Model for Improvement - Part Two: Measurement and Feedback for Quality Improvement Efforts

Greg Randolph, MD, MPH[a,b,c,*], Megan Esporas, MPH[a,b,c], Lloyd Provost, MS[d], Sara Massie, MPH[a,e], David G. Bundy, MD, MPH[f]

WHY THIS TOPIC AND WHY NOW?

The public, government, payors, and health care professionals increasingly agree that the quality of health care in the United States is in urgent need of improvement.[1] Measurement and feedback are fundamental aspects of quality improvement (QI); thus, national and local health care organizations are paying more attention to the selection and use of quality measures. To date, most of the attention and effort has been directed at developing measures at the national level to compare the performance of physicians and health care organizations, ie, what is referred to as national performance measures or standards. Little attention has been directed at measurement and feedback to guide QI projects, which represents the frontline of QI work.

[a] Department of Pediatrics, North Carolina Children's Center for Clinical Excellence, North Carolina Children's Hospital, CB# 7230, Chapel Hill, NC 27599-7230, USA
[b] Department of Pediatrics, University of North Carolina at Chapel Hill, CB# 7230, Chapel Hill, NC 27599-7230, USA
[c] Public Health Leadership Program, University of North Carolina at Chapel Hill, CB# 7230, Chapel Hill, NC 27599-7230, USA
[d] API –Austin, 115 East Fifth Street, Suite 300, Austin, TX 78701, USA
[e] Child Health Research Program, N.C. Translational and Clinical Sciences Institute, University of North Carolina at Chapel Hill, CB# 7230, Chapel Hill, NC 27599-7230, USA
[f] Department of Pediatrics, Johns Hopkins University School of Medicine, CMSC 2-121, 600 N Wolfe Street, Baltimore, MD, USA
* Corresponding author. Department of Pediatrics, University of North Carolina at Chapel Hill, CB# 7230, Chapel Hill, NC 27599 7230.
E-mail address: randolph@unc.edu (G. Randolph).

A substantial knowledge gap exists among health care professionals about how to select and use measures to guide QI projects.

In this article, the authors attempt to address this gap by providing evidence-based strategies from health care and other industries, augmented with practical examples from the authors' collective years of experience designing measurement and feedback strategies for frontline health care improvement teams. The article's focus is on the design and development of measurement and feedback strategies for QI projects in local settings, such as health care systems, hospitals, and clinical practices. The article also briefly discusses organizational measures, such as those included in dashboards or scorecards, which guide decision-making and priority setting for system-wide improvements.

The authors use a broad definition of quality measures, including not only clinical, but also satisfaction and financial measures, which are critical to health care organizations.[2] This focus aligns with the Institute of Medicine's definition of quality: "The degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge."[1]

This article does not address measurement for QI research, which is typically too slow, too expensive, too much of a data collection burden, and too complex for clinicians and staff engaged in QI efforts in local settings.[3] In addition, the authors do not address national performance standards used for quality assurance and accreditation, such as those established by the Joint Commission and the Center for Medicare & Medicaid Services (CMS). These types of data are regularly collected and reported by hospitals, but they are infrequently used, and rarely sufficient, to guide improvement projects. For example, accountability measures are insufficient for QI teams because they are designed for external groups (eg, payors and regulators for comparison or public reassurance) and typically focus on only outcomes, not processes, making it impossible to link any process changes to changes in outcomes.[3]

BACKGROUND AND HISTORY

The current use of measurement and feedback in health care QI originates from both medicine and industrial/business QI. In the mid-1800s, two pioneers, Florence Nightingale and John Snow, introduced the use of data in health care. Nightingale used measures of mortality to document the low quality of care injured British soldiers received during the Crimean War in Turkey. These data provided a compelling argument for reform by demonstrating that better sanitation in the field hospital could prevent deaths. Subsequent field hospital reforms reduced the death rate from greater than 40% to less than 5%. Nightingale also developed graphical methods to present data, clearly demonstrating that measures provide an organized way of approaching improvements in medical and surgical practice. Nightingale's achievements led to the adoption of formal record keeping in British hospitals.

Also in the mid-1800s, John Snow tracked the incidence and geographic location of cholera in London. By mapping the clusters of cholera cases, Snow provided convincing evidence that homes supplied by the Broad Street water pump had a much higher incidence of cholera than homes supplied with water from other sources. Despite the lack of a clear causal explanation (Snow postulated a "cholera poison"), the data demonstrating the pattern of disease outbreak was convincing enough to persuade authorities to remove the handle from the Broad Street pump, which greatly reduced the incidence of cholera.

The first American to publish on performance measurement in health care was Ernest Codman, a surgeon at Massachusetts General Hospital. Codman advocated

for hospital reform and systematic health care performance assessment, and in 1910, proposed the "end result system of hospital standardization,"[4] whereby a hospital would track every patient it treated long enough to determine whether treatment was effective. By tracking patient outcomes, adverse events could be identified and changes could be made to improve the care of future patients.

Modern QI science originated in the 1920s and is often attributed to the work of Walter Shewhart of Bell Telephone Laboratories.[5] Shewhart's work made measurement fundamental to QI activities. He published principles and techniques basic to the use of measurement, including concepts of statistical control, operational definitions, and visual display of data. He postulated two rules for presentation of data:

1) Data has no meaning apart from its context.
2) Data contains both signal and noise; to be able to learn, one must separate the signal from the noise.

One of Shewhart's colleagues, W. Edwards Deming, an American statistician and physicist, popularized his ideas in the industrial production and management sector beginning in the 1940s.[6] Deming applied Shewhart's concepts in government and industry, described the Plan-Do-Study-Act (PDSA) Cycle as an approach to QI, and taught these concepts and methods to thousands of people, primarily in industries other than health care.

In the 1970s, physician Avedis Donabedian proposed a model for assessing health care quality, describing seven pillars of quality: efficacy, efficiency, optimality, acceptability, legitimacy, equity, and cost. He posited, "*Structure* is the environment in which health care is provided, *process* is the method by which health care is provided, and *outcome* as the result of the care provided."[7] Focusing on structure, process, and outcome, he emphasized the importance of measurement and evaluation of health care quality, assuring completeness and accuracy of medical records, observer bias, patient satisfaction, and cultural preferences for health care.

The 1980s saw an emphasis on outcome measurement in the health care literature as a result of variability in medical practice, evidence-based medicine, and regulatory agency requirements. During this time, physicians Paul Batalden and Don Berwick were among many health professionals who began to study and apply Deming's ideas to health care. In 1991, Batalden and Berwick helped form the Institute for Healthcare Improvement (IHI), which has led the application of QI science to health care in the United States and internationally.[8] The IHI has been influential in promoting the adoption of measurement and feedback to improve the quality of health care.

THE FUNDAMENTALS OF QUALITY IMPROVEMENT PROJECTS

A cornerstone of QI science is the concept of a system.[9–12] Systems are complex and dynamic, and a system's nature has a profound effect on the measurement and feedback of QI projects undertaken within it. A system is "an interdependent group of items, people, or processes working together toward a common purpose". A system's identification of its common purpose aligns its parts. Those parts are interdependent, meaning that all parts of the system, and all relations between and among those parts, can influence system performance. Thus, QI projects require multiple measures, at multiple levels, to understand the effects of change on the different components of the system, and on the system as a whole.

The aim of a QI project must guide the selection of project measures. Project measures must be specific to the processes and outcomes being targeted for improvement and specific to the local target audience (within a health care system,

hospital, or clinical practice). Stakeholders, those who can affect or be affected by changes in a system, especially frontline clinicians and staff, should receive feedback on the QI effort to understand how the work is helping to achieve the improvement aim. This type of measurement and feedback answers the fundamental question, "Are the changes we are making (via QI projects) leading to improvement?"

In the following sections, the authors describe the process of developing, testing, and implementing measurement and feedback strategies in a typical QI project. **Fig. 1** presents a summary of this process, and a glossary of QI terms used in our discussion is available at the end of the article.

PURPOSES OF MEASUREMENT AND FEEDBACK

In QI, measurement and feedback are used to: (1) identify problems and establish baseline performance; (2) inform and guide QI projects; (3) select and test changes for improvement; and (4) assess progress toward organizational goals.

One strategy for evaluating performance and identifying potential areas for improvement is careful review of patient feedback from complaint systems, feedback forms, written and telephone surveys, and focus groups. Analysis of these data can help define gaps and create the case for conducting QI projects to address those gaps. Another mechanism for evaluating current performance is through continuous monitoring of system-level measures. For example, if performance gaps are detected in the hospital-wide infection rate, the leadership team can initiate a portfolio of improvement projects with the objective of improving this overall measure.

The second purpose of measurement in QI, which is the primary focus of this article, is to inform and guide QI projects. At the beginning of a QI project, a balanced set of measures, including outcome, process, and balancing measures, is established to support the team's aim statement (ie, the project goals).[13] These measures are then reported graphically (typically monthly) on run charts (see previous article on the model for improvement for an example of a run chart). Statistical process control charts[14] can also help monitor the progress of the project in accomplishing the improvement team's aim.

The third purpose of measurement and feedback is to develop, test, and implement changes. One common approach is to use PDSA cycles,[13] where each cycle is designed to answer specific questions about changes that the improvement team is testing. The "plan" step of the PDSA cycle involves specifying these questions and developing measures and a data collection plan to that will answer them. The previous article about the model for improvement shows several examples of PDSA cycles. Usually the measures used in these small tests of change are specific process measures related to the change(s) being evaluated in the cycle; sometimes, they are the project outcome measures stratified for the clinicians or patients who were part of the cycle. Often, the most important part of feedback in a PDSA cycle comes from qualitative data, particularly from comments staff or patients make about the changes being tested.

The fourth purpose of measurement is to assess progress toward organizational goals. The authors briefly discuss the use of organizational scorecards and dashboard in a later section of this article.

OPTIMAL ATTRIBUTES OF QUALITY IMPROVEMENT MEASURES

Selecting measures for QI projects can be challenging. Fortunately, there is increasing knowledge and experience to help improvement teams select useful measures. In this section, the authors describe the most important attributes of QI measures based on
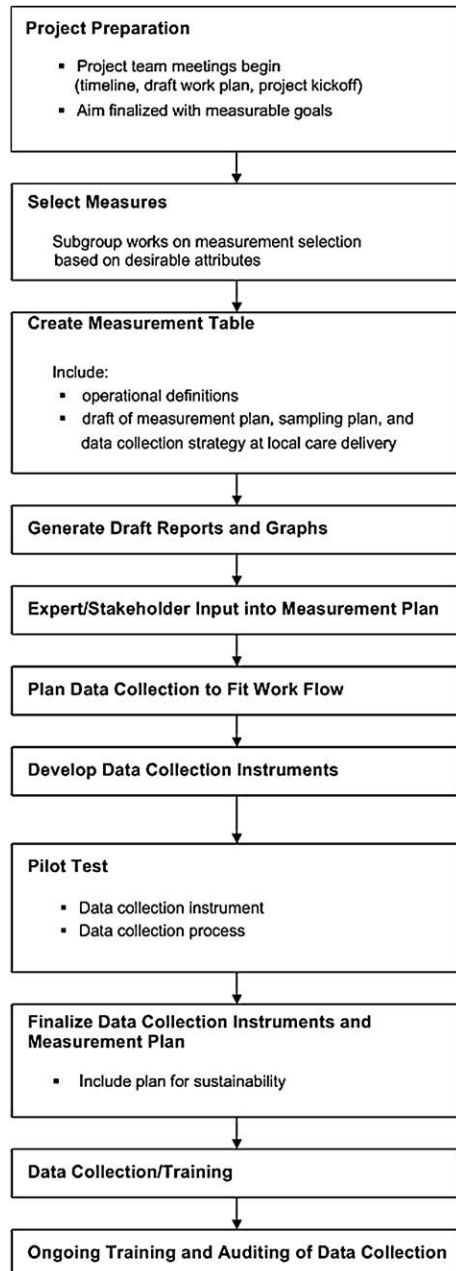
Project Preparation

- Project team meetings begin
  (timeline, draft work plan, project kickoff)
- Aim finalized with measurable goals

↓

Select Measures

Subgroup works on measurement selection
based on desirable attributes

↓

Create Measurement Table

Include:
- operational definitions
- draft of measurement plan, sampling plan, and
  data collection strategy at local care delivery

↓

Generate Draft Reports and Graphs

↓

Expert/Stakeholder Input into Measurement Plan

↓

Plan Data Collection to Fit Work Flow

↓

Develop Data Collection Instruments

↓

Pilot Test

- Data collection instrument
- Data collection process

↓

**Finalize Data Collection Instruments and
Measurement Plan**

- Include plan for sustainability

↓

Data Collection/Training

↓

Ongoing Training and Auditing of Data Collection

Fig. 1. Flow chart demonstrating the stages of development, testing, and implementation of a measurement and feedback plan during a typical QI project.

the literature and collective experience: tailored to the target audience, comprehensive, carefully defined, and involving minimal measurement burden. These attributes are summarized in **Table 1**.

Tailoring to the Target Audience

When selecting measures, it is critical to consider the target audience (ie, those who will be viewing, using, and interpreting the data).[15–17] The target audience will invariably include clinicians and clinical staff, so it is important that measures address high-impact clinical targets for the population of patients affected by the project (eg, using a common chronic condition like asthma for a chronic disease management improvement project).[15,16] The target audience should always include system leaders; therefore, measures should link to high-level organizational priorities such as strategic and financial goals.[16,17] Although there are some advantages to borrowing measures used by peer organizations, teams should be aware that some measures, especially those originating from research, can be overly complex or difficult to understand for the local target audience. An example is the Continuity of Care index $(\text{coc} = (\Sigma_{j=1}^{8} n_j^2 - N)/N(N-1))$ used in many pediatric continuity of care research studies.[18] Using a simpler measure, such as the percentage of patients seen by their primary care physician, would be much more intuitive for most clinicians.

Quality improvement measures must also be credible to the target audience. Measures from national organizations should be used when applicable, as long as they align well with the interests of the local target audience. Good resources for publicly available, evidence-based QI measures across a wide variety of clinical topics include the National Quality Measures Clearinghouse sponsored by the Agency for Healthcare Research and Quality,[16] the National Quality Forum,[19] the Institute for Healthcare Improvement,[20] and the National Initiative for Children's Healthcare Quality.[21] However, it is worth noting that the availability of pediatric-specific national measures lags that of adult-specific measures.

Including Comprehensive Measures

Three types of measures are essential to QI: (1) outcome measures, or those that address how the health care services provided to patients affect their health, functional status, and/or satisfaction; (2) process measures, which address the health care services provided to patients; and (3) balancing measures, which evaluate unintended consequences or the stability of the system being changed in the project. A balanced set of measures for a QI effort should include at least one outcome, process, and balancing measure.[16,17,22]

Outcome measures are significant for clinicians, as well as for leaders, who want to know the ultimate impact of a project. However, these measures can be slow to change over time, so inclusion of one or more process measures allows the team to see the effects of a QI effort more quickly. In addition, at the end of the project, process measures demonstrate that the intended changes were indeed implemented.

Process measures also illustrate the link between the changes made to the system and changes in the outcome. A common approach to calculating improvement is by an item-by-item measurement, such as the proportion of patients who receive each individual component of a health care service. Another approach, all-or-none measurement, is potentially advantageous in some situations. In this approach, the numerator is the number of patients who receive all of the measured components and the denominator is the total number of patients. All-or-none measures are ideal for a process that includes a series of critical steps, all of which must be completed to produce desirable outcomes, such as the process of inserting a central line.[23] If

all-or-none measurement is not appropriate, other types of composite measurement can be used.[24]

Any improvement project, which by definition involves making changes to one or more processes, can have unintended consequences. Balancing measures assess these potential unintended consequences and assure teams that they have indeed improved their overall system, rather than optimizing one part of the system at the expense of another.[17] Balancing measures can also be important in helping address the concerns of those who are resistant to the proposed changes. For example, in a project to improve immunization delivery in a private pediatric practice, a balancing measure might be the effect of changes to this high-volume process on office visit wait times.

Carefully Defining Measures

It is important to define measures in such a way that the baseline levels (ie, the level of performance before the QI project starts) are neither too high (difficult to detect improvement) nor too low (may be deflating or seem implausible to the target audience). Measures must also be responsive (show improvement in a timely fashion) to the planned changes in the system, and minimize the impact of variation unrelated to process changes, thereby improving the "signal-to-noise ratio" for detecting improvements to the system.[23,25] For example, a difference in wait times on Mondays versus Wednesdays related to differences in patient volumes on those days would be "noise" in an initiative attempting to lower overall wait times.

During a QI project, a measure's responsiveness to detecting change is affected by the time between piloting a change on a small scale and having the change fully implemented, as well as the time to have the change affect a patient's care and the time to sampling the affected patients. The last issue is particularly important in pediatrics because researchers often sample by age group (eg, sampling 24-month to 30-month olds for immunization status for vaccines given between 0 to 15 months). Thus, months may elapse between the introduction of changes and the measurement of their effects. Optimizing how a measure is defined and sampled can often minimize these delays. For example, changing a measure's denominator to "all children seen in the past month with asthma" rather than "all children seen in the past quarter with asthma" would allow for more rapid detection of the results of a change to asthma care processes.

Minimizing Measurement Burden

Finally, and most importantly, teams should minimize the burden of measurement for their QI project.[17,25,26] Effort spent creating, collecting, displaying, analyzing, and interpreting data, although absolutely critical to improvement, is nonetheless time taken away from making changes to the system. The measurement burden is much like the burden of carrying water on a hike up a mountain: hikers must carry water on the way to the top, but carrying too much water will slow them down unnecessarily, or worse, tire them to the point of not being able to reach the top at all. A small, balanced set of five to eight measures will usually suffice for most improvement projects. **Table 1** lists several strategies to minimize the burden of measurement, such as using data that already exist in your organization.

Putting it All Together: A Balancing Act

Measurement selection involves making tradeoffs among these desirable measure attributes. For example, a measure that is more responsive to changes in a system, such as measuring immunizations by 18 months rather than by 24 months, may

Table 1
Key attributes of measures to support QI projects

Measure Attributes	Considerations
Tailored to the target audience	
Meaningful, important and relevant to target audience	Address areas with substantial effect on the health of population
	High burden of illness, high volume, problem-prone process, poor quality and/or high variation
	Relevance to financial and strategic issues
Understandable to target audience	Avoid unnecessary complexity
Credible for target audience	Use nationally recognized practice guidelines when possible
	If using nationally recognized practice guidelines, measures should account for patient preferences and clinician judgment
Comprehensive (ie, includes outcome, process, and balancing measures)	
Outcome measures	Outcome measures address how the health care services provided to patients affect their health, functional status, and/or satisfaction
Process measures	Process measures address the health care services provided to patients
	Consider all or none measures
Balancing measures	Balancing measures address potential unintended consequences of changes to processes
Carefully defined	
Definition assures baseline levels are neither too high nor too low	If too high, difficult to show improvement
	If too low, may be discouraging or cause disbelief
Definition assures measures are responsive to changes in the system	Minimize delay between improvements and measuring the effects of improvements

Definition minimizes the impact of variation	Reduce known causes of variation through stratification (eg, time of day, location/unit, staff or clinicians) to allow for more sensitivity in detecting improvement
	Seasonal variation - avoid measures with this when possible or handle with rolling averages or year over year comparisons
Measurement burden minimized	
A small, balanced set of measures	Strive for a set of measures that describes a system as much as possible with as few measures as possible
	Use or adapt existing measures when possible
Data collection built into the flow of work	Understand work flow (eg, through process mapping) and find best place to collect data in work flow and best person(s) to collect
	Use existing data where possible
Small sample sizes	20–40 observations when collected frequently are often adequate (eg, medical record abstraction on 25 patients collected monthly)
Simple data collection instruments and methods	Use simple and quick instruments like check sheets, checklists
	Leverage technology (eg, email surveys, scannable forms)

Data from Refs. [2,16,17,22,23].

have a very low baseline level in a particular practice. Improvement teams are well served if they are aware of these attributes and can weigh these tradeoffs as they strive to create useful, though never perfect, measures for their project. **Table 2** presents an example of one project team's measure set to illustrate a product of this balancing act.

THE MEASUREMENT DEVELOPMENT PROCESS

After identifying robust candidate measures with the attributes described above, it is time to commence the measurement development process. The process of measurement development involves establishing operational definitions, collecting data, pilot testing, establishing baselines, and setting goals. During this process, it is also important to plan for sustainability after the project is complete. The next section describes the measurement development process in detail.

Establishing Operational Definitions

The first step of the measurement development is establishing operational definitions. Well-intentioned measures often fail to yield actionable information because of a lack of clear operational definitions.[6] For example, a QI team aiming to improve the emergency department (ED) care of children with pneumonia might select the measure: "Percentage of children with pneumonia who receive their first dose of antibiotics within four hours of arrival to the ED." Operational definitions for this measure would need to include, at a minimum, how target children are defined (eg, age restrictions, chronic disease exclusions); how the presence of pneumonia is defined (eg, clinical versus radiologic versus administrative [eg, billing] diagnosis); and how children are accounted for if they received antibiotics before their arrival in the ED or left the ED in less than four hours.

Table 2
Example of project measurement set used by the UNC Division of Pediatric Gastroenterology for improving care for children with inflammatory bowel diseases

Measure	Project Goal
Process measures	
% patients seen this month with disease activity recorded	95%
% patients seen this month with active disease that have a documented plan to escalate therapy	95%
% patients seen this month with classified steroid status	95%
% chronic steroid users (per clinician) seen this month with documented plan for tapering ($\pm$ maintenance medication)	95%
% patients seen this month with growth and nutritional status classified at visit	95%
% patients seen this month classified as "at risk" or "failure" for nutrition with documented intervention	95%
Outcome measures	
% of patients with active disease (mild, moderate, severe) measured by physician global assessment this month	< 20%
% patients classified as satisfactory growth status	95%
Balancing measure	
Waiting time	N/A

When defining measures, many QI teams have found it helpful to create a table that summarizes the relevant features of each measure. Measure name, numerator, denominator, source, frequency of collection, and inclusion and exclusion criteria are a minimal set of features, but more detailed specification is often necessary. In addition to defining individual measures, it is important to consider whether to track measures independently (eg, whether each of five steps in a ventilator care bundle were performed daily) or in an all-or-none fashion (eg, whether all five steps in the ventilator care bundle were performed daily), as stated earlier.[10]

Collecting Data

A critical step in developing a data collection strategy is careful consideration of the utility and accessibility of existing data sources. In many health care settings, the electronic medical record (EMR) serves as a valuable source of data. For example, an immunization improvement effort might involve querying the EMR monthly to assess the immunization status for all children of a certain age. These electronic systems are often a robust source of data, but they can be difficult to access. Using an EMR can be a challenge unless improvement team members have both administrative access to the data and the technological expertise required to extract and analyze the data.

Another strategy involves analyzing data collected in an existing patient registry. For example, patient registries of children with chronic health conditions can be used to monitor performance on outcome, process, and even balancing measures. One advantage of registry data is that they typically provide information on the entire population of interest (eg, all children with sickle cell disease seen in a hematology clinic). A disadvantage of registry data is that data specifications are typically already in place before the initiation of an improvement project, limiting the scope of what can be tracked to what already exists in the registry.

Regardless of the means of data collection, it is essential to develop a collection strategy that yields a maximal quality of actionable data with a minimal amount of extra effort. Data collected for other purposes (eg, clinic billing) can often be used in improvement efforts, and may involve substantially less time and effort to collect than data collected specifically for a QI project. Nonetheless, it is often worth comparing a small sample of billing data with more robust data sources (eg, EMR) to ensure that billing data accurately reflect the results they are intended to measure. For example, a review of 20 visits for which the billing data indicate a principal diagnosis code of bronchiolitis could be compared with EMR notes from each of the 20 visits to confirm the accuracy of the billing data as a means of identifying bronchiolitis visits.

Pilot Testing

Measures that appear conceptually strong often fail in the implementation process. For example, using a validated questionnaire is a logical approach to measuring and tracking patient satisfaction; however, if patient flow in the setting of interest is such that patients rarely have time to complete the survey, data collection will be problematic. By pilot testing measures and the data collection process on a small scale, improvement teams can determine whether the measurement system functions as planned. Shortening the survey, administering it at a different point in the visit, conducting it as an interview instead of a survey, or mailing it to patients might all be reasonable alternatives. These strategies would also need to be pilot tested to directly observe potential barriers to their effectiveness. Often, correctable errors (eg, ''which provider did you see today?'' inadvertently omitted from the patient satisfaction survey) can be identified after pilot testing the survey with a few patients.

Establishing Baselines and Setting Goals

Establishing baseline performance and setting performance goals are essential steps in developing the measurement strategy. In simple terms, measuring baseline performance answers the question, ''where are we now?'' and setting a goal establishes ''where do we want to go and how soon do we want to get there?'' The period of baseline measurement should be long enough to provide convincing data to recipients but short enough to avoid impeding initiation of QI efforts.

In the authors' experience, QI efforts are frequently derailed by attempts to collect ''perfect'' baseline data when more concise (or approximate) data would suffice. Depending on the frequency of the events being measured, prospective versus retrospective baseline data collection may be considered. For example, in a project aimed at reducing catheter-associated bloodstream infections in a pediatric intensive care unit (ICU), historical (ie, retrospective) data from the preceding year may be the best approach for establishing an accurate baseline, given the rarity of the event. In contrast, a 1-week, prospective hand-washing observation period may provide sufficient baseline data for a hand-hygiene QI project.

Goal setting is typically focused on process and outcome measures and can be conceptualized in several ways. First, QI teams can set absolute goals (eg, ''90% receiving a preventive service''), which may be consistent with external standards (such as Healthy People 2010). Absolute goals may help team members see linkages between their local efforts and larger regional and national efforts to improve care, as well as compare their performance directly against a standardized benchmark. Some teams choose to set across-the-board, though arbitrary, absolute goals for their process measures (eg, ''95% completion of all processes'') while linking absolute goals for outcome measures to national benchmarks. Second, teams can set goals relative to baseline performance (eg, ''reduce the time required to process refill prescription requests by 50%''). This can be particularly effective when baseline performance is poor. When the outcomes being measured are rare (eg, ''pediatric codes outside the ICU''), goals can be conceptualized in terms of time between events (eg, ''reach 6 months with no pediatric codes outside the ICU''). Finally, goals can be a combination of the above strategies, as well as staged over time (eg, ''reduce wait time for well child appointments to half of present wait time within 6 months, and to less than 2 days within 12 months''). In the authors' experience, whether they are staged or not, it is usually preferable to use absolute goals rather than relative goals because they are more tangible and meaningful to teams and stakeholders.

Planning for Sustainability

Sustainability is a key component of measurement development, and should be considered from the earliest phases. Quality improvement efforts that are sustained over long periods are usually those in which sustainability is considered early in the program's design and implementation, including the measurement strategy. What may seem like a reasonable measurement strategy when viewed in the lifespan of a QI project (eg, ''review 10 medical records each week to determine the percentage of children receiving recommended hemoglobin screening'') may seem burdensome when extended beyond the project period. If data collection is considered ''extra'' work to the involved staff or is performed by an external research team or a team that will not operate at the completion the project, long-term sustainability of the measurement process will be unlikely. Conversely, if data collection is integrated into the everyday responsibilities of frontline staff, the likelihood of sustained data collection will increase. Successful teams often reduce the frequency of measurement

(eg, from weekly to monthly or monthly to quarterly) after goal performance has been reached, as well as strive to "hard-wire" data collection into standard practice operations. The authors provide further discussion about implementing the sustainability plan later in this article.

PRINCIPLES OF QUALITY IMPROVEMENT FEEDBACK

Quality improvement measures not only document progress in QI efforts but also serve as strong motivators for improvement. However, in order for these data to resonate with and motivate frontline clinicians and staff, both the message and the messenger must be thoughtfully considered. Bradley and colleagues[27] recently identified seven themes essential to effective feedback of data in QI projects (**Box 1**). Six of the seven themes relate directly to the message of the data feedback, and the remaining theme relates to the messenger.

The Message: Is it Believable?

The importance of the perceived validity of the feedback is highlighted in three themes: (1) data must be viewed as valid to motivate change; (2) it takes time to develop data credibility; and (3) the source and timeliness of data feedback are critical to perceived validity. Discharge diagnosis codes, for example, can be an excellent means of identifying cases of a given illness in administrative data. However, if clinicians are skeptical about the accuracy of discharge diagnosis coding, inferences drawn from such data may garner little traction for improvement. Likewise, year-old data leave room for clinicians to assert, "things have changed since then," whereas real-time data are more difficult to dismiss.

The Message: Is it in the Proper Context and can it be Sustained?

As previously discussed, feedback can be framed in ways that enhance its effectiveness. Comparison of local data to national or other benchmarks can provide motivation for improvement. Similarly, clinician-level data documenting each individual's performance in relation to the group's performance can be valuable, though considerable care is needed to avoid creating a punitive atmosphere. Lastly, just as the sustainability of any measurement strategy must be considered in its development, the sustainability of the feedback system is equally important and must be formally integrated into the standard operating procedure of an organization if it is to persist.

Box 1
Key themes for effective quality improvement measurement feedback

Data must be perceived by [clinicians] as valid to motivate change

It takes time to develop the credibility of data within [an organization]

The source and timeliness of data are critical to perceived validity

Benchmarking improves the meaningfulness of the data feedback

[Clinician] leaders can enhance the effectiveness of data feedback

Data feedback that profiles an individual [clinician's] practices can be effective but may be perceived as punitive

Data feedback must persist to sustain improved performance

Data from Bradley EH, Holmboe ES, Mattera JA, et al. Data feedback efforts in quality improvement: lessons learned from US hospitals. Qual Saf Health Care 2004;13(1):26–31.

The Message: Does it Effectively Convey the Underlying Information?

Valid and actionable data are most effective when they are presented to frontline users in visually compelling ways.[28,29] A complex table of data may provide all the necessary information, but this display of data may make it difficult to determine whether improvements are being made. In contrast, an annotated run chart, which plots data over time, can more readily provide answers to questions such as, "how far have we progressed toward our goal?", "how much further do we have to go?", "have we sustained our gains?" and "what change, if any, resulted in an improvement?" Annotating run charts to indicate key events (eg, time the project team started or new change introduced) and applying statistical process control limits to distinguish random (noise) from nonrandom variation (signal) add useful detail to the visual display of performance data. These enhancements help distinguish true changes in processes or outcomes from the expected variation seen in stable systems over time.

The Messenger

Respected leaders, often described as "champions," are frequently the most effective messengers for delivering performance data to frontline clinicians and staff.[27,30] For example, when physician leaders provide feedback to other physicians, it provides validation that the data came from colleagues with shared interests in promoting, as well as shared barriers to providing, optimal clinical care. In contrast, when data originate from an external body, such as a QI department or payer, clinicians may be more likely to view the data as invalid, punitive, or simply irrelevant. The importance of identifying messengers who can convey the feedback in honest and motivational ways cannot be overstated and is consistent with the literature on the effectiveness of academic detailing (the use of clinical expert outreach visits to physician practices).[31]

ASSURING SUSTAINABILITY OF ONGOING MEASUREMENT

Achieving sustainability of QI projects depends on continuing to measure key processes and outcomes. For QI teams, measurement provides a source of learning during implementation and a method of maintenance after implementation. Some of the measures developed and used in testing and implementation should be considered for ongoing use in the organization when the QI team disbands. However, the measures may be reported and analyzed less frequently during this ongoing monitoring phase. Measuring over time allows an organization to determine whether it is continuing to get the desired results and whether those results can be predicted to continue in the future. This process addresses the question, "Is there a need to update the process or make new changes?"

To prepare for this continued use of key measures, the QI team should consider strategies such as these:

- Incorporate measurement with another existing work activity;
- Use existing data collection systems or develop easy-to-use data collection forms;
- Build measurement into EMRs, registries, or other data systems so that it is easy for administrative staff and clinicians to use as part of usual clinical care;
- Clearly define roles and responsibilities for ongoing data collection after the QI project is complete;
- Set aside time to review data with those who collect it before completion of the QI project. This will help everyone understand how the data is being used, and provide reinforcement for their efforts.

Although run charts of the key measures are effective for learning during an improvement project, plotting key measures on a statistical process control chart is more useful for maintaining a change. These charts provide signals to detect whether the process is beginning to deteriorate (eg, staff reverts to practices used before the improvements were made). In addition, control charts allow teams to predict future expected performance. Statistical process control charts can be maintained for the key measures used in any improvement effort. **Fig. 2** shows an example of a control chart[32] for a measure of harm during hospital admission. These maintenance charts will provide signals if the process ever begins to deteriorate, or, in this example, if data exceeds the upper control limit.

MEASUREMENT FOR SUPPORTING ORGANIZATIONAL PERFORMANCE

Just as measures are used to identify performance gaps and set priorities for improvement at the project level, health care organizations, including clinical practices, networks, hospitals, and health systems, also need a balanced set of system-level measures to track progress toward their strategic goals. Measures guide the direction and focus of QI efforts across the organization and should complement other system-level performance measures (eg, finance, use).[33] Collectively, this set of measures should serve as both a gauge of current performance and as input into the future direction of the organization.[34] Leaders and governing boards need to be actively involved in selecting and developing a dashboard of quality measures that reflects the culture and mission of their organization.[30,35] To improve quality at the system-level, senior leadership needs to be committed and accountable to improving (or maintaining) performance on these measures, and prepared to build the will needed to drive change throughout the organization.[30,36]

Organizational performance measures should help leaders understand their progress toward accomplishing their mission. Leadership should identify a limited set of approximately 10 to 20 high-level measures that are focused on what the organization
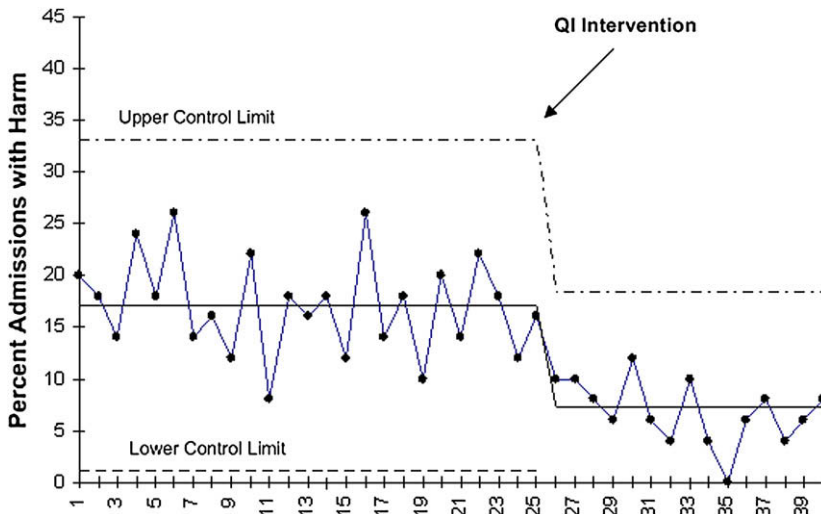
Fig. 2. Statistical process control chart showing an improved process that can be monitored for sustainability. The chart has an annotation marking the beginning of the QI intervention.

wants to accomplish and are balanced from the perspective of organization stake-holders. Dashboard measures should endure year to year; however, the strategic priorities of the organization will often focus on one or two measures each year.

This dashboard of measures should include operational definitions that can be easily understood at all levels of the organization, from the leadership to front-line clini-cians and staff (see **Table 3** for an example of a children's hospital dashboard). Like project measurement, organizational measurement must be timely (no more than a month's lag between data and review) and it should not require excessive data collection, which hinders sustainability. The data should be extracted and graphed to show patterns and trends so that improvements can be tracked over time.[30,37] Following these principles ensures that leaders and the governing board of the health care system can continuously monitor and respond to data, and keep the organization moving toward meeting its goals.

Another question that system-level performance measures should help to answer is: "How do we compare with others?" Unlike the mission question stated above, which should be asked at every leadership and board meeting, this question should be ad-dressed annually when reevaluating the organization's strategic goals.[38] Bench-marking against peer organizations can help determine how the system is comparatively performing and it can identify opportunities for learning from other best practices.[37]

Table 3		
An example of a children's hospital dashboard, which tracks the organization's performance.		
Strategic Category	**Measure**	**Operational Definition**
People	New hire nursing turnover	% nurses leaving within 12 months of hire
	Faculty overall satisfaction	% rating "very good"
	Staff overall satisfaction	% rating "very good"
Service	Patient/family overall satisfaction	% recommending UNC
Clinical Quality	Adverse events	Adverse drug events per 1000 doses
	Hospital associated infections	# infections per 1000 pt days
	Raw mortality	# inpatient deaths
	Cardiac arrests	# cardiac arrests per 1000 pt days
	Chronic care management	# patients in chronic care registry
	Readmission rate	# patients readmitted within 72 hours
Finance	Payor mix	% distribution of gross revenues across insurance plans
	Gross charges	Actual cash collection on bills
	Fundraising	Amount of gifts per year
	RVU	# relative value units (inpatient and outpatient)
Innovation	Publications	# faculty publications per quarter
	Family advisory groups	# active family advisory groups
Growth	Hours on diversion	# hours PICU unable to accept new admissions
	Inpatient admissions	# inpatients admitted to wards, PICU and NICU
	Inpatient patient days	# inpatients in a bed
	Inpatient length of stay	Average # days patients stay in hospital

Courtesy of North Carolina Children's Hospital, Chapel Hill, NC; with permission.

Tracking organizational performance on quality measures is increasingly relevant to the financial security of practices, networks, hospitals, and health systems. These measures are being incorporated into federal data reporting standards, such as The Joint Commission and the CMS Core Measures. CMS will link hospital reimbursement to organizations' performance on the core measures, a strategy known as "pay for performance," an approach that numerous payors have adopted or will soon adopt.[39,40]

SUMMARY

Measurement and feedback are fundamental aspects of QI. The authors have described a pragmatic approach to measurement and feedback for QI efforts in local health care settings, including hospitals and clinical practices. The authors included evidence-based strategies from health care and other industries, augmented by their collective practical experience designing measurement and feedback strategies. The authors also described an approach to developing, testing, and implementing measurement and feedback strategies during the stages of a typical QI project.

The process as described here will assist health care professionals in knowing how to measure the effects of their QI projects and how to use these data to improve care. Health care professionals will need to understand and know how to apply the principles that are summarized in this article in order for hospitals and clinical practices to meet the growing demand to dramatically improve their performance.

ACKNOWLEDGMENTS

REFERENCES

1. Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academy Press; 2001.
2. Nelson EC, Mohr JJ, Batalden PB, et al. Improving health care, part 1: the clinical value compass. Jt Comm J Qual Improv 1996;22:243–58.
3. Solberg LL, Mosser G, McDonald S. The three faces of performance measursement: improvement, accountability and research. Jt Comm J Qual Improv 1997; 23(3):135–47.
4. Codman EA. A study in hospital efficiency (1917). Reprinted by the Joint Commission on Accreditation of Healthcare Organizations. Illinois: Oakbrook Terrace; 1996.
5. Shewhart WA. The economic control of quality of manufactured product (1931). Milwaukee (WI): Reprinted by ASQC; 1980.
6. Deming WE. Out of the crisis. Cambridge (MA): Massachusetts Institute of Technology; 1986.
7. Donabedian A, Bashshur R. An introduction to quality assurance in health care. New York: Oxford University Press; 2003.
8. Kenney C. The best practice: how the new quality movement is transforming medicine. New York: Public Affairs; 2008.
9. Forrester J. Principles of systems. Cambridge (MA): Productivity Press; 1986.
10. Nolan TW. Understanding medical systems. Annals of Internal Medicine 1998; 128(4):293–8.

11. Deming WE. The new economics for industry, government, and education. 2nd edition. Cambridge (MA): Massachusetts Institute of Technology; 1994. p. 92–115.

12. Senge P. The fifth discipline: the art & practice of the learning organization. New York: Doubleday; 1994.

13. Langley G, Nolan K, Nolan T, et al. The improvement guide: a practical approach to enhancing organizational performance. San Francisco (CA): Jossey-Bass Pub; 1996.

14. Shewhart WA. In: Deming WE, editor. Statistical method from the viewpoint of quality control (1939). New York: Dover Press; 1986. p. 1–15.

15. Joint Commission on Accreditation of Health Care Organizations. Attributes of core performance measures and associated evaluation criteria. 2008. Available at: http://www.jointcommission.org/PerformanceMeasurement/Performan ceMeasurement/. Accessed November 18, 2008.

16. NQMC Agency for Healthcare Research and Quality. National quality measures clearinghouse. 2008. Available at: http://www.qualitymeasures.ahrq.gov/. Accessed November 18, 2008.

17. Association of Public Health Observatories. The good indicators guide: understanding how to use and choose indicators. NHS Institute for Innovation and Improvement. 2008. Available at: http://www.apho.org.uk/resource/item.aspx?RID=44584. Accessed November 18, 2008.

18. Christakis DA, Wright JA, Zimmerman FJ, et al. Continuity of care is associated with well-coordinated care. Ambul Pediatr 2003;3(2):82–6.

19. National Quality Forum. 2008. Available at: http://www.qualityforum.org/. Accessed November 18, 2008.

20. Institute for Healthcare Improvement. 2008. Available at: http://www.ihi.org. Accessed November 18, 2008.

21. National Initiative for Children's Healthcare Quality. 2008. Available at: http://www. nichq.org. Accessed November 18, 2008.

22. American Academy of Pediatrics Steering Committee on Quality Improvement and Management and Committee on Practice and Ambulatory Medicine. Principles for the development and use of quality measures. Policy statement. Pediatrics 2008;121:411–8.

23. Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. JAMA 2006;295(10):1168–70.

24. Reeves D, Campbell S, Adams J, et al. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. Med Care 2007;45(6): 489–96.

25. Nelson EC, Splaine ME, Plume SK, et al. Good measurement for good improvement work. Qual Manag Health Care 2004;13(1):1–16.

26. Pronovost PJ, Nolan T, Zeger S, et al. How can clinicians measure safety and quality in acute care? Lancet 2004;363:1061–7.

27. Bradley EH, Holmboe ES, Mattera JA, et al. Data feedback efforts in quality improvement: lessons learned from US hospitals. Qual Saf Health Care 2004; 13(1):26–31.

28. Carey RG, Lloyd RC. Measuring quality improvement in healthcare: a guide to statistical process control application. Wisconsin: Quality Press; 2001. p. 43–150.

29. Tufte ER. The visual display of quantitative information. 2nd edition. Cheshire (CT): Graphics Press; 2001.

30. Denham CR. Leaders need dashboards, dashboards need leaders. J Patient Saf 2006;2(1):45–53.

31. O'Brien MA, Rogers S, Jamtvedt G, et al. Educational outreach visits: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2007;4:CD000409. DOI: 10.1002/14651858.

32. Lloyd R. Quality health care: a guide to developing and using indicators. Boston: Jones & Bartlett Publishers; 2004.

33. Martin LA, Nelson EC, Lloyd RC, et al. Whole system measures. IHI innovation series white paper. Cambridge (MA): Institute for Healthcare Improvement; 2007.

34. Provost L, Leddick S. How to take multiple measures to get a complete picture of organizational performance. Natl Prod Rev 1993;12(4):477–90.

35. Kroch E, Vaughn T, Koepke M, et al. Hospital boards and quality dashboards. J Patient Saf 2006;2(1):10–9.

36. Conway J. Getting boards on board: engaging governing boards in quality and safety. Jt Comm J Qual Patient Saf 2008;34(4):214–20.

37. Reinertsen JL, Pugh MD, Bisognano M. Institute for Healthcare Improvement. IHI innovation series. Seven leadership leverage points for organization-level improvement in health care. Cambridge (MA): Institute for Healthcare Improvement; 2005.

38. Reinertsen JL. From the top: getting the board on board. IHI conference: boards, dashboards, and data. Boston. 2007. Available at: http://www.ihi.org/IHI/Topics/LeadingSystemImprovement/Leadership/EmergingContent/BoardsDashboardsData.htm. Accessed November 10, 2008.

39. The Joint Commission. Performance measurement initiatives. Available at: http://www.jointcommission.org/PerformanceMeasurement/PerformanceMeasurement/default.htm. Accessed November 10, 2008.

40. Kaiser Daily Health Policy Report. Medicare stops paying for 10 reasonably preventable medical errors. Available at: http://www.kaisernetwork.org/daily_reports/rep_index.cfm?hint=3&;DR_ID=54758. Accessed October 1, 2008.

GLOSSARY

Balanced set of measures: A set of measures which, taken together, reflect as much of a system as possible without duplication, overlap or gaps.[17]

Benchmark: An externally agreed-upon comparator to compare performance between similar organizations or systems.[17]

Composite indicator: An aggregation of numerous indicators that aims to give a one-figure indicator in order to summarize measures further.

Control charts: A graphical tool for displaying the results of statistical process control.

Control limits: Define the area representing random (also called "common cause") variation on either side of the centerline, plotted on a control chart.

Dashboard: A tool used for collecting and reporting data on system-level measures that demonstrate the overall quality of a health system over time. Dashboards provide a quick summary of structural, process and outcome performance.

Plan-Do-Study-Act (PDSA) Cycle: A quality improvement method consisting of four repetitive steps for learning and improvement. Plan: develop a plan to learn, test, or implement a change; Do: execute the plan; Study: compare predictions to results and document learning; Act: make changes based on learning and set up next cycle. Most changes require several PDSA cycles.

Run chart: A time series graph where the X-axis represents time longitudinally and the measure value is on the Y-axis. Run charts often include a median for the data points and can be augmented by inserting comments (annotations) at the point in time where process changes are made by the improvement team or other changes occur that could affect the data (outside of those made by the improvement team, such as a sudden loss of several staff members in a clinic).

Statistical process control (SPC): Statistical analysis and display (eg control charts), which helps distinguish normal, everyday, inevitable variation ("common cause variation") from nonrandom ("special cause variation") variation. The latter indicates something special is happening which could be caused by an improvement project or something that warrants a fuller understanding and investigation.[17]