

Lessons from the Field: Making Performance Comparisons

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

List of Acronyms	2
Lessons from the Field: Making Performance Comparisons.....	3
<i>Introduction.....</i>	<i>3</i>
<i>Challenges to Implementation</i>	<i>5</i>
<i>Grantee Approaches.....</i>	<i>6</i>
<i>Grantee Key Findings and Implications</i>	<i>8</i>
References.....	12

List of Acronyms

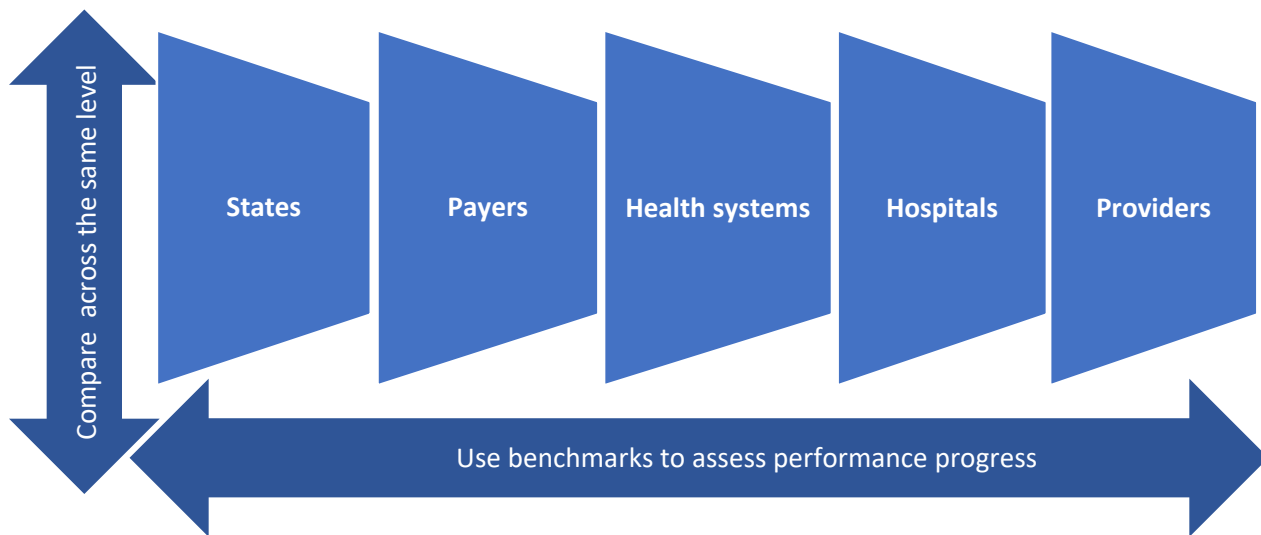
AHRQ	Agency for Healthcare Research and Quality
CEPQM	Children's Hospital Boston/Center of Excellence for Pediatric Quality Measurement
CheQ	University of Florida/Child Health Quality Partnership
FQHCs	Federally Qualified Health Clinics
HCAHPS	Hospital Consumer Assessment of Healthcare Providers and Systems
IMPLEMENT	University of California, San Francisco/IMPLEMENTing MEasures NeTwork (IMPLEMENT) for Child Health Network
KI	Key Informant
KII	Key Information Interviews
NCINQ II	National Committee for Quality Assurance/National Collaborative for Innovation in Quality Measurement: Implementing and Improving
P-HIP	Seattle Children's Hospital/Pediatric Hospital Care Improvement Project
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
RHCs	Rural Health Clinics
RF	Research Foci
SDOH	Social Determinants of Health
SES	Socio-Economic Status
TCD	Transcranial Doppler

Lessons from the Field: Making Performance Comparisons

Introduction

This lesson from the field report examines two Research Foci (RF) central to the Pediatric Quality Measures Program (PQMP) grantees' work. These RF broadly focus on performance comparisons: comparing performance between organizations at the same level, i.e., at the state, payer, health system, hospital, and provider levels, and assessing progress on performance using benchmarking, as depicted in Figure 1.

Figure 1: Making Comparisons



The specific questions are:

- How might relative performance be compared at each level, such as between different provider groups/organizations (e.g., Federally Qualified Health Centers, pediatric group practice, multi-specialty group practice) or between different Accountable Care Organizations/managed care health plans?
- What are evidence-based and scientifically sound methods for benchmarking progress on these measures? What level of improvement can be expected for measures--and in what time frame – taking into account different QI approach(es) undertaken at different levels? For example, if improvement is likely to be more rapid at the provider-level, what are the implications for performance targets established 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 relates to making performance comparisons. Each of these key considerations and supporting findings from the literature and key informants are discussed below.

Making ‘fair’ comparisons.

In order to make fair comparisons at the same level, it *may* be important to compare peers. The literature and experts suggest tailoring metrics to create peer groups of similar organization types/settings (based on key attributes, such as organization structure, size, infrastructure or even staff composition). For accountability (e.g., HEDIS® measures, payment, etc.), ‘apples-to-apples’ comparisons and concerns about fairness are critical in defining the appropriate peer group. For QI purposes, measurement comparisons can be geared to guiding improvement efforts, e.g., organizations may choose to compare to their own baseline results, or to similar organizations.

Several articles examined considerations for measurement within a given level. Frequently, the authors favored the concept of tailoring metrics for specific provider types and/or settings. In one article, the authors argued that even within a level – for example, clinics or hospitals – application of a quality measure developed for the community or primary care setting is inappropriate when used in a secondary or tertiary care facility (Naessens et al., 2017). Likewise, another paper focused on the importance of developing metrics tailored for Federally Qualified Health Clinics (FQHCs) or Rural Health Clinics (RHCs) to facilitate fair comparisons across their respective peer groups (Jones, 2017).

Several of the key informants discussed that comparing similar organizations or ‘peer groups’ (e.g., by provider type) is important. One key informant added that it might be important to *also* understand the patient characteristics of organizations when making comparisons.

Determine whether to risk adjust.

Adjustments for demographic, clinical and social risk factors may be necessary to ensure that performance comparisons are appropriate. The most common approaches used to account for these factors (demographic, clinical and social) are risk adjustment, risk stratification and exclusion. Generally, there is agreement that outcome measures should be adjusted for patient characteristics such as age, gender, and severity of illness. There is less of a consensus in terms of adjusting for social determinants of health (SDoH), but consideration of their impact on outcome measures is increasing. When adjustment for SDoH is warranted, the ability to do so relies on the availability and quality of social determinants data.

Much of the literature emphasizes the importance of making appropriate comparisons of outcome measures using risk adjustment to control for differences in patient populations across settings. For example, “... It is through comparison across providers that opportunities for improvement are identified. Providers with superior risk-adjusted outcomes set the goal for what is possible to achieve. In order for performance results to be meaningful and valid for identifying differences in performance across providers, outcome performance measures must be adjusted for different levels of risk in the patients served” (NQF Report, 2014).

At least three key informants noted that if a measure is related to meeting standards of care (a process measure), it does not make sense to risk adjust. They asserted that the same standard should apply, irrespective of practice type or patient population. Another KI cited pediatric well-child visits and dental encounters (process measures) as examples of metrics that do not need to be risk-adjusted. He also provided an example of an outcome measure where he would expect to risk adjust - the 30-day all-cause readmissions measure.

Selecting and using benchmarks.

The type of benchmark(s), and standards for establishing benchmarks, varies depending on the goal of the benchmarking activity and purpose (quality improvement or accountability) of the measure. More specifically, comparing performance through the use of benchmarks can be used to pursue multiple objectives: to foster competition, to encourage organizations to advance from a baseline level of performance or to incentivize entities to work toward a desired level.

The literature describes circumstances that may influence how benchmarks are set, for example, where performance is historically high (e.g., childhood vaccination rates in the U.S.), expectations for improvement are likely to be minimal so benchmarks may focus on maintaining performance or on an absolute goal whereas where performance is historically low, or where a state is significantly below the national average, an improvement goal or percentage increase from baseline may be most effective (Medicaid IAP, 2018). The level at which a benchmark is set should also consider whether there are opportunities for improvement and whether achieving the benchmark is feasible (NQF, 2017). Parikh, et al., (2019) provide support for benchmarks that represent “measurable and attainable goals for standardization of care.”

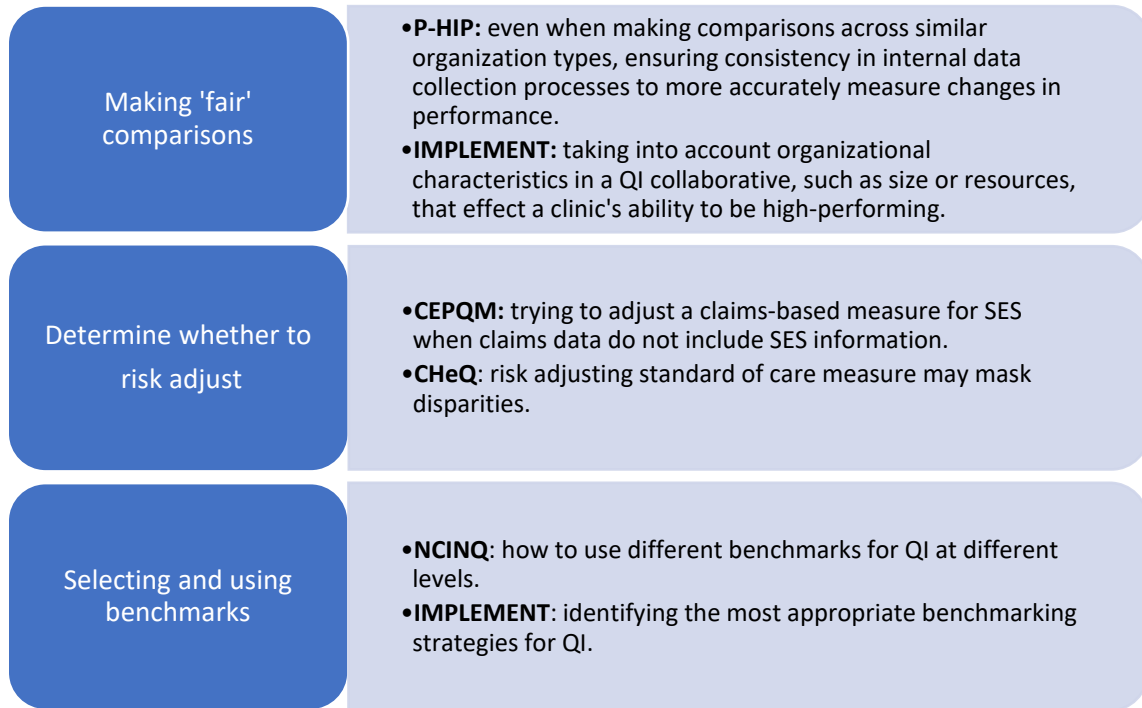
Several key informants described approaches to setting targets and expectations for the pace of improvement. One informant noted that the health system expects to see greater improvement earlier in an initiative and is careful not to set targets so high that they are not attainable. Several of the informants also spoke of the importance of setting incremental improvement targets (e.g., 20 percent improvement in the first year, 10 percent thereafter) and also of establishing different benchmarks or targets, depending on the circumstances and goals (e.g., improvement vs. achievement or by region).

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 the work of all six grantees specifically relates to the considerations described above for one or more of their pediatric quality measures. For each key considerations, 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.

Challenges to Implementation

The grantees identified a number of challenges to making performance comparisons. While not all grantees faced the same challenges, there were a number of commonalities across projects related to the key considerations. When comparing performance across entities within the same level, a central issue faced by most of the grantees’ projects was whether and how to account for differences in patient populations, with respect to demographics and illness severity as well as for socio-economic status (SES). In the context of quality improvement projects, the grantee teams had to identify strategies and data sources for selecting benchmarks and implementing approaches to their use. Several examples are presented in Figure 2.

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



Grantee Approaches

The grantees' approaches to addressing key implementation challenges related to measure comparisons and benchmarking – not surprisingly – relied heavily on quantitative data and statistical analyses. In a few cases, the grantees employed qualitative approaches to support their data analyses. Selected examples focused on these data analyses and stakeholder interactions are shown in Figure 3.

Figure 3. Grantees used varied approaches to addressing implementation challenges

Data analysis

- The **CEPQM** team combined results from three types of evaluation (strength of association with outcomes; correlation of adjusted scores with vs. without the adjuster; largest single change in scores) to determine whether each of the relevant case mix variables had sufficient impact to be retained for a final set of case-mix adjusters.
- The **IMPLEMENT** team tested the impact of whether or not to include risk adjustment for SES using different statistical approaches. They tested a baseline risk adjustment model (age, gender, chronic condition indicator) compared to baseline risk adjustment plus SES variables (using census data for the patient zip code for the following variables: % adults with less than high school education; % male unemployment; % households below federal poverty; and median household income).
- The **CHeQ** team chose not to risk adjust their antipsychotic measure; instead, they analyzed receipt of recommended care at the provider level so that disparities could be targeted through quality improvement efforts.
- Benchmarking data were established for the targeted measures by examining baseline performance across the eight participating hospitals in the **P-HIP** team's QI collaboratives. For example, for the targeted hospital-to-home transition quality measure, the highest performing institution scored 78 (on a 0-100 scale), so the eight participating hospital teams selected 85 as the goal score for the QI collaborative.

Stakeholder interactions

- To encourage QI project engagement, **Q-METRIC** shared comparative performance data on their Transcranial Doppler measure with key stakeholders (states, health plans, and providers).
- Working with the Children's Hospital Association, **P-HIP** was able to confirm the appropriateness of the peer groups they selected for making comparisons. The team's aim was to understand whether the process of care measures they targeted were sensitive/responsive to quality improvement interventions.
- NCINQ** explored the use of benchmarks for QI purposes through their Learning Collaborative with five Medicaid plans in New York. In the collaborative, they worked at the state, health plan, practice/individual provider and patient levels to make improvements on the HEDIS antipsychotic measures.

Grantee Key Findings and Implications

Based on their analytic and engagement activities, grantees produced findings for each of the considerations described above that can be used to expand evidence about making appropriate comparisons and using benchmarks to spur performance improvement. Several grantees found that it was important to use contextual factors to group similar types of organizations for making comparisons. In considering risk adjustment for socio-economic factors, grantees distinguished between process measures, where adjustments may mask disparities, and outcome measures, where adjustment may mitigate the challenges providers face in caring for vulnerable populations. Grantees selected a variety of benchmarking approaches that reflected their goals and intended uses. More of the key findings and the implications for implementation efforts are presented in the following series of tables.

Figure 4: The purpose of the measure – accountability (payment and public reporting) versus quality improvement – may impact the approach for making comparisons.

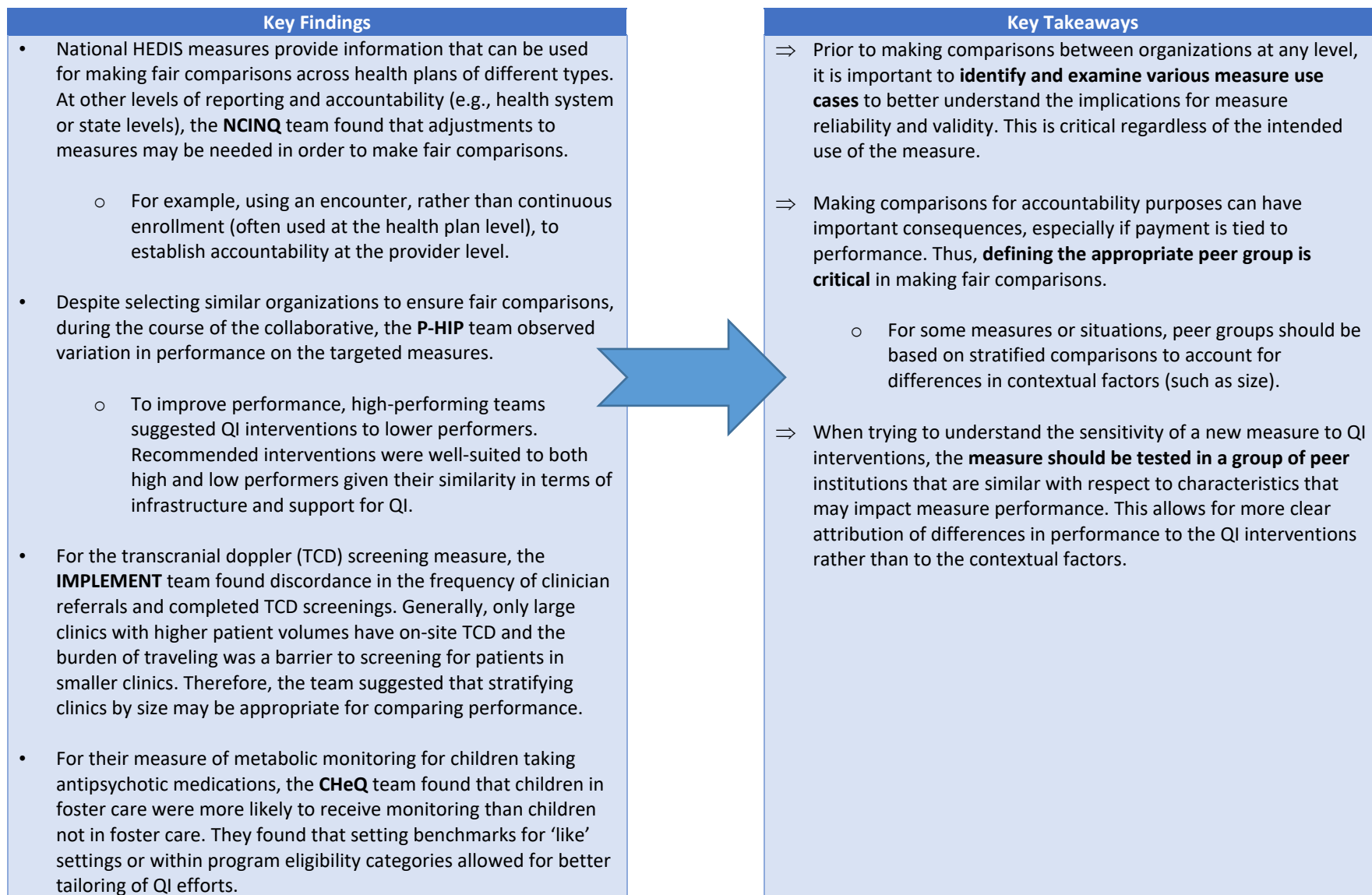


Figure 5: Adjustments for demographic, clinical and social risk factors may be necessary to ensure that performance comparisons are appropriate.

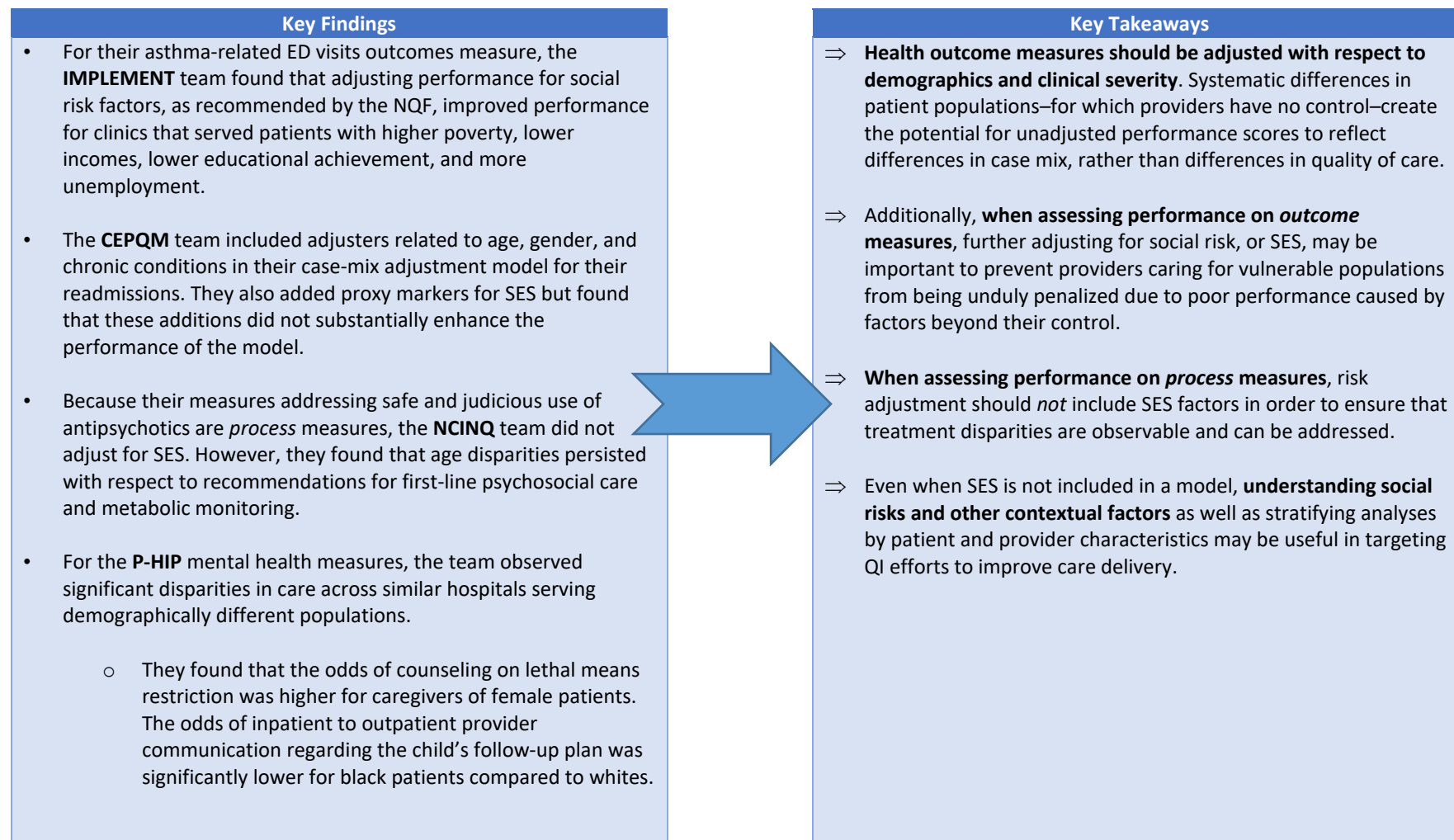
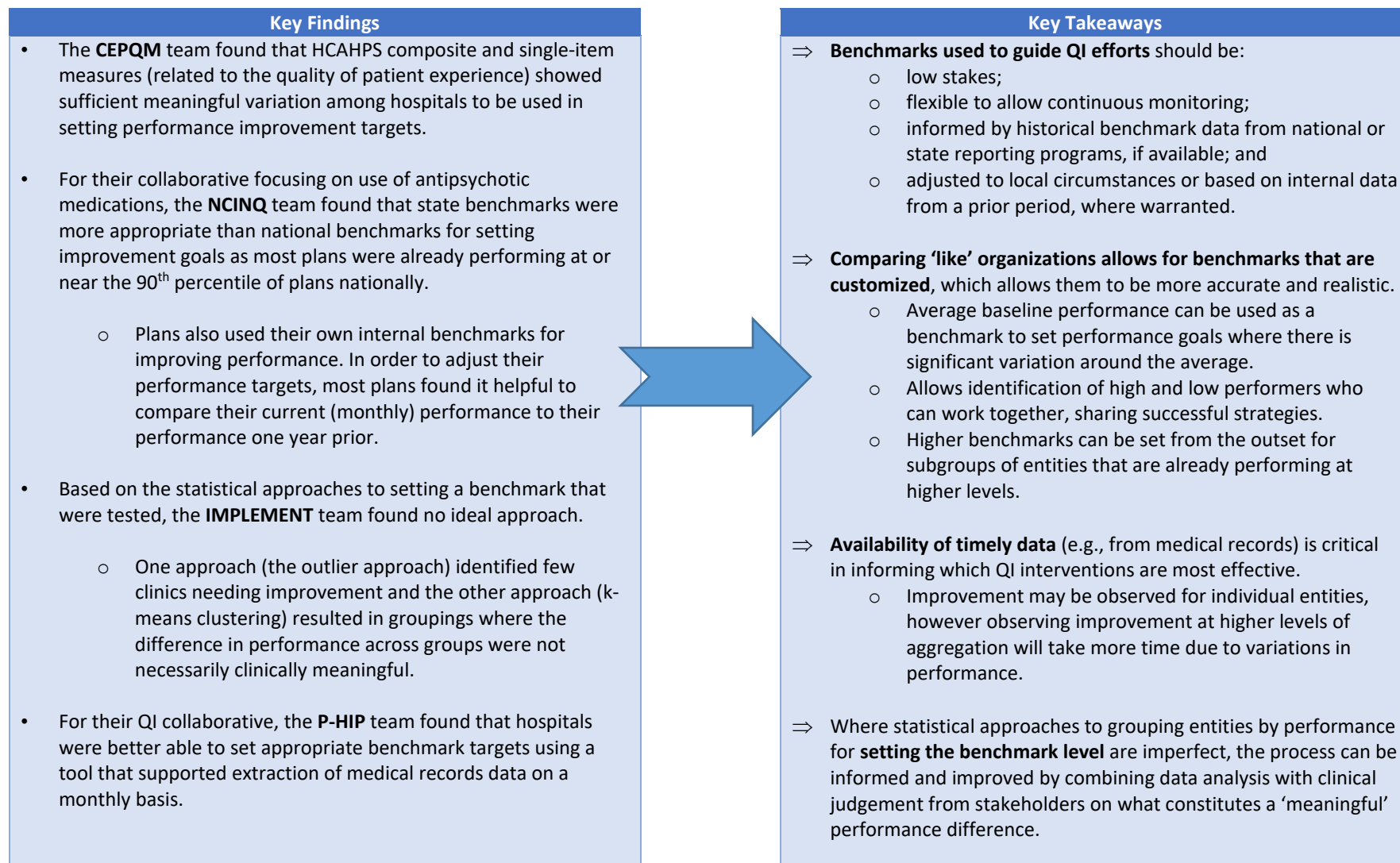


Figure 6: The type of benchmark(s), and standards for establishing benchmarks, vary depending on the goal and purpose (quality improvement or accountability) of the measure.



References

- Adams, R.J., McKie, V.C., Hsu, L., Files, B., Vichinsky, E., Pegelow, C., Abboud, M., Gallagher, D., Kutlar, A., Nichols, F.T. and Bonds, D.R., Brambilla, D., Woods, G., Olivieri, N., Driscoll, C., Miller, S., Wang, W., Hurlett, A., Scher, C., Berman, B., Carl, E., Jones, A.M., Roach, E.S., Wright, E, Zimmerman, R.A., Waclawiw, M. (1998). Prevention of a First Stroke by Transfusions in Children with Sickle Cell Anemia and Abnormal Results on Transcranial Doppler Ultrasonography. *N Engl J Med*, 339, 5-11. DOI: 10.1056/NEJM199807023390102
- Berenson, R. A., Pronovost, P. J., Krumholz, H. M. (2013). Achieving the Potential of Health Care Performance Measures. Retrieved from <http://webarchive.urban.org/publications/412823.html>
- Berry J, Zaslavsky A, Toomey SL, Chien AT, Jang J, Bryant MC, Klein DJ, Kaplan W, Schuster MA. Recognizing Differences in Hospital Quality Performance for Pediatric Inpatient Care. *Pediatrics*. 2015;136:251-62.
- Bird C.E., Seeman T., Escarce J.J., Basurto-Davila, R., Finch, B.K., Dubowitz, T., Heron, H., Hale, L., Merkin, S.S., Weden, M., & Lurie, N. (2010). Neighbourhood socioeconomic status and biological 'wear and tear' in a nationally representative sample of US adults. *J Epidemiol Community Health*, 64(10): 860-865.
- Bright MA, Kleinman L, Vogel B, Shenkman E. Visits to Primary Care and Emergency Department Reliance for Foster Youth: Impact of Medicaid Managed Care. *Acad Pediatr*. 2018;18(4):397-404. doi: 10.1016/j.acap.2017.10.005
- Ettorchi-Tardy, A., Levif, M., & Michel, P. (2012). Benchmarking: a method for continuous quality improvement in health. *Healthcare policy = Politiques de sante*, 7(4), e101–e119. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/23634166>
- Hughes, R. (2008). *Patient Safety and Quality: An Evidence-Based Handbook for Nurses*.
- Jones, A. (2017) Using Data to Measure Performance Advancing Value-Based Payment Methodologies for FQHCs and RHCs. Retrieved from https://nashp.org/wp-content/uploads/2017/11/Measuring-Performance-in-FQHC-APMs_HMA_2016.pdf
- Kumar SV, Bangar S, Neumann A, Kookal KK, Yansane A, Tokede O, Obadan-Udoh E, Mertz E, Simmons K, Even J, Mullins J, White J, Kalenderian E, Walji M. Assessing the validity of existing dental sealant quality measures. *J Am Dent Assoc*. 2018 Sep;149(9):756-764.e1. doi: 10.1016/j.adaj.2018.05.001. 2018 Jul 4. PMID: 29980245
- Leckman-Westin E, Finnerty M, Scholle E., et al. Differences in Medicaid Antipsychotic Medication Measures Among Children with SSI, Foster Care, and Income-Based Aid. *J Manag Care Spec Pharm*. 2018; 24(3):238-246. doi: 10.18553/jmcp.2018.24.3.238

- Martsof GR, Barrett ML, Weiss AJ, et al. (2016). Impact of Race/Ethnicity and Socioeconomic Status on Risk-Adjusted Readmission Rates: Implications for the Hospital Readmissions Reduction Program. *Inquiry: A Journal of Medical Care Organization, Provision and Financing*, 53, 0046958016667596.
- Medicaid Innovation Accelerator Program (2018) Determining Performance Benchmarks for a Medicaid Value-Based Payment Program. Retrieved from <https://www.medicaid.gov/state-resource-center/innovation-accelerator-program/iap-downloads/functional-areas/vbp-benchmarking-brief.pdf>; <https://www.medicaid.gov/state-resource-center/innovation-accelerator-program/iap-downloads/functional-areas/vbp-benchmarking-webinar.pdf>
- Naessens, J. M., Van Such, M. B., Nesse, R. E., Dilling, J. A., Swensen, S. J., Thompson, K. M., ... Santrach, P. J. (2017). Looking Under the Streetlight? A Framework for Differentiating Performance Measures by Level of Care in a Value-Based Payment Environment. *Academic medicine: journal of the Association of American Medical Colleges*, 92(7), 943–950. doi:10.1097/ACM.0000000000001654
- National Quality Forum. (2017). NQF Initiative to Determine the Impact of Adjusting Healthcare Performance Measures for Social Risk Factors Highlights Successes, Opportunities. Retrieved from https://www.qualityforum.org/News_And_Resources/Press_Releases/2017/NQF_Initiative_to_Determine_the_Impact_of_Adjusting_Healthcare_Performance_Measures_for_Social_Risk_Factors_Highlights_Successes,_Opportunities.aspx
- National Quality Forum. (2014). Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors. Retrieved from https://www.qualityforum.org/Publications/2014/08/Risk_Adjustment_for_Socioeconomic_Status_or_Other_Sociodemographic_Factors.aspx
- National Quality Forum (2017) *Strengthening the Core Set of Healthcare Quality Measures for Children Enrolled in Medicaid and CHIP*. Retrieved from https://www.qualityforum.org/Publications/2017/08/Strengthening_the_Core_Set_of_Healthcare_Quality_Measures_for_Children_Enrolled_in_Medicaid,_2017.aspx
- Office of the Assistant Secretary for Planning and Evaluation (U.S. Department of Health and Human Services). (2016). Report to Congress: Social Risk factors and Performance under Medicare’s Value-Based Purchasing Programs. Retrieved from <https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicare-value-based-purchasing-programs>
- Parikh. Establishing Benchmarks for the Hospitalized Care of Children with Asthma, Bronchiolitis, and Pneumonia. *Pediatrics*. 2014.
- Shenkman E, Tomar S, Manning D...Mistry K. Feasibility and usability of measuring receipt of sealants in 2 states. *JADA*, October 2019, DOI: <https://doi.org/10.1016/j.adaj.2019.05.022>

Toomey SL, Elliott MN, Zaslavsky AM, et al. Variation in Family Experience of Pediatric Inpatient Care as Measured by Child HCAHPS. *Pediatrics*. 2016;139(4):e20163372

Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation. (2019). *Overall Hospital Quality Star Rating on Hospital Compare Public Input Request*. Retrieved from <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Downloads/Overall-Hospital-Quality-Star-Rating-on-Hospital-Compare-Public-Input-Period.pdf>