

Crude versus case-mix-adjusted control charts for safety monitoring in thyroid surgery

Antoine Duclos,^{1,2} Nicolas Voirin,^{3,4} Sandrine Touzet,^{1,2} Pietro Soardo,⁵
Anne-Marie Schott,^{1,2} Cyrille Colin,^{1,2} Jean-Louis Peix,⁵ Jean-Christophe Lifante⁵

► An appendix is published online only. To view this file please visit the journal online (<http://qshc.bmj.com>).

¹Hospices Civils de Lyon, Pôle Information Médicale Évaluation Recherche, Lyon, France

²Université de Lyon, EA Santé-Individu-Société 4129, Lyon, France

³Université de Lyon 1, Laboratoire de Biométrie et Biologie Évolutive, Lyon, France

⁴Hospices Civils de Lyon, Hôpital Édouard Herriot, Service d'Hygiène, Épidémiologie et Prévention, Lyon, France

⁵Hospices Civils de Lyon, Centre Hospitalier Lyon Sud, Service de Chirurgie Générale et Endocrinienne, Pierre Bénite, France

Correspondence to

Dr Antoine Duclos, Hospices Civils de Lyon, Pôle d'Information Médicale Évaluation Recherche, Unité d'Étude des Pratiques Professionnelles, 162 avenue Lacassagne, F-69424 Lyon Cedex 03, France; antoineduclos@yahoo.fr

Accepted 24 June 2009

Published Online First

8 April 2010

ABSTRACT

Introduction Patient-safety monitoring based on health-outcome indicators can lead to misinterpretation of changes in case mix. This study aimed to compare the detection of indicator variations between crude and case-mix-adjusted control charts using data from thyroid surgeries.

Methods The study population included each patient who underwent thyroid surgery in a teaching hospital from January 2006 to May 2008. Patient safety was monitored according to two indicators, which are immediately recognisable postoperative complications: recurrent laryngeal nerve palsy and hypocalcaemia. Each indicator was plotted monthly on a p-control chart using exact limits. The weighted κ statistic was calculated to measure the agreement between crude and case-mix-adjusted control charts.

Results We evaluated the outcomes of 1405 thyroidectomies. The overall proportions of immediate recurrent laryngeal nerve palsy and hypocalcaemia were 7.4% and 20.5%, respectively. The proportion of agreement in the detection of indicator variations between the crude and case-mix-adjusted p-charts was 95% (95% CI 85% to 99%). The strength of the agreement was $\kappa=0.76$ (95% CI 0.54 to 0.98). The single special cause of variation that occurred was only detected by the case-mix-adjusted p-chart.

Conclusions There was good agreement in the detection of indicator variations between crude and case-mix-adjusted p-charts. The joint use of crude and adjusted charts seems to be a reasonable approach to increase the accuracy of interpretation of variations in outcome indicators.

INTRODUCTION

The statistical control chart concept is a method of decision support combining time-series analysis with a graphical presentation of data.¹ It is based on the use of a control chart that guides the users towards appropriate action for improvement according to the nature of the variability.² The control chart is a well-documented tool to monitor and improve the safety of healthcare processes.³ It is helpful in interpreting and reducing sources of indicator variability by distinguishing special causes of variation from common causes. Special cause variations reflect substantial disparities in patient safety that deserve further investigation and action. In contrast, common cause variations arise due to misleading factors, including random events.

There is strong interest in integrating control charts into clinical practice, particularly to alert multidisciplinary clinical teams when special causes

of variation are detected. However, there is still debate about how to use control charts to monitor the outcomes of care, specifically regarding the necessity of adjusting for changes in patient case mix. Some argue that case-mix adjustment remains complex,⁴ may not be essential for longitudinal monitoring of outcomes⁵ or can lead to the erroneous conclusion of an unbiased measure.⁶ However, a control chart adjusted for case mix could theoretically improve the detection of special cause variations by controlling for biases that result from changes in case mix over time.^{7,8} In thyroid surgery, reported complication rates may vary depending on the extent of resection and the severity of the pathological thyroid condition, as well as the patient's characteristics.^{9–11} This study aimed to compare the detection of indicator variations between crude and case-mix-adjusted control charts in monitoring the safety of thyroid surgery.

METHODS

Population and data collection

A prospective survey was conducted in the department of endocrine surgery in a large teaching hospital.¹² The study population included every patient who underwent thyroid surgery during a 29-month period between 1 January 2006 and 31 May 2008. The surgical team under observation included three surgeons. Specific information about each thyroid surgery and its related complications was collected by the surgeon in charge of the patient using a standardised form. Other information was obtained from the hospital information system, which contains standard discharge abstracts, including compulsory data on the patients' demographics and the thyroid disease diagnoses.

Indicators and control chart characteristics

Two indicators were used to monitor the safety of thyroid surgery. The first indicator was the in-hospital proportion of immediate postoperative recurrent laryngeal nerve palsy, which was systematically assessed by flexible transnasal laryngoscopy among patients who underwent a lobectomy or a bilateral thyroidectomy. The second indicator was the in-hospital proportion of immediate postoperative hypocalcaemia, which was rigorously defined by a serum calcium level lower than 2 mmol/l only among patients who underwent a bilateral thyroidectomy.

Each indicator was extracted from the hospital information system and plotted on a Shewhart p-control chart.¹ Crude and case-mix-adjusted p-charts were constructed such that each data point

expressed the observed proportion of complications per month for samples of variable size, as follows:

- The central line value of the crude p-chart was constant and was determined based on the overall proportion of complications. Exact limits were calculated for each month using the binomial-based SD of the overall proportion of complications.
- The central line value of the case-mix-adjusted p-chart varied monthly and was established based on the expected proportion of complications per month. Exact limits were calculated for each month using the binomial-based SD of the expected proportions of complications.

When a special cause variation was detected by the p-charts, after checking the completeness and quality of the data, a systematic investigation was undertaken to identify the cause using two complementary methods. We first looked at the logbook, in which all changes in care processes were continually reported by surgeons. Every 4 months, a multidisciplinary meeting was also conducted to discuss the observed variations in monitored indicators on control charts.

Statistics

Case-mix adjustment was based on the patients' gender and age, thyroid disease diagnosis and type of thyroidectomy.^{9–11} Case-mix-adjusted ORs for each postoperative complication were calculated using a multivariate logistic regression with all previous variables entered in the model. The expected proportion of each complication was then calculated for each month based on the model estimates. Final models were assessed for potential interactions. A finding of non-significance ($p > 0.05$ and the closest to 1) following the Hosmer–Lemeshow test was interpreted as an adequate fit of the model to the data.¹³ The discriminatory power of the models was measured using the c-statistics (Receiver Operating Characteristic curve score). Values less than 0.7 were considered to show poor discrimination.¹⁴

The p-charts' limits were determined according to an exact method based on the binomial distribution.¹⁵ Exact control and warning limits were set at 99.73% (3 SD from the mean) and 95.45% (2 SD from the mean), respectively (online Appendix 1). The detection of a special cause variation was defined as a single point outside the control limits or two out of three successive points between a warning limit and a control limit on the same side of the central line.¹⁶

The agreement between crude and case-mix-adjusted control charts in detecting indicator variations was measured using the weighted Cohen κ statistic.¹⁷ The positions of the data points for both postoperative recurrent laryngeal nerve palsy and hypocalcaemia were compared in terms of five ordinal levels based on warning and control limits.

Statistical analyses were conducted using SPSS (release 12.0.0; SPSS, Chicago), and control charts were generated using Microsoft Office Excel 2007 (Microsoft Corporation, Redmond, Washington).

RESULTS

Study population

We evaluated the outcomes of 1405 thyroidectomies. Among these procedures, 1036 were total thyroidectomies, 195 were lobectomies, 97 were thyroidectomies with lymph node resection, and 77 were completion thyroidectomies. The makeup of the thyroid diseases that were operated on was 758 non-toxic multinodular goitres, 301 carcinomas, 176 non-toxic solitary nodules, 100 Graves disease and 70 other diagnoses. The median patient age was 53 years (range 9–93), and 78.7% of patients were women (1106/1405).

The overall proportions of immediate postoperative recurrent laryngeal nerve palsy and hypocalcaemia were 7.4% (103/1388) and 20.5% (243/1186), respectively. Among the 309 patients who had an immediate complication consecutive to bilateral thyroidectomy, 215 had hypocalcaemia, 67 had recurrent laryngeal nerve palsy, and 27 had both complications.

Crude versus case-mix-adjusted control charts

The type of thyroid surgery ($p = 0.001$) was associated with a risk of postoperative recurrent laryngeal nerve palsy, whereas patient gender ($p < 0.001$), age ($p < 0.001$) and the type of thyroid surgery ($p < 0.001$) were associated with a risk of postoperative hypocalcaemia (table 1).

There was one difference between the crude and case-mix-adjusted control charts related to postoperative recurrent laryngeal nerve palsy (figure 1): in May 2006, the indicator reached the adjusted warning limit but not the crude warning limit. There were two differences between the crude and case-mix-adjusted control charts related to the monitoring of postoperative hypocalcaemia (figure 2). In July 2007, the indicator just reached the control limit of the crude control chart, whereas it crossed over the control limit of the case-mix-adjusted control chart. In January 2006, the indicator reached the crude but not the adjusted warning limit. Investigation of the single special cause variation that was detected in July 2007 revealed two elements. First, operating-room renovations took place from July to September 2007, leading to a reduction in available operating time for the same number of patients (from 42 to 32 h per week). Second, one surgeon was away in July 2007, although the number of patients undergoing thyroid surgery remained constant, so in July almost all procedures were performed by the only other surgeon available.

Table 1 Factors associated with immediate postoperative complications

	Adjusted ORs†(95% CI)	
	Recurrent laryngeal nerve palsy (n = 1384)	Hypocalcaemia (n = 1182)
Patient gender		
Female	0.78 (0.48 to 1.25)	2.74 (1.74 to 4.34)
Male*	1.0	1.0
Patient age (years)		
Age ≤ 40	0.83 (0.47 to 1.46)	2.57 (1.70 to 3.88)
40 < age ≤ 50	0.72 (0.39 to 1.32)	2.03 (1.32 to 3.14)
50 < age ≤ 60	0.80 (0.46 to 1.39)	1.02 (0.65 to 1.61)
Age > 60*	1.0	1.0
Thyroid disease		
Malignant neoplasm	1.38 (0.47 to 4.08)	1.35 (0.55 to 3.30)
Non-toxic solitary nodule*	1.0	1.0
Non-toxic multinodular goitre	1.49 (0.54 to 4.12)	1.25 (0.53 to 2.95)
Graves disease	0.62 (0.13 to 2.83)	1.56 (0.60 to 4.08)
Others‡	0.86 (0.20 to 3.78)	1.05 (0.32 to 3.37)
Type of surgery		
Lobectomy	0.46 (0.15 to 1.36)	
Total thyroidectomy*	1.0	1.0
Completion thyroidectomy	1.50 (0.68 to 3.28)	1.20 (0.65 to 2.22)
Thyroidectomy with lymph-node resection	3.35 (1.67 to 6.75)	2.90 (1.69 to 4.97)

*Reference group.

†No interactions were observed in the models. The Hosmer–Lemeshow goodness-of-fit statistics for the recurrent laryngeal nerve palsy and hypocalcaemia models were 0.93 and 0.52, respectively. The c-statistics for the recurrent laryngeal nerve palsy and hypocalcaemia models were 0.65 (95% CI 0.60 to 0.71) and 0.68 (95% CI 0.64 to 0.72), respectively.

‡Autoimmune thyroiditis and thyrotoxicosis with multinodular goitre or single thyroid nodule.

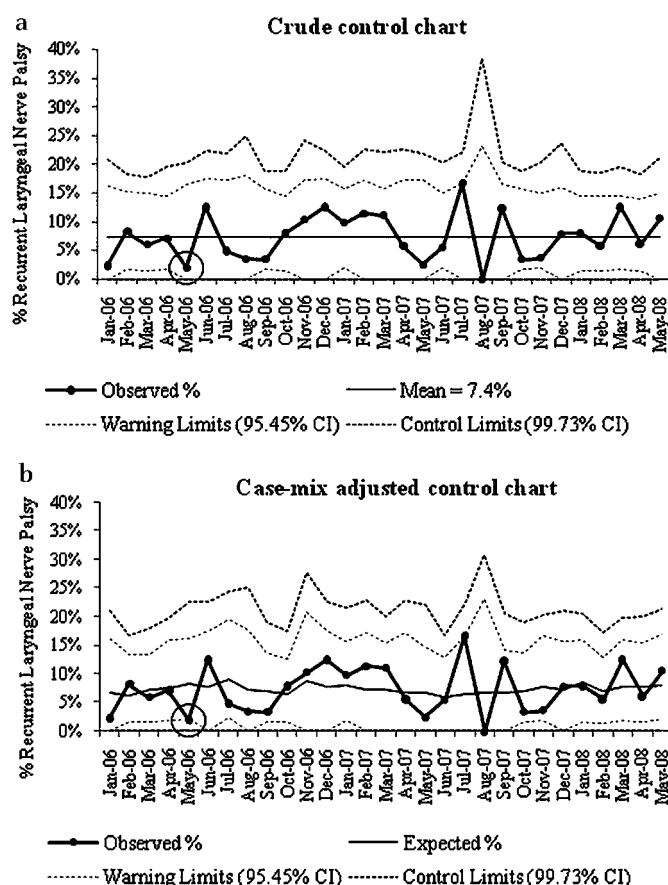


Figure 1 Crude versus case-mix adjusted p-control charts of immediate postoperative recurrent laryngeal nerve palsy.

The agreement between the crude and case-mix-adjusted control charts in detecting indicator variations (table 2) was 95% (95% CI 85% to 99%). The strength of agreement was $\kappa=0.76$ (95% CI 0.54 to 0.98).

DISCUSSION

There was good agreement between crude and case-mix-adjusted p-charts in detecting indicator variations for the monitoring of thyroid surgery safety. However, the single special cause of variation that occurred in July 2007 for hypocalcaemia monitoring was only detected by the case-mix-adjusted p-chart, suggesting that this method was slightly more sensitive than the crude control chart. Investigation of this cause revealed that the overactivity of a single surgeon was probably detrimental to the quality of thyroid surgery. Consequently, the primary concern should be to guarantee that surgeons can operate under satisfactory conditions and that they are not given excessive workloads.¹²

Performance monitoring using outcome indicators requires that the observed variations reflect true variations in patient safety. Variability in medical outcomes typically arises from a combination of four key elements: data quality, patient case-mix, quality of care and random effects.¹⁸ First, we assume that the observed variation in thyroid surgery outcomes is not influenced by the quality of data, since our monitoring system was implemented in a hospital where standardised modalities for detecting postoperative complications and for data collection were not changed during the study period. Second, provided that the adjustment was sufficient, and in view of the satisfactory agreement

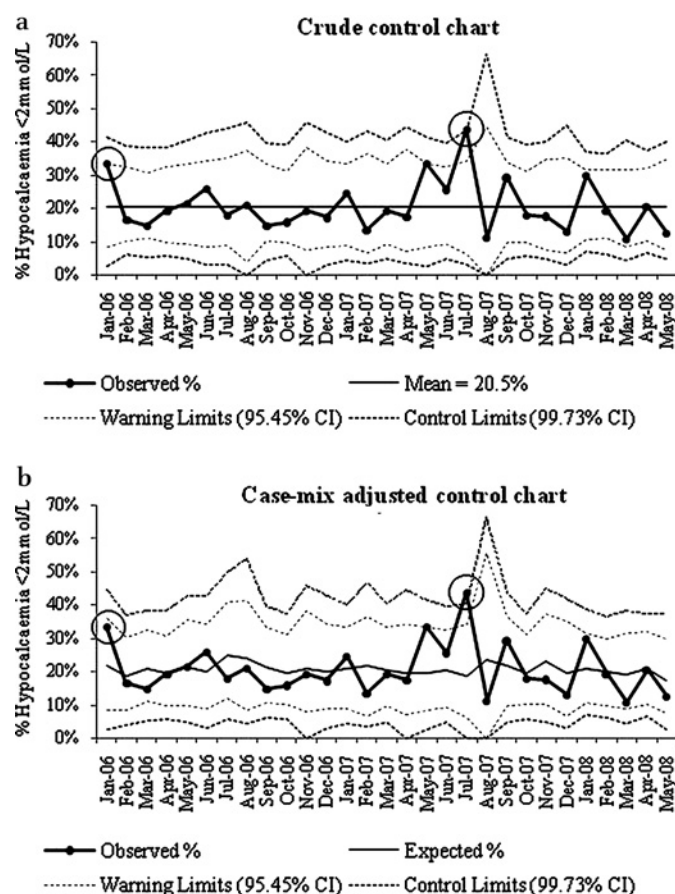


Figure 2 Crude versus case-mix adjusted p-control charts of immediate postoperative hypocalcaemia.

between crude and case-mix-adjusted p-charts, we also presume that the occurrence of complications was poorly influenced by the patient case-mix. This may be biased if adjustment is inadequate or if it does not include some unmeasured (eg, patient weight, large or thoracic goitre, and invasive cancer) or unknown case-mix factors, which might affect the outcome irrespective of quality of care.¹⁹ To assess whether the choice of variables for case-mix adjustment changed the conclusions of this study, we tested models based on different adjustment methods and observed similar results (data not shown). Third, the rationale for using case-mix-adjusted control charts is that the residual unexplained variation in outcomes and the detection of special causes are more likely to be attributable to the quality of care. The remaining variance is presumed to be due to the characteristics of the individual surgeon (such as technical, physical and

Table 2 Agreement between crude and case-mix-adjusted p-charts in detecting indicator variations

Case-mix adjusted limits*		<WL	=WL	WL-CL	=CL	>CL	Total
Crude Limits*	<WL	53	1	0	0	0	54
	=WL	1	2	0	0	0	3
	WL-CL	0	0	0	0	0	0
	=CL	0	0	0	0	1	1
	>CL	0	0	0	0	0	0
	Total	54	3	0	0	1	58

*Each unit in the table represents the position of a data point on a recurrent laryngeal nerve palsy or hypocalcaemia control chart, according to five ordinal levels based on exact warning limits (WL, 95.45%) and control limits (CL, 99.73%), as follows: <WL (below the 95.45% limit), =WL (equal to the 95.45% limit), WL-CL (between the 95.45% and the 99.73% limits), =CL (equal to the 99.73% limit) and >CL (above the 99.73% limit).

mental conditions) or to other organisational factors related to the surgical team (including equipment availability, team coordination or workload).²⁰ It would be useful to further adjust for variables regarding surgeons' compliance with evidence-based practices in order to restrict the sources of variability in medical outcomes to uncontrollable random effects. In thyroid surgery, adjustments could include the systematic visualisation¹¹ of recurrent laryngeal nerves in order to avoid injuries, as well as routine identification of at least two parathyroid glands in order to avoid permanent hypocalcaemia.²¹ Furthermore, to evaluate how one surgeon's performance influences the outcomes of the entire team, we could adjust for surgeon profile. Real-time monitoring of activity for each surgeon is also feasible using cumulative sum charts (CUSUM) to detect any defects in surgical processes as soon as possible.^{22 23} Focusing on individual performance allows each surgeon to check the quality of their daily practices and facilitates the identification of special cause variations. Nevertheless, compared with the simplicity of an adjusted p-chart, such an alternative tool can give less straightforward interpretations of variations in outcome indicators.

Though relatively simple and user-friendly, the value of the p-chart lies in its intuitive nature and the fact that the data are displayed graphically. Certainly, crude p-charts are fast and easy to produce, which provides quick performance feedback to the surgical staff in the context of real-time monitoring. One main limitation of these charts concerns the potential relationship between the occurrence of complications and case-mix characteristics. The clinical heterogeneity introduced by considering the outcomes of all procedures together in a broad grouping without case-mix adjustment is often argued to be clinically inappropriate. Due to the few differences observed between crude and case-mix-adjusted p-charts, our findings suggest that the expected gain from routine use of a case-mix-adjusted p-chart would be relatively minimal. Yet, as the great majority of data points were positioned between the lower and upper warning limits, statistical significance tests were likely to be underpowered and therefore a mere formality. Despite an acceptable fit to the data, the discriminatory power of the models was relatively poor. This might suggest that case-mix adjustment only controlled for a small part of the variability in thyroid surgery outcomes, which may simply demonstrate consistent care. This is perhaps not surprising, considering that thyroidectomy is a highly standardised procedure compared with other healthcare processes.²⁴ Additional confounding variables (such as other patient characteristics, surgeon profiles, organisational factors or adherence to evidence-based practices) could more precisely represent the variability in thyroid surgery outcomes and thus could be useful adjustments.

More research is needed to provide empirical evidence to corroborate the usefulness of the case-mix adjusted control chart in other settings.⁵ Although case-mix adjustment requires training in data analysis and is relatively time-consuming to integrate into the team's daily routine, it may be useful when indicator is positioned close to the limits of the crude control

chart, which requires careful interpretation. The joint use of both crude and adjusted p-charts seems to be a reasonable approach to increase the accuracy of interpretation of outcome indicator variations in clinical practice.

Acknowledgements We are grateful to P Messy, for his helpful contribution to the study, and to A Favre, for the English revision of the manuscript.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Previous communication The study's main results were presented at the 24th Patient Classification Systems International Working Conference (Lisbon, Portugal, 8–11 October 2008).

REFERENCES

1. **Shewhart WA.** *Economic control of quality of manufactured product.* New York: D Van Nostrand Company, 1931.
2. **Mohammed MA, Cheng KK, Rouse A, et al.** Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001;**357**:463–7.
3. **Thor J, Lundberg J, Ask J, et al.** Application of statistical process control in healthcare improvement: systematic review. *Qual Saf Health Care* 2007;**16**:387–99.
4. **Iezzoni LI.** Using risk-adjusted outcomes to assess clinical practice: an overview of issues pertaining to risk adjustment. *Ann Thorac Surg* 1994;**58**:1822–6.
5. **Marshall T, Mohammed MA.** Case-mix and the use of control charts in monitoring mortality rates after coronary artery bypass. *BMC Health Serv Res* 2007;**7**:63.
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**:1147–54.
7. **Hart MK, Lee KY, Hart RF, et al.** Application of attribute control charts to risk-adjusted data for monitoring and improving health care performance. *Qual Manag Health Care* 2003;**12**:5–19.
8. **Alemi F, Oliver DW.** Tutorial on risk-adjusted P-charts. *Qual Manag Health Care* 2001;**10**:1–9.
9. **Thomusch O, Machens A, Sekulla C, et al.** The impact of surgical technique on postoperative hypoparathyroidism in bilateral thyroid surgery: a multivariate analysis of 5846 consecutive patients. *Surgery* 2003;**133**:180–5.
10. **Dralle H, Sekulla C, Haerting J, et al.** Risk factors of paralysis and functional outcome after recurrent laryngeal nerve monitoring in thyroid surgery. *Surgery* 2004;**136**:1310–22.
11. **Thomusch O, Machens A, Sekulla C, et al.** Multivariate analysis of risk factors for postoperative complications in benign goiter surgery: prospective multicenter study in Germany. *World J Surg* 2000;**24**:1335–41.
12. **Duclos A, Touzet S, Soardo P, et al.** Patient safety monitoring in thyroid surgery using the Shewhart control chart. *Brit J Surg* 2009;**96**:171–4.
13. **Hosmer DW, Lemeshow S.** *Applied logistic regression.* New York: Wiley, 1989.
14. **Aylin P, Bottle A, Majeed A.** Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ* 2007;**334**:1044.
15. **Mohammed MA, Worthington P, Woodall WH.** Plotting basic control charts: tutorial notes for healthcare practitioners. *Qual Saf Health Care* 2008;**17**:137–45.
16. **Benneyan JC, Lloyd RC, Plsek PE.** Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care* 2003;**12**:458–64.
17. **Cohen J.** Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bulletin* 1968;**70**:213–20.
18. **Mant J.** Process versus outcome indicators in the assessment of quality of health care. *Int J Qual Health Care* 2001;**13**:475–80.
19. **Pitches DW, Mohammed MA, Lilford RJ.** 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.
20. **Vincent C, Moorthy K, Sarker SK, et al.** Systems approaches to surgical quality and safety: from concept to measurement. *Ann Surg* 2004;**239**:475–82.
21. **Pattou F, Combemale F, Fabre S, et al.** Hypocalcemia following thyroid surgery: incidence and prediction of outcome. *World J Surg* 1998;**22**:718–24.
22. **Woodall WH.** The use of control charts in health-care and public-health surveillance. *J Qual Technol* 2006;**38**:89–104.
23. **Rogers CA, Reeves BC, Caputo M, et al.** Control chart for monitoring cardiac surgical performance and their interpretation. *J Thorac Cardiovasc Surg* 2004;**128**:811–19.
24. **Hegner CF.** A history of thyroid surgery. *Ann Surg* 1932;**95**:481–92.