Performance monitoring in Australia and England: from scandals to action

Mohammed A Mohammed and Andrew J Stevens

The Queensland approach may deliver empirical evidence on whether performance monitoring leads to improved quality of care

Several high profile medical disasters have occurred recently in Australia and England and have raised concern about quality of care and patient safety. In Australia, the major disasters have included those at the Bundaberg Hospital in Queensland,[1] and the Campbelltown and Camden Hospitals in New South Wales; and in England, the Bristol Royal Infirmary scandal[2] and the Shipman Affair.[3] Although the specifics in each instance are unique, a common outcome from the subsequent inquiries has been a recommendation for some form of centralised performance monitoring using routinely collected data. Why? Because retrospective desktop analysis of routinely collected data signalled the medical disaster before the "whistle blew".

What is interesting about Queensland Health's response to the Bundaberg scandal, as described by Duckett et al in this issue of the Journal (*page 571*),[4] is that these workers have combined their preferred casemix-adjusted statistical analysis (variable life-adjusted display [VLAD] plots) with an explicit protocol for action. In so doing, Duckett et al have recognised that the ultimate purpose of data (and analysis) is to guide action[5] (for improvement) while avoiding the commonly held casemix adjustment fallacy.[6]

The casemix adjustment fallacy begins with a deceptively simple equation that informally relates the variance in outcomes (eg, mortality) to a combination of three variables — chance, patient casemix factors, and quality of care. Statistical techniques are used to account for chance and patient casemix factors. The residual unexplained variance in outcome is then prematurely assigned to quality of care.[6] This is naïve, because it assumes too much and is an oversimplification. For example, it assumes that the outcome variance equation is of a closed form (three variables only, when we have no basis to conclude this); it assumes perfect data (no errors in data, no measurement error) and perfect casemix adjustment (when we know that important casemix variables are not as yet measured or measurable); and, finally, it fails to recognise that all statistical models are but simplified approximations subject to error. This casemix fallacy — where the residual unexplained

variance in outcome is automatically attributed to quality of care — is to be diligently avoided[7] if the process of monitoring is to remain credible and useful in improving quality of care.

The protocol for action developed by Duckett et al has two key characteristics: use of a pyramid model of investigation,[6,7] which recognises that, in searching for assignable causes of variation, a useful strategy (which mitigates the casemix adjustment fallacy) is to systematically check five variables (data, patient casemix, structure, process of care, and carers), underpinned by an explicit clinical governance framework. A key feature of the pyramid model of investigation is the prior degree of belief, for which there is empirical evidence,[8] that the vast majority of explanations for unusual variation will be located at the base (data, patient casemix) and the least likely explanations will be located at the apex (quality of care and carers). Indeed, as part of the Shipman Inquiry, two general practitioners were flagged as having "unacceptably" (the casemix adjustment fallacy) high casemix-adjusted death rates. Application of the pyramid model of investigation aided the discovery that this excess mortality was credibly explained by the numbers of patients in nursing homes — place of death was not included in the casemix adjustment method.

Furthermore, the inclusion of data (at the base of the pyramid) ensures that data improvement is an integral part of any monitoring process, as this is usually a highly cost-effective way of reducing variation, to the benefit of all concerned. So it is a matter of concern that Duckett et al found that data improvement was frequent and "frustrating to hospitals and clinicians". One reason for this frustration is because concepts such as "true positives" and "false positives" from diagnostic testing are applied to performance monitoring — but this is unhelpful and misleading.

Consider, for example, a signal of high mortality which is due to a data error. If we predicate the definition of signals of special cause variation on malpractice or quality of care,[9] then we are required to concede that a false-positive signal has occurred. But the notion of a false positive in the evaluation of a diagnostic test is based on a comparison with a reference "gold standard" test, by which the "true" disease state of each patient is known (albeit using the gold standard). In contrast, in performance monitoring, there is no gold standard and the true state of each data point is unknown, so the notion of false-positive signals makes little sense. Furthermore, in diagnostic testing, a false-positive result can mislead us into taking the wrong action, whereas finding and fixing a data error actually moves us in the right direction — towards understanding and eliminating special causes of variation.[10] The possible confusion with diagnostic testing should be remedied through clarifying the purpose of monitoring and providing training on the method and its limitations.

Duckett et al have adopted VLAD plots, which are simple visual tools for displaying variation. They show the indicator of interest one patient at a time, while accounting for the prior risk (casemix adjustment) and the play of chance as indicated by upper and lower limits. It is important to appreciate that, wherever the limits are set (except for 0 or infinity), we will occasionally commit an error of commission (when we wrongly identify a signal from the noise) or an error of omission (when we fail to detect a true signal from the noise). The choice of limits is, as Duckett et al state, a matter of judgement that is informed by statistical theory, economic consideration, and empirical evidence. What is exciting about the Queensland approach is that it is well poised to deliver the oft-lacking empirical evidence (because most monitoring systems do not include a specific action protocol) on the use of a performance monitoring system. This empirical evidence is the ultimate criterion by which to judge the usefulness of the performance monitoring system itself.

As Duckett et al recognise, performance monitoring is the search for signals of variation from background noise in the presence of imperfect data, imperfect casemix adjustment, and imperfect methods of action. Although the performance monitoring system must be alert to "bad apples" (and this is noted in the pyramid model for investigation), care needs to be taken to ensure that the search for "bad apples" does not become the primary aim, as this will (most likely) increase fear and thereby hinder genuine continual improvement.[9] Ultimately, performance monitoring needs to embrace the wider aspects of the science of improvement,[9] in which statistical methods for analysing variation are but one component. The other major components involve "appreciation of a system", "theory of knowledge" and "psychology of humans", in which all stakeholders in the performance monitoring system cooperate towards the aim of continual improvement. The challenge for performance monitoring in general[6,10] has been that it has itself been at variance with key principles of the science of improvement, and so the extent to which the Queensland model delivers continual improvement remains to be seen.

Author details

Mohammed A Mohammed, PhD, Senior Research Fellow
Andrew J Stevens, Professor
Department of Public Health and Epidemiology, University of Birmingham, Birmingham, UK.
Correspondence: m.a.mohammed@bham.ac.uk

References

1 Van Der Weyden MB. The Bundaberg Hospital scandal: the need for reform in Queensland and beyond [editorial]. *Med J Aust* 2005; 183: 284-285.
2 Smith R. One Bristol, but there could have been many [editorial]. *BMJ* 2001; 323: 179-180.
3 Dyer C. Shipman inquiry to investigate 466 deaths. *BMJ* 2001; 322: 1201.
4 Duckett SJ, Coory M, Sketcher-Baker K. Identifying variations in quality of care in Queensland hospitals. *Med J Aust* 2007; 187: 571-575.
5 Mohammed MA, Cheng K, Rouse A, Marshall T. Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001; 357: 463-467.
6 Lilford R, Mohammed MA, Spiegelhalter D, et al. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004; 363: 1147-1154.
7 Mohammed MA, Rathbone A, Myers P, et al. An investigation into high mortality general practitioners flagged up via the Shipman Inquiry. *BMJ* 2004; 328: 1474–1477.
8 Pitches D, Mohammed MA, Lilford R. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. *BMC Health Serv Res* [Internet] 2007; 7: 91. http://www.biomedcentral.com/1472-6963/7/91 (accessed Oct 2007).
9 Aylin P, Best N, Bottle A, Marshall C. Following Shipman: a pilot system for monitoring mortality rates in primary care. *Lancet* 2003; 362: 485-491.
10 Deming WE. The new economics. 2nd ed. Cambridge, Mass: MIT, 1994. ❑