

March 6, 2006

Tap Into the Power of Comparative Performance Data

By Patrice L. Spath, BA, RHIT

For The Record

Vol. 18 No. 5

The increased public attention afforded to healthcare quality and safety issues over the past 15 to 20 years has resulted in a corresponding demand for comparative performance data. Early efforts by the healthcare industry and the federal government involved using existing claims data for comparative purposes. Later initiatives led to more refined efforts to define and apply performance measures that could enable practitioners and providers to compare performance among hospitals. Payers, regulators, practitioners, and institutions used the data produced from these measures to identify unnecessary practice variation—a significant source of both cost and quality issues.

These and other activities spurred by the federal government's healthcare reform efforts in the mid-1990s paved the way for the performance report cards we have today. Report cards represent the industry's attempt to package performance data and present them to the public in an accessible way to enable more informed healthcare choices. They also serve providers by giving them comparative information that can be used for performance improvement activities.

There has and continues to be an almost frenzied effort to develop the ideal healthcare report card in both the private and public sectors. The demand for publicly available comparative performance measurement data continues to increase steadily, and the development and dissemination of healthcare provider report cards remains on the rise. The generation of comparative performance data has proliferated as more stakeholder attention has been placed on measuring, quantifying, and reporting performance. It seems there's a virtual sea of available comparative data, which raises valid questions about its value. Do the reports contain accurate and reliable measures of performance? Are the data useful for improving performance?

To gain the improvement benefits of comparative performance activities, the data must be transformed into meaningful information that promotes understanding, acceptance, and action. Data analysis is what enables this transformation. Examining comparative data and presenting the results of your analysis in a way that can lead to patient care improvements is neither a matter of luck nor mystery. There are three basic steps you can take to analyze the data and report the results to your organization's administration, physician leaders, and other relevant groups for discussion and action. The flowchart in Figure 1 provides an overview of the three steps and the decisions that eventually lead to a determination for action.

Step 1: Verify the Accuracy of Your Data

Although tedious and painstaking, you must check the data you submitted against the data reflected in the comparative report. There's no way of getting around this task. The most straightforward way to accomplish this check involves verifying the numbers from your database or other input documents against the values in the report. If you find data quality problems, be careful about presenting the data to anyone in your organization. If possible, hold off on presenting the data until it's corrected. Report the accuracy problem to the organization to which you submitted the data and request that corrections be made and an amended report issued.

Even when the data you submitted match the information provided in the comparative report, data reliability can still be a problem. Data with little reliability have limited value for comparison purposes. The reliability of data is determined by how data are produced. For example, if you are comparing your rate of various types of patient incidents with the rates reported in other organizations, it is difficult to guarantee the accuracy of the information. First, you cannot be 100% certain that all patient incidents are being reported in your organization. Studies indicate that less than 10% of some types of incidents are regularly reported in any given organization. Even if you know the approximate percentage of incidents actually being reported by people in your facility, the rate of incident reporting in the other organizations submitting data is unknown. Comparative reports of voluntarily reported patient incidents are highly suspect for data reliability problems and should be used cautiously.

Data derived directly from administrative databases used for billing purposes may have greater reliability due to the financial implications of inaccurate data. In addition, vendors of measurement systems often have data quality control systems in place. However, the techniques used to ensure data accuracy vary among vendors. Check with

your vendor to determine how it evaluates data accuracy and what its analyses reveal. Data have high reliability if repeated measurements of the same phenomenon provide consistent results.

Step 2: Acquaint Yourself With the Comparison Group

Ideally, your organization participates in comparative measurement systems that include data from facilities similar to yours. Nonetheless, you may find that your performance rates are being compared with groups that have quite dissimilar characteristics. This may represent a reporting problem—the measurement system reports do not stratify results by important or influential facility characteristics or patient characteristics.

In some instances, the problem may originate in the data collection process—information about important and influential facility characteristics or patient characteristics are not entered into the database. For example, your organization may have a higher (or lower) number of indigent patients and yet that data element is not part of the comparative database. Without valid comparisons, the measurement results may be faulty. A valid comparison is one in which the comparative data actually measure the concept they purport to measure.

Valid comparison groups are difficult to define. It is virtually impossible to create peer groups that are identical in all ways: clinically, demographically, and therapeutically. There are literally hundreds of variables that may impact clinical outcomes and quality indicators. These variables range from early childhood experiences to current stressors and comorbid medical illnesses. Risk adjustment of index and benchmark populations as well as population stratification can help enhance the validity of the comparisons, although the data necessary for these adjustments are hard to come by. Comparison performance reports can guide your improvement efforts, but such comparisons will never be perfect.

When reviewing comparative performance data for your organization, ask yourself: Is this a valid comparison? Are there facility or patient characteristics not accounted for? Some data systems use wage- and case-mix adjustments to account for unique facility characteristics and risk-assessment systems to account for differences in patient severity and acuity. If you are not satisfied that the results represent a valid comparison for your facility, it may be possible to request special reports from the data system vendor.

Many comparative measurement systems can produce custom reports in various formats. A participating facility can request the subset of aggregate data based on the characteristics of the peer group of interest. Any number of peer group characteristics (facility and patient) can be combined to create reports; however, the variables of interest must be contained in the database. For example, participants in the National Trauma Registry System (TRACS), sponsored by the American College of Surgeons and its Committee on Trauma, can select from many facility and patient variables. Examples of the facility-specific data elements include the following:

- number of adult and pediatric hospital beds;
- number of intensive care unit (ICU) beds available for trauma patients and burn patients;
- hospital teaching status; and
- trauma level designation.

There are numerous TRACS data elements related to patient characteristics, including the following:

- age;
- gender;
- race/ethnicity;
- principal payment source;
- ICD-9-CM codes of diagnoses and procedures;
- Glasgow Coma Scale total in the emergency department (ED);
- Revised Trauma Score in the ED;
- first temperature score in the ED;
- length of stay in the hospital;
- total days in the ICU;
- presence or absence of ventilator support; and
- total Functional Independence Measure score.

Participants in the Quality Indicator Project, sponsored by the Maryland Hospital Association, can produce custom reports based on a single measure, on all the measures in an individual indicator, or for all measures in an entire indicator set. The report can cover an individual quarter or a range of quarters. Custom reports for an individual measure can be displayed either as a table of statistics or one of two graphs. The graph options include a comparison chart or a box plot. Custom reports based on all measures in an individual indicator or on all measures in an entire

indicator set are available in table format only. The table can be displayed on screen by participants, or may be downloaded as either an Adobe Acrobat file or a tab-delimited file that can be exported to another application.

Before you determine whether a peer group comparison is valid for your purposes, you need to have an understanding of the facility or patient characteristics that affect measurement results.

What if you were considering the question: How does our rate of long stays (greater than six hours) in the ED compare with other hospitals? If you are an urban teaching hospital, you will want to be compared with other urban teaching hospitals. If you also have a high volume of ED patients, you'll want the peer group to include only urban teaching hospitals with high ED patient volumes. This does not preclude the possibility of comparing your ED long stay rates with hospitals that don't meet the urban, teaching, high-volume criteria. Your facility may find it useful to see how "everyone else" compares with your experience.

You may be unable to manipulate the peer group comparisons through production of customized reports or other means. In these instances, it is important to understand the potential limitations of the comparative data and communicate these limitations to users. For example, say your skilled nursing facility is shown to have a higher-than-expected rate of patients with pressure ulcers. Why? Because your facility has a first-class wound care program and patients at high risk for pressure ulcers are more likely to be admitted to your facility. Don't be too quick to attribute the cause of significant variations. An in-depth analysis should be done to determine what factors are affecting performance rates at your facility.

It is unlikely that perfect peer group comparisons will ever be possible. Either unique facility and/or patient characteristics will not be adequately accounted for. However, perfection is the enemy of improvement—understand what the data are (and are not) telling you. Don't expect comparison data to be perfect before actions can be taken to investigate variations. When you're satisfied with the accuracy of the data reflected in the report and the relative validity of peer comparison groups, you're ready to move on to Step 3.

Step 3: Uncover the Story in the Data

Now it's time to compare your performance with that of other institutions or programs. If it is significantly worse than desired, further analysis must be undertaken to determine why your performance is not at the level of other organizations. On the other hand, if a comparison chart shows your performance is comparable or better than the peer group, you can look at the stability of your individual performance to identify opportunities for improvement.

Control charts are one technique for evaluating stability of your performance. Different types of control charts are used to display different types of data. Shown in Figure 2 is an example of one type of control chart, an X-bar chart. The chart displays the surgical wound infection rate in a hospital for a 30-month period. In two of the months (sixth and 29th), the rate exceeded the upper control limit. What occurred during these months definitely needs to be investigated. However, even if the upper control limit had not been exceeded, it would be important to determine why the rates are showing considerable variation from month to month. Do these variations represent improvement opportunities?

Comparative performance data can tell many different stories. Reactions to the information will vary according to the significance of the results and the organization's improvement priorities. The reason your organization chose to be involved in the comparative project will also influence the reaction to the data. For example, if information is being submitted to the comparative database primarily to satisfy accreditation requirements, then clinicians may be less apt to investigate seemingly small differences in performance. However, if clinicians are greatly interested in how patient outcomes for a particular condition compare with the outcomes achieved by other organizations, even small variations may be thoroughly evaluated and acted upon.

It is important to understand the motivations of your audience when looking for the stories in your data. To determine what the comparative performance data is telling you, three factors must be evaluated:

- the difference between your performance rate and the comparison group;
- the amount and type of variation in trend reports of your performance data; and
- the gap between your performance rate and internal expectations.

The first question to ask about the data is whether the performance rate reported by your organization is different from that of other organizations. In some instances, this question can be answered through routine observation; however, more information is needed to determine whether the differences are statistically significant.

The concept of statistical significance is derived from the central limit theorem and the probabilistic properties of a normal distribution. For example, suppose 1,000 patients are asked four hours after surgery to rate their pain on a scale of 1 to 8. Theoretically, when the number of times each reported pain rating is plotted on a graph, the results would approximate a bell-shaped curve, referred to as a “normal distribution.” Once the data are gathered, the mean and standard deviation (a mathematically determined measure of variance in a normal distribution) are calculated.

One would expect that the pain scores reported most often (approximately 68%) would fall within one standard deviation from the mean in either direction. Two standard deviations from the mean would account for roughly 95% of the reported pain scores. And roughly 99% of all the pain scores reported would fall within three standard deviations from the mean. The chance of a pain score being reported beyond four standard deviations is very small. The theoretical normal distribution concept with standard deviations added is depicted in Figure 3.

So how does this relate to comparing performance? There is only a 5% chance that your rate will fall more than two standard deviations from the mean of the comparison group and only a 1% chance that your rate will exceed three standard deviations. The probability of exceeding two or three standard deviations is low—meaning that exceeding these parameters is statistically significant. How significant? It depends somewhat on the opinion of the people involved in the analysis. Some will suggest that any rate that exceeds two standard deviations should be looked at more closely for the cause of variation. Others establish three standard deviations as the cut-off point for significance.

An arsenal of statistical methodology is available to evaluate the significance of performance results reported by an organization. These methodologies range from simple analyses (eg, t-test, chi-square test) to sophisticated (eg, multiple logistic regression, Cox proportional hazards model). If you do not have sufficient expertise to judge the relevance of the results, the performance measurement system vendor may provide some form of assessment of the statistical aspects of the comparative reports. In all likelihood, any finding of unusually high or low performance is probably not due to chance alone.

Tests of statistical significance and other data analysis techniques are merely tools that can help point to potential problem areas. Don't rely solely on these tools to identify improvement opportunities. Your facility's performance may be consistent with other organizations and yet improvements are still desirable. Why? Because people in your organization have established stretch goals. Average or slightly better than average performance is not the stopping point in an organization with stretch goals—people aspire to achieve optimal performance.

Quality goals, reflective of a commitment to performance excellence, may be established as part of your organization's strategic planning process. Medical staff departments may select certain conditions or procedures for which they have higher-than-average performance expectations. Nursing and other clinical departments can launch improvement projects that are intended to achieve lofty goals, such as a restraint-free environment. For any number of reasons, it may be necessary to look beyond statistical significance and process variation and decide whether your organization is realizing internally defined expectations.

The purpose of sharing the story in the data is to obtain support for a more in-depth investigation of the findings. If there appears to be opportunities to improve your organization's performance, careful presentation of the results will motivate the right people to take the first steps: acknowledge the existence of significant performance variation and the need for further investigation.

Figure 1: Steps of Comparative Performance Data Analysis

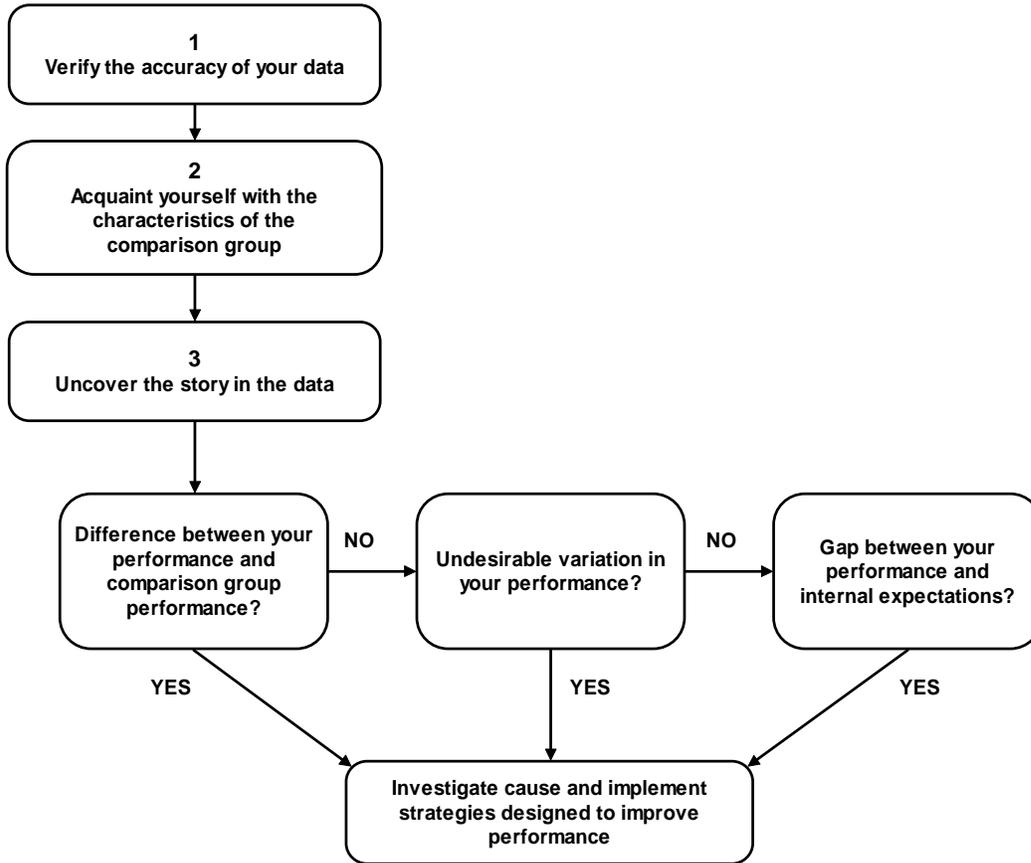


Figure 2: Surgical Wound Infection Rates Plotted on an X-bar Control Chart

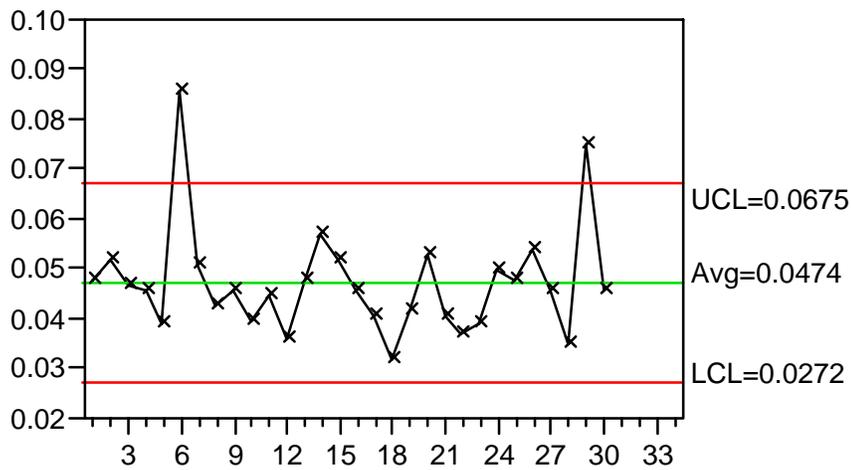


Figure 3: Theoretical normal distribution with standard deviations

