

Counterpoint

'Counterpoint' is an occasional feature presenting discussion of a topic that is currently under debate in quality of care circles. We invite readers to submit Letters to the Editor adding their opinion to the topic.

Statistical process control tools for monitoring clinical performance

The quality assurance of medical practice in most countries is effected through a mixture of informal assessment and peer review, and through more formal accreditation, credentialing or delineation of privilege. The process of assessment and review is often subjective and without explicit reference to pre-determined standards of practice. It has been argued that comparative treatment outcome data on an individual doctor's performance is required to make quality assurance credible [1]. Equally, many would add that objective and quantitative methods to monitor the quality of a doctor's performance based on treatment outcome data could be more widely applied and would lend credence to the quality assurance process.

Statistical process control (SPC) tools such as control charts have been widely used in the manufacturing industry for a long time. Also, the case for their application in health care has long been made [2–4]. Resistance to their more widespread adoption in health care may stem from lack of evidence or lack of conviction regarding their power or utility. This issue of the *International Journal for Quality in Health Care* contains yet another demonstration of the utility and power of such techniques. The application of risk-adjusted sequential probability ratio tests (SPRT) retrospectively to two high profile examples, the Bristol Royal Infirmary pediatric cardiac surgery data and Harold Shipman's data, demonstrate convincingly how in both cases the technique would have provided early warning of poor performance. This should argue cogently for the SPC case and add to its evidence base.

The application of SPC in the clinical context poses special challenges. The best-known control charts are those pioneered by Walter Shewart [5], such as his Xbar and R charts. Shewart charts, however, are designed to detect large but transient shifts in the process mean, typically in large volume manufacturing processes. This limits their application in clinical practice for several reasons:

- (1) The throughput of clinical practice is typically very slow; for example, a surgeon may perform no more than 1–5 procedures a day. It is both undesirable and inconvenient for a performance monitoring system to require sample sizes of greater than one to accumulate before analysis.
- (2) In clinical practice, even a small shift in process mean may be of concern; for example, it may indicate adverse deterioration in mortality rate or complication rate.

This issue is of obvious concern for monitoring clinical practice, which requires early warning of poor performance before too many patients are harmed.

- (3) Case-mix varies among practices, and adjustment for case-mix is required to improve the accuracy of the signal. Such regression adjustment filters out the component of variability in the outcome measurement induced by factors outside the process being monitored, such as case-mix. This enhances the traceability of the source of the problem, and thus avoids efforts being wasted on investigating non-existent problems or, worse still, unfair accusation of poor performance toward practices with a large number of high-risk patients.

Experience in applying such SPC techniques as SPRT, as demonstrated in the paper by Spiegelhalter *et al.* in this issue [6], and CUSUM [7–14] have shown that these challenges could be met successfully. These experiences in applying SPRT and CUSUM in clinical performance monitoring should also help to dispel the perhaps common misconception that such techniques could only detect extreme aberration in performance [15]. Indeed, one may argue that current practices that are based on subjective review or a simple monitoring scheme that does not exploit the power of modern SPC techniques have yet to demonstrate their capability in providing early warning of poor performance in clinical practice.

More statistical techniques are being added to our toolbox for use in clinical monitoring [16,17] and existing techniques are being refined [18]. The plethora of techniques may be confusing. While all these techniques are demonstrably effective in signaling poor performance—an obviously fundamental requirement we must ask of them—they are not necessarily equally effective. There is one simple criterion for picking the optimal technique and its design for a given monitoring situation. Monitoring schemes such as SPRT, CUSUM and others are designed to detect small changes in the treatment process. Unfortunately, this also means they would allow an extremely tiny shift of no clinical importance to eventually produce a signal. In other words, a false alarm may occur. Any monitoring system that is prone to frequent false alarm will be quickly discredited, and not be used. This is the inherent trade-off between sensitivity and false alarm in any monitoring system. We want the system to be sensitive

so that it can provide early warning of an adverse trend in performance. On the other hand, we do not want the system to be so sensitive that it is prone to false alarms. Among all techniques designed to have the same false alarm rate, the optimal technique is that which detects adverse deterioration in outcomes the quickest.

By current published evidence, the case for applying SPCs in monitoring clinical performance, in my opinion, is very strong. More research, both theoretical as well as empirical, should be devoted to refining these techniques to improve their sensitivity, and more comparative research of the relative performance of these techniques in typical monitoring situations is warranted.

T. O. Lim

Clinical Research Centre,
Hospital Kuala Lumpur,
Jalan Pahang, Kuala Lumpur, Malaysia

References

1. Johnson JN. Making self-regulation credible. *Br Med J* 1998; **316**: 1847–1848.
2. Laffel G, Blumenthal D. The case for using industrial quality management in health care organizations. *JAMA* 1989; **262**: 2869–2873.
3. Blumenthal D. Total quality management and physicians' clinical decisions. *JAMA* 1993; **269**: 2775–2778.
4. Kritchevsky SB, Simmons BP. Continuous quality improvement: concepts and applications for physician care. *JAMA* 1991; **266**: 1817–1823.
5. Shewart WA. *Economic Control of Quality of Manufactured Product*. New York: Van Nostrand Reinhold Co., 1931.
6. Spiegelhalter D, Grigg O, Kinsman R, Treasure T. Risk-adjusted sequential probability ratio tests (SPRTs): applications to Bristol, Shipman and adult cardiac surgery. *Int J Qual Health Care* 2003; **15**: 7–13.
7. De Leval, MR. Analysis of a cluster of surgical failures: Application to a series of neonatal arterial switch operations. *J Thorac Cardiovasc Surg* 1994; **107**: 914–924.
8. Parry BR, Williams SM. Competency and the colonoscopist: A learning curve. *Aust NZ J Surg* 1991; **61**: 419–422.
9. Williams SM, Parry BR, Schlup MM. Quality control: an application of the cusum. *Br Med J* 1992; **304**: 1359–1361.
10. Van Rij AM, McDonald JR, Pettigrew RA *et al*. Cusum as an aid to early assessment of the surgical trainee. *Br J Surg* 1995; **82**: 1500–1503.
11. Kestin IG. A statistical approach to measuring the competence of anaesthetic trainees at practical procedures. *Br J Anaesth* 1995; **75**: 805–809.
12. Bolsin S, Colson M. The use of the Cusum Technique in the assessment of trainee competence in new procedures. *Int J Qual Health Care* 2000; **12**: 433–438.
13. Lim TO, Soraiya Azmi, Zaki Morad, Ding LM. Assessing Doctors' performance: Application of CUSUM technique in monitoring doctor's performance. *Int J Qual Health Care* 2002; **14**: 251–258.
14. Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000; **1**: 441–452.
15. Frankel S, Sterne J, Smith G. Mortality variations as a measure of general practitioner performance: implications of the Shipman case. *Br Med J* 2000; **320**: 489.
16. Goldstein H, Spiegelhalter DJ. Statistical aspects of institutional performance: league tables and their limitations (with discussion). *J Royal Statistical Soc A* 1996; **159**: 385–444.
17. Christiansen C, Morris C. Improving the statistical approach to health care provider profiling. *Ann Intern Med* 1997; **127**: 764–768.
18. Steiner SH, Cook RJ, Farewell VT. Monitoring paired binary surgical outcomes using cumulative sum charts. *Stat Med* 1999; **18**: 69–86.