QI	Quality Improvement
Q-METRIC	University of Michigan/Quality Measurement, Evaluation, Testing, Review, and Implementation Consortium
RF	Research Foci

## Lessons from the Field: Using Pediatric Quality Measures at Multiple Levels

### Introduction

These lessons from the field report examines two Research Foci (RF) central to the Pediatric Quality Measures Program (PQMP) grantees' work. These RF broadly focus on how quality measures can be used across multiple levels—including the state, health plan, hospital, and provider, as depicted in Figure 1.

Figure 1. Using the Same Measure at Multiple Levels, or Folding Up and Down



\*Payers include health plans.

\*\*Value based payment (VBP) contractors include accountable care organizations and other alternative payment arrangements.

\*\*\*Providers include individual providers, practices, and medical groups.

The specific questions are:

- What are the appropriate uses for each measure and each level of measurement, given a measure's "intended use" by developer/steward? What are the different standards and criteria that should be applied to the development and use of measures used for payment versus quality improvement (QI)? How do we determine measures can be appropriately used/aggregated at multiple levels (state, health plan, and provider levels) and be "folded up or down<sup>1</sup>"?
- How can the same measure be used to evaluate quality of care between multiple levels (i.e., state, health plan, and provider levels) to ascertain how improvement at one level drives overall improvement at the state level?

In examining these questions across two distinct data sources—literature reviews and key informant interviews (conducted by the PQMP Learning Collaborative)—a set of key considerations emerged that

relate how to use quality measures at multiple levels. Each of these key considerations and supporting findings from the literature and key informants are discussed below.

---

#### Availability of levers across levels.

---

Levers available to influence measure outcomes vary across the different levels (state, health plan, hospital, and provider). In order to affect performance, the accountable entity must have a way of influencing the outcome, yet the specific levers at each level often vary depending on the quality measure under consideration. These levers should be carefully considered when selecting and/or modifying measures to be folded up and down.

The literature provides specific examples of influencing measure outcomes at different levels:

- In one study, the authors ascribed variance in patient experience primarily to the physician and practice site levels, rather than to the physician network or health plan level (Safran et al., 2006).
- In another study, the authors found that health plans have an effect on HEDIS® scores that is independent of the effect of providers and provider groups with which they contract (Baker et al. 2004).

One state Medicaid Medical Director shared *“I think it’s super important that if we have a measure, for example, at the plan level and holding accountability, that there’s something they can do about it. ... if you’re going to hold someone accountable for follow up after mental health stay but you’re not telling the entity that they had a mental health stay, you’re in trouble. So, the entire thing falls apart.”*

---

#### Attribution of patients across levels.

---

When using a measure at different levels from the level for which it was developed, alternate attribution models and validation may be necessary to specify the denominator populations (Ryan et al., 2016). Measures intended to be used at the health plan level generally target an enrolled population, which becomes the denominator for measure calculation. At the health system, hospital, or provider practice level, however, there is no standardized enrolled population. Patients often see multiple providers who may practice at various hospitals or in several systems, creating challenges when trying to link the patient to a particular provider for purposes of defining the denominator; the measure denominator may need to be adapted to ensure the appropriate population is included at each level.

The literature points to two approaches for adapting a measure denominator for folding measures down from the health plan to the provider practice level:

1. Instead of using the specified denominator (an enrolled health plan population based on claims data), researchers used a visit-based approach to define the population for the denominator of several pediatric quality measures. The researchers included children with at least one visit recorded in the electronic medical record from a safety-net clinic and a network of community outpatient centers, respectively (Casciato et al., 2012; Gold et al., 2012).

One of the key informants (KI) – a state director of quality improvement – emphasized the importance of defining the relationships between patients and providers in order to construct accurate denominators at multiple levels. For a pilot program, the state had health plans create a patient-level data flag linking patients to practices and value-based purchasing contractors; the flag enabled the state to observe performance across multiple levels as part of their state quality reporting.

2. These same authors stressed the importance of alternatives to enrollment-based approaches in defining denominators for value-based payment models. To define denominators for these types of arrangements, a patient can be ‘attributed’ to a provider based on an algorithm that tracks and counts the various providers seen (Casciato et al., 2012; Gold et al., 2012).

---

#### Variation in specifications and data sources across levels.

---

For a measure concept, the specifications that define what to measure and the data sources that define how to measure it may vary across levels. The ability to maintain measure consistency and fidelity across levels can depend in large part on what data sources are available to populate the measure inputs. For example, measures developed for use at the state or health plan level often rely on program administrative or claims data. At the health system, hospital, or practice level, however, these claims data are often unavailable and there is a greater reliance on medical records data.

One study found that clinics that serve many uninsured and discontinuously insured patients had a significant number of self-paid visits—documented in electronic medical records (EMRs) but not in claims data—and, therefore, yield inaccurate measure rates when evaluating these clinics with administrative data alone (Gold et al., 2012). These authors also noted that, while many quality measures are based on medication dispensing indicators from claims data, medical record data only provide information on prescriptions written, rather than those filled.

Several key informants described strategies for overcoming the differences in availability of data sources across levels. One key informant – a state Medicaid medical director – noted that pharmacy claims, which are available in a timely manner and are becoming more widespread, could be used at the provider, ACO, and plan levels. Another state Medicaid medical director indicated that the state was able to access and link vital statistics data between mothers and babies to measure quality of care.

---

#### Small sample sizes when folding down to lower levels.

---

When using a measure at levels lower than what was intended – in other words, when “folding a measure down” – sample sizes may be too small to support precise calculations. The number of patients for which a measure is calculated varies across levels and may pose challenges when assessing performance at the individual physician and physician practice level, or even for smaller health plans. This issue with small numbers may be exacerbated at lower levels for measures targeting less common health conditions. At these lower levels, or for conditions affecting smaller numbers of patients, lack of precision means that the measure estimates may have large variances, which would make it difficult to determine whether differences in performance across providers or over time are ‘real,’ i.e., not due to random variation in the composition of the provider’s patient population.

Safran et. al. (2006) developed physician-specific samples of an appropriate size by pooling patient samples across multiple payers and calculating measures using a physician’s entire panel. These authors noted that—particularly for condition-specific indicators—a given physician or practice may have a small number of patients fitting the criteria and it may not be feasible to measure care at the individual physician level.

Key informants also indicated that the approach of pooling patients for provider practices across health plans to increase sample size for quality measurement is being used in state Medicaid programs.

---

### Lack of consensus on measure standards.

---

There is a lack of consensus on whether standards and criteria should vary with the intended use of a measure (payment or accountability versus quality improvement). There is no general agreement on whether standards and criteria used to calculate and to assess the soundness of a measure should vary depending on whether the measure is used for payment or accountability or for quality improvement purposes. Below are two very different perspectives on measure standards found in the literature:

- In the late 1990s, Solberg, Mosser, and McDonald suggested that quality improvement measures rely on data at the process level and the resulting metric can be approximate, since it only needs to be directionally correct for an organization to recognize where processes need to be refined. These authors also pointed out that measures for purposes of accountability are meant to create comparisons between like entities, are used to allow patients or entities to make health care choices, and can be used to encourage change (e.g., pay for performance). Since accountability measures are used for comparisons between like entities, they should be precise, valid, and reliable.
- On the other hand, the National Quality Forum (NQF), Intended Use Advisory Panel (2016b), there was agreement that there are “qualitative differences” between measures meant for quality improvement *only* and those focused on accountability, i.e., for use in public reporting or payment. The panel noted that QI-only measures are important but are used primarily by provider organizations for internal improvement projects and not for public purposes.

The KIs had differing opinions about the standards that should be applied when using a measure for QI and/or payment. A few of the KIs strongly believed that the same standards and criteria should be applied regardless of the goal of the measure. One KI expressed concern 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*.” However, several KIs shared that measures used solely for QI do not need to meet the same standards as those used for payment. They stated that QI measures need to be reported on a faster timeline and that the measures should be tailored to the entity conducting the QI, both of which will make it easier to use the information to take action.

---

### Reporting is used to engage stakeholders across levels.

---

Measure reporting and dissemination is used to set priorities, to impact measure performance, and to engage relevant stakeholders in driving quality improvement efforts across multiple levels. Driving quality improvement at the state level seems to hinge, at least in part, on reporting and sharing measure outcomes.

Findings from the evaluation of the CHIPRA Quality Demonstration indicate that six of the participating states used quality reports to drive statewide quality improvement, implementing projects as part of the demonstration that involved using measures collected from multiple levels to drive performance at the state level (Anglin & Hossain, 2015).

States participating in the CHIPRA Quality Demonstration also used measurement or quality reporting from lower levels to drive improvement in a number of ways, broadly grouped as follows: (i) analyzing and reporting results to stakeholders; (ii) engaging stakeholders for broad-based and specific QI efforts; and (iii) implementing policy or programmatic changes (“Spotlight on Alaska”, 2018; “Spotlight on Florida”, 2018; Brach et al., 2015; “Spotlight on North Carolina”, 2018).

A few of the KIs emphasized the importance of generating broad stakeholder engagement across multiple levels to identify priorities, confirm potential levers and barriers to change, and coordinate action to achieve improvement across multiple levels. Multi-stakeholder activities were sometimes used to discuss and obtain consensus on priorities and create a shared imperative for broad-based action. These collaborations involve using data comparing performance to drive improvement and discussions of the best way to implement programmatic changes in response to challenges identified.

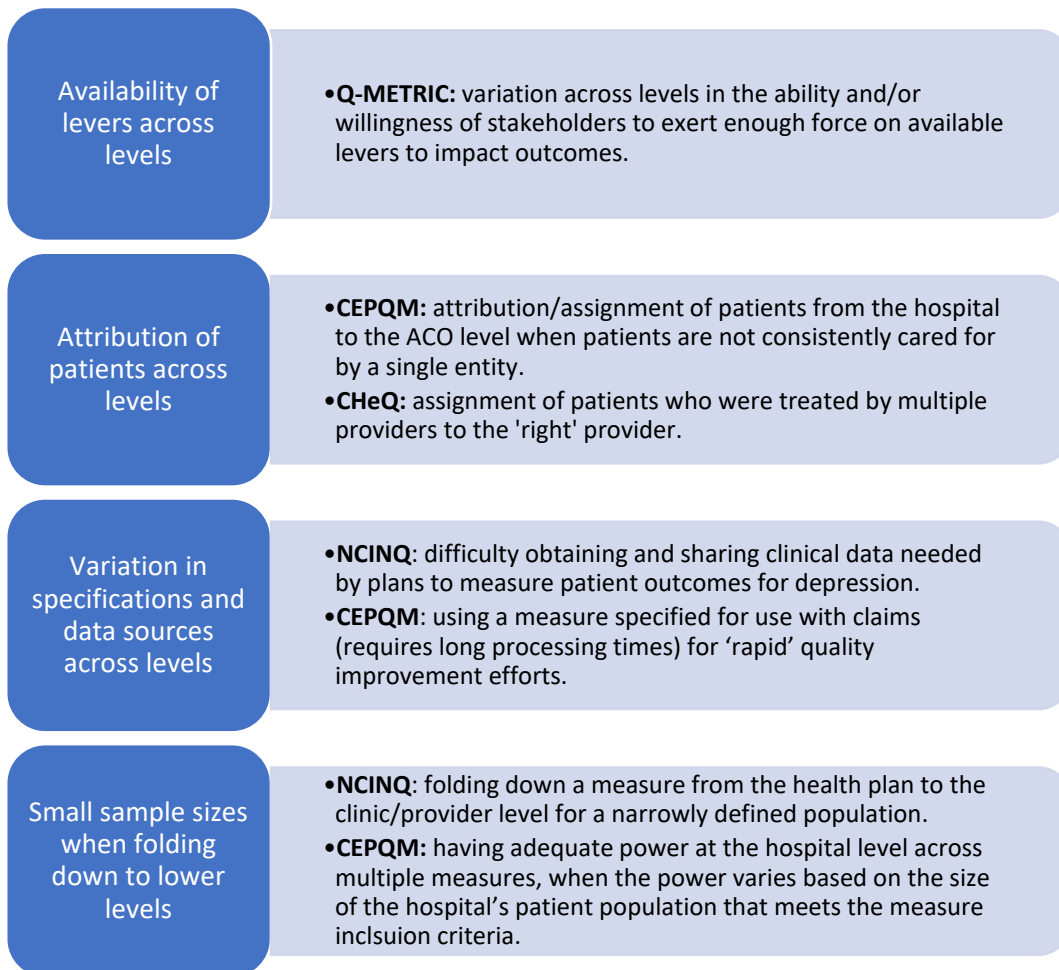
Four of these key considerations are further illustrated below using the findings from the PQMP grantee demonstration projects. The remainder of this *Lessons from the Field* provides examples of how grantees’ work – when implementing one or more of their pediatric quality measures – reflect and contributed to these key considerations. For each consideration listed above, the grantees described: (1) the challenges they faced during implementation, (2) the approach(es) they took to address the challenges, and (3) their team’s specific findings and implications for measure implementation.

### Grantee Challenges to Implementation

The grantees identified a number of challenges to implementing measures at multiple levels. While not all grantees faced the same challenges, there were a number of commonalities across projects related to the key considerations. In using measures across levels, several projects had to develop algorithms to attribute patients from, for example, a health plan to a health system or provider. Others experienced challenges in adapting measure specifications to fit the different data sources available across levels. Several specific examples are presented in Figure 2.



Figure 2. Examples of Grantee Implementation Challenges, by Key Consideration



### Grantee Approaches

The grantees used a wide array of approaches to support their efforts to implement measures at different levels. Both quantitative and qualitative approaches were used, separately and in combination with each other. Selected examples focused on data analysis, tool development and stakeholder interactions are shown in Figure 3.

Figure 3. Grantees used varied approaches to address implementation challenges

#### Data analysis

- To assess health plan levers, **NCINQ** analyzed high-performing plans to identify the characteristics and contextual factors associated with performance on two measures focused on antipsychotic use.
- Two of the grantee teams used attribution approaches to address challenges:
  - The **CHeQ** team analyzed claims data to support attributing patients to the provider where the child had received the most services/care in the previous 12 months, and
  - The **Q-METRIC** team reviewed existing attribution models to support designing a set of standards to determine health system attribution for children with sickle cell disease.
- The **CEPQM** team conducted a retrospective analysis of hospital discharges – by hospital and state – to assess whether there was adequate power to measure all-condition or condition-specific readmissions.
- The **CEPQM** team respecified readmissions measure to be calculated with EMR data.

#### Tool development

- The **CEPQM** team tested administering Child HCAHPS using tablets rather than a paper-based format to improve the timeliness of the data collection in order to provide near real-time performance measurement and to increase response rates.

#### Stakeholder interactions

- **NCINQ** asked participating health plans to categorize key data elements of a measure by data source, allowing the team to examine which types and the range of data used to report the measure.
- **NCINQ** convened a learning collaborative to make improvements on antipsychotic measure performance across levels - state, health plan, practice/provider and patient level.
- The **Q-METRIC** team met with key stakeholders at each level to better understand which levers could be used to support improved outcomes and hosted a design meeting with key stakeholders focused on quality improvement for patients with SCA.

### Grantee Key Findings and Implications

Based on their information gathering and analytic activities, grantees produced key findings related to each consideration described above that can be used to expand evidence about implementation of quality measures and using measures in quality improvement efforts. Generally, grantees were able to use and implement most of their measures at multiple levels with some adaptations of specifications or data sources that took into account contextual factors at each level. Across grantees and measures, the level that was most difficult to make adaptations for was the provider level – often the grantees could not overcome sample size or data availability issues. More of the key findings and the implications for implementation efforts are presented in the following series of figures.

Figure 4: Levers available to influence measure outcomes vary across the different levels (state, health plan, hospital, and provider).

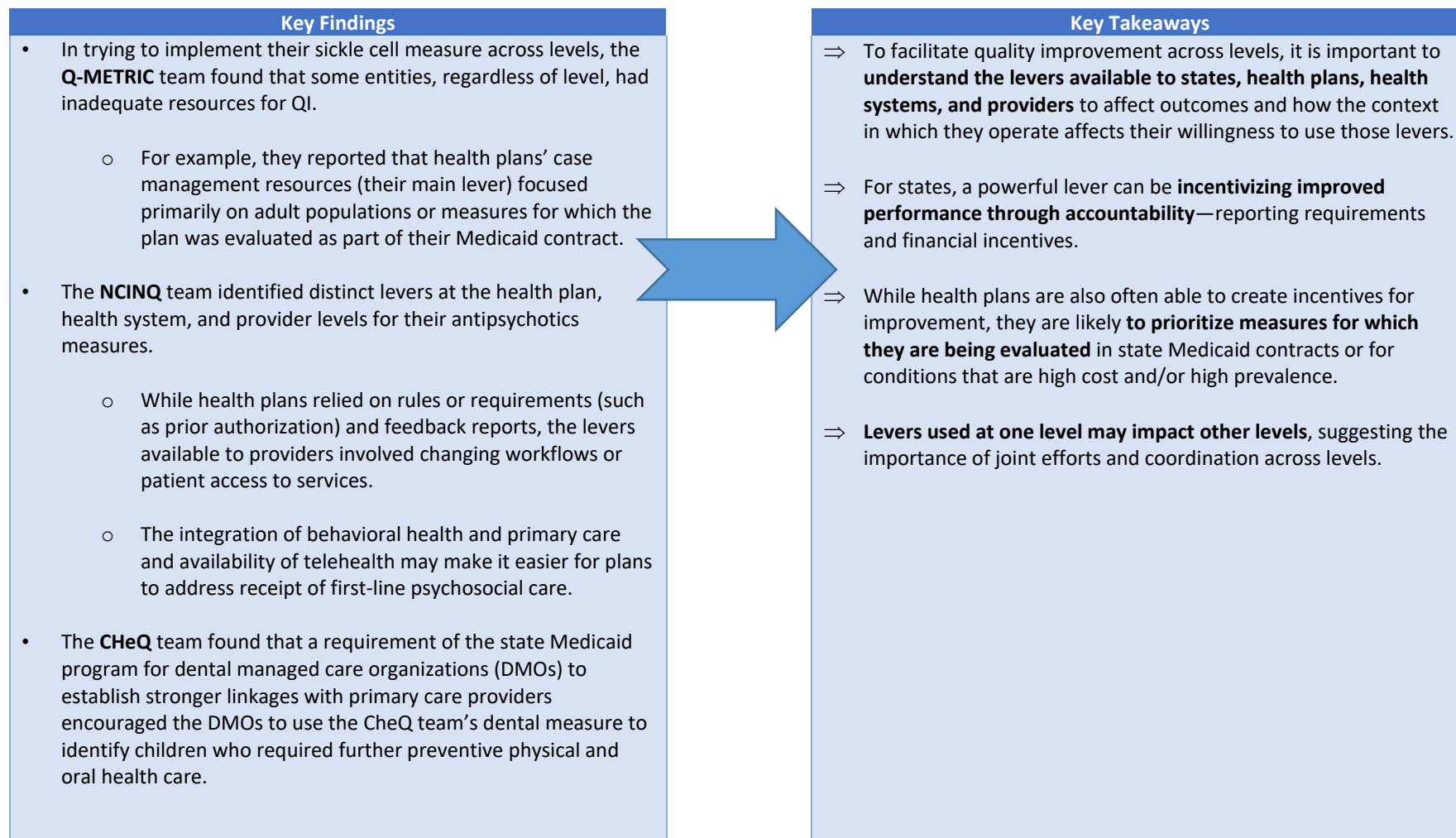


Figure 5: When using a measure at different levels from what was intended, alternate attribution models and validation may be necessary to specify the denominator populations.

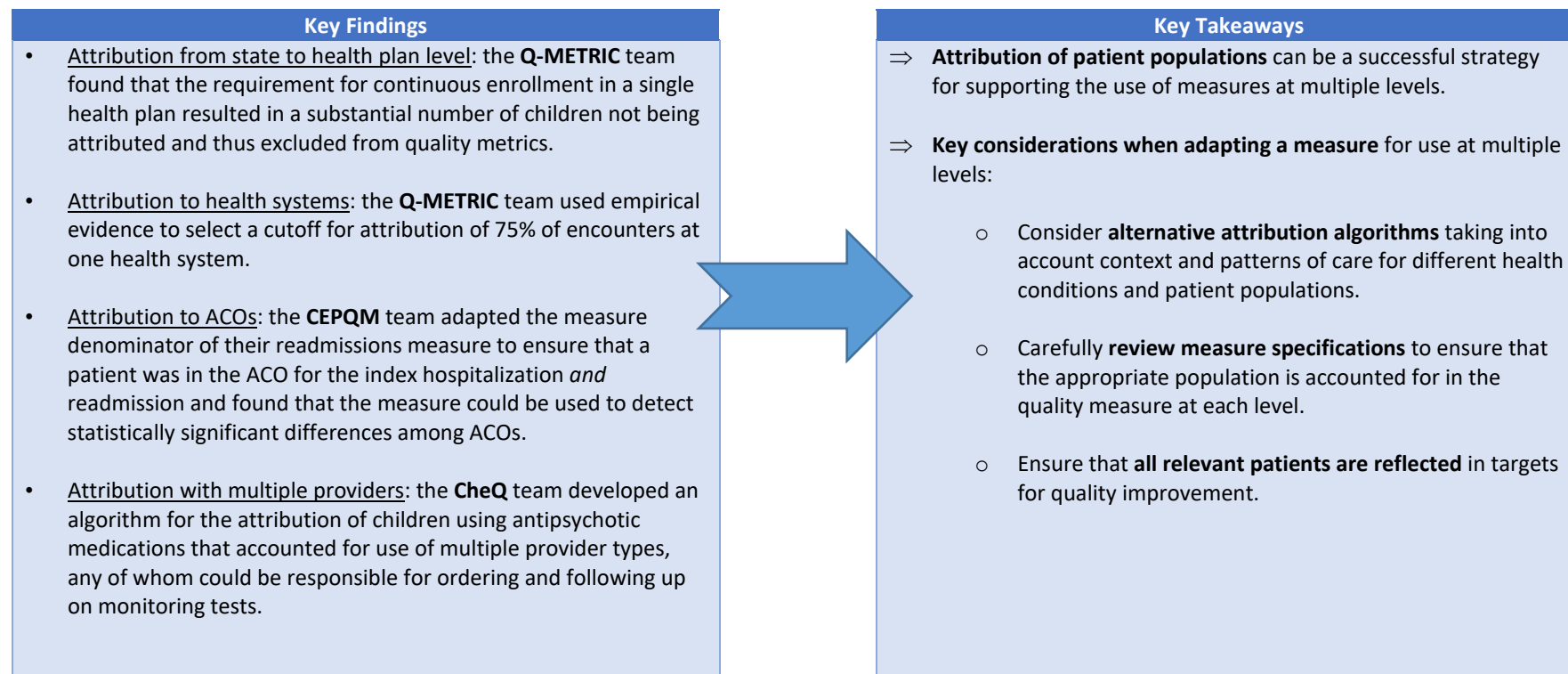


Figure 6: For a measure concept, the specifications that define what to measure and the data sources that define how to measure it may vary across levels.

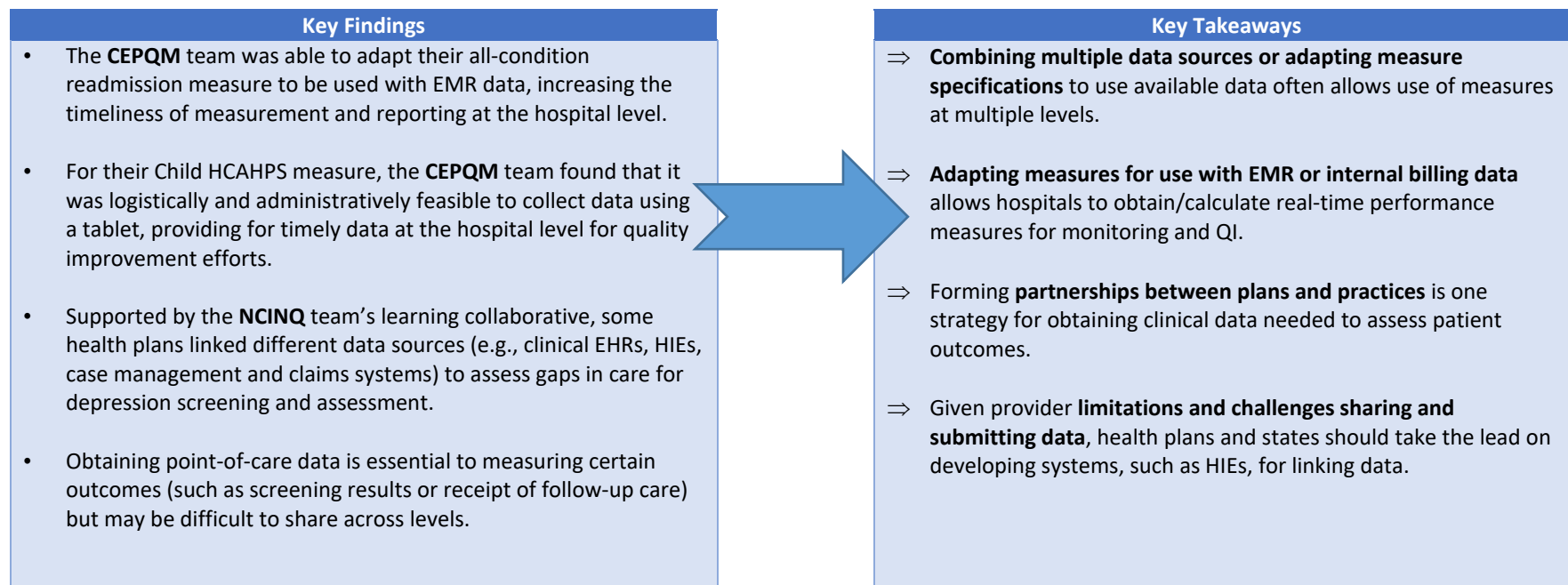
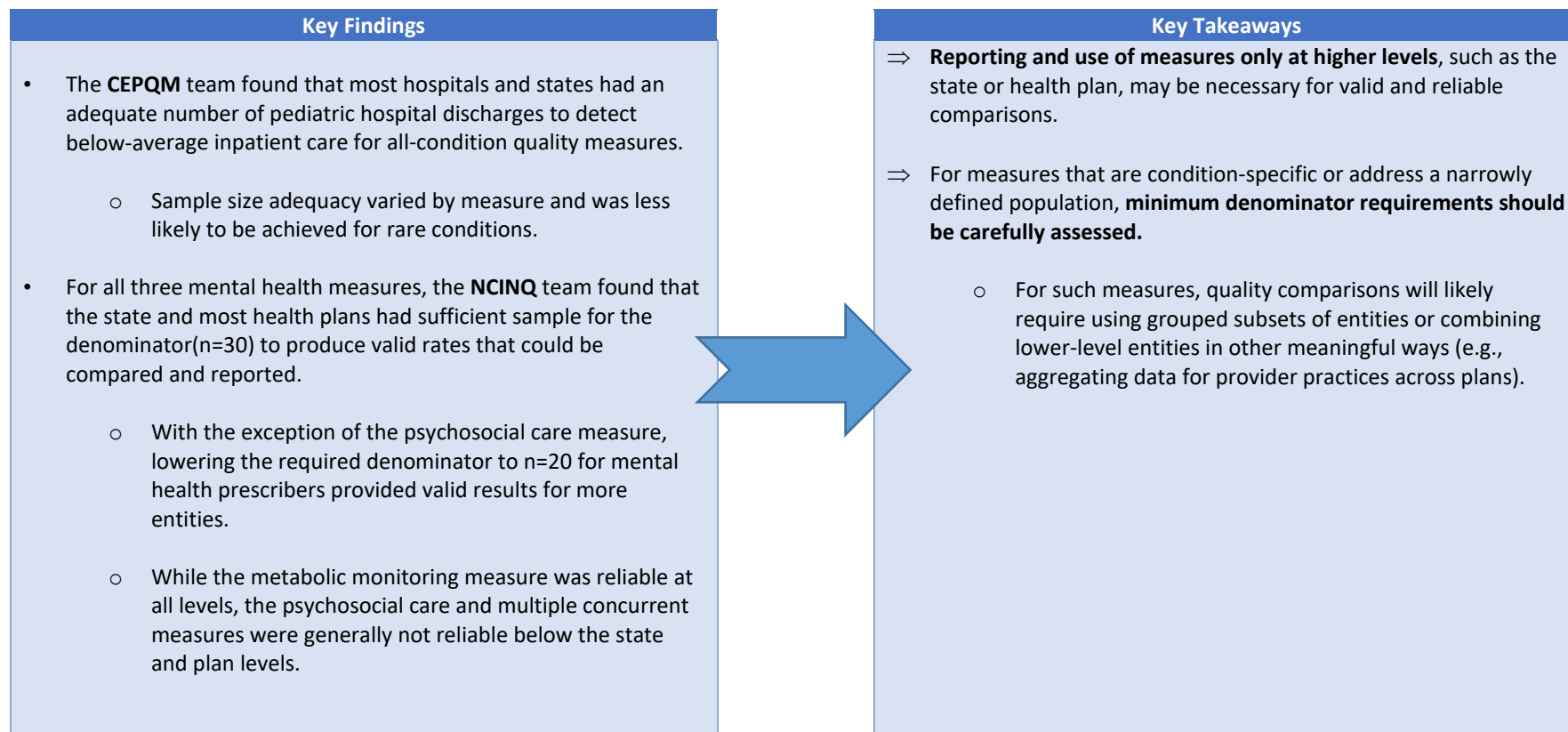


Figure 7: When using a measure at levels lower than what was intended – in other words, when “folding a measure down” – sample sizes may be too small to support precise calculations.



## References

- Agency for Healthcare Research and Quality. (2018, January). Spotlight on Alaska [PDF file]. *National evaluation of the CHIPRA quality demonstration grant program*. Retrieved from: [https://www.ahrq.gov/sites/default/files/wysiwyg/policymakers/chipra/statesummaries/cchipra-750-alaska-state-snapshot.pdf](https://www.ahrq.gov/sites/default/files/wysiwyg/policymakers/chipra/statesummaries/chipra-750-alaska-state-snapshot.pdf)
- Agency for Healthcare Research and Quality. (2018, January). Spotlight on Florida [PDF file]. *National evaluation of the CHIPRA quality demonstration grant program*. Retrieved from: <https://www.ahrq.gov/sites/default/files/wysiwyg/policymakers/chipra/statesummaries/chipra-750-florida-state-snapshot.pdf>
- Agency for Healthcare Research and Quality. (2018, January). Spotlight on North Carolina [PDF file]. *National evaluation of the CHIPRA quality demonstration grant program*. Retrieved from: <https://www.ahrq.gov/sites/default/files/wysiwyg/policymakers/chipra/statesummaries/chipra-750-north-carolina-state-snapshot.pdf>
- Anglin, G., & Hossain, M. (2015). How are CHIPRA quality demonstration States using quality reports to drive health care improvements for children? *National evaluation of the CHIPRA quality demonstration grant program*.
- Baker, L. C., Hopkins, D., Dixon, R., Rideout, J., & Geppert, J. (2004). Do health plans influence quality of care?. *International Journal for Quality in Health Care*, 16(1), 19-30.
- Brach C., Anglin G., Chamberlain K., & Kelley D. (2015, September 21). Using Quality Measures to Drive Improvement: Lessons from the CHIPRA Quality Demonstration - Presentation for the National Academy of State Health Policy's Vanguard and CHIP Directors. *National evaluation of the CHIPRA quality demonstration grant program*. Retrieved from <https://www.ahrq.gov/policymakers/chipra/demoeval/what-we-learned/nashp-webinar-slides.html>
- Casciato, A., Angier, H., Milano, C., Gideonse, N., Gold, R., & DeVoe, J. (2012). Are pediatric quality care measures too stringent? *J Am Board Fam Med*, 25(5), 686-693.
- Gold, R., Angier, H., Mangione-Smith, R., Gallia, C., McIntire, P. J., Cowburn, S., ... & DeVoe, J. E. (2012). Feasibility of evaluating the CHIPRA care quality measures in electronic health record data. *Pediatrics*, 130(1), 139-149.
- National Quality Forum. (2016). *Report From The Intended Use Advisory Panel: Recommendations to Enhance the Consensus Development Process (CDP)*. Retrieved From: [https://www.qualityforum.org/Projects/i-m/Intended\\_Use/Final\\_Report.aspx](https://www.qualityforum.org/Projects/i-m/Intended_Use/Final_Report.aspx)
- Ryan A, Linden A, Maurer K, Werner R, Nallamotheu B. Attribution methods and implications for measuring performance in health care. Washington, D.C.: National Quality Forum 2016

- Safran, D. G., Karp, M., Coltin, K., Chang, H., Li, A., Ogren, J., & Rogers, W. H. (2006). Measuring patients' experiences with individual primary care physicians: results of a statewide demonstration project. *Journal of General Internal Medicine*, 21(1), 13-21.
- Solberg, L. I., Mosser, G., & McDonald, S. (1997). The three faces of performance measurement: improvement, accountability, and research. *The Joint Commission journal on quality improvement*, 23(3), 135-147.