



Sequential analysis of uncommon adverse outcomes

A. Morton^{a,b,*}, K. Mengersen^b, M. Waterhouse^{b,c}, S. Steiner^d, D. Looke^a

^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 30 April 2010

Available online 24 July 2010

Keywords:

Uncommon adverse events
Complex surgical site infections
MRSA bacteraemias
Statistical process control
Evidence-based systems

SUMMARY

Sequential analysis of uncommon adverse outcomes (AEs) such as surgical site infections (SSIs) is desirable. Short postoperative lengths of stay (LOS) result in many SSIs occurring after discharge and they are often superficial. Deep and organ space (complex) SSIs occur less frequently but are detected more reliably and are suitable for monitoring wound care. Those occurring post-discharge usually require readmission and can be counted accurately. Sequential analysis of methicillin-resistant *Staphylococcus aureus* bacteraemia is also needed. The key to prevention is to implement systems based on evidence, e.g. using 'bundles' and checklists. Regular mortality and morbidity audit meetings are required and these may need to be followed by independent audits. Sequential statistical analysis is desirable for data presentation, to detect changes, and to discourage tampering with processes when occasional AEs occur in a reliable system. Tabulations and cumulative observed minus expected ($O - E$) charts and funnel plots are valuable, supplemented in the presence of apparent 'runs' of AEs by cumulative sum analysis. Used prospectively, they may enable staff to visualise and detect patterns or shifts in rates and counts that might not otherwise be apparent.

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

Introduction

Morbidity and mortality (M&M) audits are required for adverse events (AEs) that are uncommon. Complementary statistical process control (SPC) analysis can assist in the detection of detrimental changes, while discouraging tampering with satisfactory processes in response to an essentially random AE.^{1,2}

Methods for SPC analysis of uncommon AEs are illustrated using Queensland Health Centre for Healthcare Related Infection Surveillance and Prevention (CHRISP) data.³ CHRISP conducts standardised surveillance of healthcare-acquired infection (HCAI) across 23 major Queensland hospitals.

Complex (deep and organ space) surgical site infections (SSIs) complicating 2575 orthopaedic procedures performed between 2001 and 2006 at one hospital (X) are employed to demonstrate the analysis of binary data. Surveillance involved five procedures:

partial (PHR) and total (THR) hip replacements, total knee replacements (TKR), and revisions of total hip (RTHR) and total knee (RTKR) replacements. Complex SSIs occurring during the patient's stay or within 30 days following surgery were recorded. Hospital X recorded 30 complex SSIs (1.17%).

Methicillin-resistant *Staphylococcus aureus* (MRSA) bacteraemia data collected from all wards of one hospital (Y) between 2001 and 2008 are used to illustrate the analysis of count data. In that period (2921 days), 87 bacteraemias were observed, giving a daily rate of 3%. Hence, for this institution, approximately one case was expected every month. This rate is too low to use standard SPC methods such as Shewhart charts.

Methods for calculating an expected probability

Estimated probabilities were derived from 136 complex SSIs complicating 12 838 orthopaedic procedures performed in 18 hospitals and recorded in the 2001–2006 CHRISP database. Separate estimates were obtained for relatively homogeneous subgroups. The 12 838 records were initially stratified by procedure type and the US Centers for Disease Control and

* 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).

Prevention Nosocomial Infections Surveillance System (NNIS) risk index.⁴ This index has four levels but only three (labelled 0, 1 and 2) were used with these data. Consequently, the initial stratification used 15 categories. Amalgamation of categories having small sample sizes led finally to the use of seven categories labelled A–G (Table I).

The expected probability for a category was estimated by the number of complex SSIs divided by the number of records in that category. For example, from a total of 364 RTHR procedures with any risk index, 11 complex SSIs were observed, giving an estimated probability of $11/364 = 0.03022$ for category G (Table I).

Hospital X performed 2575 procedures and 30 of these resulted in complex SSIs. The expected number was estimated by:

$$0 \times (0.018987) + 873 \times (0.007496) + 176 \times (0.01533) + 3 \times (0.025594) + 1311 \times (0.007968) + 212 \times (0.009459) + 0 \times (0.03022) = 21.77,$$

where the first number in each term is the number of procedures in the category, and the number in parentheses is the corresponding expected probability from Table I (hospital X undertook no surveillance of PHR or RTHR procedures).

Logistic regression provides an alternative means for calculating expected probabilities. For our example, the response was complex SSIs, and the explanatory variables were procedure type and NNIS risk index (Appendix 1). The expected number of complex SSIs for hospital X was 22.48. Logistic regression is usually required when there are more than two explanatory variables.

Sequential analysis

Tabulations

Table II shows observed and expected counts of AEs within half-yearly time periods; there may have been an increase in the numbers of AEs in 2004 and the second half of 2005.

Cumulative observed minus expected ($O - E$) charts with cumulative sum signals

These charts are similar to cumulative $E - O$ variable life-adjusted display (VLAD) charts.⁷ Two-standard-deviation (2 SD) equivalent control limits are included (Appendix 2).⁸ The cumulative $O - E$ line is approximately horizontal when the rate/count is stable, rises when it deteriorates, and falls when performance improves. The control limits indicate whether the accumulated AEs may differ from what is expected. Figure 1 suggests that there were two runs of SSIs. The first started on 3 November 2003 at 2.3 fewer SSIs than expected and ended on 12 November 2004 with 5.1

Table I

Estimated expected probability of the incidence of a complex SSI for categories based on stratification by procedure type and NNIS risk index

Category	Procedure	Risk index	Complex SSI	Procedures	<i>P</i>
A	Partial hip replacement	All	3	158	0.01899
B	Total hip replacement	0	25	3335	0.0075
C	Total hip replacement	1, 2	26	1696	0.01533
D	Revision of total hip replacement	All	14	547	0.02559
E	Total knee replacement	0	36	4518	0.00797
F	Total knee replacement	1, 2	21	2220	0.00946
G	Revision of total knee replacement	All	11	364	0.03022

SSI, surgical site infection; NNIS, National Nosocomial Infection Surveillance System.

Table II

Observed and expected numbers of complex surgical site infections (SSIs) for hospital X

Year	Half	No. of procedures	Observed no. of complex SSIs	Expected no. of complex SSIs
2001	1	159	2	1.37
	2	194	1	1.63
2002	1	164	2	1.33
	2	183	1	1.51
2003	1	229	0	1.89
	2	238	1	2.08
2004	1	195	5	1.65
	2	224	6	1.90
2005	1	231	1	1.97
	2	270	7	2.31
2006	1	200	2	1.66
	2	288	2	2.46
Total		2575	30	21.77

excess SSIs. The second run started on 8 August 2005 (4.2 excess SSIs) and ended on 13 March 2006 (9.1 excess SSIs). The upper limit was reached on 13 August 2006. Figure 2 shows a run of MRSA bacteraemias between November 2001 (8 fewer than expected) and October 2002 (O and E similar). The upper limit was not reached. The incorporation of a cumulative sum (CUSUM) test into the cumulative $O - E$ chart allows a hospital to identify a significant run of AEs in a timely fashion.¹⁰ A CUSUM test works by replacing each datum by a weight (w). Weights are then added sequentially. If their sum falls below zero, it is reset to zero. If it reaches a pre-specified control limit, denoted h , a signal is said to occur, indicating that a run of AEs has reached statistical significance. The choice of h involves a trade-off. Small values ensure that problems are detected quickly. Conversely, large values limit the frequency of false alarms. It is usual to set h to achieve a desired average time between false alarms, called average run length (ARL). This can be done by simulation as described by Ng *et al.*¹¹

The log-likelihood ratio CUSUM test can be used to determine when the odds of a binary AE have increased by a factor of $r > 1$.^{12,13} In this case, the weight for the i th observation is

$$w_i = O_i \times (\log r) - \log(1 + (r - 1) \times E_i),$$

where $O_i = 1$ if the AE occurred and 0 otherwise, and E_i is the estimated expected probability of the AE occurring. The CUSUM is frequently set to detect a doubling ($r = 2$) of the odds of an AE such as a complex SSI occurring. The CUSUM statistic was updated after each orthopaedic procedure. With $h = 2.75$ we expect approximately 5000 procedures between false alarms. For monitoring count data, the weight is

$$w_i = O_i \times \log(r) - (r - 1) \times E_i,$$

where O_i and E_i are the observed and expected counts, respectively. The CUSUM was set to detect a doubling in the MRSA bacteraemia rate ($r = 2$) and was updated using daily counts of MRSA bacteraemias. By setting $h = 3$, the average time between false alarms was approximately 150 months.

CUSUM signals are marked by prominent arrows in Figures 1 and 2. With the complex SSI data, signals occurred on 23 September 2004 and 12 December 2005. There was a CUSUM signal for the MRSA bacteraemias on 26 March 2002. Following a signal, investigation of the relevant systems should occur, the CUSUM value is reset to zero, and monitoring is recommenced.

Cumulative funnel plots

Cumulative funnel plots have been used for presenting percutaneous coronary intervention data that typically display low AE

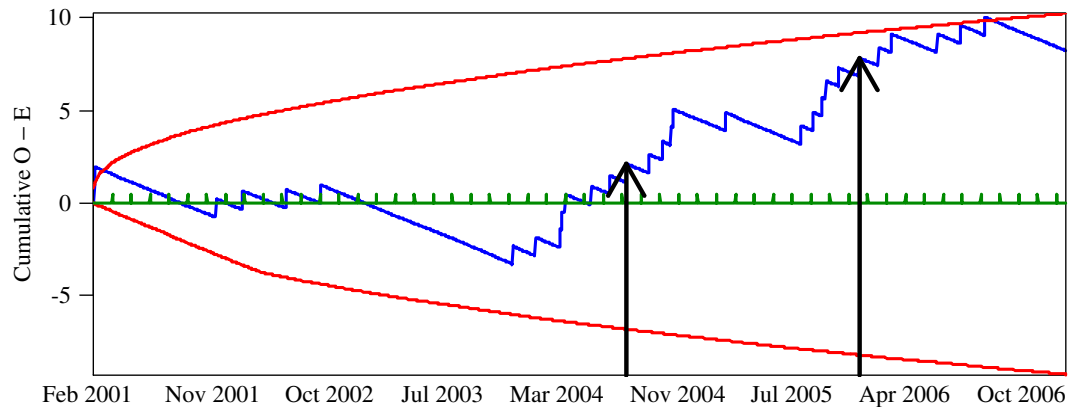


Figure 1. Orthopaedic complex surgical site infection, observed minus expected ($O - E$) plot from February 2001 to December 2006; total: 2575. Blue: observed; red: 95% limits; arrows: cumulative sum signals; tick marks every 50 units. Signals on 12 August 2004, 9 November 2005.

rates.¹⁴ This chart is similar to the cumulative $O - E$ chart except that the cumulative AE rate is shown (Appendix 2). Figure 3 shows the cumulative rate for the complex SSI data. A CUSUM test is incorporated. Prediction limits around the expected rate were obtained as described for cumulative $O - E$ charts (Appendix 2).

Discussion

In addition to indicating when excess AEs may be occurring, SPC methods help prevent tampering with reliable systems, and can aid in data presentation. When AEs are uncommon, e.g. fewer than two per month, standard Shewhart charts are unlikely to respond in a timely fashion to increases in AEs that warrant attention. Tabulations provide a simple means of summarising data and can indicate the presence of a cluster of AEs. When using tabulations, attention should be paid to counts/rates in contiguous time periods since clusters may span divisions.

The standard CUSUM chart can detect significant runs of AEs but displays CUSUM weights rather than data values. Sherlaw-Johnson incorporated CUSUM signals into VLAD charts.¹⁰ The cumulative $O - E$ chart is excellent for displaying a series of observations as it is possible to see the number of AEs differing from expected at any time in the series. Runs of AEs that produce a CUSUM signal are indicated with an arrow. Similarly, the funnel plot can display CUSUM signals.

It may be necessary to select a new starting point for a control chart if it is becoming cluttered with sequences of data that are no longer relevant. If possible, the new starting point should

correspond to a time at which the process is stable, as suggested by a horizontal cumulative $O - E$ line.

The CUSUM limit h should be selected to avoid excessive delays or frequent false signals. The latter can result in 'tuning out' so that a genuine change is missed, or there may be tampering with otherwise reliable systems. Scientists in infection control departments may need professional help when monitoring is commenced.

The log-likelihood CUSUM test has been criticised when used with rare events on technical grounds.¹⁵ However, it remains in regular use in clinical units to monitor mortality rates that are between 1% and 2%, and a recent report has suggested that it is suitable for monitoring uncommon AEs.¹⁶ An alternative is to use the closely related sequential probability ratio test.¹³ However, the CUSUM is widely employed and its resetting to zero following a signal may be an advantage.

Alemi has shown how monitoring increasing intervals between uncommon AEs can indicate success following system changes.¹⁷ However, we have found that some clinical workers have trouble relating to interval charts. It is also difficult to incorporate risk adjustment.

Although these methods can help to recognise problems they do not identify their causes, and delay in signalling can occur when AEs are rare. All serious AEs should be studied, for example in M&M audit meetings, followed, if necessary, by independent audits.^{18,19}

With early discharge, post-discharge SSIs are becoming increasingly important. However, there can be problems with the identification of superficial post-discharge SSIs.²⁰ Concentrating on complex SSIs that can be counted more accurately has been

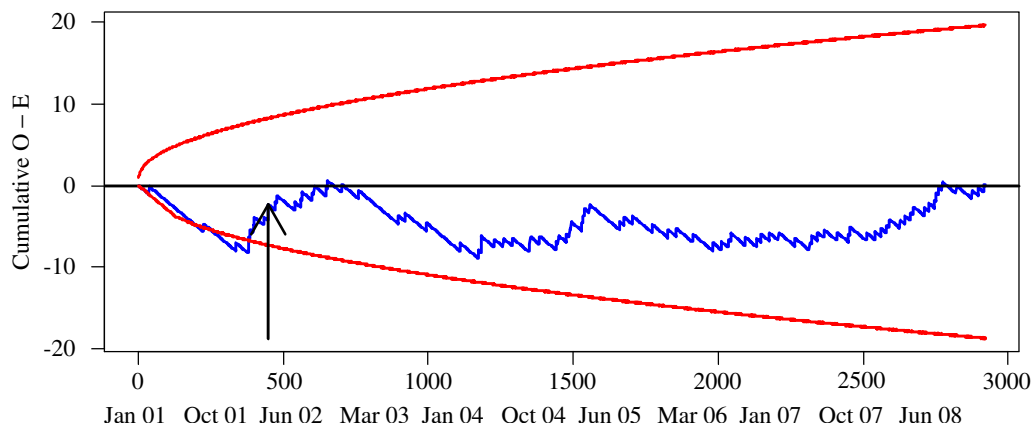


Figure 2. MRSA bacteraemias, observed minus expected ($O - E$) chart from January 2001 to December 2008; $N = 2922$ days. Signals on 26 March 2002.

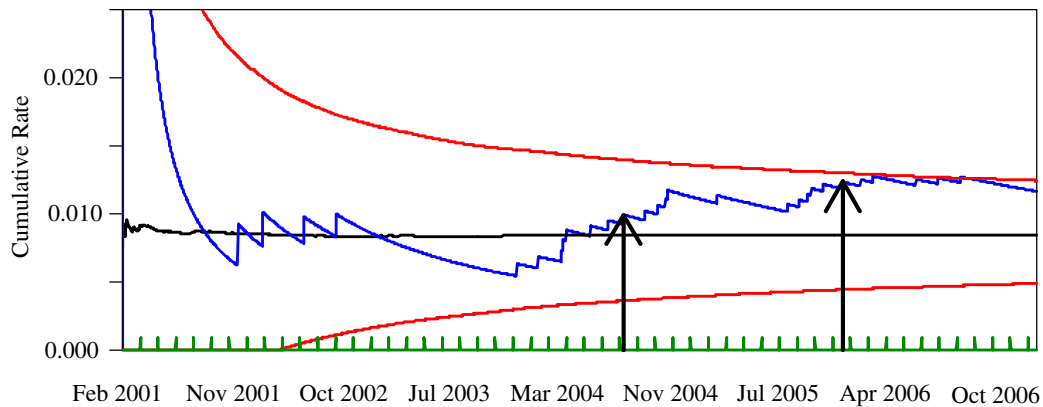


Figure 3. Orthopaedic complex surgical site infection, funnel plot from February 2001 to December 2006; total: 2575. Blue: observed; black: expected; red: 95% limits; arrows: CUSUM signals; tick marks every 50 units. Signals on 12 August 2004, 9 November 2005.

recommended although a mechanism is needed to count readmissions to other hospitals. Changes in their frequency can be used to understand changes in overall SSI rates.²¹

Whereas 30 day follow-up for procedures involving prostheses can be inadequate for detecting late complex SSIs, available CHRISP resources made longer follow-up impracticable. Since these outcomes are clinically and economically important, hospital staff are encouraged to present these cases at M&M audit meetings, and, if necessary, subject them to independent audit.^{18,19}

The bacteraemia data came from a hospital in which transmission of MRSA is usually well-controlled and historically an average of one case of MRSA bacteraemia occurred each month. Counts were sufficiently low for monitoring to involve all wards. If an increase were to occur, it is probable that the care of intravenous devices had been compromised and/or the evidence-based system for reducing transmission had broken down. M&M audit complemented by cumulative $O - E$, funnel plot and CUSUM analysis can serve as a warning system should these occur.

The occurrence of a signal should be used to indicate that a systems analysis is required, not that there is definitely substandard performance.²² Hasty judgment, e.g. when there is data error, can waste time, damage morale and weaken a reliable system.²

Finally, it must be emphasised that prevention of AEs is most important. This involves the implementation of systems based on evidence, for example in the form of 'bundles'.²³ Gawande has described how the use of checklists can help to ensure their application.²⁴

Conflict of interest statement

None declared.

Funding sources

None.

Appendix 1

In R (S) notation the logistic regression formula is⁵:
 $g < -\text{glm}(\text{complex SSI} \sim \text{group 2} + \text{group 3}, \text{family} = \text{binomial})$

	Estimate	SE	z-value	Pr(> z)
(Intercept)	-4.7822	0.1089	-43.908	<2e-16
Group 2	0.5249	0.2453	2.140	0.0324*
Group 3	1.2052	0.2148	5.611	2.01e-08

Null deviance: 1507.5 on 12837 degrees of freedom.

Residual deviance: 1480.2 on 12835 degrees of freedom.

The fitted values were: group 1, 0.0083; group 2, 0.014; and group 3, 0.0272. Each of these groups differed from the other two. Group 1 included PHR and TKR all NNIS Risk Indices (RI) and THR RI zero, group 2 THR RI one, and group 3 RTHR and RTKR all RI and THR RI two.

A random effects analysis would be required when groups of hospitals have differing expected outcome rates, e.g. bacteraemia rates would be higher in large hospitals performing complex procedures than in community hospitals.⁶

Appendix 2

For the upper and lower control limits for binary data (e.g. complex SSIs), Grunkemeier *et al.* recommend prediction limits based on the variance of the expected values.⁸ If $\Sigma_i \pi$ is the cumulative expected value at the i th procedure, the variance is $\Sigma_i (\pi \times (1 - \pi))$ so the upper 2 SD limit is $2\sqrt{\Sigma_i (\pi \times (1 - \pi))}$ since for cumulative $O - E$ the E is subtracted from both the O and E so that the cumulative $O - E$ line is at zero when they are equal. When samples are small, the distribution of binary data is skewed and limits based on the exact binomial distribution are more accurate.⁹ For a 2 SD approximate limit, the smallest value of $X(X_u)$ is found such that

$$\sum_{X=X_u \rightarrow N} C(N, X) \pi^X (1 - \pi)^{(N-X)} \leq 0.02275.$$

In practice it is simpler to use the beta distribution. For the upper 2 SD limit, one finds X such that, in S notation,

$$1 - \text{pbeta}(\pi, X + 1, N - X) = 0.02275.^5$$

Similar calculations using the gamma distribution apply with count (e.g. MRSA) data.

Cumulative funnel plots are similar to cumulative $O - E$ charts except that cumulative rates are displayed. For the i th procedure or day the above values are divided by N_i , the corresponding cumulative number of procedures or days.

References

1. Winkel P, Zhang N. *Statistical development of quality in medicine*. Chichester: Wiley; 2007.
2. Deming W. *Out of the crisis*. Cambridge: Cambridge University Press; 1982. p. 327.
3. 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.

4. Culver D, Horan T, Gaynes R, et al. Surgical wound infection rates by wound class, operative procedure, and patient risk index. *Am J Med* 1991;**91**(Suppl. 3B):S152–S157.
5. R Development Core Team. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing; 2009.
6. 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.
7. Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S. Monitoring the results of cardiac surgery by variable life-adjusted display. *Lancet* 1997;**350**:1128–1130.
8. Grunkemeier G, Jin R, Wu Y. Cumulative sum curves and their prediction limits. *Ann Thorac Surg* 2009;**87**:361–364.
9. Hart M, Hart R, Schmaltz S. Control limits for p control charts with small subgroup sizes. *Qual Manag Health Care* 2007;**16**:123–129.
10. Sherlaw-Johnson C. A method for detecting runs of good and bad clinical outcomes on variable life-adjusted display charts. *Health Care Manag Sci* 2005;**8**:61–65.
11. Ng J, Morlet N, Brenner A, et al. Techniques to monitor endophthalmitis and other cataract surgery complications. *Ophthalmology* 2008;**115**:3–10.
12. Steiner S, Cook R, Farewell V, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000;**1**:441–452.
13. Spiegelhalter D, Grigg O, Kinsman R, Treasure T. Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery. *Int J Qual Health Care* 2003;**15**:7–13.
14. Kunadian B, Dunning J, Roberts A, et al. Cumulative funnel plots for the early detection of interoperator variation: retrospective database analysis of observed versus predicted results of percutaneous coronary intervention. *BMJ* 2008;**336**:931–934.
15. Webster R. Development of statistical methods for the surveillance and monitoring of adverse events which adjust for differing patient and surgical risks. PhD thesis, Queensland University of Technology; 2008.
16. Sego L, Woodall W, Reynolds M. A comparison of surveillance methods for small incidence rates. *Stat Med* 2008;**27**:1225–1247.
17. Alemi F. Probabilistic risk analysis is practical. *Qual Manag Health Care* 2007;**16**:300–310.
18. Singer A. Mandatory regular meetings of hospital staff would complement medical audit and revalidation. *BMJ* 2000;**320**:1072.
19. Thompson A, Stonebridge P. Building a framework for trust: critical event analysis of deaths in surgical care. *BMJ* 2005;**330**:1139–1142.
20. Whitby M, McLaws M-L, Collopy B, et al. Post-discharge surveillance: can patients reliably diagnose surgical wound infections? *J Hosp Infect* 2002;**52**:155–160.
21. Anderson D, Chen L, Sexton D, Kaye K. Complex surgical site infections and the devilish details of risk adjustment: important implications for public reporting. *Infect Control Hosp Epidemiol* 2008;**29**:941–946.
22. Mohammed M, Rathbone A, Myers P, et al. An investigation into general practitioners associated with high patient mortality flagged up through the Shipman inquiry: retrospective analysis of routine data. *BMJ* 2004;**328**:1474–1477.
23. 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.
24. Gawande A. *The checklist manifesto*. London: Profile Books; 2010.