Temporal Variation in Surgical Mortality Within French Hospitals

Antoine Duclos, MD, PhD,†‡ Stéphanie Polazzi, MPH,* Stuart R. Lipsitz, ScD,‡ Sandrine Couray-Targe, MD, MPH,* Atul A. Gawande, MD, MPH,‡ Cyrille Colin, MD, PhD,*† William Berry, MD, MPH,‡ John Z. Ayanian, MD, MPP,‡§ and Matthew J. Carty, MD‡*

Background: Surgical mortality varies widely across hospitals, but the degree of temporal variation within individual hospitals remains unexplored and may reflect unsafe care.

Objectives: To add a longitudinal dimension to large-scale profiling efforts for interpreting surgical mortality variations over time within individual hospitals.

Design: Longitudinal analysis of the French nationwide hospital database using statistical process control methodology.

Subjects: A total of 9,474,879 inpatient stays linked with open surgery from 2006 through 2010 in 699 hospitals.

Measures: For each hospital, a control chart was designed to monitor inpatient mortality within 30 days of admission and mortality trend was determined. Aggregated funnel plots were also used for comparisons across hospitals.

Results: Over 20 successive quarters, 52 hospitals (7.4%) experienced the detection of at least 1 potential safety issue reflected by a substantial increase in mortality momentarily. Mortality variation was higher among these institutions compared with other hospitals (7.4 vs. 5.0 small variation signals, $P<0.001$). Also, over the 5-year period, 119 (17.0%) hospitals reduced and 36 (5.2%) increased their mortality rate. Hospitals with improved outcomes had better control of mortality variation over time than those with deteriorating trends (5.2 vs. 6.3 signals, $P=0.04$). Funnel plots did not match with hospitals experiencing mortality variations over time.

Conclusions: Dynamic monitoring of outcomes within every hospital may detect safety issues earlier than traditional benchmarking and guide efforts to improve the value of surgical care nationwide.

Key Words: surgical mortality, hospital performance, outcome variation, statistical process control, control chart

(*Med Care* 2013;51: 1085–1093)

From the *Hospices Civils de Lyon, Pôle Information Médicale Evaluation Recherche; †Université de Lyon, EA Santé-Individu-Société, Lyon France; ‡Center for Surgery and Public Health, Brigham and Women's Hospital, Harvard Medical School; and §Department of Health Care Policy, Harvard Medical School and Division of General Medicine, Brigham and Women's Hospital, Boston, MA.

Study concept and design: A.D., A.A.G., M.J.C., S.R.L.; Acquisition of data: A.D., C.C., S.C.-T., S.P.; Statistical analysis: A.D., S.R.L.; Analysis and interpretation of data: A.D., A.A.G., C.C., J.Z.A., M.J.C., S.C.-T., S.R.L., S.P., W.B.; Drafting of the manuscript: A.D., M.J.C.; Critical revision of the manuscript for important intellectual content: A.D., A.A.G., C.C., J.Z.A., M.J.C., S.C.-T., S.R.L., S.P., W.B.

The authors declare no conflict of interest.

Reprints: Antoine Duclos, MD, PhD, Center for Surgery and Public Health, Brigham and Women's Hospital, Harvard Medical School, One Brigham Circle, 1620 Tremont St., Boston, MA 02120. E-mail: antoineduclos@yahoo.fr.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website, www.lww-medicalcare.com.

ISSN: 0025-7079/13/5112-1085

Historically, systems for monitoring health care outcomes have been limited by difficulties related to data acquisition. The development of large repositories of information over the past decades has partly resolved this concern. What remains a challenge is how to optimally interpret the data in a timely manner to better inform efforts designed to improve the value of care delivery.[1]

Within the realm of surgery, traditional approaches to evaluate hospital care have included comparisons of aggregated outcomes between institutions. In-hospital mortality has frequently been selected as a primary outcome measure due to its gravity and the fact that it is more readily available in hospital databases than other complications.[2] However, inferring that differences in mortality across institutions reflect differences in the quality of surgical care delivered must be carefully considered in light of variations in data coding, patient recruitment, and clinical pathways.[3,4] Performance ranking of hospitals may be heavily influenced by the methodology used.[5,6] Lastly, cross-sectional assessments of institutions are not ideally suited to monitor hospital performance prospectively, to highlight the occurrence of sudden increases in mortality caused by safety issues, or to demonstrate that mortality has decreased or increased for a specific hospital.

A more dynamic approach would rely on the monitoring of hospital mortality longitudinally to interpret variations in surgical outcomes over time. Although wide variability across hospitals exists and has been intensively investigated,[7,8] the degree of temporal variation within institutions is still unknown and may reflect unsafe care. The biomedical literature suggests transferability of quality improvement methodologies from the manufacturing industry to the health care arena.[9] In particular, statistical process control (SPC) principles are based on the identification and reduction of variability in production processes utilizing control charts.[10] Validated through 50 years of experience, these methods have proven useful for improving quality[11] and now offer a potential foundation for the establishment of prospective outcome monitoring to detect and reduce errors in surgical care.[12]

To add a longitudinal dimension to large-scale profiling efforts, we mainly investigated surgical mortality variations over time for individual French hospitals on a nationwide basis utilizing SPC control chart methodologies. Assuming that understanding of temporal fluctuations in mortality may foster improved surgical safety, we have secondarily sought to explore the degree to which better control of small mortality variation over time was associated with less detection of potential safety issues and a trend toward better outcomes within hospitals.

METHODS

Study Population and Data Source

We used data from the French nationwide hospital database in acute care from 2006 through 2010 to identify all adults 18 years of age or older who underwent surgery and possibly died during their hospitalization. This database has the advantage that it includes information about all inpatient stays that have occurred in every French public and private hospital. In 2008, the French population amounted to 63,961,859 persons,[13] resulting in 17,017,081 acute hospitalizations.[14] Standard discharge abstracts for each of these hospitalizations contain compulsory information about the patient (sex and age), the primary and secondary diagnoses using the International Classification of Diseases, 10th revision (ICD-10 codes), as well as the procedural codes associated with the care provided.

For this study, we first identified a subgroup of 15,357,111 stays with open surgical procedures performed in 1365 hospitals (Appendix 1 illustrates the study flow chart). Ophthalmologic, dental, and obstetric procedures were excluded, as well as pediatric, ambulatory and palliative care, and organ retrieval and stays longer than 30 days. As outpatient surgery is associated with a very low rate of mortality, we further restricted our analysis to 11,218,969 inpatient stays with at least 1 night spent in the hospital. Finally, given our interest in monitoring surgical mortality longitudinally, we further limited the study population to hospitals that had 50 or more discharges per quarter, where at least 1 death was recorded per year, and with low variation in annual volume of stays (under 50%) over the 5-year period. Thus, we were expecting to exclude some hospitals whose particular activity would have jeopardized the relevance and feasibility of performance monitoring over time, and to minimize artifacts due to systematic coding errors or non-adjustable case-mix variations. With these exclusions, our study cohort included 9,474,879 inpatient stays with open surgery performed in 699 facilities.

Study Design and Outcome Assessment

The main outcome measure in our study was in hospital death within 30 days of patient admission. To fit models for outcome adjustment independently from the data to be monitored, the database was randomly split into training and testing datasets with stratification by hospital and quarter of discharge.[15] The arbitrary choice for a 20/80 splitting was made in consideration of the large study sample, allowing us to compute robust estimates from the training dataset without reducing heavily the size of the testing dataset. Accordingly, the training data included 20% of the initial dataset (1,900,588 stays) and served to identify the risk factors and estimate their effect on surgical outcome. In the final model, death was the outcome of interest, whereas patients' characteristics (sex, age, Elixhauser comorbidities,[16] and surgical procedure codes) were selected a priori as clinically important covariates, and the year of hospital discharge to control for potential secular trends[17] and coding variations.[18] We estimated parameters of the multivariate logistic regression model using generalized estimating equations with an exchangeable working correlation structure to account for clustering of surgical cases from the same hospital.[19,20] The intraclass correlation coefficient was used to estimate the design effect for each hospital, reflecting inflation in the variance due to clustering of patients within a hospital.[21] A conservative inflation factor was imputed to limits calculation on both control charts and funnel plots for avoiding over detection of outliers within hospitals and dealing with mortality over dispersion across hospitals.

The testing data included the remaining 80% of the initial dataset (7,574,291 stays), with the aim of studying variations in mortality both longitudinally for each individual hospital using control charts and cross-sectionally between all hospitals using funnel plots. For every surgical case, the expected probability of in-hospital death was computed, leading to an expected mortality rate for each hospital. To derive each hospital's risk-adjusted operative mortality, we calculated the hospital-specific ratio between the observed and the expected mortality multiplied by the overall mortality rate from the pooled hospital data.

Statistics and Charts

Aggregated funnel plots were initially generated to depict every hospital's mortality as a function of its total number of surgical cases to illustrate a traditional benchmarking approach.[22] Control and warning limits were set at 3 and 2 SDs around the central line, respectively, using a continuity correction in the Wald asymptotic confidence limits.[23] Confidence limits crossing 1 indicated whether a particular hospital's performance differed from the overall population mortality rate for the entire study period. Poor performing hospitals were positioned above the upper limits, whereas high performing hospitals with unusually good results were below the lower limits.

Adjusted mortality was then plotted on a Shewhart P-control chart to simulate prospective outcome monitoring for every individual hospital. Each data point depicted the hospital mortality per quarter. The central line reflected the mean mortality for each individual hospital. Control, warning, and small variation limits were set at 99.7% (3 SD), 95.5% (2 SD), and 68.3% (1 SD) around the mean, respectively, based on the binomial distribution.[24] The detection of a potential safety issue was defined as a single point outside the upper control limit or 2 out of 3 successive points outside the upper warning limit, reflecting a substantial increase in mortality temporarily.[25] According to these rules, the false-positive rate of a control chart for detecting a safety issue based on 20 time-ordered points was 8.9%.[26] Furthermore, small variations in mortality were monitored and quantified as the total number of signals over the upper or lower 1 SD limits for a given hospital. On the basis of the control chart interpretation, we estimated the proportion of hospitals having experienced the detection of at least 1 safety issue, and the mean frequency of small variations signals per hospital during the study period.

For each hospital, we then determined separately mortality trends over the 5-year period using a weighted propensity score approach to facilitate outcome comparisons over time. We selected this approach due to a concern about the possibility that in small hospitals some procedures were not done very often or some particularly sick people were rarely seen. Accordingly, to avoid skewing the results, we wanted to down-weight these people when examining trends. Despite adding to computational burden when implementing our approach on a large scale, this was assumed to avoid discouraging the hospital from performing surgery on these atypical patients. Each patients' characteristics (sex, age, Elixhauser comorbidities,[16] and surgical procedure codes) were weighted by the inverse probability of being in a given hospital by year combination, with the goal of balancing all patients' characteristics across hospitals and years. Computation of the adjusted mortality trend was then based on a logistic regression model. In-hospital death was the outcome of interest, whereas the year of hospital discharge was considered as the predictor. In cases in which the linear term for the predictor was noted to be significant, we determined whether the hospital had experienced a trend toward improvement (decreased mortality) or deterioration (increased mortality).

Finally, the aggregated funnel plots were updated to depict the supplemental individual institutional signaling data on control chart and trend data described above. In addition, hospitals characteristics and the frequency of small variation signals per hospital were compared between institutions having experienced the detection of a potential safety issue or not, and between those with improvement or deterioration in their mortality rate. All analyses were performed using SAS software (version 9.2; SAS Institute Inc., Cary, NC), tests were 2-sided and based on nonparametric statistics where appropriate.

RESULTS

Population

A total of 699 French hospitals had 9,474,879 hospital discharges related to open surgery during the study period, with an overall in-hospital mortality rate of 1.14%, ranging from 0.05% to 3.69% per institution (Table 1). On the basis of a multivariate analysis using training dataset, there was a reduction in the risk of postoperative death from 2006 through 2010 (supplementary Table, Supplemental Digital Content 1, http://links.lww.com/MLR/A610 that details the estimates for the logistic regression model). Patient characteristics that were independently associated with in-hospital death included male sex, increasing age, and various Elixhauser comorbidities. The more risky procedures were operations on the respiratory (odds ratio 6.2; 95% confidence interval, 5.7–6.7), digestive (4.7, 4.5–4.9), nervous (3.5, 3.3–3.8), or cardiovascular systems (2.1, 2.0–2.2), whereas procedures on the genital organs (0.3, 0.3–0.4), the endocrine system (0.4, 0.3–0.5), or on the ear, nose, mouth, or pharynx (0.6, 0.5–0.7) were associated with a lower risk of death.

TABLE 1. Hospitals and Population Studied for Fiscal Years 2006–2010 in France

Characteristics	
Hospitals (n = 699)	
Public or private non-for-profit hospital [N (%)]	365 (52.2)
Volume of stays per hospital [median (min–max)]	9,803 (1,808–439,097)
Inpatient mortality rate per hospital [median (min–max)] (%)	0.99 (0.05–3.69)
Inpatient stays (n = 9,474,879)	
Women [N (%)]	4,900,064 (51.7)
Age [mean (SD)] (y)	57.5 (18.6)
No. different Elixhauser comorbidities* [mean (SD)]	0.7 (1.1)
No. different anatomical sites operated [mean (SD)]	1.1 (0.3)
Operation on the nervous system [N (%)]	273,416 (2.9)
Operation on the ear, nose, mouth, and pharynx [N (%)]	452,589 (4.8)
Operation on the cardiovascular system [N (%)]	1,393,754 (14.7)
Operation on the hematologic and lymphatic system [N (%)]	284,757 (3.0)
Operation on the respiratory system [N (%)]	118,492 (1.3)
Operation on the digestive system [N (%)]	1,040,122 (11.0)
Operation on the urinary system [N (%)]	199,255 (2.1)
Operation on the male genital organs [N (%)]	459,697 (4.9)
Operation on the female genital organs [N (%)]	504,765 (5.3)
Operation on the endocrine system [N (%)]	242,546 (2.6)
Operation on the musculoskeletal system [N (%)]	3,808,466 (40.2)
Operation on the integumentary system [N (%)]	1,529,182 (16.1)
Inpatient mortality rate [N (%)]	107,644 (1.14)

*Elixhauser comorbidities include congestive heart failure, cardiac arrhythmias, valvular disease, pulmonary circulation disorders, peripheral vascular disorders, hypertension uncomplicated/complicated, paralysis, other neurological disorders, chronic pulmonary disease, diabetes uncomplicated/complicated, hypothyroidism, renal failure, liver disease, peptic ulcer disease excluding bleeding, AIDS/HIV, lymphoma, metastatic cancer, solid tumor without metastasis, rheumatoid arthritis/collagen vascular diseases, coagulopathy, obesity, weight loss, fluid and electrolyte disorders, blood loss anemia, deficiency anemia, alcohol abuse, drug abuse, psychoses, depression.

Mortality Variation

Individual institutional control chart generation from the testing dataset revealed that 52 hospitals (7.4%) detected at least 1 potential safety issue over the 20 successive quarters (Table 2). Among these hospitals, the mean

TABLE 2. Factors Associated With Mortality Variation Over Time Within Hospitals

	Mortality Variation				
	Stable		Safety Issue Occurrence*		
	(n = 647)	(95% CI)	(n = 52)	(95% CI)	P
Small variation signals per hospital on longitudinal control chart, mean					
1 SD lower limit	2.1	(2.0–2.2)	3.5	(2.8–4.2)	< 0.001
1 SD upper limit	2.9	(2.8–3.0)	3.9	(3.6–4.3)	< 0.001
1 SD limits	5.0	(4.8–5.2)	7.4	(6.6–8.3)	< 0.001
Hospital status [N (%)]					0.001
Public or private non-for-profit	351 (54.3)	(50.3–58.1)	14 (26.9)	(15.6–41.0)	
Private for profit	296 (45.7)	(41.9–49.7)	38 (73.1)	(59.0–84.4)	
Hospital volume of surgical cases [N (%)]					0.37
Very low volume	163 (25.2)	(21.9–28.7)	11 (21.2)	(11.1–34.7)	
Low volume	165 (25.5)	(22.2–29.0)	10 (19.2)	(9.6–32.5)	
High volume	157 (24.3)	(21.0–27.8)	18 (34.6)	(22.0–49.1)	
Very high volume	162 (25.0)	(21.7–28.6)	13 (25.0)	(14.0–39.0)	

*One point outside of the upper control limit or 2 out of 3 consecutive points outside of the upper warning limit on control chart.
CI indicates confidence interval.

frequency of safety issues detected per hospital was 1.2 (95% confidence interval, 1.1–1.3). Collectively, 691 hospitals (98.9%) signaled at least once for small mortality variations with a mean number of signals per hospital of 5.2 (5.0–5.3), including 2.2 (2.1–2.3) lower limit signals, and 3.0 (2.9–3.1) upper limit signals (data not shown in the table). The mean number of small variation signals was significantly higher among hospitals that detected a potential safety issue on control charts compared with those that did not (7.4 vs. 5.0 signals, $P<0.001$). Among 23 hospitals characterized by more than 10 small variation signals, 39.1% also detected a safety issue with a substantial increase in mortality temporarily, compared with 11.7% of 257 hospitals having had 6–10 signals and 3.1% of 419 hospitals with less than 5 signals ($P<0.001$, data not shown in the table).

The detection of a safety issue was higher among private hospitals than public hospitals (73.1% vs. 26.9%, $P=0.001$). Conversely, institutional surgical volume was not associated with substantial variation in mortality, as shown by similar rates of safety issue occurrence across hospitals ($P=0.37$).

Mortality Trends

Concomitantly, 119 (17.0%) hospitals reduced and 36 (5.2%) increased their mortality rate over the 5-year study period (Table 3). Hospitals with improved performance demonstrated better control of small mortality variation over time than those with a deteriorating trend (5.2 vs. 6.3 signals, respectively, $P=0.04$). Among hospitals having experienced less than 5 quarters with small variation signals, 82.3% reduced their mortality, compared with 72.3% of hospitals having had 6–10 signals and 40.0% of hospitals with more than 10 signals ($P<0.05$, data not shown in the table).

Moreover, improving hospitals were more likely to be public institutions than private institutions (69.7% vs. 30.3%, $P<0.001$). Institutional surgical volume was not associated with trends in mortality, as shown by similar rates of safety

TABLE 3. Factors Associated With Mortality Trend Over Time Within Hospitals

	Mortality Trend				
	Improvement		Deterioration		
	(n = 119)	(95% CI)	(n = 36)	(95% CI)	P
Small variation signals per hospital on longitudinal control chart, mean					
1 SD lower limit	2.2	(2.0–2.5)	2.9	(2.2–3.5)	0.07
1 SD upper limit	3.0	(2.7–3.2)	3.5	(3.1–3.9)	0.02
1 SD limits	5.2	(4.8–5.6)	6.3	(5.4–7.3)	0.04
Hospital status [N (%)]					< 0.001
Public or private non-for-profit	83 (69.7)	(60.7–77.8)	6 (16.7)	(6.4–32.8)	
Private for profit	36 (30.3)	(22.2–39.4)	30 (83.3)	(67.2–93.6)	
Hospital volume of surgical cases [N (%)]					0.82
Very low volume	28 (23.5)	(16.2–32.2)	10 (27.8)	(14.2–45.2)	
Low volume	29 (24.4)	(17.0–33.1)	10 (27.8)	(14.2–45.2)	
High volume	30 (25.2)	(17.7–34.0)	9 (25.0)	(12.1–42.2)	
Very high volume	32 (26.9)	(19.2–35.8)	7 (19.4)	(8.2–36.0)	

CI indicates confidence interval.

issue occurrence between improving and deteriorating hospitals ($P = 0.82$).

Aggregate Versus Longitudinal Interpretations

Concordance in the interpretation of mortality variation across and within hospitals was limited. A total of 102 hospitals (14.6%) were identified as high performers and 50 (7.2%) as poor performers on traditional funnel plot analyses, whereas 547 (78.3%) hospitals were average performers and plotted within the warning limits (Fig. 1). When comparing outlier detection on aggregated funnel plots against longitudinal control charts, 11.8% (12/102) of hospitals that were identified as high performers on funnel plot also had detected a potential safety issue during the study period compared with 6.6% (36/547) among the average performers and 8.0% (4/50) among the hospitals identified as poor performers ($P = 0.18$). Also, the corresponding number of signals for small mortality variation the 20 successive quarters were, on an average, 5.7 for high performers, 5.1 for average performers, and 4.8 for poor performers ($P = 0.63$). In contrast, 56.5% (13/23) of hospitals that were identified as high performers on funnel plot also improved their mortality during the study period, against 76.4% (84/110) of hospitals plotted within the warning limits and 100% (22/22) for poor performers (Fig. 2, $P < 0.01$).

DISCUSSION

Main Findings and Comparison With Other Studies

Our study yields 3 notable findings. First, the application of industrial quality control principles to health care offers a solution for improving the reliability of surgical processes as evidenced by temporal fluctuations in mortality

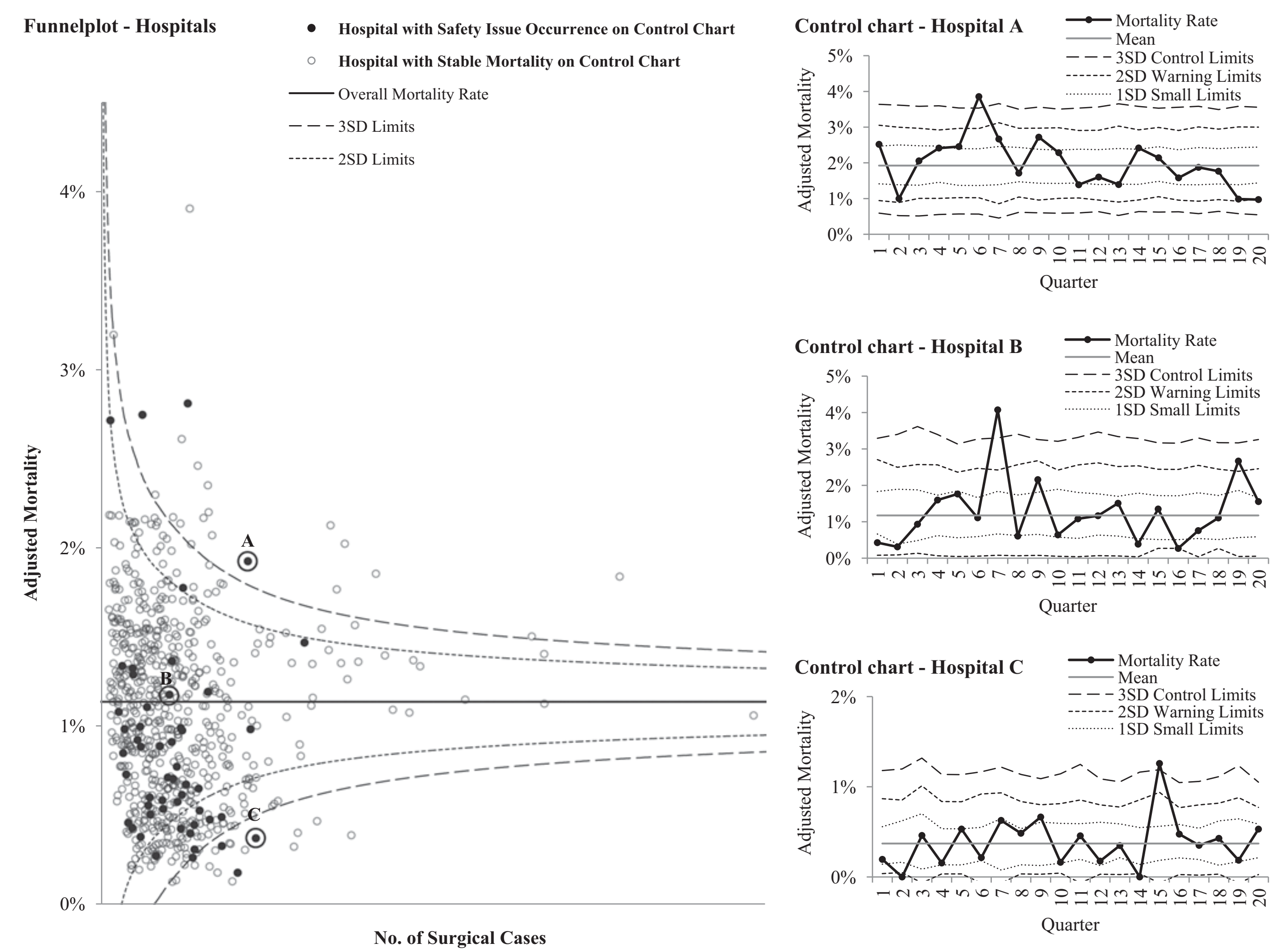


FIGURE 1. Aggregated funnel plot of adjusted mortality rates in French hospitals depending on longitudinal mortality variation (2006–2010). The distribution of surgical mortality rates versus surgical volumes in French hospitals is depicted in a standard funnel plot. Individual representative hospitals are indicated at points A, B, and C. By aggregated approaches, these would be considered poor, average, and good performers, respectively. Complementary utilization of longitudinal control charts, however, demonstrates that each of these 3 institutions has actually detected a potential safety issue during the 5-year period as reflected by substantial increase in mortality rates temporarily.

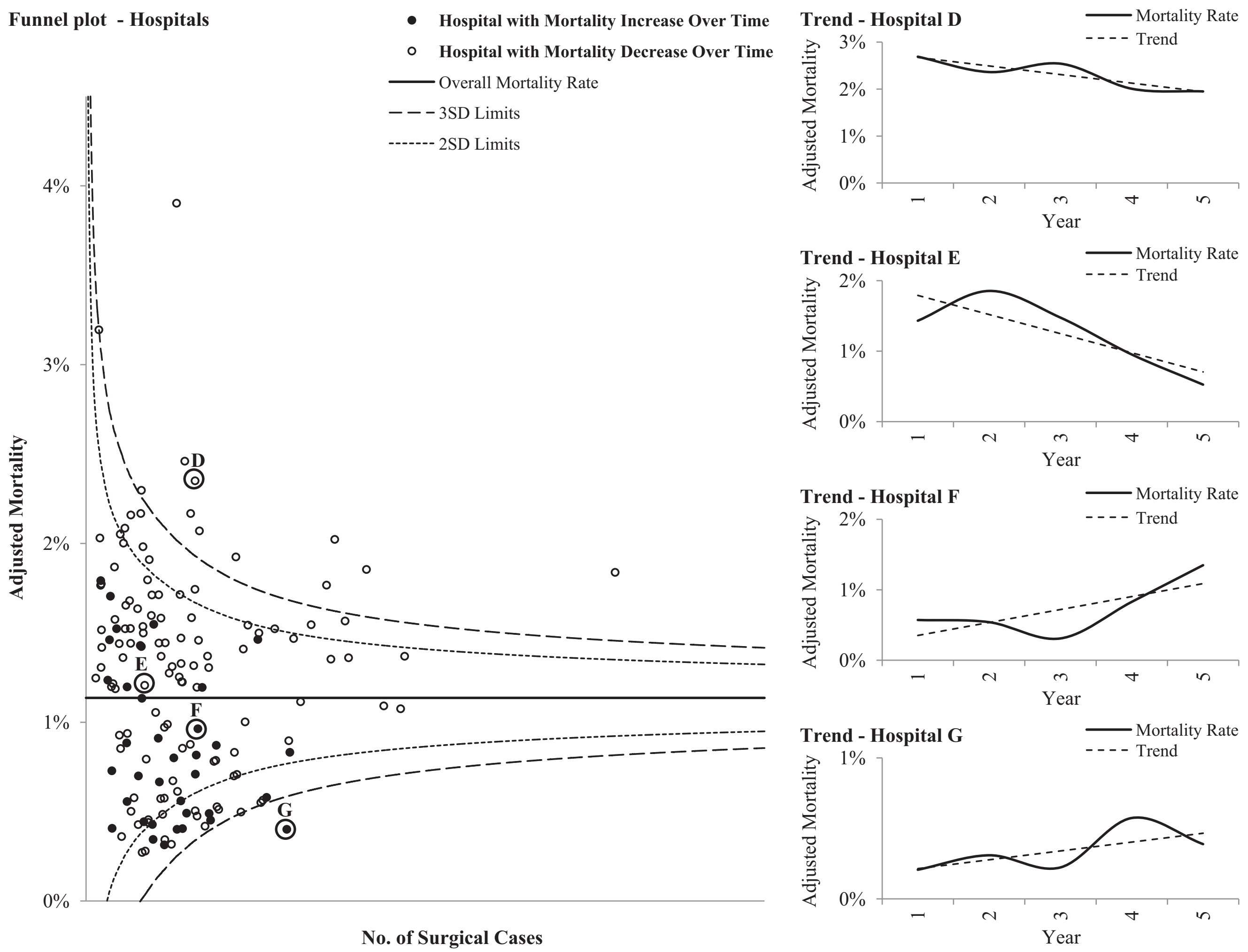


FIGURE 2. Aggregated funnel plot of adjusted mortality rates in French hospitals depending on longitudinal mortality trends (2006–2010). The distribution of surgical mortality rates versus surgical volumes in French hospitals is depicted in a standard funnel plot. Individual representative hospitals are indicated at points D, E, F, and G. By aggregated approaches, these would be considered poor, average, average, and good performers, respectively. Complementary assessment of temporal trends, however, demonstrates that institutions D and E actually manifested a trend toward gradual performance improvement, whereas institutions F and G demonstrated a decline in performance over time.

within individual hospitals. Second, it illustrates the advantages of using dynamic, as opposed to static, analytic methodologies to assess institutional performance monitoring efforts as the latter are unable to capture temporal trends. Finally, it suggests that top performing institutions would have better control of small mortality variations over time, whereas those experiencing the detection of potential safety issues or with a deteriorating trend in performance are characterized by an increased variability in their surgical mortality.

The past decade has been marked by an emphasis in comparing institutions based on their aggregate outcomes or their volume of procedures,[27] with the assumption that these metrics would reflect quality of care and that market competition would lead to improved surgical care.[28] In addition to methodological flaws and lack of evidence for most procedures,[29,30] such assumptions may not apply to the dynamic monitoring of outcomes within institutions. SPC and derived approaches have been previously applied to surgical outcomes monitoring.[9,12,31] Nevertheless, most of these experiences were conducted locally, and no formal investigation of the magnitude of mortality variations over time within hospitals has been published to date at the national level. Our study demonstrates how the depiction of performance data over time allows concerning spikes in mortality to be identified that may be masked by more traditional, static approaches. These findings not only include sentinel spikes in quarterly mortality, but also lower-grade mortality variations that would otherwise be undetectable. Identification of such variations is the cornerstone of the SPC methodology, as correction of unwarranted variability is typically correlated with improved outcomes.[10] Our data provide further evidence supporting this point, as optimally performing hospitals that experienced fewer potential safety issues or a trend toward reduced mortality were also noted to demonstrate the lowest variability in quarterly mortality.

Policy Implications

We have utilized large national databases to elucidate meaningful institution-specific performance analyses. Assuming adequate data integrity and granularity, this study provides a model for efforts to achieve improvements in health care equivalent to those achieved in other industries. It reveals the value of implementing control charts to track temporal variation in surgical mortality. This methodology lays the foundation for prospective monitoring of institutional performance, and can subsequently be used to identify appropriate ways of improving patient safety.[32,33] Iterative assessment permits more ready identification of aberrant patterns of mortality and may therefore trigger more timely investigations and interventions to correct them routinely.[34] Although control limits detect potential alteration in patient safety that should be addressed promptly, the repetition of small variation signals over time period may reflect poor control of surgical processes that could be gradually resolved for avoiding safety issues and achieving better outcomes.

In considering each hospital as its own performance benchmark, control chart provides a means to control for institution-specific confounding variables in a manner not offered by standard comparisons to a sample-wide mean. Furthermore, the depiction of trended data provides an indication as to the performance trajectory of individual institutions compared with traditional cross-sectional benchmarking of institutional outcomes that fails to identify improving or worsening outcomes. One could argue that league tables of hospitals performance should combine temporal variability in their mortality over time based on small variation signals, the detection of potential safety issues with mortality exceeding control limits temporarily, and the trend toward improving or deteriorating mortality.

Limitations of Study

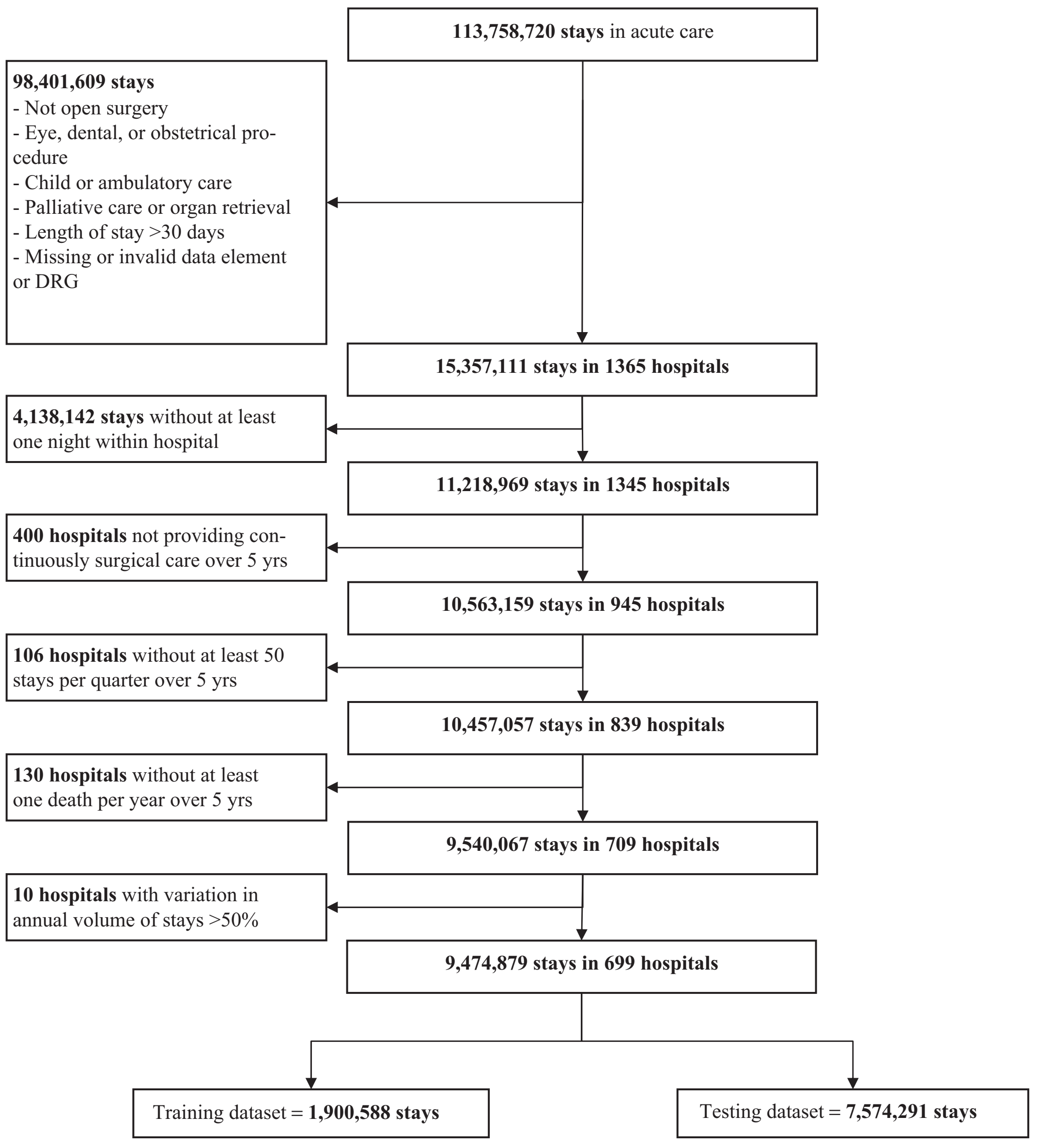
We acknowledge several limitations to our study. First, we were only able to monitor 30-day in-hospital mortality for a broad range of procedures without knowing the prognosis of patients outside hospital or separating preventable from inevitable deaths.[3] This point is critical, as we assumed that hospital-wide mortality was a valid metric of the quality of hospital care.[5] The next step to corroborate our findings would consist of focusing on specific surgical procedures and measurement of related complications. Second, even if the risk of misinterpretation due to variability in data coding and patient case-mix over time within the same hospital may be reduced compared with what is expected from one institution to another,[35] risk adjustment can only account for factors that can be identified and measured accurately.[36,37] Adjusting for surgical complexity was difficult due to a broad range of surgical procedures with a high heterogeneity related to death risk and the fact that some patients may have a combination of procedures during hospitalization. A few deaths clustering together in time due to rarely occurring conditions carrying a high surgical risk and difficult to consider rigorously using our administrative dataset may have resulted in a spike in the temporal control chart. “Do not resuscitate” designations and admitting diagnoses were not available in the French national database for adjusting mortality.[38] Also, we did not consider emergency admissions[6] and were not able to control for providers characteristics such as surgeons’ experience or surgical team familiarity.[39,40] Third, there was a discordance between the within-hospital and between-hospital analyses. Although the between-hospital comparison could be confounded in ways the within-hospital comparison is not, this lack of agreement may raise a potential concern about the method’s internal validity. Fourth, we have retrospectively simulated a prospective monitoring of outcomes, supposing that the detection of a substantial mortality increase on a control chart would reflect a potential safety issue in the past. However, only the investigation and identification of a special cause of variation in the hospital would provide confirmation that a safety issue actually occurred.[24]

CONCLUSIONS

In summary, our study demonstrates the feasibility of a nationwide institutional monitoring of surgical outcomes using existing data resources. This methodology can be reproduced worldwide based on inpatient discharge abstracts that are routinely collected in many countries with a common set of variables.[41] By interpreting variations in surgical outcome within and across hospitals, this approach may provide a more comprehensive assessment of institutional performance and lead to earlier flagging of outliers to make surgical care safer and more reliable.

APPENDIX

Flow diagram of hospitals and surgical cases retained in final dataset.



REFERENCES

1. Duclos A, Carty MJ. Value of health care delivery. *JAMA*. 2011;306:267.
2. Rosen AK, Itani KM, Cevasco M, et al. Validating the patient safety indicators in the Veterans Health Administration: do they accurately identify true safety events? *Med Care*. 2012;50:74–85.
3. Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ*. 2010;340:c2016 doi:10.1136/bmj.c2016.
4. van Gestel YR, Lemmens VE, Lingsma HF, et al. The hospital standardized mortality ratio fallacy: a narrative review. *Med Care*. 2012;50:662–667.
5. Shahian DM, Wolf RE, Iezzoni LI, et al. Variability in the measurement of hospital-wide mortality rates. *N Engl J Med*. 2010;363:2530–2539.
6. Mohammed MA, Deeks JJ, Girling A, et al. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ*. 2009;338:b780 doi: 10.1136/bmj.b780.
7. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med*. 2009;361:1368–1375.
8. Osler M, Iversen LH, Borglykke A, et al. Hospital variation in 30-day mortality after colorectal cancer surgery in Denmark: the contribution of hospital volume and patient characteristics. *Ann Surg*. 2011;253:733–738.
9. Nicolay CR, Purkayastha S, Greenhalgh A, et al. Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare. *Br J Surg*. 2012;99:324–335.
10. Shewhart WA. *Economic Control of Quality of Manufactured Product*. New York: Van Nostrand; 1931.
11. Berwick DM. Controlling variation in health care: a consultation from Walter Shewhart. *Med Care*. 1991;29:1212–1225.
12. Thor J, Lundberg J, Ask J, et al. Application of statistical process control in healthcare improvement: systematic review. *Qual Saf Health Care*. 2007;16:387–399.
13. Institut National de la Statistique et des Etudes Economiques (INSEE). Résultats du recensement de la population 2008: Population par sexe et âge regroupé [National Institute of Statistics and Economic Studies. General population census in 2008]. 2009. Available at: http://www.recensement.insee.fr/tableauxDetailles.action?zoneSearchField=FRANCE&codeZone=1-FE&idTheme=12&idTableauDetaille=48&niveauDetail=1. Accessed December 31, 2012.
14. Agence Technique de l'Information sur l'Hospitalisation (ATIH). Statistiques PMSI: Ensemble des établissements 2008 [Technical Agency of Hospital Information. Inpatient stays statistics in 2008]. 2009. Available at: http://stats.atih.sante.fr/cgi-bin/broker?annee=2008&base=0&typett=1&ok=Lancer+le+traitement&_program=mcoprog.affiche_present.sas&_service=default&_debug=0&tab=1. Accessed December 31, 2012.
15. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York: Springer; 2009.
16. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43:1130–1139.
17. Weiser TG, Semel ME, Simon AE, et al. In-hospital death following inpatient surgical procedures in the United States, 1996-2006. *World J Surg*. 2011;35:1950–1956.
18. Lindenauer PK, Lagu T, Shieh MS, et al. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003-2009. *JAMA*. 2012;307:1405–1413.
19. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73:13–22.
20. Alexandrescu R, Jen MH, Bottle A, et al. Logistic versus hierarchical modeling: an analysis of a statewide inpatient sample. *J Am Coll Surg*. 2011;213:392–401.
21. Kish L. *Survey Sampling*. New York: John Wiley & Sons; 1965.
22. Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med*. 2005;24:1185–1202.
23. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. New York: John Wiley & Sons; 2003.
24. Duclos A, Voirin N. The p-control chart: a tool for care improvement. *Int J Qual Health Care*. 2010;22:402–407.
25. Montgomery DC. *Statistical Quality Control: A Modern Introduction*. Hoboken, NJ: Wiley; 2008.
26. Champ CW, Woodall WH. Exact results for shewhart control charts with supplementary runs rules. *Technometrics*. 1987;29:393–399.
27. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med*. 2002;137:511–520.
28. Mayer EK, Bottle A, Rao C, et al. Funnel plots and their emerging application in surgery. *Ann Surg*. 2009;249:376–383.
29. Shahian DM, Normand SL. The volume-outcome relationship: from Luft to Leapfrog. *Ann Thorac Surg*. 2003;75:1048–1058.
30. 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.
31. Biau DJ, Resche-Rigon M, Godiris-Petit G, et al. Quality control of surgical and interventional procedures: a review of the CUSUM. *Qual Saf Health Care*. 2007;16:203–207.
32. Mohammed MA, Cheng KK, Rouse A, et al. Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet*. 2001;357:463–467.
33. Norberg A, Christopher NC, Ramundo ML, et al. Contamination rates of blood cultures obtained by dedicated phlebotomy vs intravenous catheter. *JAMA*. 2003;289:726–729.
34. Duclos A, Touzet S, Soardo P, et al. Quality monitoring in thyroid surgery using the Shewhart control chart. *Br J Surg*. 2009;96:171–174.
35. Bottle A, Jarman B, Aylin P. Strengths and weaknesses of hospital standardised mortality ratios. *BMJ*. 2010;342:749–753.
36. Mohammed MA, Stevens AJ. A simple insightful approach to investigating a hospital standardised mortality ratio: an illustrative case-study. *PLoS One*. 2013;8:e57845.
37. Lilford R, Mohammed MA, Spiegelhalter D, et al. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet*. 2004;363:1147–1154.
38. Pine M, Jordan HS, Elixhauser A, et al. Enhancement of claims data to improve risk adjustment of hospital mortality. *JAMA*. 2007;297:71–76.
39. Duclos A, Peix JL, Colin C, et al. Influence of experience on performance of individual surgeons in thyroid surgery: prospective cross sectional multicentre study. *BMJ*. 2012;344:d8041 doi: 10.1136/bmj.d8041.
40. Xu R, Carty MJ, Orgill DP, et al. The teaming curve: a longitudinal study of the influence of surgical team familiarity on operative time. *Ann Surg*. 2013. [Epub ahead of print].
41. Organisation for Economic Cooperation and Development (OECD). Health at a Glance 2011: OECD Indicators. 2011. Available at: http://dx.doi.org/10.1787/health_glance-2011-en. Accessed December 31, 2012.