Another View on How to Measure Health Care Quality

by James C. Benneyan and Frank C. Kaminsky

Note: This column, written by two professionals who specialize in health care quality improvement, provides an alternate view on some of the issues discussed by David Birnbaum in his April 1994 guest column on applying statistical methods in health care. I hope you will be as interested as I am in the discussion of how statistical tools are applied—that is, the context in which tools are chosen and used—and how this impacts their effectiveness. As is customary in such discussions, I have given Birnbaum an opportunity to respond to the points raised by James C. Benneyan and Frank C. Kaminsky.

I know I speak for all authors when I urge interested readers to comment on this or other columns by contacting me directly or through letters to the editor. Health care quality improvement and cost reduction are vital national issues to which Quality Progress readers should contribute. I look forward to learning about your views.—Bert Gunter

As a guest contributor to Statistics Corner, David Birnbaum, an epidemiologist, discussed traditional epidemiological approaches to health care quality, their similarities to total quality management (TQM), and some difficulties in measuring health care quality. In general, we agree with many of Birnbaum’s statements about the similarities that exist between TQM and epidemiology and the view that both disciplines are valuable for studying and improving health care quality. We do wish to comment, however, on several other statements in Birnbaum’s column, especially those that might lead Quality Progress readers to believe that health care systems are different from manufacturing systems and, consequently, are less likely candidates for the successful application of statistical process control (SPC).

In addition, while Birnbaum advocates the increased use of outcome-based measurements for the assessment of health care quality, readers should be aware that not all TQM practitioners support the present use of such measures. In fact, we believe that using these measures, especially as aggregate metrics, represents a step in the wrong direction that will, in the long run, hinder efforts to improve health care systems using W. Edwards Deming’s methodology.

These concerns are based on our views of Deming’s philosophy, TQM objectives, and the importance of statistical methods to achieve these objectives. These views are reflected by the following statements:

1. In a preface to an article on health care quality improvement by D.A. Bergman in Pediatrics in Review, the editor, a pediatrician named L.F. Nazarian, made the following comment: “Simply stated, quality improvement refers to organized, rational, scientifically valid programs that analyze what people are doing—in industry, medicine, or any other area—and then devise ways in which the job can be done even better.”

2. Deming, who was adamant about the importance of studying process variation over time, repeatedly stated: “If I had to reduce my message for management to just a few words, I’d say it all had to do with reducing variation.”

If a primary objective of TQM is to improve health care by reducing process variation, the behavior and variability of the present system must be understood. Statistical thinking in developing this process understanding is vital, and SPC should be used to study the statistical behavior of the process over time. Despite increasing attention within health care on the value of quality management principles, the use of basic SPC has yet to become widespread.

Applicability of SPC to nonclinical, clinical, and laboratory processes

Birnbaum described the history and current state of hospital epidemiology as one that has progressed from “numerical analysis by hypothesis testing to a preference for confidence intervals, and has yet to fully embrace exploratory data analysis by graphical methods.” The term “graphical methods” refers to diagrams and charts, such as those typically used by SPC practitioners. Birnbaum also discussed some concerns about the effectiveness and appropriateness of using SPC in epidemiological investigations. For example, Birnbaum discussed the differences in applying statistical tools for measuring quality in health care vs. other industries, with particular emphasis on some complex issues related to infection control. While differences certainly exist between all industries, such statements, taken out of context, might unintentionally mislead health care quality professionals to the opinion that SPC is less applicable in health care than in other industries.

Others have also suggested that hospital systems are somehow different from manufacturing systems and are therefore unlikely candidates for SPC implementation. There are many documented cases, however, that should dispel the misconception that SPC techniques are ineffective in analyzing health care systems. While many of Birnbaum’s comments were specific to the study of the complex topic of infection control, SPC methods have wide application and have been used extensively to analyze nonclinical, clinical, and laboratory systems. In addition, experimental design has been used successfully to reduce the natural variation of nonclinical and clinical processes once they were brought into states of statistical control.

SPC is equally applicable to clinical and nonclinical health care processes. For example, in the area of nosocomial infections, Birnbaum stated that, “Since 90% to 98% of nosocomial infections are random events rather than part of adverse trends (endemic rather than epidemic), SPC alone is not a sufficient tool.” The phrase “random events” implies a state of statistical control. This statement is associated with later statements that “outliers and runs do not invariably signal quality defects” and that the goal is continual improvement of health care and “not simply managing a system in statistical control.” The study of random events is precisely the objective of SPC and, in the case of nosocomial infections, the geometric distribution has been used to detect increases in infection rates.

In fact, in earlier papers, Birnbaum discussed issues related to the incidence of nosocomial infections and illustrated a surveillance approach for monitoring in-
fection rates. This surveillance approach is quite similar to the concept of a p control chart, with an adjustment for hospital census and patient severity mix. In any case, without examining these data over time with respect to appropriate statistical control limits, the infection rate could increase and the potential to improve processes, reduce costs, and save lives could go undetected. For example, classical methods might report only an average rate for the entire time period, perhaps summarized and compared by year or quarter.

In clinical laboratories, a variety of techniques—including basic probability theory, histograms, statistical control charts, mathematical optimization, and computer simulation—have been used to develop an understanding of variability in the Pap smear screening process used in cervical cytology. As in any industry, sometimes more advanced methods need to be used or developed to account for process complexity. For example, a new type of statistical control chart was developed to monitor the use of enzyme-linked immunosorbent assays to test for HIV antibodies. A similar study used SPC to develop an understanding of the variability of a different process for testing for HIV antibodies. It is interesting to note that the use of SPC in clinical laboratories was suggested at least as early as 1950.

**Value of SPC in traditional hospital epidemiology**

Many traditional statistical analysis methods used by both statisticians and hospital epidemiologists can be thought of as static, in the sense that the chronological order and variability of the process data over time are not of primary interest. For example, using classical chi-square tests and one-way analysis of variance (ANOVA) to test for significant differences in means are examples of static (time-independent) tests. When variability over time is important, these tests might not be appropriate. Many SPC techniques—such as statistical control charts, reliability engineering, autocorrelation analysis, and time-series analysis—examine the dynamic nature of process variability across time and should be used whenever possible to understand health care systems. More important, understanding variability over time provides more process information; the use of statistical control charts can result in completely different conclusions from those reached through traditional static methods.

For example, the accuracy of laboratory readings of Pap smears as a method for detecting cervical cancer (in particular, the rate of false negatives) is an important example of using SPC to analytically understand and improve health care processes. This situation is one in which traditional static methods have proved less diagnostic than SPC. As Table 1 shows, the results of a one-way ANOVA fail to reject the hypothesis that all results came from the same process (p = 0.271). The trial control chart in Figure 1 contains the same data, but it clearly indicates that the laboratory process was not in statistical control and had significantly deteriorated over time. Control charts can also be used to describe and improve laboratory processes or to identify particular individuals in need of retraining. In fact, in situations in which the goodness of a process can have enormous liability consequences, the use of SPC and related methods to understand and take appropriate management action has been a deciding factor in determining legal responsibility.

**Use of outcome metrics, standards, and report cards**

Early in his article, Birnbaum stated that health care is assuming more of an outcome focus "after decades of focus..."
Many of the "differences" that complicate data analysis in health care also exist in other industries. For example, autocorrelation, cyclic or trending behavior, significantly correlated multivariate, censored data, retrospective data, raw data unadjusted for mixed populations, and aggregated data can negate standard SPC assumptions of independent and identically distributed random variables in any industry.

Of course, when mixtures of populations and other types of heterogeneity exist, more advanced methods should be used to study these processes. For example, although patients certainly do vary, several comparison approaches are possible, such as analysis of patients' lengths of stay adjusted for the severity of admitting conditions. The issue of evaluating the competence of individual doctors also illustrates the important concept of severity-adjusted analysis. For example, the basic idea of the much-debated coronary-artery bypass graft mortality data is that, by publishing survival rates by individual heart surgeons, the public can somehow identify the "good" vs. "bad" doctors. Arguably, raw unadjusted data can make the best heart surgeons at the best hospitals appear the worst, as they might primarily operate on the toughest surgical cases.

In his article "Robust Design: A New Tool for Health Care Quality?", A. Blanton Godfrey suggests that robust methods can also be used to achieve more consistent quality of health care processes despite uncontrollable variation, such as patient populations, admitting conditions, or appointment timeliness. Godfrey acknowledges that, "In health care, as in complex manufacturing or design processes, there are many process control variables but even more noise variables, or variables that cannot be controlled." A more advanced mathematical data analysis method, called data envelopment analysis, is being used successfully to examine multivariate, heterogeneous health care processes, such as the identification and transfer of best physician practice styles.

The future of quality measurement in health care

In "Measuring Health Care Quality," Birnbaum presents an epidemiological approach to quality in health care systems. He gives considerable food for thought and encourages those interested in improving health care quality to look
closely at standard techniques used by statisticians in other fields. One important example is the use of statistics to investigate variation in clinical quality applications, such as Pap smear, HIV, and hepatitis testing. We hope that publication of more articles of this type will continue to bring together TQM practitioners from various fields and encourage others to develop a wider range of tools that can be applied to improving health care systems.

SPC can be used effectively to complement traditional methods of hospital epidemiology, particularly when the time sequence of data can reveal important information. While subtle differences certainly exist in every industry, health care is not nearly as distinct as many suggest. SPC and other quality engineering methods can be as useful in health care as in other industries. A more relevant question concerns who should advocate and apply SPC to ensure the greatest and quickest effect on improving health care. Quality engineers and epidemiologists can benefit from greater interaction with each other. Their related methods and focus provide a solid foundation to expand the application of SPC for the continual improvement of health care.

Birnbaum’s Reply

I’m in substantial agreement with James Benneyan and Frank Kaminsky and thank them for amplifying key points that might not have been clear in my brief article. SPC is underutilized in health care, especially in the “hotel service” aspects and defined clinical support processes in which experiments can yield clear choices for SPC variables. (This applies to laboratory procedures such as those described by Benneyan and Kaminsky.) These aspects are no different from those in other industries. Techniques advocated by such luminaries as W. Edwards Deming and John Tukey are useful additions to hospital epidemiologists’ analytical tools.

Clinical outcomes, however, tend to be incompletely defined, stochastic, multifactorial events beyond the realm of true experimentation. They are best studied by proven epidemiologic methods. The unwary analyst can be mislead by casual adoption of Deming’s and others’ advice to simply count infections or by assuming that industries’ familiar ±3 standard deviation limits provide optimal sensitivity and specificity.

Selecting the wrong target event (which might be as subtle as counting all infections or all falls rather than first infections or first falls or counting the number of patients who suffer one or more repetitive adverse outcomes) can obscure the most meaningful risk markers or causative risk factors, thus rendering techniques such as life-table analysis or SPC less powerful. Further, the predictive value of many clinical factors or measurements short of case outcome is insufficient to warrant their adoption for SPC charting. Proliferation of administratively simple but misleading report card metrics is a prime example of the worst in today’s practices. It results from applying simple statistical techniques without sufficient understanding of the epidemiology involved.

In the past, my health care experience and background led me to conclude that:

• “Control committees” in hospitals frequently contained content but not process knowledge.
• Those appointed to monitor processes usually had insufficient knowledge of epidemiology and biostatistics.
• Divergent groups often spent more time belittling each other’s methods than studying methodology together.

Recent events, however, might be leading to improvement. It is significant that the Society for Healthcare Epidemiology of America and the Joint Commission on Accreditation of Health Care Organizations have just initiated a collaborative project to critically examine quality indicator systems. The project goal is to use rigorous epidemiology to identify the processes that drive quality indicators and, through ongoing research, to understand the sources of indicator variation, identify which sources of variation can be controlled, determine how indicators can be improved to better reflect good hospital care, and to disseminate knowledge to guide other institutions’ quality improvement efforts.

Health care is different, but it is neither so unique nor too complex to benefit from SPC. Readers should not assume that simple methods like SPC do not apply to health care. On the other hand, readers should also not assume that these methods can be transplanted from other industries to health care without validation and refinement.

I strongly endorse Benneyan and Kaminsky’s final comments. More frequent interaction between quality engineers, epidemiologists, health care providers, and others is not an option for mutual benefit—it is a requirement for survival in the interdisciplinary area of health service evaluation.

References

Statistics Corner cont.


James C. Benneyan is a senior quality engineer at Harvard Community Health Plan in Brookline, MA. He received a master's degree in industrial engineering and operations research from the University of Massachusetts in Amherst. Benneyan is a student member of ASQC. Any questions or comments about this column may be sent to him through e-mail: benneyan@eecs.umass.edu.

Frank C. Kaminsky is the principal of Productivity Sciences Inc. in Greenfield, MA. He received a doctorate in industrial and systems engineering from Northwestern University in Evanston, IL. Kaminsky is a member of ASQC.

David Birnbaum is a consultant and author of software for hospital epidemiology and infection control in Sidney, British Columbia, Canada. He holds an interdisciplinary doctorate from the University of British Columbia in Vancouver. Birnbaum is a member of ASQC.

Bert Gunter is a contributing editor of *Quality Progress*, a member of ASQC, and a statistical consultant. Any questions or comments about this column may be sent to him at PO. Box 9, Hopewell, NJ 08525 or through e-mail: bgunter@mcmail.com.

What did you think about this article?

*Quality Progress* needs your feedback. On the postage-paid reader service card inserted toward the back of this magazine, please circle the number that corresponds with your opinion of the preceding article.

- Excellent Circle #381
- Good Circle #382
- Fair Circle #383
- Poor Circle #384