

We are increasingly realizing not only how critical measurement is to the quality improvement we seek but also how counterproductive it can be to mix measurement for accountability or research with measurement for improvement.

PERFORMANCE MEASURES AND MEASUREMENT

The Three Faces of Performance Measurement: Improvement, Accountability, and Research

LEIF I. SOLBERG, MD

GORDON MOSSER, MD

SHARON McDONALD, RN, PhD

Currently there is a great deal of pressure for public accountability of health care organizations, especially for managed care plans and even for medical groups and individual clinicians. Purchasers, legislators, and consumer advocates are all calling for public disclosure of patient satisfaction and other health care outcomes, on the theory that the comparative information will be used in choosing providers and thereby will force attention to quality issues.

In such an atmosphere, it should not be surprising that clinicians are wary. Although they recognize

that the high cost of health care has understandably led to an insistence on the assurance of quality, they also recognize better than anyone else how hard it is to develop and collect valid and reliable quality measures. As Dennis S. O'Leary, president of the Joint Commission on Accreditation of Healthcare Organizations (Oakbrook Terrace, Ill), has stated, "the problem with measurement is that it can be a loaded gun—dangerous if misused and at least threatening if pointed in the wrong direction."¹(p. 354)

Simultaneously with this pressure for measurement

Leif I. Solberg, MD, is Clinical Director of Research, Group Health Foundation/HealthPartners, Minneapolis; a member of the Board of Directors for the Institute for Clinical Systems Integration, Minneapolis; and a member of the Editorial Advisory Board of The Joint Commission Journal on Quality Improvement. **Gordon Mosser, MD**, is Executive Director, and **Sharon McDonald, RN, PhD**, is Manager, Health Care Measurement, the Institute for Clinical Systems Integration/HealthPartners. The work described in this article was partially supported by HealthPartners and U.S. Agency for Health Care Policy and Research Grant RO1 HS08091.

Copyright © 1997 by the Joint Commission on Accreditation of Healthcare Organizations

The concepts described in this article are the result of work with hundreds of clinicians and clinic staff involved with the Institute for Clinical Systems Integration (ICSI) and the IMPROVE trial. We are very grateful for those experiences and the insights of innumerable people. We extend special thanks to Eugene C. Nelson, DSc, MPH, at Dartmouth University (Hanover, NH) for his leadership in measurement issues and for his review of and comments on this article.

Please send requests for reprints to Leif I. Solberg, MD, Group Health Foundation, 8100 34th Ave S, PO Box 1309, Minneapolis, MN 55440-1309; phone 612/883-5017; fax 612/883-5022; e-mail Leif.I.Solberg@HealthPartners.com.

Article-at-a-Glance

Background: In the current climate of public accountability, many clinicians have become uncomfortable with any efforts to create measurement systems. This is unfortunate because measurements are absolutely essential to efforts for improving the processes of medical care. In Minnesota, work has been conducted with clinicians on measurement pursued for accountability, improvement, and research.

Measurement in the improvement process: There are at least three steps in process improvement where measurement is likely to be important: when identifying which problems, or opportunities for improvement, need attention; when the process improvement team is obtaining baseline measurements; and after a new improved process has been implemented.

Contrast with measurement for accountability: Data for accountability, which are data on outcomes or results, do not usually illuminate how

the outcomes were achieved or how processes might be changed to improve them. The measures selected for accountability will be measures that matter to external parties, for example, outcome data on complication rates or costs of care.

Contrast with measurement for research: Although objectives and methods of measurement for research make it different from measurement for improvement in many respects, its familiarity to physicians—and its attractiveness to them as scientists—poses a problem for measurement for improvement in health care. Measurement for research is typically too slow, too expensive, and too elaborate to be useful for improving processes in single clinics or hospitals.

Summary and conclusions: Experience in guideline implementation and measurement efforts has yielded lessons on how to understand the differences in purposes of measurement.

for accountability, there has been increasing recognition of the need for health care measurement with a different aim—improvement. Although accountability measures may identify areas and organizations that need improvement, these results are necessarily so far downstream that they are rarely of much help to the process of improving the delivery of health care.² Knowing that your health plan or medical group has, for example, below-average rates of providing mammograms* does not tell you anything useful about why that is so or where to begin efforts to change that rate. In the turmoil of health care change going on in Minnesota, we have had an unusually early and valuable opportunity to work with clinicians on measurement pursued for accountability and measurement pursued for improvement. On the one hand, the state-legislated Health Data Institute is gathering public accountability quality data about both health plans and medical groups, while the Buyers Health Care Action Group and our own managed care organization (HealthPartners) are doing the same for care systems and medical groups. On the other hand, the clinicians in our unique collaboration of 19 medical groups associated with HealthPartners in the Institute for Clinical Systems Integration (ICSI; Minneapolis) are measuring for improvement purposes as they develop and implement clinical practice guide-

lines.³⁻⁵ We have also been working with more than 40 clinics on a randomized trial of an intervention to use continuous quality improvement (CQI) to implement preventive services guidelines (the IMPROVE [IMproving PREvention through Organization, Vision, and Empowerment] Project).⁶

ICSI is a quality improvement (QI) organization that bridges a managed care organization (HealthPartners) and the following medical groups: Park Nicollet Clinic/Health System Minnesota, HealthPartners Medical Group, Mayo Clinic, and 16 others in the Minneapolis-St Paul area. It is governed by a board of physicians from the member medical groups, as well as three purchaser representatives and the HealthPartners medical director. In addition to health care guideline development and implementation, ICSI pursues population health measurement, medical technology assessment, health care information systems development, and training in the use of continual improvement methods.

When the ICSI medical groups started to talk about their measurement of guideline implementation in 1993, there was a great deal of confusion and emotion caused by widely differing understandings of the purposes and means of measurement. Not only was there confusion about whether measures were for accountability or improvement, but clinicians' notions of measurement for research also added to the confusion. This article is designed to describe what we have learned

* Per the National Committee for Quality Assurance's (Washington, DC) Health Plan Employer Data and Information Set (HEDIS) 2.5.

through painful and laborious efforts as we have clarified these confusions about measurement. It is especially addressed and dedicated to the clinicians for whom this clarification is particularly important. We are increasingly realizing how critical measurement is to the QI we seek, yet how counterproductive it can sometimes be to mix measurement for accountability or research with measurement for improvement. Considered one by one, measurement for each purpose can be good and very important. If done poorly, it can be bad. If the measurements are mixed together in inappropriate ways, they can indeed become harmful or destructive, with the mixed purposes interfering with one another.

This article should serve as a cautionary tale for health care organizations embarking simultaneously on a program of guideline implementation and a program of performance reporting to purchasers. Commonly the measurement activities needed to support these two programs are conceived as a single endeavor. A unified series of sampling procedures, data collection routines, and data display methods is anticipated to meet both the improvement and the external reporting needs. Our experience suggests that this approach, while appealing for its apparent efficiency, is a pitfall. Specifically, measurements collected for improvement purposes typically are not useful for external reporting and, if used for external reporting, may poison the improvement effort.

These lessons are particularly important for managed care plans. As they increasingly strive for both accountability and improvement from the contracted private groups serving their members, plans need to be very sensitive to the risk that those groups will be forced to "game" their data rather than collect data to be used for real improvements.

Improvement as a Process

Before directly comparing measurement for improvement with that for accountability or research, it is important to convey the process of improvement and just how measurement fits into that process. After all, modern QI concepts had their origins in the statistical process control measurements developed by Walter Shewhart at the Bell Telephone Laboratories in the 1920s.⁷ The marriage of those techniques with *an overall* management

Figure 1a. Example of a Simple Process

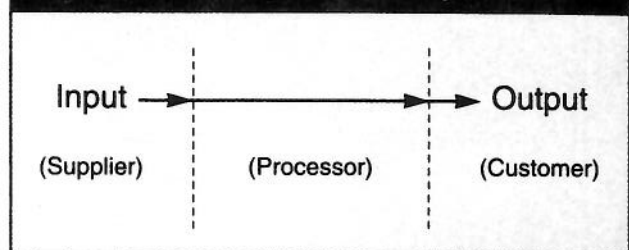


Figure 1a. In a simple process, an action or series of actions (by a processor) converts an input from a supplier to an output for a customer.

philosophy by W. Edwards Deming, Joseph Juran, and others has resulted in the quality movement as it is known through various terms and acronyms (TQM [total quality management], CQI, and so on). Although arriving later in health care than in other fields, QI concepts have rapidly proliferated here through the efforts of Berwick, Batalden, and others.⁸⁻¹⁰

Understanding the term *process* itself and developing process thinking is fundamental to an understanding of how to improve anything. All work of any kind can be thought of as a process, that is, an action or a series of actions (by a processor) that converts an input from a supplier to an output for a customer (Figure 1a, above). Although this may be most clear when one thinks of physically constructing something, both services and mental actions are also processes. Most processes are complex (Figure 1b, below), involving a series of linked steps in which the customer of one step in a process becomes the supplier to another step as well as a processor in yet others. The important concept here is to see how these linked actions or tasks are focused ultimately on producing outcomes for customers. Everything we do in health care involves processes, whether it is the steps involved in making an appointment (Figure 2, p 138) or

Figure 1b. Example of a Complex Process

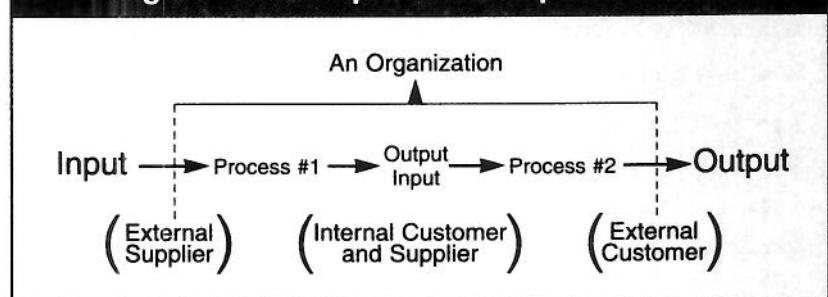
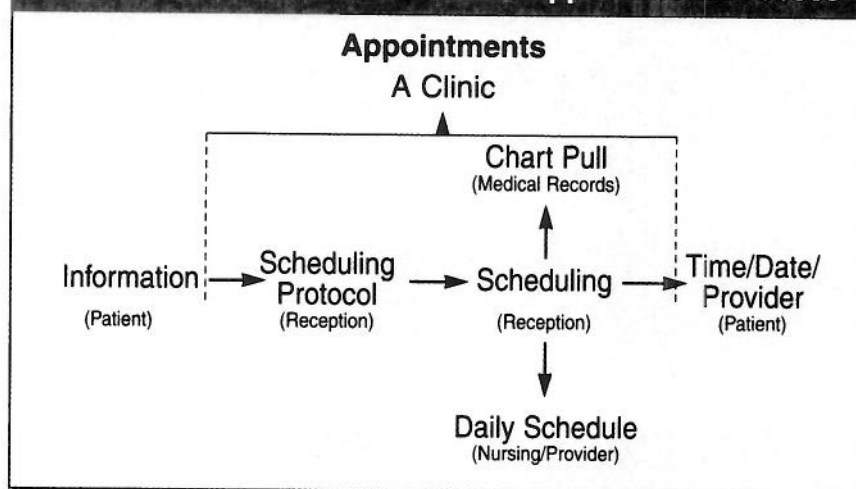


Figure 1b. In a complex process, the customer of one step in a process becomes the supplier to another step, as well as a processor in yet others.

Figure 2. A Medical Care Process: Appointment Process**Figure 2.** The appointment process is depicted in a series of steps.

in diagnosing and treating a urinary tract infection.

In addition, the work that is done to improve any of these health care processes is, itself, also a process. In this case, the input is a work process needing improvement and the output is a new, improved version of the work process. In between, the team is using an organized improvement process to create that output (Table 1, below). As Berwick has pointed out, physicians are involved in almost all important health care processes and they often fill all three roles: supplier, processor, and customer.¹¹ Thus, it is wasteful or foolish to try to improve health care processes without physicians' input, either by choosing to work on processes in which they have a lesser role or by not involving them in improvement.

One of the benefits of concentrating on processes rather than on people in improving care is that it takes the fear and blaming out of the equation. This allows

everyone to concentrate on improvement rather than on defensiveness. Another benefit is that the most powerful improvements usually come from an understanding of processes and from efforts to systematize them, not from education or exhortation of individuals. A focus on process also forces one to pay more attention to the desires of the customer and to the use of a data-driven scientific approach to change rather than a reliance on hunches and tampering.

The first step in improving medical care involves identifying and defining boundaries around

the process to be improved. To illustrate the role of measurement in the improvement process, we can use an example from one of the earliest ICSI guidelines, uncomplicated urinary tract infection (UTI) in women.¹² This is a particularly relevant example because it was the guideline that first caused us to become aware of the need to clarify and separate different purposes for measurement. Figure 3 (p 139) represents the possible steps in the multiple ways a clinician in a traditional clinical setting might diagnose and treat a woman complaining of acute urinary tract symptoms. The actual pathway followed for a particular patient will vary widely by clinic, clinician, and even from one patient to the next patient with a similar condition for an individual clinician. An important goal for most improvement efforts is to reduce this wide variation to only those variations that are desirable, that is, those important to the needs of a particular patient.¹³

Once the process has been identified along with descriptions of the goal and reasons to work on it (that is, the mission and problem statements), it is possible to specify who should be involved in improving it. If the process is simple or confined to one type of worker, one person can improve it. If the process is more complex, it is usually better to set up a team, with members from each work area that has significant involvement in the process. That has been a central concept being tested in the IMPROVE Project.

A review of Figure 3 suggests that an improvement team for this UTI process should probably include a physician, a nurse, a receptionist, and a laboratory

Table 1. The Seven-Step Process Improvement Model

1. Identify the problem
2. Collect data to understand your current process
3. Analyze the data to understand root causes
4. Choose an approach
5. Develop the process(es)
6. Implement
7. Evaluate and improve in an iterative cycle through these steps

Figure 3. Traditional Care Process for Urinary Tract Infection (UTI) in Women

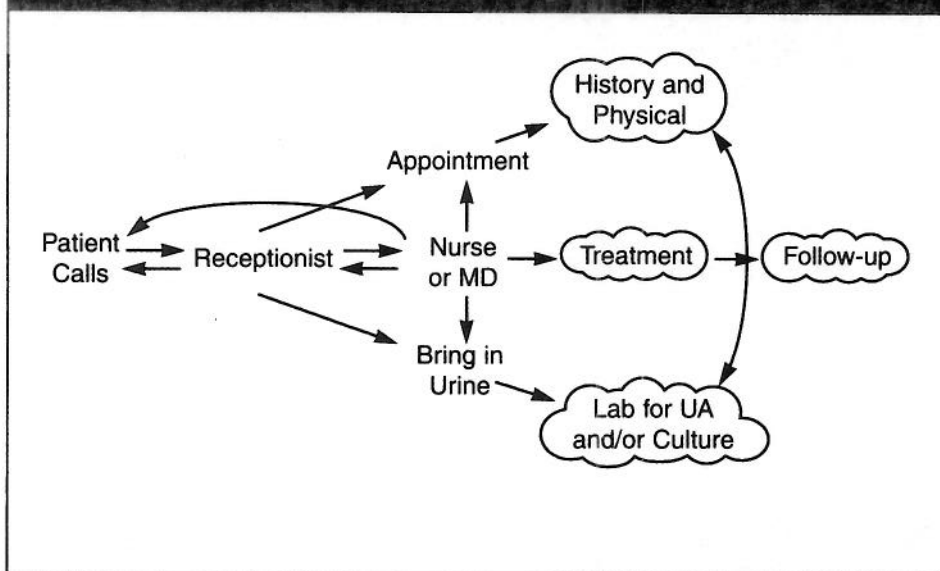


Figure 3. The process of diagnosing and treating UTI is depicted in a series of steps. The cloud shapes identify areas that are impossible to define because many different types of actions can occur within them. From the clouds and the multiple pathway arrows in the figure, it is clear that there is no well-defined way to diagnose and treat such a problem in the setting depicted. MD, physician; lab, laboratory; UA, urinalysis.

technician, since each of them knows a part of the process and probably will need to be involved in making any changes to it. From here on, the actual process improvement steps may vary among the various CQI approaches, but a structured progression of steps is usually followed. (The model used at HealthPartners is depicted in Table 1.)

The Role of Measurement in the Improvement Process

There are at least three steps in process improvement where measurement is likely to be important. The first is in step 1 (Table 1), that is, identifying which problems, or opportunities for improvement, need attention. In the case of the process for managing UTI, this identification may come from tallies of complaints from patients about the difficulties in getting treatment or from billing data that show UTI to be a common and costly area of care. If a care system or clinic wishes to improve quality, it will also need to regularly obtain and analyze data about the frequency and cost of services.

A need for measurement also arises in step 2, when the process improvement team is obtaining baseline measurements. These data about the current care process are collected for several reasons:

1. To better understand the extent and nature of the problem.

2. To provide motivation to change by documenting the extent of the problem.

3. To provide points of comparison with remeasurements obtained after changes are made.

Even when there is no systematic process in place, such as that depicted in Figure 3, it is still desirable and possible to make measurements for these three purposes. In the management of UTI, for example, one can measure how long it takes for a patient to receive treatment, how often a urinalysis or

culture is obtained, and how often various antibiotics are used in treatment. Alternatively, baseline measurements can be done when one knows more about what changes are desired.

Finally, measurement becomes particularly useful at step 7, after a new, improved process has been implemented. At this point it is important to learn the extent to which the new process is being used and what the impact has been on patients, clinic personnel, costs, and so on. Thus, these measurements are made soon after implementation as a basis for deciding whether further actions are needed to modify or implement the new process more fully. The measurements should also be repeated periodically to monitor the new process, especially if further changes are made.

Figure 4 (p 140) illustrates the UTI care process in one medical group that has implemented the ICSI guideline. This diagram combines the key care decisions defined in the guideline with specific information about who is to do what in that medical group. Without this second implementation addendum, which organizes the care process, a guideline is simply another set of cognitive recommendations that is unlikely to lead to changes in behavior.^{14,15}

The main explicit goals for the UTI guideline are

Figure 4. Systematized Guideline Process for Urinary Tract Infection (UTI) in Women

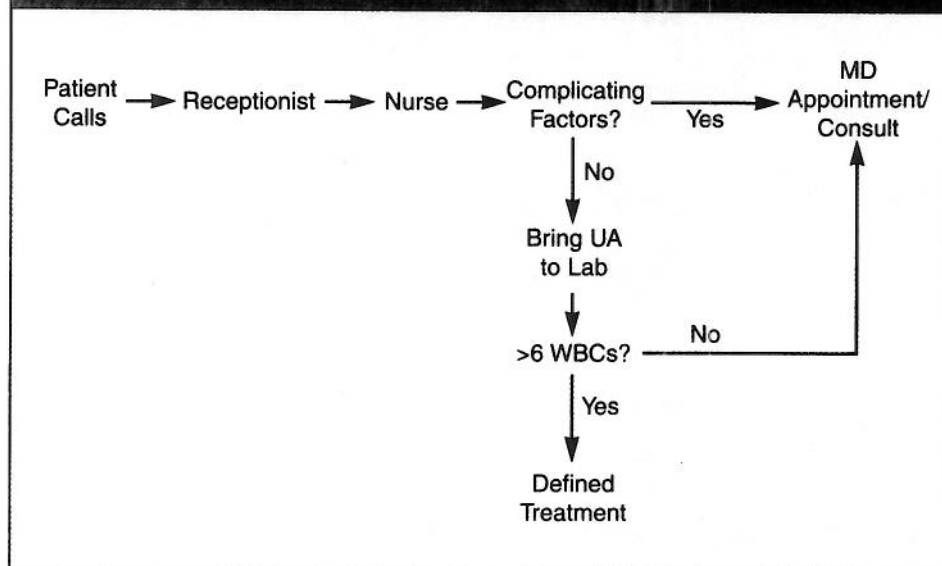


Figure 4. A systematized guideline can be used to depict the process for diagnosing and treating UTI. MD, physician; UA, urinalysis; lab, laboratory; WBC, white blood cell count.

to reduce unnecessary urine cultures and to provide three-day treatment with selected antibiotics for women with uncomplicated UTIs. Thus the key process measurements to obtain before and after implementation are the frequencies with which these events are part of the care process. An additional implicit outcome goal is to improve patient satisfaction by reducing the time and barriers to treatment. To make these changes more likely, the medical group implementation team that devised Figure 4 decided to institute a systems change—providing for standing orders and a standardized data collection form that permits and facilitates direct nurse management of patients who fit the criteria for standard treatment. Thus another desirable process measure in this setting is the frequency with which appropriate patients are managed by a nurse rather than by a physician.

Characteristics of Measurement for Improvement

Now that the improvement process and its measurement needs have been defined for implementation of the UTI guideline, we can use them as examples of the characteristics of measurement for improvement. Table 2 (p 141) summarizes this information in the left-hand column.

Who needs the measurement data? The medical group's UTI process improvement team is the main

and may make it impossible to focus on systematic improvements of the process.

Why do they need data? The reasons follow:

1. *To understand the current process of care for UTIs.* Without information about the usual roles, tasks, and extent of variation and inefficiency in the current process, it is hard to know how to improve on that process. If most patients are already receiving three-day treatment and avoiding cultures, there is no need for those changes.
2. *To understand the attitudes of key process workers and customers—patients, physicians, and staff.* Information about their satisfaction with the current process and any contemplated changes may be critical to success.
3. *To motivate the improvement team, clinic management, and all relevant personnel to want to improve (that is, change).*
4. *To provide a baseline against which to compare remeasurement after any changes are made.*
5. *To learn how well the new process is working and how key processors and customers feel about it.*

What measurements are needed? The individual measurements must be specific to the process being improved and should involve key process steps. Similarly, the data must be specific to a medical group and even to each site or division that has a separable implementation. Data for an overall multisite group is usually of little improvement value because the data subsume units

customer of the data. However, the team will share selected results with other people in the medical group whose support is needed for changing the process. These other people will include at least the group's management and those whose work affects the new care process. The team will not share these data publicly outside the group because doing so adds nothing and may be counterproductive. If, for example, the measurements show poor compliance or care deficiencies, public display causes fear and blame

Table 2. Characteristics of Measurement for Improvement, Accountability, and Research

	Improvement	Accountability	Research
Who?			
Audience (Customers)	Medical group Quality improvement team Providers and staff Administrators	Purchasers Payers Patients/members Medical groups	Science community General public Users (clinicians)
Why?			
Purpose	Understanding of a. process b. customers Motivation and focus Baseline Evaluation of changes	Comparison Basis for choice Reassurance Spur for change	New knowledge, without regard for its applicability
What?			
Scope	Specific to an individual medical site and process	Specific to an individual medical group and process	Universal (though often limited generalizability)
Measures	Few Easy to collect Approximate	Very few Complex collection Precise and valid	Many Complex collection Very precise and valid
Time period	Short, current	Long, past	Long, past
Confounders	Consider but rarely measure	Describe and try to measure	Measure or control
How?			
Measurers	Internal and at least involved in the selection of measures	External	External and usually prefer to control both process and collection
Sample size	Small	Large	Large
Collection process	Simple and requires minimal time, cost, and expertise Usually repeated	Complex and requires moderate effort and cost	Extremely complex and expensive May be planned for several repeats
Need for confidentiality	Very high (Organization and people)	None for objects of comparison—the goal is exposure	High, especially for the individual subjects

with separate implementation needs. Data tied to individual workers are also usually not needed or not helpful because the unit of analysis is too small; implementation typically involves a whole care team. However, when the data pertain to decisions made by individual physicians operating essentially alone, data on individuals can be help-

ful, for example, in judging whether particular patients with chest pain should be hospitalized.

The measures chosen must be few and easy to collect. They should cover short periods, since there are few resources and little time to collect them. To be most useful, they will need to be able to be re-collected

periodically. Nelson has provided a helpful insight into the problem of how to obtain repeated measurements of an event that occurs infrequently.² He suggests that one can simply count the number of cases between events, for example, emergency visits after initial guideline treatment of a UTI, and create a run chart of these intervals to track changes in occurrence rates. Finally, if there are confounding factors, such as details of past infections or severity, it is usually not possible to measure them.

How are the measurements selected and performed? It is important that those persons doing the improving choose the measurements or at least be involved in the choice; otherwise the measures may not be relevant or used.¹⁶ Because a high degree of precision is not necessary for improvement purposes and because data collection needs to be simple and repetitive, small samples, for example, 10 to 20 cases per sample, are usually appropriate. Ideally, the data can be collected during the care process by clinic staff to minimize effort and maximize meaningfulness to those affected. The potential drawback to this approach is the difficulty in obtaining consistent and reliable data unless care is exercised in setting up the data collection process. In any case, the data will usually be collected internally by the clinical site staff or those who actually carry out the health care process.

To continue with the UTI example, a QI team likely will want to measure at least the percentage of cases that exhibit the following:

- *Initial urine culture;*
- *Use of recommended antibiotics for three days; and*
- *Nurse versus physician management.*

These data points can be obtained from medical records selected consecutively from the log for appointments or triage phone calls. The 10–20 cases needed to produce “good enough” data should not require more than one hour’s attention once the charts are found. In addition, a CQI team would probably want to collect some information about patient and clinician and staff satisfaction with the process. Ideally, all measures would be taken as baseline measures before any change and then remeasured as often after the change as is necessary to ensure that adequate improvement occurs and is maintained.

Contrast with Measurement for Accountability

Table 2 highlights the ways in which measurement for accountability compares to measurement for improve-

ment. Although the principal audience for accountability data is external to a medical group (purchasers and actual/potential patients), the group will also be interested in how its data look compared to those for other groups. In this way, there is some overlap with measurement for improvement since the comparison may serve as a stimulus for internal improvement work.

Data for accountability, which are data on outcomes or results, do not usually illuminate how the outcomes were achieved or how processes might be changed to improve the outcomes. The measures selected for accountability will be measures that matter to external parties, for example, outcome data on complication rates or costs of care for UTI. However, since these will be difficult to measure, a surrogate of patient satisfaction with UTI treatment may be used instead. Thus there is a tendency for the data to become more and more remote from data that might be used to change processes of care.

Because data for accountability are intended to reveal and to compare the performance of medical groups or other health care institutions, they must be precise, reliable, and valid. These requirements have several consequences. First, the samples must be sufficiently large to achieve the desired precision. It may be difficult to obtain enough cases to describe small- to medium-sized medical groups, and it will rarely be possible to provide enough data about an individual clinic site. Thus again the value of these data for improving processes is limited. Second, to obtain a sufficient sample size, data must often be collected for long periods, with the consequence that by the time the data are available, the processes that produced the results have changed and the data are not useful for improvement purposes. Third, since comparison across different organizations is of the essence in measuring for accountability, any number of factors affecting the comparative performance of different institutions must be measured and taken into account in reporting the results from the various organizations, for example, severity or population differences. This measurement of potential confounders results in substantial complexity, increasing the cost and delaying the output relative to measurement that is useful for improving processes.

Accountability data are intended to be nonconfidential. They are intended to be used for judgment. The generation and reporting of these data will commonly result in fear and defensiveness (see Sidebar

Sidebar 1. Data Collection Methods: The Fear of Misuse

When the Institute for Clinical Systems Integration (ICSI) was developing its first measurement plans, in early 1993, a proposal was made to gather information about compliance with the urinary tract infection (UTI) guideline by asking clinics to collect the desired data elements in the course of visits with patients. This would be done on a special recording sheet for each case presenting with appropriate symptoms, and these forms would then be sent to ICSI, which would aggregate and report on the data. The argument made in favor of this approach was that it was too hard to identify all cases in other ways and that this would allow ICSI to assess and assure the methodologic quality of the information. By including information on case follow-up on these forms, it was also thought that it would be possible to do some research with the data, that is, to identify associations between case characteristics and outcomes.

Some clinicians took great exception to this ap-

proach, both on the grounds that it was an intrusive time effort in a busy practice day and that it seemed like a way to track individual clinician behavior. Frankly, there was no trust that ICSI would have the power to keep the data confidential and would not be forced to reveal it by the purchasers or by the state. They feared misuse of the information.

Consequently, this measurement method was never used; instead, reliance was placed on clinics' aggregation of their own chart reviews. However, the clinician fear of misuse was so great that ICSI and the Buyers Health Care Action Group had to obtain special legislation at the state level that would protect this quality improvement information from disclosure. In addition, insofar as the effort to mix research with improvement and accountability purposes was clearly not going to satisfy any of those aims, it was dropped from the approach.

1, above, and Sidebar 2, below). Heavy emphasis within a medical group on accountability data will be counterproductive if time and energy are diverted into a defensive effort to show that the data are wrong or that a particular medical group's patients are different from those treated by other medical groups. At the extreme, if all data collected for improvement are potentially open to public scrutiny for accountability purposes, the medical group or other organizations will be tempted to shut down or delay data collection or distort the results to protect its interests. For this reason, data collected internally for improvement purposes should be protected

from public scrutiny unless purchasers, regulators, or other external parties have declared in advance that data on the issues in question are required for accountability. The price of requiring that all improvement data be public data would be damaging to genuine improvement efforts.

To meet the requirements of data reliability and to avoid distortions arising from understandable self-interest, the only sound solution to obtaining accountability measurements is to have the data collected by a central external group or, if the data are to be collected internally, to have them collected according to standard-

Sidebar 2. Data Display Issues: The Need for Anonymity

In 1993, the early days of the Institute for Clinical Systems Integration (ICSI), it was not yet clear how the guideline measurement data collected from each medical group should be displayed. One of the purchaser representatives from the Buyers Health Care Action Group (BHCAG) suggested that the summary reports at ICSI from the participating groups be labeled explicitly by group name. It was not because he wanted to be able to hold a particular group accountable for the degree of guideline compliance, but simply because he thought that this would help ICSI and BHCAG to better

understand and interpret the results. He argued that it would also help the medical groups to use the data for benchmarking.

The response from the member groups was "If you do that, we are not going to participate in this activity." The BHCAG representative immediately dropped his suggestion, seeing the fear that it had produced. Ever since, all data displays show only alphabet pseudonyms (which are varied from report to report so no medical group is always represented by any particular letter).

Sidebar 3. Measurement and Implementation: Measure As You Go

Both the Institute for Clinical Systems Integration (ICSI) and the IMPROVE projects have run into problems with clinics' efforts to incorporate measurement into their efforts at guideline implementation. (IMPROVE is a scientific trial of an external intervention to help clinics use continuous quality improvement [CQI] to improve delivery of preventive services.)

In training clinic leaders in how to understand their current processes and to implement systems for prevention, project staff tried to emphasize the critical importance of measurement for improvement and how it was different from measurement for research. One problem was that physicians in the clinics had the research paradigm in mind and either wanted to collect data from

large samples or simply resisted measurement because of the assumed effort involved. Another problem was that few clinics have staff with much experience with measurement tools and analysis, so there was a need for considerable skill and confidence development. Since few, if any, clinics have incorporated repeated evaluation data (for example, run charts) into their management of care of patient populations, it has taken a lot of effort to encourage them to use data as they make process changes. Clinics tend to make a change in some work process and then go on to the next problem, without using measures to determine whether the new process is actually being used and whether it is accomplishing the desired results.

ized, detailed specifications, for example, as with HEDIS data. Unless the data are collected in the same way at each location, comparisons are suspect and there is too much incentive for the objects of comparison to manipulate or game the data.

Contrast with Measurement for Research

Physicians are trained and continually updated on the basis of research data, so they are usually better informed than many other persons about what is involved in these measurements. Yet we have found that this experience is actually a drawback for improvement measurements because physicians tend to think such measures need to be as complex and precise as those for research. They may not understand the necessity for research to be complex and expensive because its intent is to produce new knowledge of widely generalizable or universal value. They may also forget the implicit assumption in science that research findings do not need to have a practical use or application (see Sidebar 3, above, and Sidebar 4, p 145).

To continue with the UTI example, research studies are more likely to be used to investigate basic etiology or pathophysiology. If the studies evaluate therapy at all, they are less likely to pursue its practical details. However, recent pressures for more practical outcomes from research have led to clinical investigations of the comparative value, for example, of three- versus ten-day antibiotic treatment regimens. It is, in fact, these

studies that have led to one of the main goals of the ICSI UTI guideline—to reduce treatment costs and problems by encouraging three-day treatment.

To provide such potentially universal new information, it becomes critical for those persons measuring for research to put enormous effort into the preliminary task of verifying that the measures and data systems to be used are precise, reliable, and valid.

An additional aspect of measurement for research which usually distinguishes it from measurement for improvement and for accountability is the need to control all possible variables other than those being studied. Therefore, subjects are selected by strict adherence to explicit criteria for inclusion and exclusion—and are usually a small subset of all potential subjects. Furthermore, research staff are usually on hand to ensure that every aspect of study methods is conducted exactly as planned.

The consequence of all this control is that the results of most clinical research investigations can only be strictly applied to those patients who fit the research criteria and in settings where the intervention can be carried out in the same controlled and (usually) expensive way. Even the National Institutes of Health (Bethesda, Md) has recognized this problem by identifying heavily controlled research experiments as a special type of research called *efficacy* studies, while those studies conducted under conditions closer to real life constraints are called *effectiveness* trials. However, even the relatively few effectiveness trials conducted are performed in a different way from measurement for improvement.

Sidebar 4. Types of Measurements: Make It Simple

One of the early guidelines developed at the Institute for Clinical Systems Integration (ICSI) was designed for management of intrapartum fetal heart rate. The measurement plan that the guideline development group devised entailed 14 separate measures—ranging from the rate of high-risk labors, through the rates of labors where external fetal monitoring was used, to the rates of oxygen or amnioinfusion use.

Within weeks of the measurement plan's implementation, it became apparent that most of the medical groups were having difficulty collecting all the required data. This was caused by the sporadicity of cases and their occurrence in hospitals, where the group had no control over work processes. In addition, there was usually an insufficient number of cases for most of the events (for example, the number of early amniotomies in cases diagnosed as failure to progress in labor) for any one medical group to be able to interpret, much less act on, data of uncertain statistical significance. It was usually impossible for indi-

vidual medical groups to come to any conclusion, then, on the basis of any reports or data. In any case, individual groups would not have known what any particular rate meant since there are no benchmarks for the rate—and most rates have obvious face value.

ICSI's experience with the management of intrapartum fetal heart rate guideline, along with others that entailed collecting too much data about issues of no great interest or importance to clinicians, led it in 1995 to modify its original approach to measurement in guideline implementation. Each guideline group was now told that an absolute maximum of two measurements could be requested per guideline. However, many of these scaled-down measurements still failed the clinician's "so what" attitude about being of any importance to patient care. In 1997 the guideline implementation and measurement approach at ICSI is being revamped to allow more detailed focus on what is important and improvable.

Although objectives and methods of measurement for research make it different from measurement for improvement in many respects, its familiarity to physicians—and its attractiveness to them as people of science—poses a problem for measurement for improvement in health care. Measurement for research is typically too slow, too expensive, and too elaborate to be useful for improving processes in single clinics or hospitals. In certain unusual circumstances, measurement for improvement and for research can be pursued hand-in-hand; but the pace of improvement is then slowed to meet the needs of research. Typically process improvement can and should be pursued more quickly since most process improvement consists of applying scientific conclusions already established elsewhere.

Similarities and Overlaps

Despite all the differences described, there are also some important similarities and overlaps among measurement for the three different purposes. For example, each requires that the user

- *be aware of the unique characteristics of measurement for the purpose at hand;*
- *select the items to be measured with great care;*
- *carefully design the sampling strategy and data collection techniques; and*

■ *interpret the findings cautiously, avoiding conclusions that exceed the limitations of the methods.*

In addition, for each purpose, measurement involves comparisons. Improvement measurements usually must be repeated to be useful, so the important comparisons are those over time. However, there may also be cautious comparisons with similar data from other organizations in a technique called benchmarking as part of the effort to learn from others.¹⁷ In measuring for accountability, the whole point is to compare organizations, while researchers prefer designs that include a comparison or control group of subjects to minimize bias.

There are also situations in which measurement for two purposes can be both possible and helpful. One obvious example is simultaneous measurement for improvement and for internal (as opposed to external) accountability. For example, some processes do require that individual clinicians change the way they have behaved with regard to a major management decision point, for example, avoiding repeat cesarean sections. Providing those clinicians with data comparing their own rates to those of others can encourage clinicians, who typically are reluctant to behave differently from their colleagues, to change their practices. This approach can be done anonymously or, in unusual situations, in

an identifiable way when a medical group's leadership needs to take action regarding specific individuals.

Neither is the distinction between measurement for improvement and measurement for research absolute. Despite the distinctions made in Table 2, data collected for improvement purposes can often be used to generate new hypotheses for subsequent research. Furthermore, once a process has been systematized and undesirable variation has thereby been reduced, meaningful research can be conducted on the process or outcome in some situations. This can be done using quasi-experimental designs,¹⁸ in which a new way of performing one step in the defined process is compared with the older approach, for example, using a different antibiotic for treating UTIs. More simply, descriptive data may be collected in the course of routine care to better understand the attitudes and behavior of clinicians or patients. For example, physicians can be asked to document why they prescribe antibiotics in situations where the drugs do not seem warranted or women can be asked how they feel about undergoing cervical cancer screening less frequently than they have experienced in the past. In these ways, improvement data can be used to study and improve care with wide generalizability.

Nelson and his colleagues at Dartmouth University (Hanover, NH) have been particularly interested in exploring these ways of using measurement data for multiple purposes.¹⁹ One example of multiple data uses is their notion of an *instrument panel*, which is used to track important areas for improvement opportunities. For example, the Northern New England Cardiovascular Disease Study Group^{20,21} is using an instrument panel to monitor and improve coronary artery bypass grafting. The Dartmouth quality group is also using patient-completed questionnaires for geriatric patients to provide both immediately usable information to their clinicians and aggregate feedback information for monitoring and designing improved care systems.²²

Note, however, that these efforts to use measurement for more than one purpose scrupulously avoid pressing improvement data routinely into service for external accountability. In the early days of ICSI, many of the disputes and problems around measurement plans arose from the fear of such usage. For example, as discussed in Sidebars 1 and 2, many medical groups did not want to provide measurement results to ICSI because they thought the measurements did not accurately

Table 3. Recommendations on How to Make More Use of Measurement in Improvement Efforts

1. Limit the number of measurements.
2. Pick measurements that are important to clinicians (and patients), ideally by having the users do the selection.
3. Make the data collection easy enough and the time frames short enough so that data collection can be repeated frequently to allow for trending changes over time.
4. Do not try to have the measures serve accountability or research purposes at the same time as improvement.
5. Build in baseline measures before implementing any changes.
6. Provide training, tools, and examples to those in clinical settings who are not used to data and this type of measurement.

reflect their care and would be used against them. Now that we have a better understanding of this problem and of the critical importance of measurement to improvement efforts, measurement efforts are both more effective and more harmonious.

Fortunately, the organizations with which ICSI works are understanding and supportive in keeping these measurement types separated. Our sponsoring health plan, HealthPartners, is collecting its own information for its Consumer Choice system of information about participating medical groups to provide health plan members with more information as they make their selections of medical groups and physicians. Similarly, the BHCAG does not even want to know which medical groups have provided which improvement data to ICSI. They know from their own quality programs that making such data public or using it for medical group accountability would simply threaten to make it less useful for improvement.

We and our colleagues at ICSI have learned a lot about measurement during the past three years. Based on this learning, we have some recommendations to others interested in making more use of measurement in improvement efforts (Table 3, above).

In addition, we have learned that it is important to have data for improvement, that these data do not need to meet research standards, and that we could turn

our original bad data into good data with better and more practical methods. However, the most important thing we have learned is to avoid the misuse and potential harm of measurement by understanding and acting on the differences among the purposes for which measurement is intended.

Summary and Conclusions

In the current climate of public accountability, many clinicians have become uncomfortable with any efforts to create measurement systems. That is unfortunate because measurements are absolutely essential to efforts

for improving the processes of medical care. In their guideline implementation and measurement efforts, ISCI and the IMPROVE Project in Minnesota have gradually learned how to distinguish between measurement for improvement and that for accountability. Both approaches are different from the approach that physicians are used to in their encounters with medical research. Understanding these differences and respecting the confidentiality of individual medical groups has been crucial to moving past confusion and suspicion to genuine improvement actions involving multiple medical groups and their contracting managed care plans. ■

References

1. O'Leary DS: Measurement and accountability: Taking careful aim. *Jt Comm J Qual Improv* 21:354-357, 1995.
2. Nelson EC: Measuring for improvement: Why, what, when, how, for whom? *Quality Connection* 4(2):1-3, 1995.
3. Reinertsen JL: Collaborating outside the box: When employers and providers take on environmental barriers to guideline implementation. *Jt Comm J Qual Improv* 21:612-618, 1995.
4. Mosser G: Clinical process improvement: Engage first, measure later. *Qual Manag Health Care* 4(4):11-20, 1996.
5. Mosser G, Reinertsen JL: Healthcare guidelines as engines for improvement. *Postgrad Med* 100(3):15-18, 1996.
6. Solberg LI, et al: Using continuous quality improvement to increase preventive services in clinical practice: Going beyond guidelines. *Prev Med* 25:259-267, 1996.
7. Shewhart WA: *Statistical Method from the Viewpoint of Quality Control*. Washington, DC: U.S. Department of Agriculture, 1939.
8. Berwick DM: Continuous improvement as an ideal in health care. *N Engl J Med* 320:53-56, 1989.
9. Berwick DM, Godfrey AB, Roessner J: *Curing Health Care*. San Francisco: Jossey Bass, 1990.
10. Batalden PB, Stoltz PK: A framework for continual improvement of health care. *Jt Comm J Qual Improv* 19:424-447, 1993.
11. Berwick DM: The clinical process and the quality process. *Qual Manag Health Care* 1(1):1-8, 1992.
12. Institute for Clinical Systems Integration: Uncomplicated urinary tract infection in women. In *Health Care Guidelines*. Bloomington, MN, 1996, pp 525-539.
13. Berwick DM: Controlling variation in health care: A consultation from Walter Shewhart. *Med Care* 29:1212-1225, 1991.
14. Brook RH: Implementing medical guidelines. *Lancet* 346:132, 1995.
15. Wall EM: Practice guidelines: Promise or panacea? (Editorial) *J Fam Pract* 37(1):17-19, 1993.
16. Mosser G: What information do clinicians need? In Schoenbaum SC (ed): *Measuring Clinical Care: A Guide for Physician Executives*. Tampa, FL: American College of Physician Executives, 1995, pp 63-70.
17. Gift RG, Mosel D: *Benchmarking in Health Care: A Collaborative Approach*. Chicago: American Hospital Publishing, Inc, 1994.
18. Marvel MK, Staehling S, Hendricks B: A taxonomy of clinical research methods: Comparisons of family practice and general medical journals. *Fam Med* 23(3):201-207, 1991.
19. Nelson EC: Using outcomes measurement to improve quality and value. *New Dir Ment Health Serv* 71:111-124, Fall 1996.
20. Nugent WC, et al: Designing an instrument panel to monitor and improve coronary artery bypass grafting. *J Clin Outcomes Manag* 1(2):57-64, 1994.
21. O'Connor GT, et al: Multivariate prediction of in-hospital mortality associated with coronary artery bypass graft surgery. *Circulation* 85:2110-2118, 1992.
22. Nelson EC, Wasson JH: Using patient-based information to rapidly redesign care. *Healthc Forum J* 37(4):25-29, 1994.