Effect of monitoring surgical outcomes using control charts to reduce major adverse events in patients: cluster randomised trial

Antoine Duclos,[1,2,3] François Chollet,[2] Léa Pascal,[2] Hector Ormando,[4] Matthew J Carty,[3] Stéphanie Polazzi,[1,2] Jean-Christophe Lifante,[1,5] on behalf of the SHEWHART Trial Group

[1]Health Services and Performance Research lab (HeSPeR, EA 7425), Université Claude Bernard Lyon 1, France
[2]Health Data Department, Hospices Civils de Lyon, France
[3]Center for Surgery and Public Health, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[4]Human Resource Department, EM Lyon Business School, France
[5]Department of Digestive and Endocrine Surgery, Hospices Civils de Lyon, France

Correspondence to: A Duclos, Service des Données de Santé des Hospices Civils de Lyon 162 avenue Lacassagne 69424 Lyon cedex 03, France; antoineduclos@yahoo.fr (ORCID 0000-0002-8915-4203)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2020;371:m3840
http://dx.doi.org/10.1136/bmj.m3840

Accepted: 3 September 2020

ABSTRACT

OBJECTIVE

To determine the effect of introducing prospective monitoring of outcomes using control charts and regular feedback on indicators to surgical teams on major adverse events in patients.

DESIGN

National, parallel, cluster randomised trial embedding a difference-in-differences analysis.

SETTING

40 surgical departments of hospitals across France.

PARTICIPANTS

155 362 adults who underwent digestive tract surgery. 20 of the surgical departments were randomised to prospective monitoring of outcomes using control charts with regular feedback on indicators (intervention group) and 20 to usual care only (control group).

INTERVENTIONS

Prospective monitoring of outcomes using control charts, provided in sets quarterly, with regular feedback on indicators (intervention hospitals). To facilitate implementation of the programme, study champion partnerships were established at each site, comprising a surgeon and another member of the surgical team (surgeon, anaesthetist, or nurse), and were trained to conduct team meetings, display posters in operating rooms, maintain a logbook, and devise an improvement plan.

MAIN OUTCOME MEASURES

The primary outcome was a composite of major adverse events (inpatient death, intensive care stay, reoperation, and severe complications) within 30 days after surgery. Changes in surgical outcomes were compared before and after implementation of the programme between intervention and control hospitals, with adjustment for patient mix and clustering.

RESULTS

75 047 patients were analysed in the intervention hospitals (37 579 before and 37 468 after programme implementation) versus 80 315 in the control hospitals (41 548 and 38 767). After introduction of the control chart, the absolute risk of a major adverse event was reduced by 0.9% (95% confidence interval 0.4% to 1.4%) in intervention compared with control hospitals, corresponding to 114 patients (70 to 280) who needed to receive the intervention to prevent one major adverse event. A significant decrease in major adverse events (adjusted ratio of odds ratios 0.89, 95% confidence interval 0.83 to 0.96), patient death (0.84, 0.71 to 0.99), and intensive care stay (0.85, 0.76 to 0.94) was found in intervention compared with control hospitals. The same trend was observed for reoperation (0.91, 0.82 to 1.00), whereas severe complications remained unchanged (0.96, 0.87 to 1.07). Among the intervention hospitals, the effect size was proportional to the degree of control chart implementation witnessed. Highly compliant hospitals experienced a more important reduction in major adverse events (0.84, 0.77 to 0.92), patient death (0.78, 0.63 to 0.97), intensive care stay (0.76, 0.67 to 0.87), and reoperation (0.84, 0.74 to 0.96).

CONCLUSIONS

The implementation of control charts with feedback on indicators to surgical teams was associated with concomitant reductions in major adverse events in patients. Understanding variations in surgical outcomes and how to provide safe surgery is imperative for improvements.

TRIAL REGISTRATION

ClinicalTrials.gov NCT02569450.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Modern surgery still has a high incidence of adverse outcomes, with important consequences for patients

Control charts to monitor outcomes have been implemented in a wide range of settings and specialties, suggesting a broad applicability to healthcare

Tangible evidence of the impact of a nationwide system for monitoring outcomes using control charts to reduce inpatient adverse events occurrence is lacking

WHAT THIS STUDY ADDS

A statistically significant reduction in major adverse events and patient death after surgery was found after implementation of a programme using control charts with regular feedback on indicators to surgical teams

The findings support the routine use of control charts to monitor variations in surgical outcomes over time to help prevent major adverse events

This affordable tool based on commonly available hospital data can be a cornerstone in the continuous improvement of patient safety

Introduction

Healthcare related adverse events are a leading cause of mortality.[1] Worldwide, the large numbers and complexity of surgeries expose patients to a risk of substantial harm.[2] One in 10 patients who undergo surgery experiences a preventable complication.[3] As most cases of surgical morbidity and mortality seem to be avoidable, improving surgical safety is a priority.[4]

Monitoring indicators with the intent of improving surgical outcomes is becoming increasingly important. To provide safer care, surgical teams must engage in quality measurements and understand how their performance fluctuates over time, which requires a strong foundation for fostering discussion on data

and feedback within teams. The ideal method for considering the huge amount of data available would be to allow the identification of safety issues in a timely and accurate fashion. Such a methodology has already been developed within the industrial world. In his theory of variation, the American physicist Walter Shewhart postulated that quality is inversely proportional to variability in production processes and that understanding the variation of some indicators could teach the operator when and how to reduce variation. To categorise variation according to the action needed to reduce it, Shewhart designed a graphical tool known as the control chart.[5] This decision support tool combines a time series analysis with a visual presentation of data, plotting successive indicator measurements in chronological order, with control limits demarcating the expected variations.[6]

By converting data into knowledge, the control chart offers a way to establish prospective outcome monitoring to guide continuous quality improvement initiatives in surgery.[7 8] The chart has been validated through more than 50 years of usage in industry and has proved effective at improving the quality of manufactured products or services.[9] Previous experience suggested its transferability to healthcare[10]; however, rigorously designed studies with a low risk of bias are needed to determine whether this affordable intervention has the potential to benefit surgical care.[11] In this cluster randomised trial, we evaluated the impact of introducing prospective monitoring of outcomes with regular feedback of indicators to surgical teams. We hypothesised that the control chart randomly implemented at the hospital level would reduce the rate of major adverse events (inpatient death, intensive care stay, reoperation, and severe complications) after digestive tract surgery.

Methods

Study design and participants

We prospectively conducted a nationwide parallel cluster randomised trial. After a pre-implementation period of two years (1 January 2014 to 31 December 2015), the surgical departments in 40 participating hospitals were randomised into two cluster groups: intervention and control hospitals. We then introduced the monitoring of surgical outcomes using control charts in hospitals allocated to the intervention group (n=20) over the next two year implementation period (1 January 2017 to 31 December 2018). The control hospitals continued with usual care (n=20). We compared the observed levels in surgical outcomes from the pre-implementation to implementation period between the intervention and control hospitals to determine any improvement attributable to implementation of the control chart.

The study was conducted in the surgical departments of 40 hospitals across France (supplementary figure S1). We screened eligible departments in all public and private hospitals performing digestive tract surgery and combined the criteria related to volume of inpatient stays (≥600 per year), rate of major adverse events (≥3.5%), and data coding quality (≥2 secondary diagnoses on average for each hospital stay). Among 134 eligible surgical departments, we enrolled the first 40 responders. All adults who underwent one of the following operative procedures in those departments were considered for inclusion: hernia repair, cholecystectomy, appendectomy, bariatric, colorectal, hepatopancreatic, or oesophageal and gastric surgery (see operations and procedures codes in supplementary appendix). We excluded patients if they were younger than 18 years, underwent ambulatory surgery, or were admitted for invasive peritoneal cancer, pre-existing adverse events, polytrauma, palliative care, or organ transplantation or retrieval.

The study was conducted according to the study protocol and data were analysed according to the statistical analysis plan (for the protocol see http://shewhart.univ-lyon1.fr).

Randomisation and masking

After the pre-implementation period, the health data department of Hospices Civils de Lyon used a computer generated randomisation schedule to assign 20 hospitals to the intervention group and 20 to the control group. To achieve comparability between the groups, randomisation was stratified on the median proportion of the primary outcome (rate of major adverse events (inpatient death, intensive care stay, reoperation, and severe complications) after digestive tract surgery) recorded in each hospital during the pre-implementation period (≤8.5% *v* >8.5%) with a difference for the overall primary outcome of 0.5% or less, and the number of inpatients between the hospital groups (difference ≤5000 patients).[12] Because this trial concerned an open label intervention and involved local investigators, it was not possible to mask hospital staff, although patients were masked to study group allocation.

Intervention

Hospitals allocated to the intervention group implemented the monitoring of surgical outcomes. A set of Shewhart p-control charts was provided for each operative procedure with indicators of postoperative death, intensive care stay, reoperation, and severe complications. The p-control chart, where p stands for proportion, is useful for the routine monitoring of a binary outcome, such as the occurrence of an adverse event. The data points on the charts depict variation in the indicators for each quarter, and the central line represents the mean indicator value for each hospital. Control and warning limits were set at 99.7% (3 SD) and 95.5% (2 SD) around the central line, respectively, based on binomial distribution.[6] Variation in a special cause was defined as a single point outside the control limits or 2 of 3 successive points outside the warning limits. Therefore, special cause was characterised by substantial changes in patient outcomes (eg, clustering of several complications) caused by unanticipated phenomena within care delivery that deserved further investigation. The control charts were displayed each

doi: 10.1136/bmj.m3840 | *BMJ* 2020;371:m3840 | the bmj

quarter as wall posters in the operating room, and variations in surgical outcomes were discussed during team meetings (see supplementary figure). In cases of worsening outcomes, special attention was paid to the identification and resolution of the causes, and actions for improvements in care were tested and implemented.

To facilitate implementation of the control chart, study champion partnerships were established at each site, comprising a surgeon and another member of the surgical team (surgeon, anaesthetist, or nurse). Each of these duos was responsible for conducting meetings to review the control chart and maintaining a logbook in which changes in care processes were recorded. In addition, the duos from each hospital met during three one day training sessions held at intervals of eight months. Simulated role play and feedback from participants at these sessions were aimed at providing the skills needed to use the control charts appropriately, leading review meetings for effective cooperation and decision making, identifying variations in special causes, and devising plans for improvement.

In parallel, the control hospitals continued with usual care—that is, no specific intervention was implemented within the surgical departments as part of the study.

Intervention is described in more detail in our protocol and in the supplementary material (see components of control chart based programme). A tutorial to reproduce the control chart programme, with slideshows, videos, and logbook is also available (http://shewhart.univ-lyon1.fr) as well as key elements on how to develop and interpret a p-chart for clinical practice and how to successfully integrate this tool within a comprehensive approach (https://academic.oup.com/intqhc/article/22/5/402/1786749).

Outcomes

The primary outcome measure was a composite of major adverse events occurring at any hospital within the 30 days after each surgical procedure. The composite outcome comprised inpatient death, intensive care stay (at least two nights in intensive care or five nights in critical care), reoperation (open or laparoscopic digestive tract procedure), or severe complications (cardiac arrest, pulmonary embolism, sepsis, or surgical site infection). The composite was based on the Clavien-Dindo classification, which ranks a surgical complication in an objective and reproducible manner.[13] For secondary endpoints, we considered each of these outcomes separately. We also assessed the frequency of signal detection related to deterioration or improvement in surgical outcomes on the control charts, and we measured the compliance of each hospital with implementation of the programme based on a previously designed six item scoring method: formation of duo partnerships, participation in training sessions, maintenance of the logbook, display of the poster, meetings of the control chart team, and the implementation of an improvement plan. Compliance with those items was mainly determined by external review of information recorded in the logbook, in addition to pictures of team meetings or posters in operating rooms and formal presentation of the improvement plan by on-site leaders during training sessions.

We obtained data from the French Medical Information System, Programme de Médicalisation des Systèmes d'Information (PMSI). The PMSI is a nationwide database routinely used for care reimbursement and updated weekly with data from all hospitals in France. The data are prospectively collected, and the database relies on a coding system with strict definitions for variables. A subset of records is audited regularly to avoid coding errors. Because of the accuracy and exhaustive data collection of the PMSI database, no patients were lost to follow-up during the study period. Inpatient stays were recorded as standard discharge abstracts containing compulsory information about patients and their primary or secondary diagnoses using ICD-10 (international classification of diseases, 10th revision) codes as well as detailed procedural codes associated with the care provided. From the PMSI database, we extracted the personal characteristics of the patients; comorbidities according to the Elixhauser algorithm, which has acceptable validity[14]; emergency admission; date and type of operative procedure; main diagnosis; complexity of the surgical procedure; and median household income based on residential codes.

Statistical analysis

According to data from the pre-implementation period, a statistical power of 80%, an α value of 0.05, an average of 1978 patients per hospital at each period, 20 hospitals in each group, an intracluster correlation coefficient of 0.056, and observed rate of 10.9% for the primary outcome, we expected a ratio of odds ratios of 0.91 between the intervention and control hospitals from the pre-implementation to implementation period.[15]

For the main analysis, we computed mixed effect logistic regression models to estimate the impact of implementing the control chart on surgical outcomes while accounting for patient clustering within hospitals. Odds ratios were used to compare surgical outcomes between pre-implementation (2014-15) and implementation periods (2017-18) in intervention and control hospitals. Utilising a difference-in-differences approach, we used the interaction between hospital groups and period to estimate the ratio of odds ratios and corresponding 95% confidence intervals that compared changes in outcomes from the pre-implementation to implementation period between intervention and control hospitals. Using estimated parameters obtained from these models and a marginal standardisation method, we determined a difference of absolute risk difference and difference of relative risks difference for every outcome, along with corresponding 95% confidence intervals computed from non-parametric bootstrap based on 1000 replications.[16] The number of avoided cases among

surgical patients in intervention hospitals during the implementation period was estimated from the difference of relative risks difference. The models were adjusted for potential confounders from a patient risk score that predicted the probability of adverse events and was previously developed from patient data derived during the randomisation period (2016). We established a specific risk score separately for each outcome and operative procedure considering all variables extracted from PMSI and hospital status in multivariable logistic regression.

In the secondary analyses, we first split the intervention hospitals into a high compliance group and a moderate to poor compliance group to investigate whether compliance with the control chart implementation programme was associated with better outcomes. Then we evaluated the impact of the intervention on signal frequency for detection of variation in special causes, considering all control charts of participating hospitals whatever the outcome and procedure type. Deterioration or improvement in surgical outcomes was counted as the number of upward or downward signals, and this was modelled using mixed effect Poisson regression models with ratios of rate ratios estimations.

Sensitivity analyses included patients who underwent ambulatory surgery. Operative procedures were stratified into minor (hernia, cholecystectomy, appendectomy, bariatric) and major (colorectal, hepatopancreatic, oesophageal and gastric) surgery. Missing household incomes in the dataset were imputed based on the mean value for each study group and period.

Analysts were not blinded as a result of the modelling of intervention impact according to its level of implementation, which deliberately introduced a distinction between surgical departments among study groups. All reported P values were two sided and we considered a value of less than 0.05 to be significant. Data were analysed using SAS version 9.4 (SAS Institute, Cary, NC).

Patient and public involvement

In this intervention study focusing on healthcare professionals, it was not appropriate or possible to involve patients or the public in the design, conduct, or reporting of our research. However, we disseminated the control chart utilisation, methodology, and results of the research to healthcare professionals and the relevant patient community based on a dedicated website (http://shewhart.univ-lyon1.fr).

Results

Of the 159 688 patients who underwent surgery in the 40 participating hospitals during the pre-implementation and implementation periods, 156 133 (97.8%) were eligible for inclusion in the trial, of whom 155 362 (99.5%) were analysed (fig 1). After the hospitals had been randomised, 75 047 patients were analysed in the intervention hospitals and 80 315 in the control hospitals. Table 1 and supplementary material (tables S1 to S3) present the hospital and patient characteristics between study groups and periods.

Partnership duos were established in all intervention hospitals; of these, 90% (18/20) participated in all training sessions. During the implementation period, each duo reported an average of 20 changes in care process in the logbook, five posters displayed in the operating room, seven team meetings (mean duration, 54 minutes; mean number of participants, 9); 95% (19/20) introduced at least one improvement plan. Nine hospitals were considered highly compliant with implementation of the control chart (implementation score 5-6), nine were considered moderately compliant (score 3-4), and two were considered poorly compliant (score 2) (see supplementary appendix for details).

During the study period, 17 469 (11.2%) patients experienced a major adverse event, 2407 (1.5%) died, 8814 (5.7%) experienced intensive care stay, 7023 (4.5%) underwent reoperation, and 6575 (4.2%) had a severe complication. Table 2 shows changes in surgical outcomes from the pre-implementation to implementation period by hospital group. A significant decrease in major adverse events (adjusted ratio of odds ratios 0.89, 95% confidence interval 0.83 to 0.96; P=0.001), patient death (0.84, 0.71 to 0.99; P=0.04), and intensive care stay (0.85, 0.76 to 0.94; P=0.001) was found in intervention compared with control hospitals. The same trend was observed for reoperation (0.91, 0.82 to 1.00; P=0.06). Those results were even more noticeable in hospitals highly compliant with implementation of a control chart based programme (fig 2), whether for patient major adverse events (0.84, 0.77 to 0.92; P<0.001), death (0.78, 0.63 to 0.97; P=0.02), intensive care stay (0.76, 0.67 to 0.87; P<0.001), or reoperation (0.84, 0.74 to 0.96; P=0.009). Similar results were found in the sensitivity analysis including ambulatory surgery (supplementary table S4 and figure S2) and patients with missing household income (supplementary table S5).

The absolute risk of a major adverse event was reduced by 0.9% (95% confidence interval 0.4% to 1.4%) in intervention hospitals compared with control hospitals, corresponding to 114 patients (70 to 280) who needed to receive the intervention to prevent one major adverse event (supplementary table S6). Also, after implementation of the control chart, 362 (95% confidence interval 141 to 573) major adverse events and 93 (2 to 183) deaths were avoided in the intervention hospitals (supplementary table S7). Supplementary figures S3 to S6 present the rates of surgical outcomes during the pre-implementation and implementation periods by study group and hospital.

The ratio of rates ratios for variations in surgical outcomes on control charts showed a significant reduction in the frequency of deterioration signals (0.60, 95% confidence interval 0.37 to 0.96; P=0.03) and a significant increase in the frequency of improvement signals (3.89, 1.40 to 10.83; P=0.009) within intervention versus control hospitals (fig 3). Among intervention hospitals, a median number of

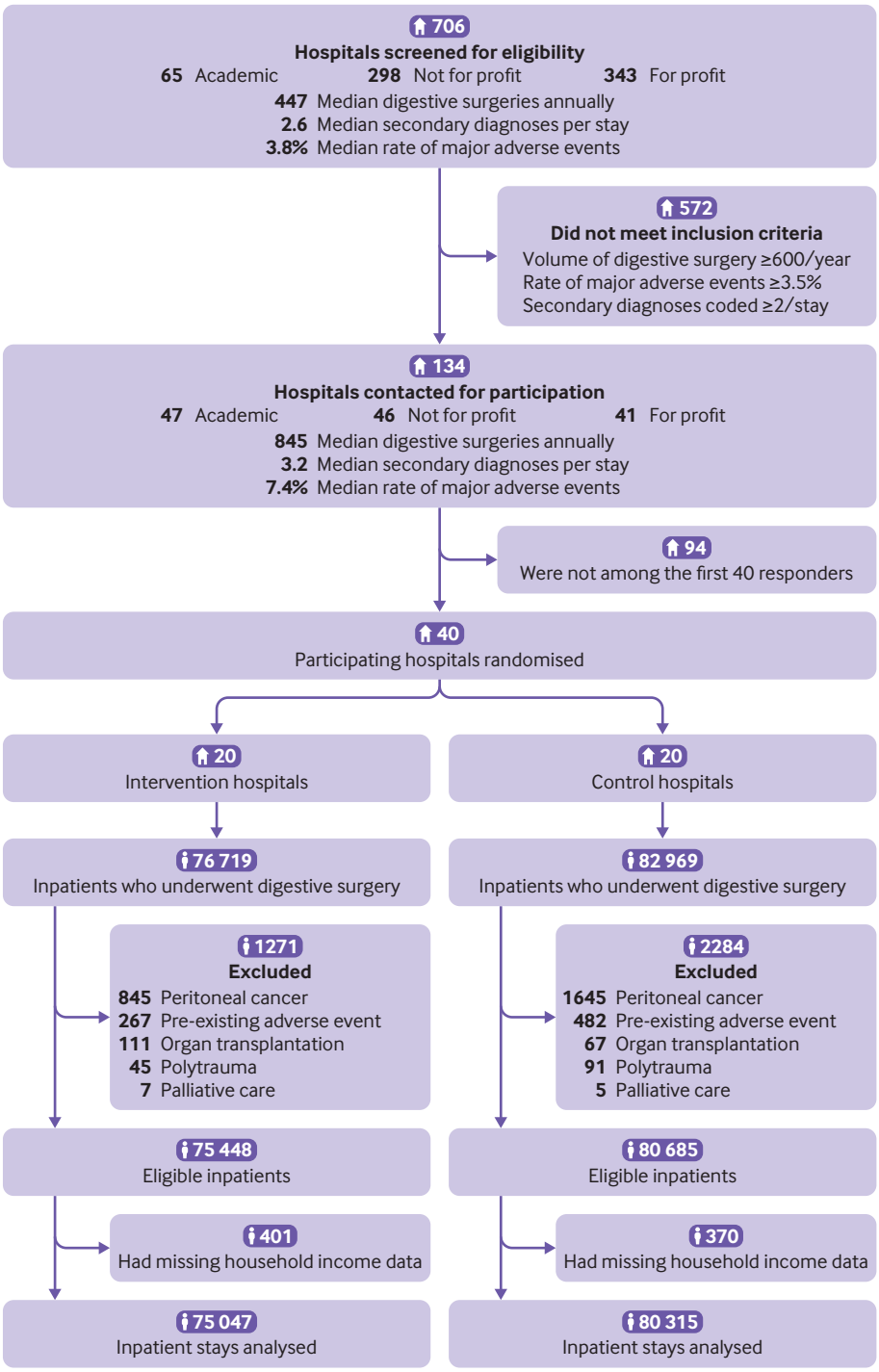

Fig 1 | Trial profile. Of the 155 362 included patients, 79 127 were assigned to the 2014-15 pre-implementation period (37 579 patients in intervention hospitals, 41 548 in control hospitals) and 76 235 to the 2017-18 implementation period (37 468 and 38 767, respectively)

3 (range 1-11) variations in special causes in each hospital were detected over eight consecutive quarters of the implementation period.

Discussion

This national cluster randomised trial showed that introducing prospective monitoring of outcomes with regular feedback of indicators to surgical teams using control charts was associated with a noticeable reduction in major adverse events after digestive tract surgery. The effect size was robust on the sensitivity analyses and even more noticeable in highly compliant hospitals that had implemented control charts. Furthermore, safety issues decreased while improvement signals increased on control charts in the intervention compared with control hospitals. These findings provide evidence that the routine use of control charts has a favourable effect on outcomes in surgical patients.

Table 1 | Hospital and patient characteristics by study group: prospective monitoring of outcomes using control charts with regular feedback on indicators (intervention group) or to usual care (control group). Values are numbers (percentages) unless stated otherwise

Characteristics	Total	Intervention group	Control group
Hospitals	n=40	n=20	n=20
Geographical region:			
South east	12 (30.0)	5 (25.0)	7 (35.0)
North east	9 (22.5)	4 (20.0)	5 (25.0)
Paris area	7 (17.5)	6 (30.0)	1 (5.0)
South west	6 (15.0)	2 (10.0)	4 (20.0)
North west	6 (15.0)	3 (15.0)	3 (15.0)
Status:			
Academic	17 (42.5)	7 (35.0)	10 (50.0)
Not for profit	14 (35.0)	10 (50.0)	4 (20.0)
Private, for profit	9 (22.5)	3 (15.0)	6 (30.0)
Median (range) No of beds	500 (170-1 081)	489 (170-960)	542 (176-1 081)
Median (range) No of surgical beds	146 (31-335)	143 (68-295)	166 (31-335)
Median (range) volume of digestive tract surgery	3697 (2092-7211)	3498 (2579-5705)	3963 (2092-7211)
Median (range) rate of ambulatory procedures	20.8 (7.2-51.2)	21.9 (9.8-51.2)	20.3 (7.2-39.1)
Median (range) No of participating surgeons	6 (2-11)	6 (3-10)	6 (2-11)
Median (range) age of participating surgeons (years)	45 (35-59)	47 (40-53)	44 (35-59)
Patients	n=155 362	n=75 047	n=80 315
Mean (SD) age (years)	56.8 (18.4)	56.6 (18.5)	56.9 (18.2)
Women	81 257 (52.3)	38 853 (51.8)	42 404 (52.8)
Median household income quartiles (€):			
Very low (11 727-18 926)	38 324 (24.7)	14 519 (19.3)	23 805 (29.6)
Low (18 927-20 206)	39 558 (25.5)	15 946 (21.2)	23 612 (29.4)
High (20 209-22 332)	38 776 (25.0)	19 758 (26.3)	19 018 (23.7)
Very high (22 332-43 350)	38 704 (24.9)	24 824 (33.1)	13 880 (17.3)
Elixhauser comorbidities*:			
0	76 652 (49.3)	36 819 (49.1)	39 833 (49.6)
1	35 597 (22.9)	17 067 (22.7)	18 530 (23.1)
2	20 684 (13.3)	9883 (13.2)	10 801 (13.4)
≥3	22 429 (14.4)	11 278 (15.0)	11 151 (13.9)
Emergency admission	36 304 (23.4)	19 757 (26.3)	16 547 (20.6)
Surgical procedure during July/August	21 760 (14.0)	10 522 (14.0)	11 238 (14.0)
Surgical procedure:			
Hernia repair	36 567 (23.5)	17 617 (23.5)	18 950 (23.6)
Colorectal	32 919 (21.2)	15 830 (21.1)	17 089 (21.3)
Cholecystectomy	30 765 (19.8)	14 872 (19.8)	15 893 (19.8)
Bariatric	18 553 (11.9)	9181 (12.2)	9372 (11.7)
Appendectomy	17 572 (11.3)	9718 (12.9)	7854 (9.8)
Hepatopancreatic	10 648 (6.9)	4585 (6.1)	6063 (7.5)
Oesophageal and gastric	8338 (5.4)	3244 (4.3)	5094 (6.3)

€ 1.00 (£0.91; $1.18).
Data relative to pre-implementation period (1 January 2014 to 31 December 2015) and implementation period (1 January 2017 to 31 December 2018) were pooled in intervention and control hospitals. Numbers might not sum to 100 because of rounding. See supplementary table S1 for details about additional patient characteristics by operative procedure.
*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 tumour without metastasis, rheumatoid arthritis/collagen vascular diseases, coagulopathy, obesity, weight loss, fluid and electrolyte disorders, blood loss anaemia, deficiency anaemia, alcohol misuse, drug misuse, psychoses, and depression.

Comparison with other studies

For decades, control charts have been tested to monitor adverse events in a wide range of settings and specialties, suggesting a broad applicability to healthcare.[10] They have also been used for other purposes, such as public health surveillance,[17] evaluating hospital performance,[18] and monitoring individual patient variables,[19] with heterogeneous adherence to methodological principles in their construction.[20 21] Although these quality control tools could be key to enhancing surgical safety synergistically with checklists[22] or process improvement engineering,[23] tangible evidence of the benefits for care management was lacking.[11]

In the broader context of improvement in health services performance based on strategies to feedback on indicators, inconsistent findings resulted from non-randomised studies. Pay-for-performance models that reward or penalise hospitals for meeting predefined indicator targets showed no or at most modest improvement in patient outcomes.[24] Although hospital enrolment in a national surgical outcomes benchmarking programme did not result in performance improvement over time,[25] alerts notification from a surveillance system using control charts might be associated with a reduction in inpatient mortality.[26] A plausible explanation is that cross sectional comparison of aggregated outcomes between

Table 2 | Comparison of primary and secondary outcomes by hospital group between pre-implementation period (2014-15) and implementation period (2017-18) in hospitals assigned to prospective monitoring of outcomes using control charts with regular feedback on indicators (intervention) or to usual care (control). Values are numbers (percentages) unless stated otherwise

	Intervention hospitals			Control hospitals			Intervention *v* control hospitals		
Surgical outcomes	Pre-implementation	Implementation	Implementation *v* pre-implementation: adjusted odds ratio (95% CI)	Pre-implementation	Implementation	Implementation *v* pre-implementation: adjusted odds ratio (95% CI)	Adjusted ratio of odds ratios (95% CI)	P value	Intraclass correlation coefficient
All surgery (n=155 362)									
Major adverse event*	4080 (10.9)	4163 (11.1)	0.95 (0.90 to 1.00)	4582 (11.0)	4644 (12.0)	1.07 (1.02 to 1.13)	0.89 (0.83 to 0.96)	0.001	0.033
Death	584 (1.6)	528 (1.4)	0.85 (0.75 to 0.96)	648 (1.6)	647 (1.7)	1.01 (0.90 to 1.13)	0.84 (0.71 to 0.99)	0.04	0.028
Intensive care stay	2147 (5.7)	2080 (5.6)	0.89 (0.83 to 0.96)	2319 (5.6)	2268 (5.9)	1.05 (0.98 to 1.13)	0.85 (0.76 to 0.94)	0.001	0.151
Reoperation	1584 (4.2)	1682 (4.5)	1.00 (0.93 to 1.07)	1817 (4.4)	1940 (5.0)	1.10 (1.03 to 1.18)	0.91 (0.82 to 1.00)	0.06	0.026
Severe complication	1466 (3.9)	1610 (4.3)	1.04 (0.96 to 1.12)	1687 (4.1)	1812 (4.7)	1.08 (1.01 to 1.16)	0.96 (0.87 to 1.07)	0.46	0.018
Minor surgery† (n=103 457)									
Major adverse event*	1243 (4.8)	1182 (4.7)	0.91 (0.84 to 1.00)	1321 (4.8)	1335 (5.4)	1.10 (1.01 to 1.20)	0.83 (0.74 to 0.94)	0.003	0.031
Death	157 (0.6)	137 (0.5)	0.85 (0.67 to 1.08)	190 (0.7)	186 (0.8)	1.01 (0.82 to 1.25)	0.85 (0.62 to 1.16)	0.30	0.035
Intensive care stay	531 (2.0)	474 (1.9)	0.87 (0.76 to 1.00)	522 (1.9)	485 (2.0)	1.03 (0.90 to 1.19)	0.84 (0.69 to 1.02)	0.08	0.099
Reoperation	575 (2.2)	551 (2.2)	0.91 (0.81 to 1.03)	643 (2.3)	679 (2.8)	1.11 (0.99 to 1.24)	0.82 (0.70 to 0.97)	0.02	0.028
Severe complication	420 (1.6)	412 (1.6)	0.95 (0.82 to 1.09)	472 (1.7)	503 (2.0)	1.13 (0.99 to 1.29)	0.84 (0.69 to 1.02)	0.07	0.027
Major surgery‡ (n=51 905)									
Major adverse event*	2837 (24.8)	2981 (24.4)	0.98 (0.92 to 1.05)	3261 (23.1)	3309 (23.4)	1.05 (0.99 to 1.12)	0.93 (0.85 to 1.02)	0.14	0.046
Death	427 (3.7)	391 (3.2)	0.84 (0.73 to 0.98)	458 (3.2)	461 (3.3)	1.00 (0.87 to 1.15)	0.85 (0.69 to 1.03)	0.10	0.028
Intensive care stay	1616 (14.1)	1606 (13.2)	0.90 (0.83 to 0.98)	1797 (12.7)	1783 (12.6)	1.06 (0.98 to 1.15)	0.85 (0.75 to 0.95)	0.006	0.180
Reoperation	1009 (8.8)	1131 (9.3)	1.05 (0.96 to 1.15)	1174 (8.3)	1261 (8.9)	1.09 (1.00 to 1.19)	0.96 (0.85 to 1.09)	0.57	0.031
Severe complication	1046 (9.1)	1198 (9.8)	1.08 (0.99 to 1.18)	1215 (8.6)	1309 (9.3)	1.06 (0.97 to 1.15)	1.02 (0.90 to 1.15)	0.77	0.022

Odds ratios were estimated using mixed effect logistic regression models to compare surgical outcomes between pre-implementation and implementation periods in intervention and control hospitals. Ratio of odds ratios captured the impact of the control chart by comparing the change in outcomes from the pre-implementation to implementation periods between the intervention and control hospitals based on a difference-in-differences approach. A ratio of odds ratios value less than unity indicated improvement as a result of control charts in intervention versus control hospitals. Estimates with corresponding 95% confidence interval considered clustering of patients at the hospital level. Outcomes were adjusted for a risk score that considered age, sex, presence of comorbidities, emergency admission, date and operative procedure, main diagnosis, complexity of surgical procedure, median household income for patient level covariates, and status for hospital level covariates.
*Composite of surgical complications occurring at any hospital within 30 days after the surgical procedure based on Clavien-Dindo classification, including patient death, intensive care stay (at least two nights in intensive care or five nights in critical care), reoperation (open or laparoscopic digestive tract procedure), or severe complication (cardiac arrest, pulmonary embolism, sepsis, or surgical site infection).
†Hernia repair, cholecystectomy, appendectomy, bariatric surgery.
‡Colorectal, hepatopancreatic, oesophageal and gastric surgery.

institutions is not ideally suited to assess hospital performance prospectively or to highlight the sudden occurrence of improving or worsening outcomes.[7] Instead, performance monitoring using control charts lays the foundation for a more dynamic approach to interpreting variations in surgical outcomes over time.[27] Performance monitoring also permits the identification of appropriate ways to improve patient safety, as iterative assessment more readily permits the identification of aberrant outcome patterns, potentially triggering more timely and routine investigations and interventions to correct them. The consideration of each hospital as its own performance benchmark also reduces methodological concerns about confounding variables that might compromise a fair and transparent interpretation of performance between hospitals.[28 29]

Strengths and limitations of this study

Because the cluster randomisation design aims to reduce contamination, and the same institution (identified as a separated geographical entity) could not commit more than one participating surgical department to the study, we assume the risk of crossover among control hospitals is unlikely. If such a scenario had occurred, the true impact of our intervention might potentially be greater.[30] Although a stepped wedge design or complier average causal effects might represent convenient solutions to reduce crossover, methodological concerns remain for adequately controlling secular trends and heterogeneous duration of exposure to the intervention among clusters.[31]

The comparability between study groups might turn out to be uncertain with a limited number of clusters. By embedding a difference-in-differences analysis within a cluster randomised trial, we sought to balance both observed and unobserved patient and hospital characteristics across study groups and to ensure hospital comparability between the pre-implementation and implementation periods. To account for underlying differences in performance among hospitals, the randomisation was stratified according to baseline postoperative outcome rates. In addition, we used risk adjustment to control for potential differences in patient populations.

Despite the careful study design, several limitations remain. Although the potential influence of a Hawthorne effect on study findings was largely compensated by the presence of a contemporary control group, feedback on indicators in the intervention group might have reinforced vigilance among surgical teams under observation. However, the attention paid to quality improvement rather than the use of charts in themselves could have improved outcomes. Also, owing to potential inaccuracies inherent in medico-administrative data, we cannot rule out the possibility that residual confounders influenced our findings. Of greater concern would be the presence of bias, in which hospital behaviour about coding accuracy of inpatient comorbidities or complications would change between study groups and periods.[32] Because that information is critical for billing purposes, inpatient deaths, intensive

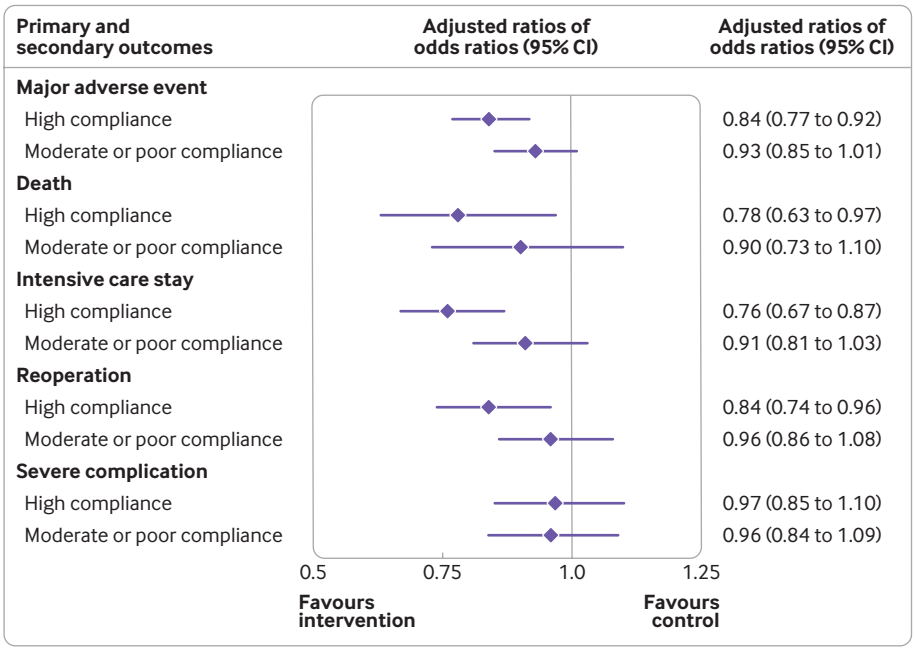

Fig 2 | Primary and secondary outcomes by implementation of control charts. Highly compliant with implementation of control charts=scores 5-6, moderately compliant=scores 3-4, and poorly compliant=score 2. The adjusted ratio of odds ratios (ROR) captured the effect of the control chart on outcomes from the pre-implementation to implementation period between highly compliant intervention and control hospitals, and between moderate to poor compliance intervention and control hospitals. A ROR less than unity indicated an improvement caused by control chart use. Bars denote 95% confidence intervals that considered patient risk score and clustering at the hospital level

care stays, and reoperations are accurately collected in hospitals claims databases. This is not necessarily the case for identifying specific adverse events based on diagnoses codes, the validity of which is debatable and might explain the absence of differences observed for severe complications between hospitals groups. Study outcome measurements within one month after surgery were also limited to occurrences during in-patient stay and might reflect what happened in the operating theatre and the quality of perioperative care, including the success or failure to rescue during a postoperative stay in the intensive care unit.[33] Furthermore, implementing the intervention in a specific country limits the generalisability of our findings. We only included the 40 first responders among eligible French departments of digestive surgery in this study. Thus, the feasibility and impact of implementing the control chart programme in a different context remain unknown.

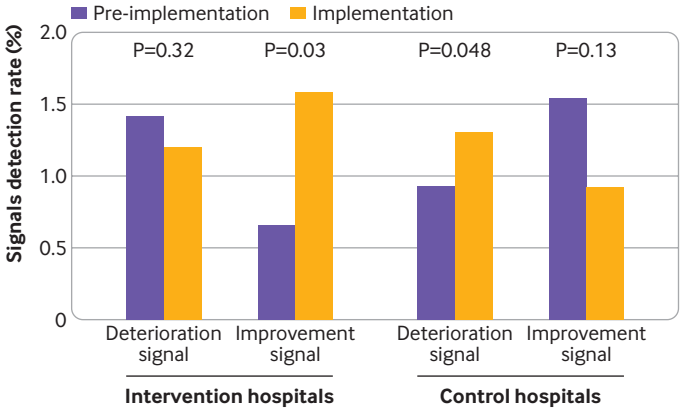

Fig 3 | Signal detection on control charts between hospitals groups. A total of 640 quarters corresponding to 16 quarters for each of the 40 hospitals were included in the analysis. Rates of signal detection were calculated as the total number of signals detected for all quarters divided by the number of interpretable indicator variations on control charts provided for all hospitals. Indicator variation was considered interpretable when the warning or control limits were not equal to 0% or 100%. The signal detection of variation in a special cause was defined as a single point outside the control limits or two of three successive points outside the warning limits. Deterioration and improvement signals were studied separately. Deterioration (improvement) signals were counted as the number of upward (downward) signals regardless of the surgical outcome and operative procedure. P values are for rates ratios estimated using mixed effect Poisson regression models to compare rates of signals between the pre-implementation and implementation periods in intervention and control hospitals

Policy implications and conclusions

Modern surgery still has a high incidence of adverse outcomes, sometimes resulting in important consequences for patients.[3] Understanding variations in surgical outcomes and how to deliver surgery safely is imperative for improvements in this area. It requires the utilisation of a tool for tracking, interpreting, and controlling outcome indicators. The control chart methodology captures knowledge as a product of care and integrates related evidence in the delivery process of a learning health system. Its implementation in routine practice might encourage surgical teams to continuously critically examine the care they deliver. Regular feedback on performance can motivate behavioural and organisational changes, leading to safer surgery. In addition to computerised decision support systems based on clinical practice guidelines, the integration of control charts into electronic health records for triggering monitoring alerts for

surgical outcomes could change the way surgeons manage patients. Considering that the probability of adverse events might vary across patients undergoing a particular surgical procedure, incorporation of risk adjustment or stratification represents an important contribution for further improving the tool's performance to enable correct interpretation of variations in outcome indicators from heterogeneous populations.[34 35] Furthermore, the cumulative sum chart might be helpful in overcoming the limitations related to sensitivity of control charts in routine practice, as it performs fairly well in detecting small changes in surgical outcomes.[36] Instead of indicators measured over a long aggregation period, which can mask and delay reactions to some process changes, the cumulative sum chart allows real time monitoring of activity, procedure after procedure, to detect any defects in surgical safety as soon as possible.[37]

Measurement alone does not, however, result in improvement; rather, the reduction of adverse events relies on dedicated champions, regular interdisciplinary meetings, and the pursuit of every cause in issues regarding patient safety. Despite the satisfactory compliance of participating hospitals with programme implementation, those were especially motivated in following the trial protocol and we cannot exclude potential barriers to widespread dissemination of this intervention. To strengthen compliance in refractory teams, solutions might come from active engagement of surgical staff for leading the process of implementation.[38] To be effective, the implementation of control charts requires constant interactions among healthcare professionals. Beyond enabling awareness of patient outcomes, this quality control tool promotes commitment and communication for sharing better practices. For this reason, our programme was not focused on surgeons alone but aimed to involve medical and non-medical team members in the operating room.

Conclusions

The value of control charts and sharing ideas within surgical teams designed to eliminate patient harm has been mostly underappreciated. In this study, the implementation of a control chart based programme was associated with a statistically significant reduction in major adverse events in patients after surgery. Such a finding shows that prospective monitoring of indicators using available data sources is feasible and can improve surgical outcomes. This methodology can be reproduced worldwide based on inpatient abstracts using a common set of data that are routinely collected in many countries.[39] A method of investigating variations in patient outcomes over time within hospitals, might augment the capacity of surgical teams to improve performance and prevent major adverse events.

We thank the members of the SHEWHART Trial Group.

SHEWHART Trial Group: Françoise Bourgoin, Housseyni Holla, and Sandra Steunou (Agence Technique de l'Information sur l'Hospitalisation); Clotilde Naudot (Assistance Publique Hôpitaux Paris - Louis Mourier); Isabelle Lacombe and Jérémie Lefevre (Assistance Publique Hôpitaux Paris - Saint Antoine); Matthew J Carty (Brigham and Women's Hospital); Jean-Marc Arimont and Charles Foulkes (Centre Hospitalier Alpes-Leman); Patrice David and Laurence Neyer (Centre Hospitalier Colmar - Louis Pasteur); Clément Gayet, Sandrine Hemet, and Loïc Le Menn (Centre Hospitalier Cornouaille - Laennec); Valérie Serra-Maudet (Centre Hospitalier Le Mans); Emeric Abet (Centre Hospitalier La Roche-sur-Yon); Matthieu Poussier (Centre Hospitalier Montélimar); Jérôme Broli, Domenico Papaleo, and Jan Martin Proske (Centre Hospitalier Sud Francilien - Jean Jaurès); Valérie Filippi and Davide Mazza (Centre Hospitalier Toulon - Sainte Musse); Bénédicte Fraleu Louër, Dominique Gratien, and Hélène Poirier (Centre Hospitalier Versailles - André Mignot); Béatrice Alves-Neto, Kévin Fixot and Matthieu Hournau (Centre Hospitalier Régional Metz Thionville - Bel Air); Jean-Marc Regimbeau (Centre Hospitalier Universitaire Amiens Picardie - Nord); Nicolas Bouviez (Centre Hospitalier Universitaire Besançon - Jean Minjoz); Yoann Marion (Centre Hospitalier Universitaire Caen Normandie - Côte de Nacre); Anne Dubois, Sophie Perret-Boire, and Denis Pezet (Centre Hospitalier Universitaire Clermont Ferrand - Estaing); Christophe Mariette (Centre Hospitalier Universitaire Lille - Huriez); Laurent Brunaud and Adeline Germain (Centre Hospitalier Universitaire Nancy - Brabois); Juliette Podevin and Edwige Riegler (Centre Hospitalier Universitaire Nantes - Hôtel Dieu); Tarek Debs (Centre Hospitalier Universitaire Nice - Archet); Riccardo Gauzolino (Centre Hospitalier Universitaire Poitiers - La Miletrie); Reza Kianmanesh (Centre Hospitalier Universitaire Reims - Robert Debré); Amine Brek (Centre Hospitalier Universitaire Saint Etienne - Nord); Sylvain Kirzin (Centre Hospitalier Universitaire Toulouse - Purpan); Benoît Bourdet and Bertrand Suc (Centre Hospitalier Universitaire Toulouse - Rangueil); Dorothée Brachet and Dumitru Cojocarasu (Centre Médico Chirurgical du Mans - Pôle Santé Sud); Philippe Granger (Clinique Belledonne); Serban Bageacu and Michel Bourbon (Clinique Mutualiste Chirurgicale – Saint Etienne); Walter Bertolaso (Clinique Pont de Chaume); Pierre Caillon (Clinique Tonkin); Hector Ormando (EM Lyon Business School); Renato Lupinacci and Olivier Oberlin (Groupe Hospitalier Diaconesses - Croix Saint Simon); Axèle Champault and Monique Sigismond (Groupe Hospitalier Paris - Saint Joseph); Pascal Frileux and Alexandre Rault (Hôpital Foch - Suresnes); Raphaël Bourdariat (Hôpital Privé Jean Mermoz - Lyon); Antoine Lamblin and Christine Leclercq (Hôpital Privé La Louviere - Lille); Bernard Pol (Hôpital Saint Joseph - Marseille); Mathilde Adam, Gilles Poncet, and Catherine Valette-Lagnel (Hospices Civils Lyon - Edouard Herriot); François Chollet, Cyrille Colin, Antoine Duclos, Keitly Mensah, Philippe Michel, Léa Pascal, Cécile Payet, and Stéphanie Polazzi (Hospices Civils Lyon - Lacassagne); Sébastien Couraud, Jean-Christophe Lifante, Guillaume Passot, Jean-Louis Peix, and Vincent Piriou (Hospices Civils Lyon - Lyon Sud); Cédric Beau (Polyclinique Jean Villar - Bruges); Philippe Benand and Benjamin Brugiere (Polyclinique Poitiers); Dine Koriche (Polyclinique Vauban - Valenciennes).

Contributors: AD and JCL conceived and designed the study. MJC, FC, AD, JCL, HO, LP, and SP acquired, analysed, and interpreted the data. FC, AD, LP, and SP drafted the manuscript. MJC, FC, AD, JCL, HO, LP, and SP critically revised the manuscript for important intellectual content. AD and LP did the statistical analysis. AD and JCL obtained funding. AD and SP provided administrative, technical, or material support. FC, AD, JCL, and HO supervised the study. AD is the guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Funding: This study was supported by a grant from the French Ministry of Health and Hospices Civils de Lyon (Programme de Recherche sur la Performance du Système des Soins (PREPS 2013-84)). The funder of the study had no role in considering the study design or in the collection, analysis, and interpretation of the data, writing of the report, or decision to submit the article for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work beyond the grant funding; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: This trial was approved by institutional review board 11 263 (Sud-Est II ethical research committee) and the French Data Protection Authority (CNIL DR-2015-309). The requirement for written informed consent was waived because all data were blinded and made available to investigators from the Secured Data Access

Platform of the French Technical Agency of Hospital Information under formal convention. The leadership of surgical departments provided facility level consent without incentives for participation and permission for trial staff to investigate anonymous data on every eligible patient from their hospital.

Data sharing: Anonymised participant data extracted from the nationwide hospital data warehouse are available from the ATIH Institutional Data Access Platform for researchers who meet the legal and ethical criteria for access to confidential data by the French national commission governing the application of data privacy laws. To obtain this dataset for an international researcher, email demande_base@atih.sante.fr. All trial materials, including the study protocol and statistical analysis plan, are freely available on the trial website (http://shewhart.univ-lyon1.fr).

The manuscript's guarantor (AD) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned and registered have been explained.

Dissemination to participants and related patient and public communities: We plan to disseminate the control chart utilisation, methodology, and results of the research to healthcare professionals and the relevant patient community based on a dedicated website available at http://shewhart.univ-lyon1.fr.

1 Makary MA, Daniel M. Medical error-the third leading cause of death in the US. *BMJ* 2016;353:i2139. doi:10.1136/bmj.i2139
2 Weiser TG, Haynes AB, Molina G, et al. Estimate of the global volume of surgery in 2012: an assessment supporting improved health outcomes. *Lancet* 2015;385(Suppl 2):S11. doi:10.1016/S0140-6736(15)60806-6
3 Panagioti M, Khan K, Keers RN, et al. Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis. *BMJ* 2019;366:l4185. doi:10.1136/bmj.l4185
4 Gawande AA, Thomas EJ, Zinner MJ, Brennan TA. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery* 1999;126:66-75. doi:10.1067/msy.1999.98664
5 Shewhart WA. *Economic Control of Quality of Manufactured Product*. Van Nostrand, 1931.
6 Duclos A, Voirin N. The p-control chart: a tool for care improvement. *Int J Qual Health Care* 2010;22:402-7. doi:10.1093/intqhc/mzq037
7 Mohammed MA, Cheng KK, Rouse A, Marshall T. Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001;357:463-7. doi:10.1016/S0140-6736(00)04019-8
8 Duclos A, Touzet S, Soardo P, Colin C, Peix JL, Lifante JC. Quality monitoring in thyroid surgery using the Shewhart control chart. *Br J Surg* 2009;96:171-4. doi:10.1002/bjs.6418
9 Berwick DM. Controlling variation in health care: a consultation from Walter Shewhart. *Med Care* 1991;29:1212-25. doi:10.1097/00005650-199112000-00004
10 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. doi:10.1136/qshc.2006.022194
11 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;99:324-35. doi:10.1002/bjs.7803
12 Hooper R, Forbes A, Hemming K, Takeda A, Beresford L. Analysis of cluster randomised trials with an assessment of outcome at baseline. *BMJ* 2018;360:k1121. doi:10.1136/bmj.k1121
13 Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 2004;240:205-13. doi:10.1097/01.sla.0000133083.54934.ae
14 Haviari S, Chollet F, Polazzi S, et al. Effect of data validation audit on hospital mortality ranking and pay for performance. *BMJ Qual Saf* 2019;28:459-67. doi:10.1136/bmjqs-2018-008039
15 Preisser JS, Reboussin BA, Song EY, Wolfson M. The importance and role of intracluster correlations in planning cluster trials. *Epidemiology* 2007;18:552-60. doi:10.1097/EDE.0b013e3181200199
16 Austin PC. Absolute risk reductions, relative risks, relative risk reductions, and numbers needed to treat can be obtained from a logistic regression model. *J Clin Epidemiol* 2010;63:2-6. doi:10.1016/j.jclinepi.2008.11.004
17 Hanslik T, Boelle PY, Flahault A. The control chart: an epidemiological tool for public health monitoring. *Public Health* 2001;115:277-81. doi:10.1016/S0033-3506(01)00459-0
18 Woodall WH. The use of control charts in health-care and public-health surveillance. *J Qual Technol* 2006;38:89-104. doi:10.1080/00224065.2006.11918593
19 Tennant R, Mohammed MA, Coleman JJ, Martin U. Monitoring patients using control charts: a systematic review. *Int J Qual Health Care* 2007;19:187-94. doi:10.1093/intqhc/mzm015
20 Koetsier A, van der Veer SN, Jager KJ, Peek N, de Keizer NF. Control charts in healthcare quality improvement. A systematic review on adherence to methodological criteria. *Methods Inf Med* 2012;51:189-98. doi:10.3414/ME11-01-0055
21 Biau DJ, Resche-Rigon M, Godiris-Petit G, Nizard RS, Porcher R. Quality control of surgical and interventional procedures: a review of the CUSUM. *Qual Saf Health Care* 2007;16:203-7. doi:10.1136/qshc.2006.020776
22 Haynes AB, Weiser TG, Berry WR, et al, Safe Surgery Saves Lives Study Group. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med* 2009;360:491-9. doi:10.1056/NEJMsa0810119
23 McCulloch P, Morgan L, New S, et al. Combining Systems and Teamwork Approaches to Enhance the Effectiveness of Safety Improvement Interventions in Surgery: The Safer Delivery of Surgical Services (S3) Program. *Ann Surg* 2017;265:90-6. doi:10.1097/SLA.0000000000001589
24 Mathes T, Pieper D, Morche J, Polus S, Jaschinski T, Eikermann M. Pay for performance for hospitals. *Cochrane Database Syst Rev* 2019;7:CD011156.
25 Etzioni DA, Wasif N, Dueck AC, et al. Association of hospital participation in a surgical outcomes monitoring program with inpatient complications and mortality. *JAMA* 2015;313:505-11. doi:10.1001/jama.2015.90
26 Cecil E, Bottle A, Esmail A, Wilkinson S, Vincent C, Aylin PP. Investigating the association of alerts from a national mortality surveillance system with subsequent hospital mortality in England: an interrupted time series analysis. *BMJ Qual Saf* 2018;27:965-73. doi:10.1136/bmjqs-2017-007495
27 Duclos A, Polazzi S, Lipsitz SR, et al. Temporal variation in surgical mortality within French hospitals. *Med Care* 2013;51:1085-93. doi:10.1097/MLR.0b013e3182a97c54
28 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
29 Shahian DM, Wolf RE, Iezzoni LI, Kirle L, Normand SL. Variability in the measurement of hospital-wide mortality rates. *N Engl J Med* 2010;363:2530-9. doi:10.1056/NEJMsa1006396
30 Keogh-Brown MR, Bachmann MO, Shepstone L, et al. Contamination in trials of educational interventions. *Health Technol Assess* 2007;11:iii, ix-107.
31 Gruber JS, Arnold BF, Reygadas F, Hubbard AE, Colford JMJr. Estimation of treatment efficacy with complier average causal effects (CACE) in a randomized stepped wedge trial. *Am J Epidemiol* 2014;179:1134-42. doi:10.1093/aje/kwu015
32 Clarke GM, Conti S, Wolters AT, Steventon A. Evaluating the impact of healthcare interventions using routine data. *BMJ* 2019;365:l2239. doi:10.1136/bmj.l2239
33 Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med* 2009;361:1368-75. doi:10.1056/NEJMsa0903048
34 Alemi F, Oliver DW. Tutorial on risk-adjusted P-charts. *Qual Manag Health Care* 2001;10:1-9. doi:10.1097/00019514-200110010-00002
35 Duclos A, Voirin N, Touzet S, et al. Crude versus case-mix-adjusted control charts for safety monitoring in thyroid surgery. *Qual Saf Health Care* 2010;19:e17.
36 Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000;1:441-52. doi:10.1093/biostatistics/1.4.441
37 Woodall WH, Fogel SL, Steiner SH. The Monitoring and Improvement of Surgical Outcome Quality. *J Qual Technol* 2015;47:383-99. doi:10.1080/00224065.2015.11918141
38 Gillespie BM, Marshall A. Implementation of safety checklists in surgery: a realist synthesis of evidence. *Implement Sci* 2015;10:137. doi:10.1186/s13012-015-0319-9
39 OECD. (2019), Health at a Glance 2019: OECD Indicators, OECD Publishing, Paris. Accessed November 19, 2019.

Supplementary information: additional material
Supplementary figure: example of a p-chart