Journal of Hospital Infection 76 (2010) 287–291

Available online at www.sciencedirect.com

Journal of Hospital Infection

journal homepage: www.elsevierhealth.com/journals/jhin

Review

Analysis of aggregated hospital infection data for accountability

A. Morton [a,b,*], K. Mengersen [a], M. Waterhouse [a,c], S. Steiner [d]

[a] Infection Management Services, Princess Alexandra Hospital, Brisbane, Queensland, Australia
[b] School of Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia
[c] St Andrew's Medical Institute Brisbane, Queensland, Australia
[d] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

ARTICLE INFO

Article history:
Received 24 November 2009
Accepted 10 June 2010
Available online 25 September 2010

Keywords:
Audit
Hospital infection
Outcome data
Statistical analysis
Evidence-based systems
Aggregated outcome data
Accountability

SUMMARY

Analysis and reporting of among-institution aggregated hospital-acquired infection data are necessary for transparency and accountability. Different analytical methods are required for ensuring transparency and accountability for within-institution sequential analysis. In addition, unbiased summary information is needed for planning and informing the public. We believe that implementation of systems based on evidence is the key to improving institutional performance and safety. This must be accompanied by compliance, outcome audit and sequential analysis of outcome data, e.g. using statistical process control methods. Checklists can be a valuable aid for ensuring implementation of evidence-based systems. Aggregated outcome data analysis for transparency and accountability should concentrate primarily on accurately presenting the outcomes together with their precision. We describe tabulations, funnel plots and random-effects (shrinkage) analysis and avoid comparisons using league tables, star ratings and confidence intervals.

© 2010 The Hospital Infection Society. Published by Elsevier Ltd. All rights reserved.

Introduction

Safe practice is based on the premise that hospitals achieve optimum outcomes by implementing systems based on evidence. Many evidence-based systems to combat hospital-acquired infections (HAIs) are referred to as 'bundles'.[1] Checklists are a valuable adjunct to their implementation and maintenance but must be complemented by regular sequential monitoring and analysis of process and outcome data.[2,3]

It is becoming increasingly necessary for hospitals to supply HAI data to central authorities for accountability, after which these data frequently undergo aggregation. Other purposes of aggregation are to provide unbiased summary information to the public and for planning. Aggregation at each of these levels is valid but motivated by different objectives.

Data can be aggregated and analysed at different levels. *Staphylococcus aureus* bacteraemias, surgical site infections (SSIs), infections with antibiotic-resistant organisms and antibiotic usage are of particular interest, with many subject to centrally set targets that are seen by central authorities as being necessary for managing projects aimed at reducing adverse events (AEs).

Nationally aggregated data tend to be published infrequently. Hence, regardless of their quality, they are of limited value for timely local improvement. This requires within-institution sequential analysis to detect unforeseen changes in a timely manner.

This paper focuses on the aggregation of outcome data and their analysis to ensure transparency and accountability. Most of the examples we present involve multiple hospitals, but the methods can easily be adapted to local level investigations, e.g. for surgical units within a hospital that perform similar procedures such as cardiac surgery.

Methods

We describe three methods: tabulations, funnel plots and random-effects (shrinkage) analysis. These methods are illustrated with risk-adjusted SSI data from the Queensland Health Centre for Health Related Infection, Surveillance and Prevention (CHRISP).[4] The National Nosocomial Infections Surveillance system (NNIS) risk index was employed for risk adjustment (A. Morton, K. Mengersen, M. Waterhouse *et al.*, submitted data).[5] It has been shown that risk adjustment permits comparison of an institution with the

* Corresponding author. Address: 40 Garioch St, Tarragindi 4121, Australia. Tel.: +61 7 33974651; fax: +61 7 38473480.
E-mail address: amor5444@bigpond.net.au (A. Morton).

0195-6701/$ – see front matter © 2010 The Hospital Infection Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.jhin.2010.06.030

among-institution average, not comparisons between institutions that may have differing risk profiles.[6]

The methods can be applied, with minor adjustment, to count data AEs such as bacteraemias. However, risk adjustment is more difficult. It involves stratifying hospitals by their complexity and the services they provide, e.g. hospitals with large haematology/oncology services can expect more frequent bacteraemias than hospitals with large maternity services.[7]

Simple tabulations

It can be useful to have a table showing the institutions, their observed (O) and expected (E) outcomes, standardised ratios (SMRs) and approximate standardised residuals, e.g. $Z = (O - E)/\sqrt{E}$, where $Z \geq 2$ suggests a possible difference from expected and $Z \geq 3$ a definite difference. A second table dividing the time period, for example by quarters, provides a summary of the work done and may indicate clusters of AEs. Contiguous time periods are important as clusters can span divisions. For institutions of special interest, a third table cross-classifying by institutions and quarters can also be useful although care is needed to avoid cluttering.

Tabulations can be performed using a spreadsheet or a statistical analysis program such as R version 2.5.1.[8] Table I shows two years (2005–2006) of elective risk-adjusted orthopaedic SSI data for the ten hospitals with the largest orthopaedic departments from the CHRISP database.[4] Table II shows the data by quarters. The data for 2006 were also tabulated by selected hospitals and quarters (table not shown).

Funnel plots

Funnel plots are similar to Shewhart charts except that the number of procedures is on the horizontal axis. Variation in binary data funnel plots is derived from the overall mean and the square root of each institution's sample size.[9–11]

When there is no risk adjustment the funnel plot calculations are similar to those for a standard Shewhart chart.[10] With risk adjustment, the data may be plotted as risk-adjusted rates (RARs) (Appendix 1).[11] Funnel plots can be performed in a spreadsheet, making them useful for preliminary data screening. The data displayed in Table I are illustrated in Figure 1. Values outside the 2 SD equivalent limits are often referred to as possible outliers and those outside the 3 SD limits as definite outliers.

Cook *et al.* have described a funnel plot analysis using SMRs.[12]

Shrinkage-adjusted estimates and shrinkage plots[13]

Data are subject to variation that can be largely random and unavoidable, and the precision of the estimated SMRs and RARs must be addressed, e.g. by adjusting, or 'shrinking' the hospital estimates towards a mean value, with the degree of shrinkage depending on the precision of the estimates. Thus estimates that are imprecise, typically due to small numbers, are more strongly adjusted to minimise the risk of them appearing, possibly incorrectly, as outliers. Estimates that are based on large sample sizes remain relatively untouched (Appendix 2). Figure 2 is the shrinkage plot for the Table I data.

Table I
Observed and expected SSIs by hospital

Hospital	SSIs	Expected	SMR	Z	Procedures
A	20	10.43	1.92	2.96	235
B	13	15.97	0.81	−0.74	367
C	6	10.43	0.58	−1.37	267
D	8	12.16	0.66	−1.19	276
G	33	39.19	0.84	−0.99	989
I	21	11.92	1.76	2.63	248
J	15	12.40	1.21	0.74	303
K	12	10.50	1.14	0.46	241
Q	23	38.34	0.60	−2.48	923
R	5	8.66	0.58	−1.24	204
Total	156	169.98	0.92	−1.07	4053

SSIs, surgical site infections; SMR, standardised mortality ratio.

Table II
Observed and expected SSIs by quarters

Year	Quarter	SSIs	Expected	SMR	Z	Procedures
2005	1	16	17.09	0.94	−0.26	415
	2	23	21.33	1.08	0.36	512
	3	23	22.21	1.04	0.17	527
	4	24	21.41	1.12	0.56	513
2006	1	16	17.32	0.92	−0.32	416
	2	13	22.93	0.57	−2.07	539
	3	24	25.58	0.94	−0.31	599
	4	17	22.11	0.77	−1.09	532
Total		156	169.98	0.92	−1.07	4053

SSIs, surgical site infections; SMR, standardised mortality ratio.

Results

Table I shows that hospitals A and I may have had more SSIs than predicted whereas hospital Q had less. Table II indicates that there were fewer SSIs than predicted in the second quarter of 2006. In 2006 there was no hospital by quarter clustering (table not shown).

Figure 1 confirms that hospitals A and I were possible high outliers with hospital Q a borderline low outlier. Hospitals G and Q performed the largest number of procedures.

Figure 2 shows that when hospital A is considered a member of this group of hospitals, it is a possible outlier and hospital I is borderline. The shrunken value for each is nearer the average for the institutions than its value without shrinkage.

Discussion

Safe practice is crucially dependent on transparency, and public reporting appears to impose a necessary discipline on hospitals.[14] Moreover, when information is withheld, there is always suspicion that something undesirable may be occurring. To achieve transparency, public reporting should also include information about implementation of the evidence-based systems necessary to ensure safe practice.[15]

Reasons given for central reporting and analysis of outcome data include institutional comparisons, to make better decisions and improve performance, for benchmarking and meeting targets and to enable potential patients to make more informed choices. However, evidence that public reporting improves performance is controversial.[16] Comparisons tend of necessity to be before and after quasi-experiments confounded by learning curves. Moreover, people may simply want their local hospital to treat them correctly, expeditiously, safely and courteously.

Analysis of large central data sets could be important for obtaining better evidence to further improve systems, and data could be collected with this objective in mind. Bayesian Networks may be valuable to better understand problems such as the transmission of in-hospital multiple antibiotic-resistant organisms.[17]

Tabulations act as an initial data display. One should look particularly at contiguous time periods as clusters can occur across divisions. Funnel plots are probably best reserved for preliminary analysis. Smaller institutions can be labelled as outliers on dubious evidence due to sample size and unavoidable variation effects

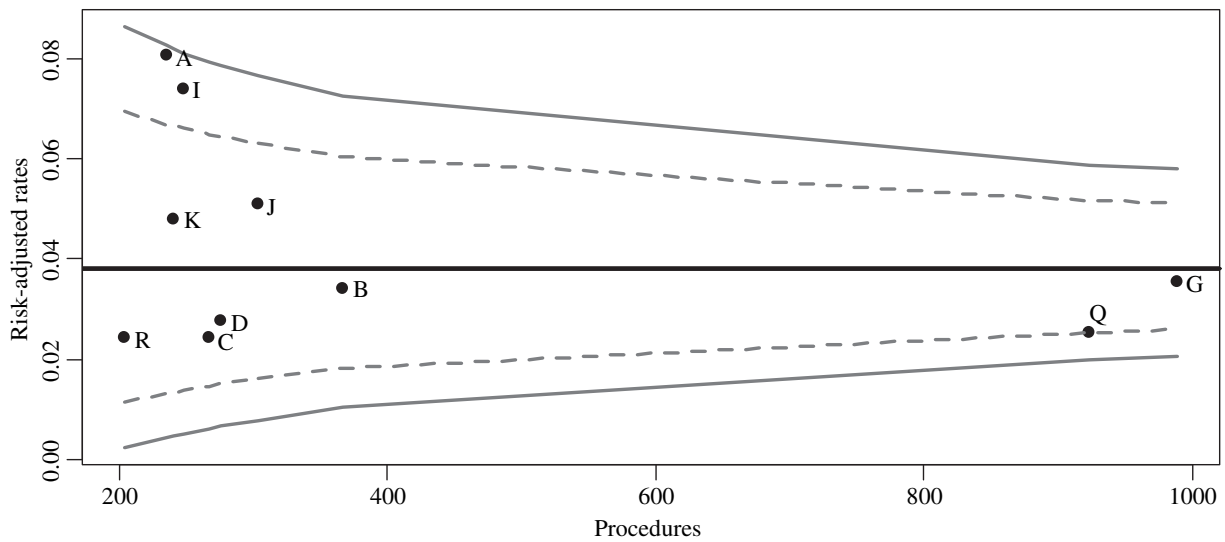

Figure 1. Orthopaedic surgical site infections funnel plot; midline is weighted average of risk-adjusted rates. The control limits are at the percentiles equivalent to 2 SD (dashed curve) and 3 SD (solid curve) from the centre line.

related to regression to the mean. Shrinkage plots are employed if funnel plots show outliers. Without this analysis, hospitals such as hospital I with relatively small numbers of procedures could be considered, possibly incorrectly, as outliers. In a judgemental environment, this might result in a disadvantage when any difference may have been due to unavoidable variation.

When there is an apparent outlier, there should be evidence of a system analysis with documentation, reporting, corrective action and follow-up. An approach such as that described by Mohammed *et al.*[18] should be followed: the data are first checked for error; the statistical methods are assessed; systems are evaluated; and finally the role of providers is considered.

It is logical for central authorities to insist that safe systems are in place, that anomalous results are investigated and that any problems identified are dealt with. However, it should be recognised that most institutions take great pride in the quality of their work and this ownership is a powerful force for sustaining quality and safety.

The limitations of risk adjustment need to be considered, eg procedure numbers in our data were dominated by two institutions. This could affect risk adjustment and SMR variation, e.g. the procedure times used to calculate the NNIS risk index would have been dominated by these two institutions.[19] Furthermore, some of the more complex surgery, for example replacement of previously inserted knee and hip prostheses, may have been referred to these centres.

Despite the need for risk adjustment, the results can be influenced by the accuracy of the expected values, and possible variation in expected values based on small datasets is typically ignored.[10,19] Imperfect risk adjustment can lead to damaging comparison of institutions, e.g. a hospital may appear to be performing poorly because it tackles difficult, inadequately risk-adjusted cases.

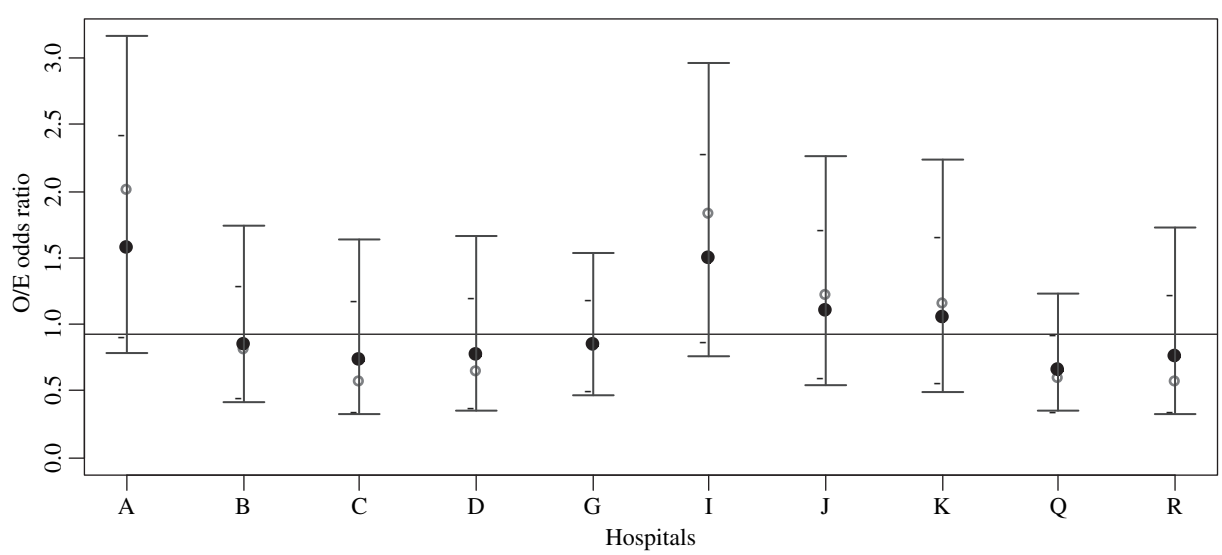

Figure 2. Orthopaedic surgical site infections shrinkage plot using lmer. The shrinkage plot has a centre line for an observed/expected (O/E) odds ratio of one. The observed odds ratio is a hollow circle and the shrinkage value a black circle. The percentile equivalent of 3 SD on either side of the shrinkage value is displayed as wide horizontal markers and the 2 SD equivalent values as short horizontal markers. These are not symmetrical because the basic measurement is in logarithms requiring exponentiation. The upper and lower 3 SD equivalent markers are joined by a vertical line. If the lower 2 SD equivalent line is above the centre line there is said to be a possible 'outlier' and if the lower 3 SD equivalent line is above it, then a definite 'outlier' is thought to be present.

Centrally performed analysis can have a number of disadvantages. There is necessarily delay with reporting of aggregated data; the possibility of data error may be increased by additional data handling; staff may be tempted to show their work in a good light by failing to report minor AEs or by gaming; central analysis may be less easy to understand; and local ownership may be lost.[20] The worst-appearing hospital may be the one that tries hardest to find every complication.

Rates of serious AEs are usually low and differences when care is or is not optimal can be small relative to unavoidable variation. Except in large institutions, it may be impossible for statistical analysis to detect real differences in a practicable time frame.[21]

Implementation of an evidence-based system, e.g. by employing 'bundles' and checklists, is the key to reducing an AE rate to a statistically predictable minimum.[2,22] If a systems defect exists, its detection, perhaps by employing systems analysis, followed by correction would similarly reduce the AE rate.[15] Local audit and monitoring, e.g. using morbidity and mortality (M&M) meetings, timely independent audit, and statistical process control (SPC), would detect unforeseen changes in AE rates.[10,23,24]

Frequently, aggregated data show excessive variability. We believe that this should be displayed and its source understood. Together with suitable risk adjustment, shrinkage plots perform the required adjustment by considering each institution as a sample from a homogeneous group of institutions having a common distribution.[10,13] This random-effects approach makes use of the data from the whole group of institutions to improve the precision of the predictions for the individual institutions. However, marked outliers may increase between-institution variation and interfere with the analysis of the remaining data, i.e. their data would not belong to the distribution of remaining institutions.[10] They should be removed for systems analysis and the statistical analysis repeated. In addition, the random-effects approach may occasionally fail to detect a small institution that is a genuine outlier, so small institution definite outliers identified by funnel plots but not by random-effects analysis should be treated with suspicion.[25] It is also important to inspect the SMRs, e.g. a small institution may have one or two bacteraemias when only 1/20 or so is expected, giving an improbable SMR of 20 or 40. These should be removed from the analysis and subjected to individual audit.

Although confidence intervals, star ratings and league tables have commonly been used to compare hospital estimates, it is now recognised that using these approaches can be unwise.[10,26] It is important that accurate data are employed. Periodic assessment of data validity and rater agreement should therefore occur.[27]

Appendix 1

Suppose that institution 'i' performs N_i procedures, of which O_i result in AEs. Then the SMR for the *i*th institution is $SMR_i = O_i/E_i$, where E_i is the expected number of AEs. This ratio is then multiplied by a reference rate (Ebar), to produce the risk-adjusted rate for institution i, $RAR_i = Ebar \times SMR_i$. The chart displays RARs for each institution, a weighted average centre line ($CL = \sum N_i RAR_i / \sum N_i$) and control limits.

Control limit calculation is controversial since the expected number may be subject to sampling error.[19] For binary data, Hart *et al.*[11] employ the normal approximation, e.g. for the upper 2 SD control limit:

$$U2 = CL + 2 \times \sqrt{(CL \times (1 - CL)/Ni)}$$

In cases where the plot displays excess heterogeneity (overdispersion), e.g. due to inadequate risk adjustment, Spiegelhalter has suggested winsorising outlier values.[28] However, it is important to understand excessive heterogeneity. With count data such as bacteraemias, false discovery rate calculation using the Benjamini–Hochberg procedure has been recommended when there are numerous institutions.[29] An alternative involves stratifying institutions by the services they provide.[7]

As the distributions of risk-adjusted rates are skewed when N_i values are small, we employ the beta and gamma distributions to calculate accurate control limits.[8] For the binary data upper 2 SD equivalent limit, one finds X such that, in S notation,

$$1 - \text{pbeta}(CL,\ X + 1,\ Ni - X) = 0.02275$$

When count (e.g. MRSA) data are risk-adjusted by hospital service, 7 indirectly standardized rates can be influenced by sample sizes and the methods described above for binary data may give unreliable results. It may be preferable to proceed directly to a random-effects (shrinkage) analysis (Appendix 2).

Appendix 2

The shrinkage value for institution 'i', denoted S_i, is approximately $(O_i + 1/V)/(E_i + 1/V)$, where V is the variation in true incidence rates among institutions.[13] We illustrate this calculation for two of the hospitals shown in Table I. The Poisson family among-hospital variance was 0.117. The largest hospital (G) had observed $O_G = 33$ and expected $E_G = 39.19$, whereas a smaller hospital (A) had $O_I = 20$ and $E_I = 10.43$. The unadjusted SMRs were then $SMR_G = 0.84$ and $SMR_I = 1.93$, with corresponding shrinkage values of $SMR_G = 0.87$ and $SMR_I = 1.5$. The change observed for the smaller institution A was greater than that for G. Thus when samples are small and variation among them is not large, the predicted estimate is drawn towards the average value.

We have employed the lmer function in the lme4/arm libraries in R to create shrinkage plots.[8] For binary data such as SSIs or cardiac surgical outcomes, lmer requires individual observations to be entered and the logit of each risk adjustment score is entered as an offset. Since the Es are entered as an offset, they are regarded as fixed.[19] Output is then the observed/expected odds ratio (Figure 2).

Occasionally, when there is very little variation among the institutions, lmer will fail as 1/V becomes very large. If a funnel plot is used for preliminary screening, it will indicate the absence of outliers making lmer use unnecessary. If desired, a random-effects analysis can still be performed using OpenBugs but specialised software and expertise are required.[30]

Conflict of interest statement
None declared.

Funding sources
None.

References

1. Beranholtz S, Provonost P, Lipsett P, *et al*. Eliminating catheter-related bloodstream infections in the intensive care unit. *Crit Care Med* 2004;**32**:2014–2020.
2. Gawande A. *The CHECKLIST manifesto*. London: Profile Books; 2010.
3. Morton A, Whitby M, McLaws ML, *et al*. The application of statistical process control charts to the detection and monitoring of hospital acquired infections. *J Qual Clin Pract* 2001;**21**:112–117.
4. Morton A, Clements A, Doidge S, Stackelroth J, Curtis M, Whitby M. Surveillance of healthcare-acquired infections in Queensland, Australia: data and lessons from the first 5 years. *Infect Control Hosp Epidemiol* 2008;**29**:695–701.
5. Horan T, Gaynes R, Martone W, Jarvis W, Emori T. CDC definitions of nosocomial surgical site infections, 1992: a modification of CDC definitions of surgical wound infections. *Infect Control Hosp Epidemiol* 1992;**13**:606–608.
6. Shahian D, Normand S-L. Comparison of "risk-adjusted" hospital outcomes. *Circulation* 2008;**117**:1955–1963.
7. Tong E, Clements A, Haynes M, Jones M, Morton A, Whitby M. Improved hospital-level risk adjustment for surveillance of healthcare-associated

bloodstream infections: a retrospective cohort study. *BMC Infect Dis* 2009;**9**:145.
8. R version 2.5.1 Development Core Team. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing; 2009.
9. Spiegelhalter D. Funnel plots for institutional comparison. *Qual Saf Health Care* 2002;**11**:390–391.
10. Winkel P, Zhang N. *Statistical development of quality in medicine*. Chichester: Wiley; 2007. pp. 62–70.
11. Hart M, Lee K, Hart R, Robertson J. Application of attribute control charts to risk-adjusted data for monitoring and improving health care performance. *Qual Mngmnt Health Care* 2003;**12**:5–19.
12. Cook D, Duke G, Hart G, Pilcher D, Mullany D. Review of the application of risk-adjusted charts to analyse mortality outcomes in critical care. *Crit Care Resusc* 2008;**10**:239–251.
13. Simpson J, Evans N, Gibberd R, Heuchan A, Henderson-Smart D. Analysing differences in clinical outcomes between hospitals. *Qual Saf Health Care* 2003;**12**:257–262.
14. Ward M, Bridgewater B. Implications of publishing surgical results. *Heart* 2007;**93**:1136.
15. Vincent C, Moorthy K, Sarker S, Chang A, Darzi A. Systems approaches to surgical quality and safety: from concept to measurement. *Ann Surg* 2004;**239**:475–482.
16. Guru V, Fremes S, Naylor D, *et al*. Public versus private institutional performance reporting: what is mandatory for quality improvement. *Am Heart J* 2006;**152**:573–578.
17. Cooper B. Confronting models with data. *J Hosp Infect* 2007;**65**:88–92.
18. Mohammed M, Rathbone A, Myers P, Patel D, Onions H, Stevens A. An investigation into general practitioners associated with high mortality flagged through the Shipman inquiry: retrospective analysis of routine data. *BMJ* 2004;**328**:1474–1477.
19. Silcocks P. Estimating confidence limits on a standardised mortality ratio when the expected number is not error free. *J Epidemiol Commun Health* 1994;**48**:313–317.
20. Pitches D, Burls A, Fry-Smith A. How to make a silk purse from a sow's ear – a comprehensive review of strategies to optimise data for corrupt managers and incompetent clinicians. *BMJ* 2003;**327**:1436–1439.
21. Mant J, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ* 1995;**311**:793–796.
22. Berwick D, Calkins D, McCannon C, Hackbarth A. The 100,000 lives campaign: setting a goal and a deadline for improving health care quality. *JAMA* 2006;**295**:324–327.
23. Singer A. Mandatory regular meetings of hospital staff would complement medical audit and revalidation. *BMJ* 2000;**320**:1072.
24. Thompson A, Stonebridge P. Building a framework for trust: critical event analysis of deaths in surgical care. *BMJ* 2005;**330**:1139–1142.
25. Austin P, Alter D, Tu J. The use of fixed- and random-effects models for classifying hospitals as mortality outliers: a Monte Carlo assessment. *Med Decision Making* 2003;**23**:526–539.
26. Adap P, Rouse A, Mohammed M, Marshall T. Performance league tables: the NHS deserves better. *BMJ* 2002;**324**:95–98.
27. Goggin L, van Gessel H, McCann R, Peterson A, Van Buynder P. Validation of surgical site infection surveillance in Perth, Western Australia. *Healthcare Infect* 2009;**14**:101–107.
28. Spiegelhalter D. Handling over-dispersion of performance indicators. *Qual Saf Health Care* 2005;**14**:347–351.
29. Jones H, Ohlssen D, Spiegelhalter D. Use of the false discovery rate when comparing multiple health care providers. *J Clin Epidemiol* 2009;**61**:232–240.
30. Thomas A, O'Hara B, Ligges U, Sturtz S. Making BUGS open. *R News* 2006;**6**:12–17.