

Good Measurement for Good Improvement Work

*Eugene C. Nelson, DSc, MPH; Mark E. Splaine, MD, MS;
Stephen K. Plume, MD; Paul Batalden, MD*

Purpose: To provide guidance on using measurement to support the conduct of local quality improvement projects that will strengthen the evaluation of results and increase their potential for publication. **Target Group:** Individuals leading quality improvement efforts who wish to enhance their use of measurement. **Procedures to Promote Good Measurement:** Eleven procedures are offered to promote intelligent measurement in quality improvement research that may become publishable:

1. Start with an important topic
2. Develop a clear aim statement
3. Turn the aim statement into key questions
4. Develop a theory about causes and effects, process changes and predictable sources of variation
5. Construct a research design and accompanying dummy data displays to answer your primary research questions
6. Develop and use operational definitions for each variable needed to make your dummy data displays
7. Design a data collection plan to gather information on each variable that will enable you to generate reliable, valid, and sensitive measures related to each research question
8. Pilot test the data collection plan, construct preliminary data displays, and revise your methods based on what you learn
9. Stay close to the data collection process as the data plan goes from idea to execution
10. Perform data analysis and display results in a way that answers your key questions

11. Review and document the strengths and limitations of your measurement work and use this knowledge to guide intelligent interpretation of the observed results.

Key words: *measurement and continuous improvement, outcomes measurement, performance measurement, quality improvement, quality improvement research, research design and improvement*

This article is written primarily for people who are leading quality improvement efforts and who not only wish to know if the actions taken result in substantial, measured improvement, but also wish to do so with sufficient rigor to publish the results of their work. It is possible, under many conditions, to transform local quality improvement “projects” into a publishable quality improvement project by paying close attention to the fundamentals—the study question, the research design, the measurement and analysis of results, and sources of bias and limitations. The purpose of this article is to provide guidance on using measurement, and related research design principles, to support the conduct of local clinical quality

From the Dartmouth Medical School (Drs Nelson, Splaine, and Plume), Quality Education, Measurement and Research (Dr Nelson), Measurement and Statistics, Dartmouth-Hitchcock Clinic (Dr Splaine), Dartmouth-Hitchcock Medical Center, Lebanon, NH; and the Pediatrics and Community and Family Medicine and Health Care Improvement Leadership Development, Dartmouth Medical School, Hanover, NH (Dr Batalden).

Corresponding author: Eugene C. Nelson, DSc, MPH, One Medical Center Drive, Lebanon, NH 03756 (e-mail: eugene.c.nelson@hitchcock.org).

This study was supported by a conference grant from the Agency for Healthcare Research and Quality (1R13HS1008601).

improvement projects in ways that will strengthen the work and, if desired, increase the potential for publication in the health care literature.

The article is divided into 2 sections. The first presents 11 procedures to promote sound measurement to support improvement work and the second provides answers to frequently asked questions about measurement and related design issues.

PROCEDURES TO PROMOTE GOOD MEASUREMENT IN QUALITY IMPROVEMENT PROJECTS

In this section we discuss 11 procedures to promote intelligent measurement in quality improvement projects. The main focus of this article is on measurement to enhance local quality improvement endeavors. However, the measures, to be useful, must be embedded in a project that is worth doing and that makes use of a research design that stands muster. Furthermore, the measures need to be used in an analysis plan that fits the conditions and answers the critical questions. It is wise, therefore, to deal with the issue of measurement—and the context in which it is used—rather than to focus only on measurement as a stand-alone topic. For the purpose of illustration, the section below makes extensive use of an illustrative clinical improvement topic—the care of end stage renal disease patients who are on hemodialysis.

Start with a topic that is clinically important

Some examples:

- The process of care for renal dialysis patients and its impact on outcomes
- The provision of evidence-based care to children in the first 3 years of life
- The care of frail homebound elderly
- The treatment of cardiovascular disease
- The treatment of low birth weight infants
- The prevention of the cigarette smoking in adolescents
- The safe delivery of healthy infants
- The identification and treatment of depression in young adults

Good improvement work and measurement thereof should be used to address important topics—things that care beneficiaries (especially patients, purchasers, employers, and community) truly value. If the focus of the improvement project is on issues that are vital to customers, it is more likely to attract the interest, support, and resources needed to see the project through. Improvement resources are perpetually limited. Because there is competition for time, and energy and staff and so forth, it is prudent to invest improvement resources on important outcomes that people value highly.

Begin the improvement work by selecting processes that produce important outcomes and where there is substantial opportunity for improvement. By “important outcomes” we mean key results: health outcomes, patient perceptions of goodness, costs incurred, and errors that result in harm. By “substantial opportunity” we mean a wide gap between the level of current performance achieved and the best possible performance that could be achieved with a redesigned process.

There is no shortage of ways to find “leads” on potentially important topics to target for improvement. They can be found by

- asking your patients, family members of your patients, your referral sources, your staff, and the employers of your patients
- reviewing data on current performance with respect to quality, satisfaction, costs, and safety and taking advantage of comparative data
- looking to professional organizations, regulatory agencies, and accrediting organizations to familiarize yourself with what they believe needs betterment
- analyzing the epidemiologic profile—sources of mortality, morbidity, and impaired functioning—of the patients/community that you serve and setting priorities based on the potential for reducing the burden of illness and injury

Develop a clear aim statement to align measures with achievement of aim

Aim statement example: We aim to improve the process of dialyzing patients with end stage renal

disease (ESRD) at Greenview Medical Center. The process begins with the decision to commence treatment on appropriately selected patients and the process ends when dialysis for ESRD is terminated. By working on this process we hope to improve mortality, morbidity, functional status, patient satisfaction with services, and to reduce the total costs of providing care for this subpopulation of patients. It is important to work on this now because comparative data reveals that the case mix adjusted mortality is greater than expected, because the costs of care exceed revenues and because staff morale is low.

After identifying an important topic, it is time to develop a finely crafted statement of your improvement aim. A good aim statement will do 3 things: (a) attract interest, (b) provide clear direction by specifying the process and the geographic locus of work, and (c) suggest yardsticks by which to measure success.

One way to construct an aim statement is to use a particular structure that covers several relevant points. The anatomy of a “structured” aim statement looks like this:

- We aim to improve (*insert the name of the process*)
- The process begins with (*insert the starting point of the process*)
- The process ends with (*insert the end point of the process*)
- By working on this process we hope to (*list the potential gains that can be made with respect to health outcomes, satisfaction, costs, safety, staff morale, etc*)
- It is important to work on this now because (*list the reasons why improvement is imperative*)

It often takes some time to craft an aim statement that is succinct, well focused, provides direction, and suggests a basis for judging success. The aim statement, when written with sufficient precision, will suggest what variables must be measured to evaluate success. As an example, the ESRD aim statement cited above calls explicitly for the measurement of certain variables in a sample of renal dialysis patients: clinical outcomes (mortality and morbidity), functional status outcomes, patient satisfaction with

renal dialysis services, and costs of medical care for renal dialysis patients (renal costs and total costs). The aim statement also suggests that it will be necessary to analyze the outcomes taking case mix variation into account (eg, through adjustment and/or through stratification).

It is helpful to test the merit of your aim statement by imagining that you will be presenting your aim statement to diverse groups—eg, medical staff, senior leadership, your colleagues, the board of trustees, the local rotary club, local employers and purchasers, editors of a professional journal—to determine if they would find your aim important, measurable, and worth pursuing.

Turn the aim statement into key questions (that people care about)

At this point the issue of measurement moves out of the background and into the foreground. Measurements in improvement projects are generally intended to answer critical questions regarding the impact of a planned intervention or test of change. The ESRD aim statement leads to some key questions that can be answered using planned tests of change. For example:

- Is the ESRD improvement process associated with a decrease in morbidity and mortality among the subpopulation of patients exposed to the intervention after controlling for case mix differences?
- Is the ESRD improvement process associated with an increase in functional health among the subpopulation of patients exposed to the intervention after controlling for case mix differences?
- Is the ESRD improvement process associated with an increase in patient satisfaction with dialysis services among the subpopulation of patients exposed to the intervention after controlling for case mix differences?
- Is the ESRD improvement process associated with a decrease in the total direct medical costs among the subpopulation of patients exposed to the intervention after controlling for case mix differences?

Develop a theory about causes and effects and process changes and predictable sources of variation

Improvement work invites scientific thinking and the use of the PDSA (plan-do-study-act) model to test changes in processes. The use of the scientific method is based on making and testing theories and replicating or sustaining results. Quantitative research methods offer a rigorous and repeatable process for learning about causes and effects and making predictions that can be tested once and replicated once again in the same location and in different locations. Consequently, the next step is to build a theory about causes and effects.

If we study the aim statement and the related research questions we can start to construct a basic theoretical framework about causes and effects. A good way to develop your theory is to construct a causal model or system diagram. The general structure of the potentially testable idea—or the theory—can be illustrated using an arrow diagram or a causal system diagram. A framework for viewing changes in clinical care processes may look like this:

**Patients with initial case mix characteristics →
Care delivery processes → Outcomes of care**

This clinical causal model suggests that patients with a set of relevant case mix characteristics at baseline, enter into a longitudinal care delivery process with multiple elements, and this produces variation in outcomes of care at a later point in time.

Using the ESRD patients as an example, we might construct a theory to indicate how patients with certain demographic and clinical characteristics at the time of initial treatment will receive a certain care delivery process with prescribed critical elements that will be associated with variation in a selected set of outcomes.

“Clinical value compass” thinking can often be used to build causal models to clarify clinical improvement work and provide a basis for selecting measures related to case mix, processes, and outcomes of care.¹

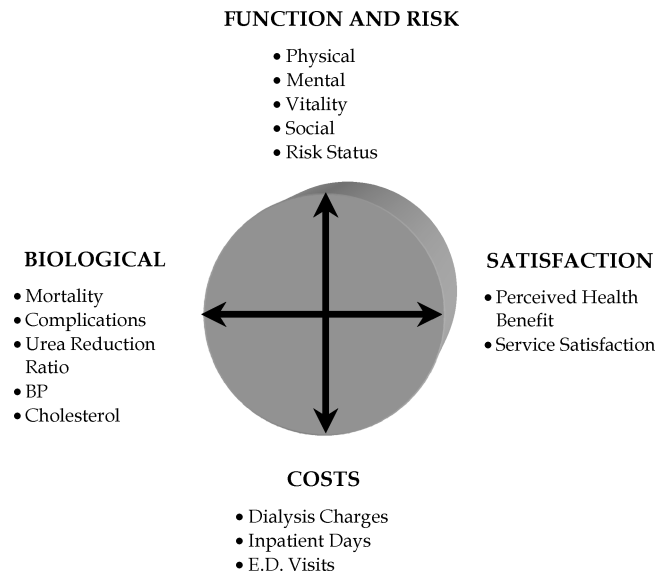


Figure 1. Value compass for end stage renal disease (ESRD) patients.

Figure 1 illustrates a clinical value compass for ESRD patients. It suggests that patients with certain baseline characteristics (such as age, gender, severity, and comorbidity) will have better outcomes (mortality, morbidity, functional status, satisfaction, and medical care costs) if their treatment process is standardized in two critical ways:

- the achievement of a specified level of dialysis for each treatment as measured by the urea reduction ratio levels achieved at the end of each dialysis session, and
- establishing a standard method to ensure that each patient’s treatment plan includes explicit matching of the patient’s major medical problems (such as hypertension, hyperlipidemia, diabetes, heart failure, etc) with a specific treatment plan that the patient complies with.

Development of a general theory about the process of dialysis can now be turned into a research question or hypothesis. A working hypothesis can be stated like this:

Dialysis patients who are treated with a protocol that ensures specified levels of urea reduction ratio for each dialysis session and specific treatment of each major clinical condition will have better functional

status and survival rates than patients who are not treated with a protocol that ensures specified levels of urea reduction ratio and tailored treatment of comorbid problems

The act of writing down a hypothesis or research question is important because it invites one to look at the statement analytically. Using value compass thinking, we can see immediately that the hypothesis does not reference 2 potentially important classes of results—patient satisfaction with care and costs of care. Additional hypotheses could be formed to address these issues.

Construct a research design—and accompanying dummy data displays—that will answer your primary research questions, and consult with a statistician

Good measurement in a poor study design makes interpretation of results difficult. Consequently, it is always smart to use the most powerful research design that can be used under local conditions.

Outside of health care, quality improvement efforts frequently make use of advanced research designs to accelerate learning.² Unfortunately, the use of experimental designs in quality improvement projects performed in health care settings is relatively rare, although there are several important exceptions to this general statement.^{3–5} Please refer to the article by Speroff in this series for a full discussion of this topic.⁶

To illustrate, we will focus on a quality improvement project designed to improve the outcomes and reduce the costs of care for dialysis patients. A specific test of change is ready to be conducted. The test of change is to use a protocol for dialysis that establishes target values of urea reduction ratio for each dialysis session for each patient. The hypothesis is that—other things being constant—the use of the protocol will be associated with better levels of urea reduction ratio and that, over time, this will produce lower rates of mortality and improved physical function.

We are now at challenging stage of the improvement project. How can the study question be em-

bedded in a research design to test the protocol based on accurate, relevant measures? The challenge is to get the best feasible research design and the best feasible measures given the constraint that the work has to be done in one’s own “real world” and with dispatch. The suggestion is to start by thinking about the strongest research design and best measures and then to “give ground” until the best feasible design and metrics are arrived at. This helps one to know the limitations inherent in the adopted methods.

As a tactical way to move forward, we first deal with the research design issue and then move to dummy data displays.

- *Research design thinking:* First, we consider a randomized controlled trial (RCT) approach but there is too much local resistance and logistical complexity. Second, we consider a time series design and begin to give thought to how this might be done in our setting. We could use a prospective cohort design where one set of 40 consecutive new-to-dialysis patients receive “customary” care (we call this group the preprotocol cohort) and a second set of 40 consecutive new-to-dialysis patients receives “protocol” care (we call this group the protocol cohort). Historical trends show that the dialysis center takes in 5 new patients per month on average; consequently, the first cohort would take 8 months to induct and the second cohort another 8 months. This is not the strongest design but it does provide substantial control and affords the team the time to phase in the “new way” and to establish the data collection procedures sequentially—first the measurement of urea reduction ratio, second the measurement of functional status at 2 months postonset of dialysis, and third, the assessment of survival rates over time.
- *Dummy data displays:* With a specific design calling for a specific number of patients in mind, we now proceed to construct the vital few dummy data displays. Dummy data displays are fictional displays of key results. Make them by acting as if the test of change has been

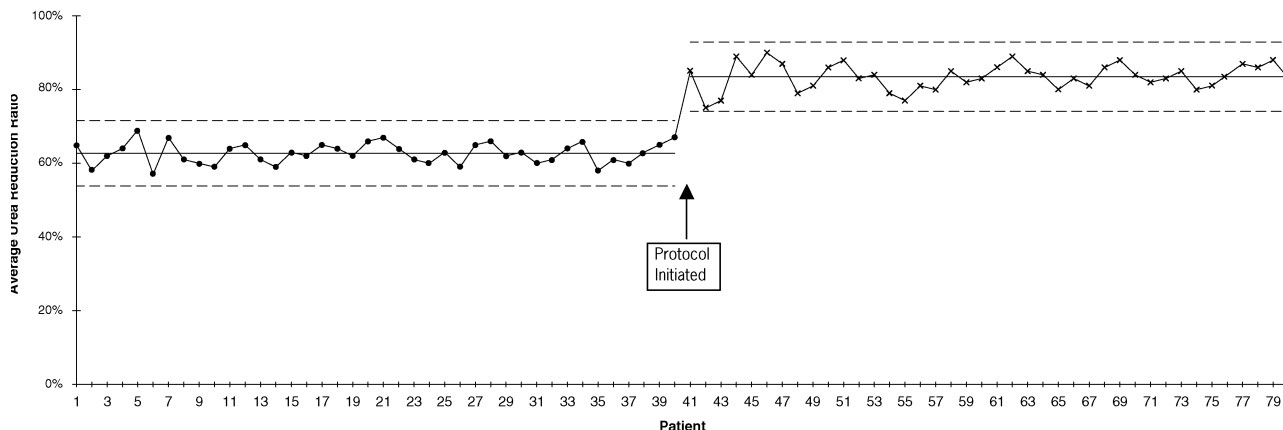


Figure 2. X-bar and R chart of data for average urea reduction ratio for dialysis patients. Each data point represents the mean of 4 dialysis treatments. Values for 40 consecutive preprotocol patients (1–40) are represented by an “O” symbol. Values for 40 consecutive protocol patients (41–80) are represented by an “X” symbol.

completed and that the project’s hypothesis is confirmed or rejected. Figure 2 shows a dummy data display for urea reduction ratio, Figure 3 provides a dummy data display for physical function score at 2 months post onset of dialysis, and Figure 4 shows a survival curve that tracks longevity in each cohort. It is useful to set up dummy data displays by posing a critical question and then making a data display to answer it.

The fictional results shown in Figure 2 are designed to suggest that the use of the protocol does achieve superior levels of urea reduction ratio. The fictional results in Figure 3 show that the physical function scores are higher for protocol patients than for pre-protocol patients. The fictional results in Figure 4 suggest that survival is greater for protocol patients at 24 months.

Once again, the act of committing ideas to paper, in this case creating dummy data displays, invites

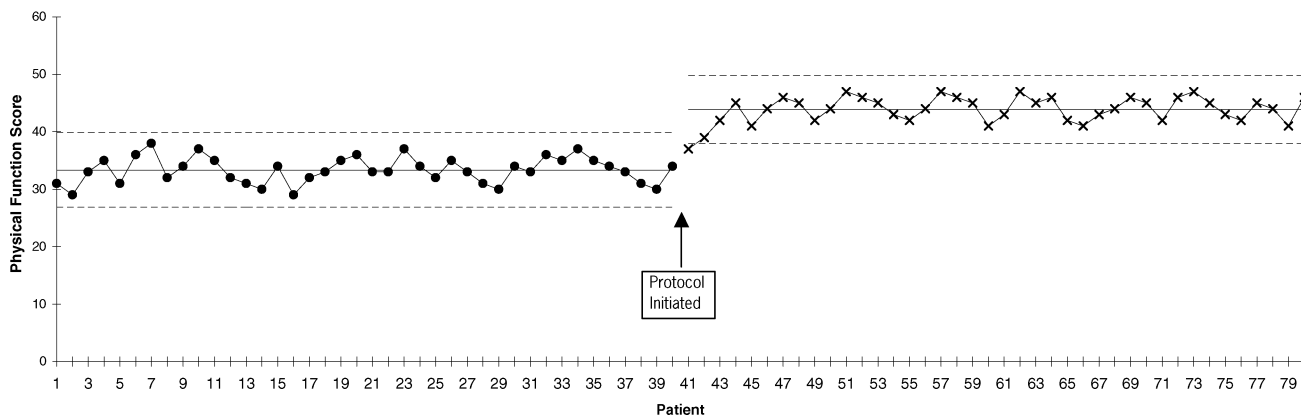


Figure 3. XmR control chart of data for physical function score from SF-36. Data on left represents 40 consecutive pre-protocol patients (1–40) at 2 months postonset of dialysis indicated by “O” symbol. Data on the right represents 40 consecutive protocol patients (41–80) at 2 months postonset of protocol indicated by “X” symbol.

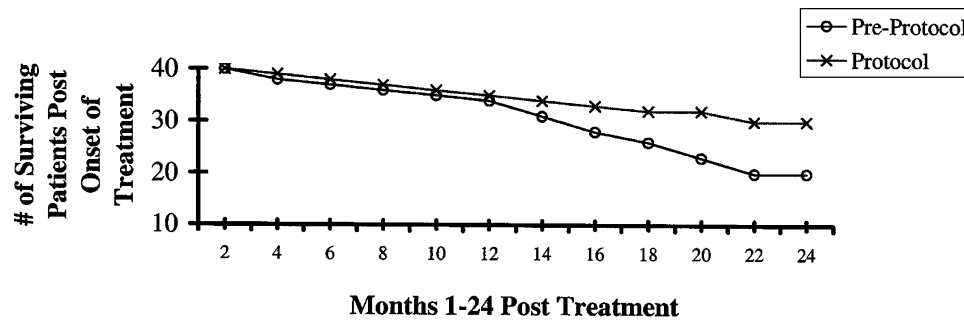


Figure 4. Survival rate of protocol patients. This display shows the number of patients out of 40 that have survived over a 24 month period post the onset of dialysis. The number of preprotocol patients surviving are denoted by the symbol “o” whereas the number of protocol patients surviving are denoted by the symbol “X”.

reflection and critical analysis. We can view these fictional results as if they are real results and ask what else would we want to know about these findings? We can ask what are the sources of variation in these results that we may wish to explore, to document and to understand. In this case we immediately recognize that the individual patient’s demographic and clinical characteristics and disease severity are likely to produce variation in these outcomes. Consequently, we wish to make plans to display the results in some different ways. First, we can display the results as actual or observed (ie, unadjusted) findings, and second as adjusted findings based on the use of multivariate analysis to control for the effects of selected variables such as age, gender, comorbidity, and disease severity. Second, we could display results based on stratification methods that could show the findings separately for 2 different subcategories of patients (ie, patients with lower case mix acuity vs patients with higher case mix acuity).

At this point it is prudent, if possible, to begin working with a statistician. A trained statistician can provide guidance on both research design issues as well as on sample size considerations and on analytical methods. It is best to find a statistician who is knowledgeable about conventional biostatistics as well as with statistical process control methods so as to take advantage of both methods of data analysis. If a statistical consultant is not available in your organization, you may be able to get assistance in your

community at a college or university or by using reference materials.^{7–10}

Develop and use operational definitions for each variable needed to make your dummy data displays

Having constructed a local theory, formulated a working hypothesis, specified a particular research design, and created dummy data displays, we are now ready to construct definitions for each variable that will be used in the improvement project. It is helpful to think about definitions of study variables at two levels of precision—conceptual definitions and operational definitions.

A conceptual definition offers a concise statement describing the variable that you wish to measure. It tells others what is meant by the name of the variable. An operational definition, on the other hand, provides a more detailed explanation of not just what the variable is but how observations will be classified to yield accurate and reproducible values.

Examples of conceptual and operational definitions for each of the 3 variables stated in our working hypothesis are listed below:

- *Urea reduction ratio*: Urea reduction ratio is an indication of the adequacy of dialysis. The urea reduction ratio is calculated with the formula $100 \times (1 - [C_t/C_0])$, in which C_t is the blood urea nitrogen measured 5 minutes after the end of dialysis and C_0 is the predialysis blood urea nitrogen. It will be measured during each dialysis

session by the dialysis nurse who records the value.^{11,12}

- *Physical function*: Physical function reflects the individual's capacity to perform physical activities that are a normal part of daily life. It will be measured based on the SF-36 physical function component score subscale. The SF-36 will be administered to dialysis patients after 8 weeks of therapy and will be scored using the protocol specified in the SF-36 procedures manual.^{13–15}
- *Survival*: Survival status indicates whether the dialysis patient is alive or dead at selected intervals after the onset of dialysis therapy. It will be recorded every 2 months by the dialysis service director based on continued participation in treatment; patients who cease treatment will be tracked to determine their status (eg, moved from area, transplant, other).

It is critically important to establish clear operational definitions and to use them. The “signals” that are emitted from each observation on each variable and that are subsequently aggregated into summary values for variables literally provide the evidence on which decisions are made about the confirmation or rejection of the hypothesis. Interpretation of results can be no better than the coding of the values that produce the results.

Design a data collection plan to gather information on each variable that will enable you to generate sufficiently reliable, valid, and sensitive measures related to each key question

The data collection plan provides the blueprints to guide the construction of the measurement system that supports the improvement enterprise. The data collection plan indicates who will gather what data under what conditions at what points in time with what tools and training. The data collection plan is constructed to minimize bias and to maximize accurate measurement. The data collection plan includes the specification of the tools and techniques for gathering data. The data collection plan indicates how the sampling plan should be carried out in concrete, operational terms. The data collection plan should take into consideration what steps should be taken to

validate the accuracy and completeness of data collection procedures.

Pilot test the data collection plan and construct preliminary data displays then revise your methods based on what you learn

The data collection plan should always include a small-scale pretest to check and refine the methods of data collection under actual operating conditions. First, gather data on 10 to 15 observations, events or people; this will usually be sufficient to determine whether or not the data collection plan will work as intended. The results of the pilot test will show whether or not data for each variable are missing or complete and whether or not people who need to be involved in gathering data are able to perform according to plan. Second, analyze this small data set. For example use the results from 10 events of interest, to check how the data for each variable are distributed (normal, skewed, bimodal, etc), and to fill in the dummy data displays with real data. This step is important because it enables you to determine precisely how each variable should be handled to create the data displays that will answer your key questions. For example, you will be able to determine cut points that should be used for variables, how to handle missing observations, whether or not coding conventions are working well or need further refinement, and so forth. This is an exciting phase of the measurement process because it is the first time that the abstract plan becomes transformed into tangible results.

Stay close to the data collection process as the data plan goes from idea to execution

When one is conducting a laboratory experiment, it is common practice to maintain a laboratory notebook to record observations as the work unfolds. So too in quality improvement it is good operating procedure to stay close to the intervention and the accompanying data collection process and to make note of what you learn. While “in the field” with a test of change it is helpful to take these actions:

- Observe the change to determine if the intervention is proceeding according to plan or deviating in certain specific ways

- Talk with the people involved in the change to learn of their impressions about what is working well, what is not working well, and ways that the process could be improved
- Clarify the intention of the work, and the methods for conducting the intervention as needed
- Study the process of recording the data to determine if observations are being recorded in a timely and accurate manner
- Continue to devise new ways to design a “new new” test of change and to improve the process of capturing data on performance

By getting close to the action, and by keeping in mind the intended intervention and the companion data collection activity vis a vis the actual intervention and data collection activity, you will benefit in 2 ways. First, you will have a better sense as to why results turn out the way they do. Second, you will gain valuable knowledge that can be used to improve the next cycle of planning and testing a change.

Perform data analysis and display the results in a way that directly answers your key questions

Before performing the data analysis, it is first necessary to edit and clean the data. This is done to ensure that missing data are minimized and that all observations are recorded using “in-range” coding conventions. For example, if a variable has 2 coding categories such as 0 = disease absent and 1 = disease present, then there should only be 2 classes of results, 0’s and 1’s; all other results are “out of range” and require reconciliation. It is common to verify that data have been recorded correctly. For example, if medical records have been reviewed to ascertain values of key variables a 10% to 20% subsample may be reviewed a second time by another individual to compare the original observations with the second set of observations to evaluate inter-rater reliability. Another method that is used to minimize error is to use redundancy in entering data to avoid data entry error creeping into the data set. If the project involves the use of qualitative data, such as observations or verbatim comments, the qualitative data will need to be coded using established coding conventions to transform observations or comments into ex-

haustive and mutually exclusive categories, which enable these findings to be summarized.

After preparing the data for analysis, one can proceed to the analysis of the data in a manner that enables you to answer your key questions based on locally gathered evidence. At this point, you are ready to make your original dummy data displays real. In general, the analysis follows a 3-step process.

- The first step is to “run the marginals” or in common parlance to tally each variable and to show the frequency distribution for each variable (ie, how many values fall into each possible response category).
- The second step is to describe the study population. The data are analyzed in a manner that answers the question, “How many people of what type took part in the project?” This step generally uses simple univariate statistics to summarize the demographic and clinical characteristics of the population.
- The third step is to test the main hypothesis or key questions. The data are analyzed in a manner that answers the question of whether “variable X is associated with variable Y” or if the beginning of an intervention corresponds with a substantial change in the level of an outcome of interest.

At this point it is wise to again involve a qualified statistician to assist with the work. The third step usually involves the use of bivariate or multivariate methods to show degree of association, and/or level of statistical significance and generally uses conventional statistical techniques such as *t*-tests, analysis of variance, multiple linear regression, logistic regression analysis, and other methods. Improvement projects often make use of statistical process control (SPC) methods that use longitudinal data display methods.¹⁶ This is a powerful family of analysis methods—using run charts, control charts, and time series analysis methods—that can show whether or not a change in the level or variation of an outcome of interest is associated with the onset of a particular intervention or planned test of change. On some occasions, longitudinal analysis techniques are used to display the influence of multiple tests of change on a key outcome of interest. Based on the pattern

of results revealed by your data displays, you will be able to accept or reject your hypothesis and to answer your key questions. Your work is not quite complete, however, because it is important to take one final action—before moving on to your next design or redesign effort.

Review and document the strengths and limitations of your measurement work and use this knowledge to guide intelligent interpretation of the observed results

The last step in the measurement process is to review your work and to diagnose the strengths and weaknesses of your planned intervention, your research design, and your efforts to measure the results. This is a time for analysis, reflection, and critical thinking. Consider sources of bias that may have entered into the work, consider unanticipated events that occurred, be they fortuitous or harmful. Discuss these issues with colleagues that participated in the improvement project and review the notes that were taken during the implementation phase of the intervention. Commit your reflections to writing and use the fruits of this exercise in 2 ways—include your observations in the “discussion” section of your paper that summarizes your quality improvement project and use the learning to improve your next project.

FREQUENTLY ASKED QUESTIONS

There are many questions about using measurement in improvement work. In this section we provide answers to some of the most frequently asked questions. Before proceeding, however, a word of warning. Most questions about measurement in quality improvement work come out of a certain, specific context that conditions a correct and useful response. Thus general questions with general answers may prove to be specifically wrong under certain actual conditions.

Do we need a control group?

Yes. It is impossible to know if a change is an improvement absent a comparison measurement of the

“prechange” state or a relevant comparison group. It follows that you will need to use a conventional pretest–posttest design or a more powerful quasi-experimental design.

If you are working to improve a complex process and if you expect that substantial change will require several iterative tests of change that will be lagged in one after the other, then you may wish to use statistical process control techniques to measure performance over time. One way to do this is to follow this approach to measurement:

- Establish a baseline set of values—using a control chart to establish the upper and lower natural process limits (UNPL and LNPL) and the mean value (\bar{x})—for key outcome measures
- Commence introducing changes into the process, one-by-one, and continue to monitor the value of outcomes over time by continuing to plot points on a control chart
- When individual changes are introduced into the process, make a “note” on the control chart to indicate the onset and name of the change and separate the points on the control chart spatially and continue plotting the subsequent run of values while maintaining the preestablished parameters for upper and lower natural process limits and mean value
- Determine if each change is associated with an improvement in key outcome values by using standard rules of detection for control charts.⁷

How big does the sample need to be?

This question cannot be answered without a specific context. It is prudent to involve a qualified statistician in the project from start to finish. If you are using conventional parametric or nonparametric statistics, you may find that you need 200 to 400 observations per group (ie, test and control group research design) or per observational period (ie, pretest posttest research design). If you are using SPC methods you will need approximately 25 “data points” to establish the parameters (upper and lower natural process limits, and average value) and each data point may represent a subgroup of values collected within a particular time period. One common sampling plan for using

SPC methods is to gather 4 observations per time period to calculate a data point; this would require 100 observations and produce 25 data points with 4 observations per data point. The time needed to gather this data depends on the rate of incoming patients into the clinical setting(s) performing the project.

What are the sources of data and types of variables most often needed to do health care improvement research?

Let’s begin with a simplifying assumption. We will posit that much health care improvement research is concerned with the delivery of care for certain types of patients who will experience a set of outcomes (clinical, functional, risk status, satisfaction) and incur an array of costs.

In prior publications we have described a framework for this common situation that we have termed “clinical value compass” thinking. This concept is illustrated in Figure 1. Using clinical value compass thinking, a subpopulation of patients enters into a treatment episode, a process of care unfolds, and at the end of the treatment episode the patient attains a set of outcomes and incurs a set of costs. Using this general model, one can analyze and measure different aspects of the episode of care—patient characteristics and expectations for care at Time 1, health care process variables associated with the delivery

of care, and patient outcomes and health care costs at Time 2. Given this method of framing the health care delivery process and episodes of care, we can then measure the process of care because this process has the capacity to “throw off” data and one has the potential to “catch” and analyze the data that the process is emitting.

What are some of the different sources of data that are being “cast off” as the process unfolds? There are 3 basic sources of data: the patient, the staff (clinical and administrative staff), and the “archives”—data that is produced as a byproduct of care delivery (eg, scheduling data, medical record data, laboratory data, diagnostic test data, financial and billing data, etc). These sources of data can provide values for the different types of variables that are required to perform improvement projects. The 3 different sources of data and examples of the types of variables that can be obtained from each source are listed in Table 1.

How can data collection be built into the flow of work?

Under some conditions data on needed variables will be available from archival sources (such as scheduling, medical records, diagnostic tests, and billing) and the data will be deemed sufficiently complete and accurate for research purposes. When this

Table 1

COMMON SOURCES OF DATA AND ILLUSTRATIVE TYPES OF VARIABLES

Source of data	Illustrative types of variables
Patient	Symptoms, pain, functional status, behavioral risk status, demographics, satisfaction with care and services, perceived health benefit, time lost from work, disruption of major role activities, verbatim comments, estimates of elapsed times, reports on care plan, reports on adherence to treatment regimens, reports on utilization of health care services such as doctor visits, emergency department visits, hospital stays, etc.
Clinician and staff	Problems, chief complaint, diagnoses, problem severity, signs, physical test results, complications, objective tests of functional status, diagnostic tests ordered, diagnostic test results, treatments ordered, treatment results, disposition, prognosis, adverse events, etc.
Archival records	Utilization of ambulatory, inpatient, and home health services, diagnostic tests ordered, diagnostic test results, appointments scheduled, appointments kept, charges for care received, demographics of patient, adverse events, etc.

is true, it makes sense to use archival data; this data can be gathered retrospectively as a “by-product” of the current delivery process. Sadly, however, the data that is needed to answer key improvement questions is oftentimes either not gathered routinely or is collected but is not sufficiently accurate for the specific purposes at hand. Under these circumstances, data must be gathered prospectively—from patients, clinical staff, or appended to archives—and this, in turn, leads to the need to build data collection into the day-to-day flow of work in the health care delivery unit that is the locus of the work. This creates a challenge—how best to “insert” a new, and acceptable data collection routine into the daily work of busy people who staff the clinical unit. To work through this challenge in a manner that is most likely to produce a successful result (ie, staff ready and able to gather needed data that is complete, reliable, and valid) it may be helpful to take these steps:

- Secure the buy-in and involvement of leadership within the organizational unit
- Visualize the work flow and patient flow in the work unit and construct a “current method” flowchart to depict the flow of patients and work within the unit
- Determine where in the flow of patient care/work it is possible to gather data on core variables
- Specify which type of individual is “there” in the flow of patient care that is in the best position to provide data on core variables
- Work with these individuals to design the best way of “trapping” data as part of the flow of patient care that is minimally disruptive and maximally accurate
- Design a “new method” flowchart to depict the flow of patients and work within the unit that reveals who will do what in the process flow to “trap” and record values for core variables
- Pilot test the “new method” to make sure that it works according to plan using quality improvement tools

A good method to “trap” the needed data is to construct “self-coding” data forms. A self-coding data form is a structured data collection sheet (or template) that lists the variables and response choices in a clear

easy-to-read and easy-to-complete format. The self-coding data sheet should include operational definitions and instructions (perhaps on the backside of the sheet) to make it easy for the person to complete the form correctly.

What’s involved in training and coordinating data collection in improvement projects?

One way to train people to gather data prospectively is to meet with the individuals who are fulfilling the data “trapping” role and to hold a discussion with them. The points that might be covered in the discussion are

- Rationale: Why this? Why now? Why is this important?
- Utility: Indicate how the data will be used and analyzed and ensure that the uses will not be threatening those who must be depended on to provide accurate data
- Current process: Discuss current work process using “current method” flowchart to illustrate
- New process: Discuss proposed way to integrate data collection into daily work using “new method” flowchart
- Data collection forms: Discuss the medium (data collection form, template, etc) that will be used to gather data on core variables
- Practice: Role play the completion of data collection forms under different conditions
- Pilot test: Conduct small scale pilot test of data collection and debrief with the individuals who’s daily work now includes data “trapping”
- Oversight: Check data at regular interval (eg, daily or weekly) to ensure that data forms are being completed in a consistent manner and discuss any problems with relevant staff

Consider working with local leadership to include data collection as part of the basic job description of staff. Work to create a supportive data collection environment that is supported by leaders, supervisors, and front line staff. Consider ways that you can make the data useful “right away” for the front line team. For example: (a) information on the patient’s viewpoint on access to services could be used to determine the need for change and could be used to

evaluate the success of efforts intended to improve access, and (b) data on hemoglobin A1c levels of diabetics under active care could be put into a visual display to show the variation in glycemic control within the subpopulation that is receiving care.

What are the sources of inaccuracy and bias and how can they be avoided?

All data are imperfect. All data contain error. The more we know about what's wrong with our data the better informed we are about how to interpret the number. In developing your research design and measurement plan, you will want to take into account ways to gather data that are sufficiently accurate for the job at hand.

If key variables are based on “objective tests” then it is prudent to make sure that the materials and instruments are in good working order, are calibrated properly, and that the results are interpreted in a standard way. If key variables are based on “reports” or “ratings” from staff or patients, then you will want to make sure that the “stimulus question” is clear, can only be interpreted in one way, and that the response choices that are offered are relevant, exhaustive, and mutually exclusive. If it is possible to use a standardized survey that has been validated properly, then it may be prudent and efficient to make use of the pre-existing survey.

There are many potential sources of bias. Here are a few of the more common forms of bias that can influence health care relevant data. One type of bias emanates from what is sometimes referred to as the social desirability effect. For example, patients may wish to appear “compliant” or “pleased” with their health care and clinicians may wish to appear “thorough” and “burdened” by very sick patients. If a substantial number of patients or clinicians are asked to report or rate performance in these areas, then, all things equal, the results may tend to favor higher compliance, higher satisfaction, fuller coding, and increased severity than would otherwise be the case. Another type of bias is associated with the effects of memory—there may be compression or expansion or selective recall of important events that the respondent is being called on to evaluate or to

report on. Another type of bias is associated with instrumentation/interpretation effects and is especially relevant when “objective” measures are being used such as blood pressure, ejection fraction, pulmonary function, pathologic specimen results, x-ray results, etc. There may be a tendency for values to vary systematically higher or lower than their “true” value.

As one considers different measures and their associated potential sources for error and bias, try to devise antidotes (specific counter actions) to minimize these problems. The best antidotes will be forward thinking and will capitalize on the up front *design* of the process of training, the process of instrumentation, the process of data checking, and the process of supervising/coaching the data collection process.

What variables should be considered as covariates or confounders or stratifiers?

After your data has been collected and you are preparing to analyze the information, you will want to make sure that you have the ability to determine if certain measurable variables “condition” the results. Make sure that you have taken the issue of confounding or covariation into account at an early stage in the research process by building dummy data displays that (a) answer key questions, and (b) anticipate the need to analyze the data in ways that can “disentangle” causes and effects or can separate different “nodes” in the “web of causation.”⁸

Some frequently occurring categories of covariates, confounders, and stratifiers are

- Patient – characteristics of individuals such as demographics, comorbidities, disease severity, and functional health status
- Time – of day, day of week, season of year
- Place – location of care
- Practitioner – clinician providing care, clinical team providing care
- Degree – magnitude of exposure to intervention or planned test of change

The great value in anticipating the need to analyze the data using covariates, confounders, or stratification variables is that you will have a better opportunity

to separate known causes of variation from unknown and the “signal” that emerges from the results will be clearer and more refined.

What statistical techniques are most likely to be useful?

Statistics are tools that we use to analyze and summarize and condense the message that is contained in the relationships among variables under study. Statistics are methods that we use to “pull out” the signals that are embedded in the structure of the data and the relationships between variables. Many useful and outstanding texts are available that will help provide an adequate question to this large and important question. The selection of statistical techniques is contingent on the type of variables used in the research (eg, nominal, ordinal, interval, ratio), the structure of the variables (ie, normal distribution, skewed, bimodal, etc.), the choice of the research design (eg, before-after, multiple contrast groups, etc.), and the nature of the analytic objective. Consider working with a qualified statistician to guide you through the selection and application of appropriate analysis methods.

Some of the most common bivariate and multivariate statistical techniques that are covered in classical statistical texts and courses involve *t*-tests, chi-square analysis, analysis of variance, analysis of covariance, multiple linear regression, and logistic regression.⁹ These methods of statistical analysis are conventional, powerful, and extremely useful.

There is another branch of statistics, however, that has emerged from the field of quality control and is often referred to as “SPC” or statistical process control.^{10,16} This family of statistical methods has proven to be extremely valuable for quality improvement projects, especially under conditions in which time-series designs are used or when, after establishing a series of time ordered baseline values, the improvement team begins to introduce sequential tests of change.⁶ One of the distinctive features of SPC techniques is that they use graphical forms of data presentation that make it easy for the analyst to visualize variation over time and to use standard rules to detect shifts, trends, special cause events, and spe-

cial patterns of causation. There are different control charts for different types of data such as individual values, subgroup averages of individual values, count data, proportion data, and so forth. Once again, relying on a qualified statistician is best to guide you through the selection and use of this set of analytic tools.

How can data be displayed in a way that is most easy to interpret?

Different people will have different preferences for displaying data. Some prefer tables and some prefer graphic displays. Both approaches are commonly used and useful. Also, different data sets will invite different methods of data display and it is helpful to allow the structure of the data to inform the selection of data display method. For example, if data are gathered in a test and control group during 3 periods of observation, then the data display should reflect values for each group for each of 3 time periods.

When using tabular data displays, it is generally important to pay careful attention to the table title, the contents of the table itself, the loss of data from the original sample to the specific table, and to providing more detailed information to help the careful reader interpret the results. A few hints on table “hygiene” follow:

- use a table title that describes the contents of the table
- provide information on sample size used in the table, ie, *N* given in the table
- provide information on the number of subjects or observations that were excluded from the original sample for specific reasons (eg, missing observations)
- use specific footnotes to clarify aspects of the table that the careful reader would want to know about (eg, describe the meaning of column labels and row labels, indicate the specific statistical tests that were used to produce *p*-values, etc)

Under many conditions, graphic data displays can be used to provide rich information that aids interpretation. One advantage of visual displays is that they can be viewed as a “picture” by the audience

Table 2

QUICK REFERENCE FOR RELATING TYPE OF DATA AND PURPOSE OF DISPLAY TO TYPE OF GRAPH

Purpose and data type	Type of graph
Display univariate distributions	Frequency bar chart
Describe parts of whole	Bar chart, divided bar chart, pie chart
Describe trends	Time series plot
Analyze individual cases	Box plot, stem and leaf plot, dot charts
Analyze 2 variable relationships	Scatterplot
Analyze multivariate relationships	Scatterplot matrix, cluster scatter plot, 3-variable plot

and as such can be interpreted quickly—if the audience has had prior experience in viewing similarly structured “pictures.” There are some excellent texts on graphic techniques that can be used as sources of reference, practical guidance, and inspiration.^{17–19} A “quick reference” list of graphical display methods related to different types of data is provided in Table 2. Information in this table is based on the book authored by Henry.¹⁷

Another helpful method is to create “process flow” data displays. A process flow data display combines a simplified flow chart of a process (eg, accessing outpatient care) with data values that measure how the process is working for key steps in the process. Figure 5 provides a process flow data display for ac-

cess to care for an outpatient visit to a primary care practice.

CONCLUSION

It is essential to build strong measurement into quality improvement projects to produce reliable results information and, when appropriate, enhance the likelihood of publication. It is important to harness measurement to serve improvement. We must find ways to make the collection and analysis of data *friendly* to the “front line” team and *useful* for many different users. This requires an understanding of the “local ecology” of health care to promote

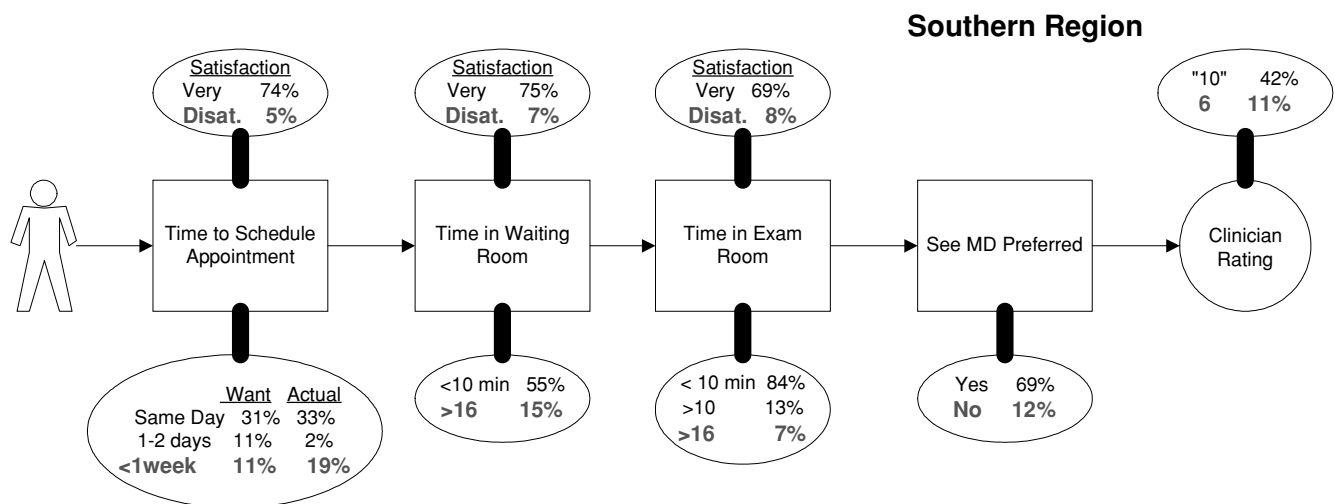


Figure 5. Process flow data display: access to care.

“friendliness” and a deep understanding of the needs of multiple users of the data to promote “utility” in the real world.

One promising strategy to conjoin friendliness with accuracy and utility is to adopt a “feed forward” approach to collecting and analyzing data. This approach designs information flows into the care delivery process to “capture” and use the data as the care process unfolds in real time and to gather the information from the most reliable source (patient, physician, nurse, scheduling system, diagnostic testing system, etc) as soon as it “happens.” The feed forward approach is being used at the Dartmouth Spine Center and in a National Institute of Health sponsored randomized clinical trial on spine surgery (ie, the SPORT Trial). Both of these make use of clinical value compass concepts, touch-screen computers, and other low technology devices to obtain real time “data inputs” from patients, doctors, nurses, and others to track processes and outcomes of care.²⁰

We look forward to the time when more and more members of the clinical team will find it increasingly feasible to blend clinical care with reflective practice, with the improvement of care using the scientific method.

REFERENCES

- Nelson EC, Mohr JJ, Batalden PB, Plume SK. Improving health care, I: the clinical value compass. *Jt Comm J Qual Improv.* April 1996;22:243–258.
- Wheeler DJ. *Understanding Industrial Experimentation.* 2nd ed. Knoxville, TN: SPC Press Inc; 1990.
- Shortell SM, Bennett CL, Byck GR. Assessing the impact of continuous quality improvement on clinical practice: what will it take to accelerate progress. *Milbank Q.* 1998;76:593–624.
- Solberg LI, Kottke TE, Conn SA, Brekke ML, Calomen CA, Conboy KS. Delivering clinical preventive services is a systems problem. *Ann Behav Med.* 1997;19:271–278.
- Wells KB, Sherbourne C, Schoenbaum M, et al. Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *JAMA.* 2000;283:212–220.
- Speroff T, O'Connor GT. Study designs for PDSA quality improvement research. *Qual Manag Health Care.* 2004;13(1):17–32.
- Wheeler DJ. *Advanced Topics in Statistical Process Control: The Power of Shewhart Charts.* Knoxville, TN: SPC Press; 1995:135–140.
- McMahon B, Pugh TP. *Epidemiology Principles and Methods.* Boston: Little Brown & Co Inc; 1970.
- Armitage P. *Statistical Methods in Medical Research.* New York: Wiley; 1973.
- Wheeler DJ, Chambers DS. *Understanding Statistical Process Control.* Knoxville, TN: SPC Press; 1992.
- Lowrie EG, Lew NL. The urea reduction ratio (URR): a simple method for evaluating hemodialysis treatment. *Contemp Dial Nephrol.* 1991;12:11–20.
- Owen WF, Lew NL, Liu Y, Lowrie EG, Lazarus JM. The urea reduction ratio and serum albumin concentration as predictors of mortality in patients undergoing hemodialysis. *N Engl J Med.* 1993;329(14):1001–1006.
- Meyer KB, Espindle DM, DeGiacomo JM, Jenuleson CS, Kurtin PS, Davies AR. Monitoring dialysis patients' health status. *Am J Kidney Dis.* August 1994;24:267–279.
- Kurtin PS, Davies AR, Meyer KB, DeGiacomo JM, Kantz ME. Patient-based health status measures in outpatient dialysis: early experiences in developing an outcomes assessment program. *Med Care.* May 1992;30(suppl 5):MS136–MS149.
- Beusterien KM, Nissenson AR, Port FK, Kelly M, Steinwald B, Ware JE Jr. The effects of recombinant human erythropoietin on functional health and well-being in chronic dialysis patients. *J Am Soc Nephrol.* May 1996;7:763–773.
- Splaine ME, Nelson EC, O'Connor GT, Batalden PB, Plume SK. Statistical measurement issues in quality improvement research. *Qual Manag Health Care.* In press.
- Henry GT. *Graphing Data: Techniques for Display and Analysis.* Thousand Oaks, CA: Sage; 1995.
- Tufte ER. *Visual Explanations: Images and Quantities, Evidence and Narrative.* Cheshire, CT: Graphics Press; 1997.
- Tufte ER. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press; 1983.
- Weinstein JN, Brown PW, Hanscom B, Walsh T, Nelson EC. Designing an ambulatory clinical practice for outcomes improvement: from vision to reality—the spine center at Dartmouth-Hitchcock, year one. *Qual Manag Health Care.* 2000;8(2):1–20.