Identifying variations in quality of care in Queensland hospitals

Stephen J Duckett, Michael Coory and Kirstine Sketcher-Baker

ABSTRACT

- Identifying and acting on variations from good practice is one of the critical tasks of clinical governance. We describe one aspect of Queensland's post-Bundaberg clinical governance arrangements: the use of variable life-adjusted displays (VLADs) to monitor outcomes of care in the 87 largest public and private hospitals in Queensland, which together account for 83% of all hospital activity.
- VLAD control charts were created for 31 clinical indicators using routinely collected data, and are disseminated monthly.
- About a third of hospitals had a run of cases in the 3-year period that flagged at the 30% level (local level investigation). For three indicators, about one in five hospitals had sufficiently cumulatively more deaths than statistically expected that the hospital was highlighted for state-wide review.
- VLADs do not provide definitive answers about the quality of care. They are used to develop ideas about why variations in reported outcomes occur and suggest possible solutions, be they ways of improving data quality, improving casemix adjustment, or implementing system changes to improve quality of care.
- Critical to the approach is that there is not just monitoring — the monitoring is tied in with systems that ensure that investigation, learning and action occur as a result of a flag.

MJA 2007; 187: 571–575

For editorial comment, see page 549

The high profile "Doctor Death" Bundaberg Hospital scandal led to public inquiries and a major shake-up of the leadership of Queensland Health at ministerial and departmental levels. The public inquiries focused attention on the management culture of Queensland Health and the need for the department to improve its transparency and openness.[1-3] In response, Queensland Health has transformed its clinical governance arrangements.[4] This article describes one aspect of the new arrangements — the use of statistical process control charts using routine data to provide a starting point for learning and subsequent action to improve the quality of care.

The method

All Queensland hospitals (public and private) regularly provide routine data to Queensland Health. These data include information on demographic characteristics of the patients, the principal diagnosis, other conditions treated, and procedures performed. Coding standards require the coded data to be provided within 35 days from the end of the month. In consultation with clinicians, 31 clinical indicators have been selected for regular monitoring of outcomes of care using statistical process control (Box 1).

Control charts are currently provided to the 87 largest public and private hospitals in Queensland, accounting for 83% of all hospital activity.[5] Public hospitals are required by administrative instruction to analyse the charts and report within the Queensland Health processes on outcomes of reviews; private hospitals are required to report to the Private Health Unit (the regulatory oversight unit within Queensland Health) on their reviews.

Rationale

In Australia and elsewhere, hospital-specific comparisons based on routine data have relied on cross-sectional analysis.[6,7] This involves aggregating data for all patients over a set period, say 12 months, and determining whether the number of adverse outcomes (eg, in-hospital deaths after admission for stroke) is higher than expected based on the average for all hospitals. By definition, these cross-sectional analyses can only occur at the end of some set period, and provide average results for all patients admitted to the hospital during that time. In contrast, statistical process control is a continuous approach, and displays data on outcomes of care of individual patients. The method can identify changes in outcomes relatively quickly and is more sensitive to such changes than less regular, cross-sectional approaches, which can obscure important patterns in the data.[8] Statistical process control also highlights the dynamic nature of health care: that patterns of care can change over time and a negative signal at some point in the past can be rectified.

Method details

Statistical process control was developed several decades ago to improve the quality of manufactured products. Its application in health is complicated by the need to adjust for risk to ensure that hospitals or doctors who see sicker patients are not unfairly penalised. Several methods have been proposed that incorporate risk adjustment;[9] the method adopted by Queensland Health follows Sherlaw-Johnson's approach,[10] and is known as variable life-adjusted display (VLAD) (Box 2). Display charts are provided to Queensland hospitals each month, with the first distribution of charts providing trend data from July 2003 to late 2006.

The first step in plotting the VLAD is to calculate, for each patient in the instant month, the expected risk (probability) of a particular outcome (death, complication, readmission, and long-stay are used in Queensland), adjusting for age, sex, and selected comorbidities specific to each indicator. This can be thought of as the average risk of an adverse outcome across all hospitals for patients with the same age, sex and comorbidity profile as the patient in question. It is estimated using a logistic regression model for the index month plus the previous 11 months of data. These data include patients as defined by the indicator admitted to Queensland hospitals (public and private) with an average of at least 20 separations a year for the relevant indicator.

Next, the expected risk is subtracted from the observed outcomes (coded as 0 or 1 for presence or absence of the outcome) and plotted sequentially. An upward movement of the chart indicates that, for the patients in question, the number of outcomes (eg, deaths) was less than that expected, while a downward movement indicates that the number of outcomes was greater than that expected.

1 Indicators used in process control charts in Queensland Health*

Medical
- Acute myocardial infarction: in-hospital mortality, readmission, long stays
- Heart failure: in-hospital mortality, readmission, long stays
- Stroke: in-hospital mortality
- Pneumonia: in-hospital mortality

Surgical, procedural
- Fractured neck of femur: in-hospital mortality, complication of surgery
- Laparoscopic cholecystectomy: complication of surgery
- Colorectal cancer: complication of surgery
- Hip replacement: complication of surgery, readmission, long stays
- Knee replacement: complication of surgery, readmission, long stays
- Prostatectomy: complication of surgery
- Abdominal hysterectomy: complication of surgery
- Vaginal hysterectomy: complication of surgery
- Paediatric tonsillectomy: readmission, long stays

Psychiatric
- Depression: readmission, long stays
- Schizophrenia: readmission, long stays

Maternity
- Selected primiparae induction of labour
- Selected primiparae caesarean section (public hospitals)
- Selected primiparae caesarean section (private hospitals)
- First births: perineal tears (3rd or 4th degree)

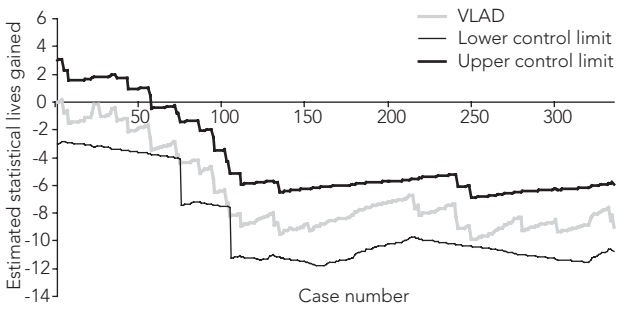
* Detailed definitions of the indicators are available at http://www.health.qld.gov.au/performance/docs/Tech_Sup.pdf (pages 24–44). ◆

Finally, thresholds are calculated where the chart is said to flag. Critical in quality improvement approaches is not just providing data, but ensuring that aberrant patterns identified by monitoring are investigated, and that practice changes occur.[9] Queensland Health has developed hierarchical flagging criteria that signal closer scrutiny, depending on the extent of variation from the state average and whether the indicator incorporates a fatal or non-fatal outcome (Box 3).

For example, if the trend line shows that the cumulative experience of outcomes of care is more than 30% worse than the state average (for an indicator with a fatal outcome), the indicator is flagged for internal hospital review. The Queensland Health VLAD policy requires identification of "clinician leads" to facilitate clinician involvement in the review process.[11] In addition to reporting through the various organisational structures of Queensland Health, public hospitals are required to report remedial action to the local consumer consultative group.

The statistical process control methods used in industry were developed to help identify special (also called assignable) cause variation,[12] which is defined as variation that warrants further investigation. Standard methods of frequentist inference (*P* values and confidence intervals) are not suitable for identifying such variation. Instead, likelihood methods, which are not affected by the problem of multiple looks at the data,[13] are used, and within this framework the characteristics of the VLAD are usually described in terms of the average run length to true or false alarm.

2 Example of a variable life-adjusted display (VLAD)



3 Flagging criteria and related action

Notification level	Fatal outcome indicator*	Non-fatal outcome indicator*	Action required
1	30%	50%	Hospital should investigate internally and report outcome to Area Clinical Governance Unit or Private Health Unit (for private facilities)
2	50%	75%	Area Clinical Governance Unit or Private Health Unit should be involved in investigation
3	75%	100%	Report to Patient Safety and Quality Board through the Area General Manager or Chief Health Officer required

* Per cent relative risk increase or reduction in outcome compared with the average for all hospitals combined. ◆

We used standard methods based on simulations[14] to identify average run lengths to true and false alarm.

The flagging criteria were set to balance the costs of investigating false alarms (where the change in outcomes is simply a statistical artefact) against the need to identify special or assignable cause variation, which might benefit from further investigation.[15] As is the case in industry,[16] this was a policy decision; the reasoning is similar to that used to decide on a balance between sensitivity and specificity for a screening test.[17]

The control limits are reset each time a trigger point is reached. For example, when monitoring using the Tier 1, non-fatal flag, if a case is flagged as hitting 50% deviation from the average, the control limits are reset and the hospital could be flagged a second time if there is a cumulative run of cases which is again 50% deviation from average (starting at the first trigger point).

Results to date

The 31 indicators currently monitored involve 17 conditions or procedures, accounting for about 6% of total discharges from Queensland public and private hospitals. Box 4 shows the indicators, and information about the dataset and the incidence of flagging of negative outcomes.

4 Incidence of adverse trend flagging in indicators used by Queensland Health by indicator and flagging criterion,* 1 July 2003 to 30 June 2006 (non-perinatal indicators), 1 January 2003 to 31 December 2005 (perinatal indicators)

	No. of admissions	Flags per 10 000 admissions, by flag level			No. of hospitals	No. of hospitals that flagged in 3 years, by flag level			% of hospitals that flagged at least once in 3 years, by flag level		
		30%	50%	75%		30%	50%	75%	30%	50%	75%
In-hospital mortality											
Acute myocardial infarction	7491	25	8	7	28	12	6	5	43%	21%	18%
Heart failure	14 975	31	23	9	57	30	23	11	53%	40%	19%
Stroke	7812	20	8	1	32	9	5	1	28%	16%	3%
Pneumonia	19 348	26	17	9	71	24	19	10	34%	27%	14%
Fractured neck of femur	5347	34	17	11	25	9	6	5	36%	24%	20%
Complication of surgery											
Fractured neck of femur	5347	17	6	4	25	6	3	2	24%	12%	8%
Laparoscopic cholecystectomy	18 526	16	10	6	53	20	12	8	38%	23%	15%
Colorectal cancer	4798	8	4	2	33	2	1	1	6%	3%	3%
Hip replacement	8490	29	16	9	43	10	5	3	23%	12%	7%
Knee replacement	13 653	19	15	9	44	11	10	8	25%	23%	18%
Prostatectomy	9854	14	6	5	36	7	4	4	19%	11%	11%
Abdominal hysterectomy	7701	13	4	3	41	7	2	1	17%	5%	2%
Vaginal hysterectomy	7551	13	11	5	37	6	6	3	16%	16%	8%
Readmission											
Acute myocardial infarction	5357	35	15	7	19	10	5	3	53%	26%	16%
Heart failure	8893	11	4	3	34	6	3	2	18%	9%	6%
Hip replacement	3073	26	13	7	16	5	3	2	31%	19%	13%
Knee replacement	4427	23	9	5	16	4	3	2	25%	19%	13%
Paediatric tonsillectomy	7868	20	8	4	10	4	3	2	40%	30%	20%
Depression	8974	40	25	17	16	5	2	2	31%	13%	13%
Schizophrenia	12 344	24	14	8	16	3	2	2	19%	13%	13%
Long stays											
Acute myocardial infarction	5410	7	4	2	19	3	2	1	16%	11%	5%
Heart failure	9005	13	7	2	34	8	5	2	24%	15%	6%
Hip replacement	3076	33	20	10	16	5	5	3	31%	31%	19%
Knee replacement	4430	25	14	9	16	5	4	3	31%	25%	19%
Paediatric tonsillectomy	7868	23	15	10	10	3	2	2	30%	20%	20%
Depression	8974	35	22	16	16	7	5	4	44%	31%	25%
Schizophrenia	12 344	34	19	14	16	6	3	3	38%	19%	19%
Maternity											
Selected primiparae induction of labour	40 821	5	1	1	53	10	4	3	19%	8%	6%
Selected primiparae caesarean section (public hospitals)	26 288	7	3	1	34	8	5	2	24%	15%	6%
Selected primiparae caesarean section (private hospitals)	14 543	6	0	0	19	5	0	0	26%	0%	0%
First births: perineal tears (3rd or 4th degree)	39 999	20	13	8	53	25	19	11	47%	36%	21%

* Per cent relative risk increase compared with average for all hospitals combined. ◆

Of the 31 indicators, five measure incidence of in-hospital mortality, eight measure complications of surgery, seven measure readmissions, and seven measure excess length of stay. A further four measure outcomes of maternity care. For each of the mortality indicators, between one-quarter and one-half of all hospitals had a run of cases in the 3-year period that flagged at the 30% (local investigation) level. Complications of care were more variable, ranging from 6% to 38% of hospitals being flagged for local investigation. For all five of the mortality indicators, at least one hospital flagged at the 75% (central investigation) level. For three of these mortality indicators, about one in five hospitals had sufficiently more deaths than statistically expected that the hospital flagged at the 75% level for central review.

5 Issues for investigation under the pyramid model

Element	Scope	Typical investigation questions
Data	Data quality issues (eg, coding accuracy, reliability of charts, definitions, and completeness)	• Are the data coded correctly? • Is the proportion of additional diagnoses coded appropriate? • Has there been a change in data coding practices (eg, are there less experienced coders)? • Is clinical documentation clear, complete and consistent?
Casemix	Differences in casemix are accounted for in the calculation of the VLAD, as much as possible given the available data. However, it is possible that some residual confounding might remain for some indicators	• Are there factors peculiar to this hospital not taken into account in the risk adjustment? • Has the pattern of referrals to this hospital changed (in a way not taken into account in risk adjustment)?
Structure or resource	Availability of beds, staff, and medical equipment; institutional processes	• Has there been a change in the distribution of patients in the hospital, with more patients in this specialty spread throughout the hospital rather than concentrated in a particular unit?
Process of care	Medical treatments of patients, clinical pathways, patient admission and discharge hospital policies	• Has there been a change in the care path being followed? • Have new treatment guidelines been introduced?
Professional	Practice and treatment methods, etc	• Has there been a change in staffing for treatment of patients? • Has a key staff member gained additional training and introduced a new method that has led to improved outcomes?

VLAD = variable life-adjusted display. ◆

The data on incidence of flagging cannot be interpreted as incidence of preventable adverse events or other measures of quality of care, nor, without analysis of the underlying causes of flagging, is it possible to make an assessment of the indicators (eg, whether mortality outcomes in acute myocardial infarction, heart failure or fractured neck of femur are inherently more variable so the higher level of flagging of these indicators represents random rather than assignable cause variation). This analysis will eventually be possible when sufficient analyses of causes of flags have been undertaken.

Analysis of outcomes

As the VLAD approach is continuous, with monthly dissemination of data, there are likely to be more frequent investigations by hospitals and clinicians than with an annual, cross-sectional approach. A key characteristic of continuous improvement is "closing the loop", ensuring appropriate investigation and actions run on.

There is an ambiguous relationship between outcome and process measures of quality of care,[18] so a pyramid model, which recognises multiple explanations for variation in recorded outcomes, is recommended as a focus for the investigation process.[19] The first investigation should be whether the data have been coded accurately. A second screen is whether there is casemix variation that has not been fully accounted for in the risk adjustment process (eg, Indigenous status is not incorporated in the risk adjustment model, but is often associated with worse clinical outcomes). Box 5 shows the stages in the pyramid model, and typical questions that should be asked as part of an investigation.

The flags are a way of standardising the process for deciding when the data are worth a closer look. A virtue of the VLAD approach is that it encourages visual inspection of data and, in many cases, a more detailed look at the data could be instigated without using flags; for example, if a downward slope appeared abruptly. The VLAD can take many possible forms, depending on the length and clustering of runs of good or poor performance. However, in terms of actions that should be taken, VLADs can be grouped into four basic patterns (Box 6).

As recommended in the pyramid model of investigation, the first round of investigations highlighted many data coding issues, and this was frustrating to hospitals and clinicians. An outcome of this review is likely to be improved data quality that will enhance the credibility of the clinical indicators.

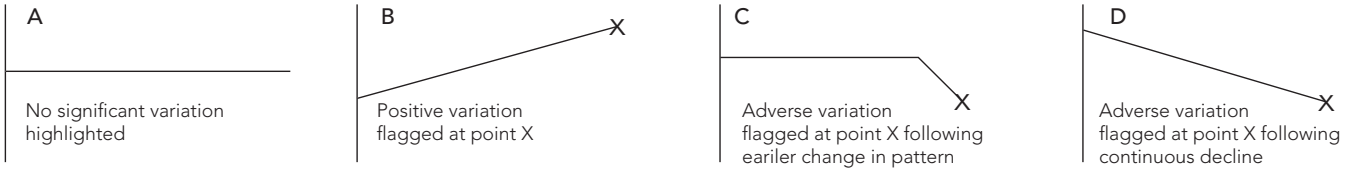
Routine data are limited and cannot provide risk adjustment for the full range of factors known to affect outcomes,[20] and this is recognised in the second stage of the pyramid model of investigation. That is, more detailed investigation at the local level might reveal that a run of poor outcomes at a particular hospital might be due to a run of sicker patients. This should not undermine the utility of statistical process control approaches: the aim is to identify causes of variation in outcomes, be they variations in data quality, casemix, or quality of care.

The standard format for reporting on the outcome of flags is still evolving. The current requirement is that the report highlights what reviews were conducted (coding audits, clinical review) and their findings, together with what management action has been undertaken. A more structural reporting format is being developed.

Discussion

As demonstrated in Box 6, statistical process control charts facilitate visual inspection of patterns of care, facilitating the task of identifying whether there has been a pattern change or a continuation of an underlying trend. However, control charts cannot provide definitive answers about the quality of care. They more closely resemble techniques from the area of statistics known as exploratory data analysis, and should be used to develop theories about why variations occur and suggest possible solutions:

6 Typical patterns of outcome variation



*The different adverse patterns (**C** and **D**) lead to different foci for investigations, with investigations following pattern **C** focusing on what changes have occurred around the time the shape of the curve changed, and investigations following pattern **D** examining systematic underlying trends. A continuation of the downward slope in either **C** or **D** may signal the need for investigations involving higher levels of the organisation.* ◆

improving data quality, improving casemix adjustment, or implementing system changes to improve quality of care.

For the same reasons, the thresholds where the chart is said to flag should not be likened to *P* values that measure the consistency of the data with the null hypothesis, leading to a decision rule about whether to accept or reject that hypothesis. Instead, they are guides, and could be calibrated so that the charts have fewer or more flags. The post-Bundaberg environment in Queensland influenced an explicit decision to have more rather than fewer flags, because we wanted to be sure of identifying true flags and were tolerant of the costs of investigating false flags.

The current approach monitors 31 indicators and is focused on the largest public and private hospitals. Over time, it is proposed to expand the indicator set to include indicators sensitive to ward care (pressure ulcers, falls) and indicators that can be used to measure outcomes in smaller hospitals (such as the incidence of possibly preventable complications).[21] This will ensure a more comprehensive monitoring of clinical outcomes across Queensland.

The new approach adopted in Queensland Health for monitoring clinical outcomes represents a significant increase in centralised monitoring and is unique in the world. What is important about this approach is not just that there is monitoring, but that the monitoring is closely linked with investigation, learning and action.

Competing interests

None identified.

Author details

Stephen J Duckett, PhD, DSc, FASSA, Executive Director[1,2]
Michael Coory, MB BS, PhD, AStat, Medical Epidemiologist[1,2]
Kirstine Sketcher-Baker, BAppSc, Principal Data Analyst[1]
1 Reform and Development Division, Queensland Health, Brisbane, QLD.
2 School of Population Health, University of Queensland, Brisbane, QLD.
Correspondence: stephen_duckett@health.qld.gov.au

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 Birrell B, Schwartz A. The aftermath of Dr Death: has anything changed? *People Place* 2005; 13: 54-61.
3 Dunbar JA, Reddy P, Beresford B, et al. In the wake of hospital inquiries: impact on staff and safety. *Med J Aust* 2007; 186: 80-83.
4 Duckett S. A new approach to clinical governance in Queensland. *Aust Health Rev* 2007; 31 Suppl 1: S16-S19.
5 Queensland Health. An investment in health. Queensland public hospitals performance report 2005–06. Brisbane: Queensland Health, 2006. http://www.health.qld.gov.au/performance/docs/Perform_rpt_05-06.pdf (accessed Oct 2007).
6 Shearer A, Cronin C, Feeney D. The state of the art of online hospital public reporting: a review of forty-seven websites. Easton, Md: Delmarva Foundation, 2004. http://www.delmarvafoundation.org/newsAndPublications/reports/documents/WebSummariesFinal9.2.04.pdf (accessed Oct 2007).
7 Robinowitz D, Dudley R. Public reporting of provider performance: can its impact be made greater? *Annu Rev Public Health* 2006; 27: 517-536.
8 Woodall W. Controversies and contradictions in statistical process control. *J Qual Technol* 2000; 32: 341-350.
9 Grigg O, Farewell V. Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. *Stat Methods Med Res* 2003; 12: 147-170.
10 Sherlaw-Johnson C. A method for detecting runs of good and bad clinical outcomes on variable life-adjusted display (VLAD) charts. *Health Care Manag Sci* 2005; 8: 61-65.
11 Queensland Health. Clinical governance implementation standard. 4. Variable life adjusted display — dissemination and reporting. Brisbane: QH, 2007. (QHEPS Document Identifier: 32547.) http://www.health.qld.gov.au/quality/publication/32547.pdf (accessed Oct 2007).
12 Deming W. The new economics. Cambridge, Mass: MIT Press, 1993.
13 Blume JD. Likelihood methods for measuring statistical evidence. *Stat Med* 2002; 21: 2563-2599.
14 Steiner S, Cook R, Farewell V, et al. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000; 1: 441-452.
15 Lim T. Statistical process control tools for monitoring clinical performance. *Int J Qual Health Care* 2003; 15: 3-4.
16 Nelson LS. Notes on the Shewart control chart. *J Qual Technol* 1999; 31: 124-126.
17 Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. *BMJ* 1994; 308: 1552.
18 Pitches D, Mohammed M, 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* 2007; 7: 91.
19 Lilford R, Mohammed M, 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.
20 Scott IA, Ward M. Public reporting of hospital outcomes based on administrative data: risks and opportunities. *Med J Aust* 2006; 184: 571-575.
21 Hughes J, Averill R, Goldfield N, et al. Identifying potentially preventable complications using a present on admission indicator. *Health Care Financ Rev* 2006; 27: 63-82.

(Received 20 Apr 2007, accepted 6 Aug 2007) ❑