OPEN

Effect of the World Health Organization Checklist on Patient Outcomes

A Stepped Wedge Cluster Randomized Controlled Trial

Arvid Steinar Haugen, MSc,† Eirik Søfteland, MD, PhD,* Stian K. Almeland, MD,‡ Nick Sevdalis, PhD,§ Barthold Vonen, MD, PhD,¶ Geir E. Eide, PhD,||** Monica W. Nortvedt, PhD,†† and Stig Harthug, MD, PhD‡‡†*

Objectives: We hypothesized reduction of 30 days' in-hospital morbidity, mortality, and length of stay postimplementation of the World Health Organization's Surgical Safety Checklist (SSC).

Background: Reductions of morbidity and mortality have been reported after SSC implementation in pre-/postdesigned studies without controls. Here, we report a randomized controlled trial of the SSC.

Methods: A stepped wedge cluster randomized controlled trial was conducted in 2 hospitals. We examined effects on in-hospital complications registered by *International Classification of Diseases, Tenth Revision* codes, length of stay, and mortality. The SSC intervention was sequentially rolled out in a random order until all 5 clusters—cardiothoracic, neurosurgery, orthopedic, general, and urologic surgery had received the Checklist. Data were prospectively recorded in control and intervention stages during a 10-month period in 2009–2010.

Results: A total of 2212 control procedures were compared with 2263 SCC procedures. The complication rates decreased from 19.9% to 11.5% ($P <$ 0.001), with absolute risk reduction 8.4 (95% confidence interval, 6.3–10.5) from the control to the SSC stages. Adjusted for possible confounding factors, the SSC effect on complications remained significant with odds ratio 1.95 (95% confidence interval, 1.59–2.40). Mean length of stay decreased by 0.8 days with SCC utilization (95% confidence interval, 0.11–1.43). In-hospital mortality decreased significantly from 1.9% to 0.2% in 1 of the 2 hospitals post-SSC implementation, but the overall reduction (1.6%–1.0%) across hospitals was not significant.

Conclusions: Implementation of the WHO SSC was associated with robust reduction in morbidity and length of in-hospital stay and some reduction in mortality.

Keywords: checklist, morbidity, mortality, randomized controlled trial, surgery

(*Ann Surg* 2015;261:821–828)

As global surgical volume increase and exceed 234 million surgical procedures annually,[1] surgical mortality has declined over the previous decades.[2] Still, crude mortality rates are reported to vary between 0.4% and 4% in high-income countries.[3–5] Increased risk of mortality is associated with major complications in hospitals with higher overall mortality.[6] In-hospital complications occur in 3% to 22% of admitted patients, with 36% to 54% related to surgery.[7–9] Prevention of complications and incidents of iatrogenic harm are deemed feasible for nearly 50% of such incidents.[3,9] Introduction of checklists in surgery can intercept and prevent such incidents[10–12] and may reduce both morbidity and mortality.[13–16]

In 2008, the World Health Organization (WHO) introduced the Surgical Safety Checklist (SSC) designed to improve consistency of care.[17] The pilot pre-/postevaluation of the WHO SSC across 8 countries worldwide, which found reduced morbidity and mortality after SSC implementation,[14] constituted the first scientific evidence of the WHO SSC effects. A number of subsequent studies to date have reported improved patient outcomes with use of checklists.[18] Furthermore, checklists have also been shown to improve communication,[19–22] preparedness,[23] teamwork,[24,25] and safety attitudes[26]—findings that have been corroborated by a recent systematic review.[27]

Although checklists are becoming a standard of care in surgery,[28] the strength of the available evidence has been criticized as being low because of (i) predominantly pre-/postimplementation designs without controls; (ii) lack of evidence on effect on length of stay; and (iii) lack of evidence on any associated cost savings. Randomized controlled trials (RCTs) are required[29]—however, in some countries or settings, they can no longer be carried out, as the WHO SSC has already become national policy (eg, United Kingdom).

We report a stepped wedge cluster RCT aimed to evaluate the impact of the WHO SSC on morbidity, mortality, and length of hospital stay (LOS). We hypothesized a reduction of 30 days' in-hospital morbidity and mortality and subsequent LOS post-Checklist implementation.

METHODS

Study Design

We conducted a stepped wedge cluster randomized controlled checklist intervention trial in 2 hospitals in Norway[30]; a tertiary teaching hospital (1100 beds) and a central community hospital (300 beds). Following the WHO implementation guidelines for the SSC,

From the *Department of Anesthesia and Intensive Care, Haukeland University Hospital, Bergen, Norway; †Department of Clinical Science, Faculty of Medicine and Dentistry, University of Bergen, Bergen, Norway; ‡Department of Surgery, Førde Central Hospital, Førde, Norway; §Centre for Patient Safety and Service Quality at the Department of Surgery and Cancer, Imperial College, London, United Kingdom; ¶Department of Surgery, Nordland Hospital, Bodø, Norway; ||Centre for Clinical Research, Haukeland University Hospital, Bergen, Norway; **Department of Global Public Health and Primary Care, Faculty of Medicine and Dentistry, University of Bergen, Bergen, Norway; ††Centre for Evidence Based Practice, Bergen University College, Bergen, Norway; and ‡‡Department of Research and Development, Haukeland University Hospital, Bergen, Norway.

Disclosure: This study received departmental support. A.S.H. was granted by the Western Regional Norwegian Health Authority (grant numbers 911635 and 911510). N.S. is affiliated with the Imperial Center for Patient Safety and Service Quality, which is funded by the National Institute for Health Research, UK. The funders had no role in the design, conduct, or analysis of this study. The authors report no conflicts of interest.

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 Web site (www.annalsofsurgery.com).

Reprints: Arvid Steinar Haugen, MSc, Department of Anesthesia and Intensive Care, Haukeland University Hospital, Jonas Liesvei 65, N-5021 Bergen, Norway. E-mail: arvid.haugen@helse-bergen.no.

ISSN: 0003-4932/14/26105-0821
DOI: 10.1097/SLA.0000000000000716

we invited 5 surgical specialties to participate; cardiothoracic, neurosurgery, orthopedic, general, and urologic surgical procedures.[31] None declined to participate. These 5 surgical specialties constituted 1 cluster each, in which patients underwent surgical procedures as usual. The order of the checklist intervention was randomized and sequentially rolled out until all 5 clusters had received the intervention. The stepped wedge design is similar to a crossover design in terms that the different clusters (here: surgical specialties) in turn switched treatment from control (no checklist) to intervention (WHO SSC) at different study time points. Furthermore, the clusters crossed over in 1 direction only (ie, introduction of the checklist) and the WHO SSC was not retracted after implementation (Figs. 1 and 2).

This stepped wedge cluster randomized design is particularly relevant where it would be considered unethical to not deliver or retract the intervention when it is predicted to do more good than harm,[30,32] as was the case here for the WHO SSC.[14,16] Furthermore, this design is useful in cases in which randomizing individual surgical procedures to control and intervention (some procedures without the WHO SSC, some with it) would be near-impossible or if possible significantly affected by contamination effects. Finally, the stepped wedge design is considered especially useful when it is not possible to deliver the intervention to all participants (individuals or clusters) at the same time for logistical, practical, or financial reasons.[30] It allows for the clusters to be their own controls and hence provide all data in both control and intervention sections of the SSC stages for comparison, thereby reducing possible contamination biases.[33] One major disadvantage of the stepped wedge design might be that it requires a large amount of data to be collected. Hence, such a drawback could be mitigated by the use of routinely collected data.[33]

This study was reviewed by the Regional Committee for Medical and Health Research Ethics (Ref: 2009/561), which advised that use of routinely collected anonymized patient data is clinical service improvement and thus no further approval or patient consent is required. Hence, the study was approved by hospital privacy Ombudsmen (Ref: 2010/413) and local hospital managers.

Checklist Intervention

The WHO SSC was first adapted to fit into the Norwegian surgical care pathway.[34] The Checklist consisted of 20 items and as per WHO guidelines was performed at 3 critical steps of the surgical procedure: the "sign in" before induction of anesthesia, the "time out" before start of surgery, and the "sign out" before the head surgeon left the operating room.[17]

The Checklist was implemented in clusters of surgical specialties consecutively numbered from 1 to 5. Randomization of the stepped wedge intervention was determined by a draw of 5 numbers into a rank order deciding the sequential roll out of the intervention. The assessor of the randomization (G.E.E.) was blinded regarding which specialty corresponded to which number. After a baseline registration period of 3 months, the SSC was implemented consecutively into orthopedic, cardiothoracic, and neurosurgery at the first hospital and then into general and urologic surgery at the second hospital. These interventions were rolled out with a time period of 3 to 4 weeks between these 5 randomized stepped wedges. The SSC was introduced in all specialties/hospitals while using an educational program with standardized lectures and information materials.[34]

Outcome Measures

The primary endpoints of this study included both major and minor complications and in-hospital mortality up to 30 days after surgery. The complications were all identified and defined using the WHO's *International Classification of Diseases, Tenth Revision* (*ICD-10*) and classified into 11 categories (Supplemental Digital Content 1, available at http://links.lww.com/BRS/A868).[35] Secondary endpoints included patients' LOS.

Data Collection

The study population included all age groups and elective or emergency surgery. Procedures not involving all 3 parts of the checklist (ie, γ knife treatment or donor surgery) were excluded. Patient

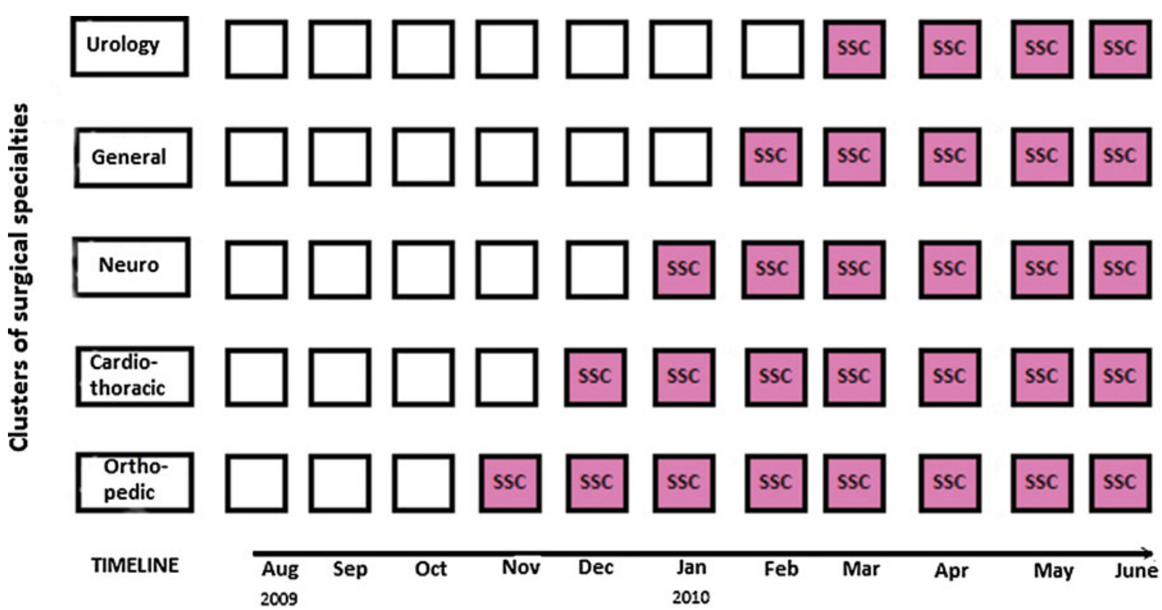

FIGURE 1. Design of the stepped wedge cluster randomized controlled SSC trial in 2 hospitals in western Norway in 2009–2010. Order of the SSC introduction to the clusters was randomized. White box indicates controls with care as usual; colored box, SSC intervention.

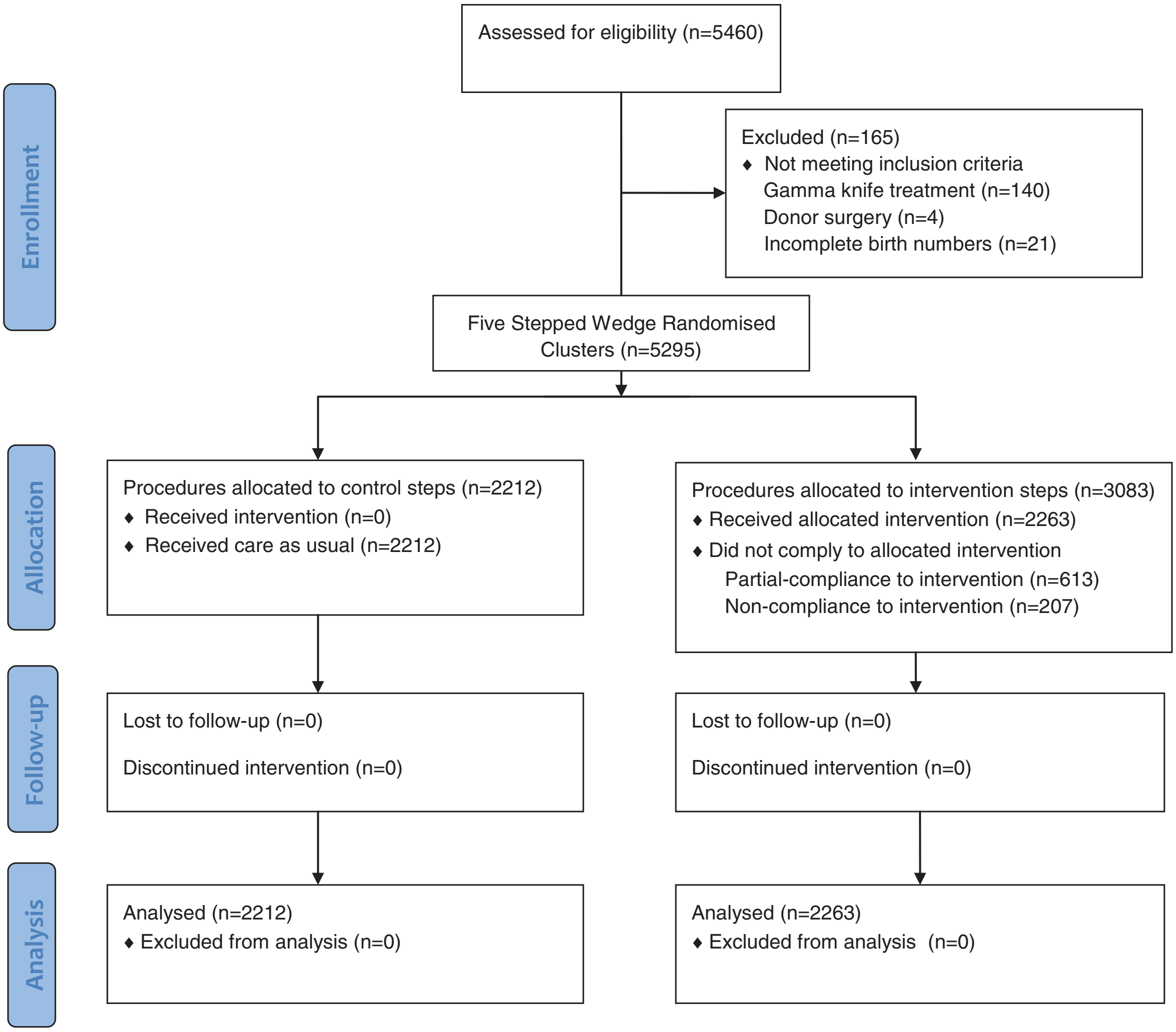

FIGURE 2. CONSORT 2010 flow diagram of the stepped wedge cluster randomized Surgical Safety Checklist intervention trial.

characteristics including age, sex, and comorbidity (American Society of Anesthesiologists score) were obtained from hospital administrative data. Types of surgery, form of anesthesia, and LOS were collected. Data were registered electronically by nurse anesthetists, operating room nurses, anesthetists, and surgeons in the operating room per regular practice (alongside other data). To reduce information bias, the clinicians were not informed as to which endpoints were measured during the study. Compliance with the SSC was assessed by nurse anesthetists and operating room nurses while performing the Checklist. This prospective assessment was performed on a pro forma and also registered in the operating room electronic database. Outcome data on all patients were extracted from the hospitals' administrative databases and collected by research assistants. Mortality was assessed from the public mortality register. All postoperative complications were prospectively assigned *ICD-10* codes by the surgeons or ward doctors as per routine practice in the hospitals at patients' discharge.

Data Handling

All research assistants were blinded to the randomization of patients into intervention and control cohorts when they handled the data and evaluated data quality. Quality check of the extracted data included a random analysis of 360 cases to check whether data concurred with the original patient administrative data. There were discrepancies in 1.7% (6/360) for names, 0.8% (3/360) for procedure and diagnostic codes, and 0.3% (1/360) for operation times and 100% match for remaining data variables. Variation was mainly due to differences between manually registered Checklist data and electronically recorded data, of which the latter was used for analysis. All *ICD-10* codes predefined as complications were identified and coded as complications and quality rechecked against the patients' medical records. The complication codes were checked for relevance to the actual surgical procedure to ensure that they were true complications and not merely an indication for surgery. The in-hospital mortality was associated to the unique patient and coded with "0" if alive at discharge/or after 30 days or "1" if mortality occurred before discharge within 30 days postoperatively.

Data from the postintervention cases (n = 3083) were handled as the total intervention group and included cases with noncompliance (intention to treat), partial compliance, and full compliance to the SSC. To investigate the SSC effects of full compliance, data from these cases (n = 2263) were handled separately in the analysis.

Statistical Analysis

The surgical specialty cohorts acted as their own controls and hence provided data in all steps of the wedge, before and after the Checklist intervention, thus reducing risk of bias. Analyses of stepped wedge RCTs involve comparing all data in the steps before (controls) with the steps after the intervention.[33] Accordingly, all patient characteristics for the controls and postimplementation of the SSC were compared using Pearson exact χ^2 test (categorical data) or with independent samples *t* test (numerical data) as appropriate. Furthermore, calculations on absolute risk reduction and relative risk reduction of morbidity and mortality with 95% confidence intervals (CIs) were based on Pearson exact χ^2 test with Bonferroni correction. For parametric analyses, effect sizes were calculated with η^2 defined as small (0.01), medium (0.06), or large (0.14).[36] Numbers needed to treat were used to calculate the preventive effect of the Checklist.[37] The sample size needed to detect a decrease in the mortality rate (in the first 30 days) from 0.08 to 0.06 at significance level 5% with 91% power was calculated to be 1110 in each group (without/with use of checklist). Intracluster correlation is considered to have minimal effect on power due to the unidirectional stepped wedge implementation of the intervention[30]; hence, it was not calculated. Binary logistic regression was used to adjust the Checklist effect on mortality and morbidity for possible confounding factors. Any difference in complication rates and procedure complexity in each surgical specialty between pre and postintervention was controlled for in the regression model with interactions. "Time" (study time points) was entered into the model to adjust for variation in complication rates at the different study time points throughout, as well as comorbidity (and other risk factors as age, surgical specialty, elective or emergency surgery, and type of anesthesia). Comparison of LOS before and after checklist implementation was tested by independent samples *t* test. For all analyses, a 2-sided *P* value of less than 0.05 was considered statistically significant. Statistical analysis was performed using SPSS version 21.0 (IBM Corp., Armonk, NY).

TABLE 1. Characteristics of Patients in the Stepped Wedge Cluster Randomized Controlled Trial in 2 Hospitals in Western Norway in 2009–2010

	Control	SSC	*P*
All procedures (n)	2212	3083	
Procedures with all parts of checklist used	2212	2263	
>1 procedure (%)	19.6	10.2	**< 0.001**
Unique patients (n)	1778	2033	
Mean age in years (SD)	54.1 ± 23.0	54.3 ± 23.3	0.869
Male sex (%)	55.6	55.9	0.718
Comorbidity by ASA (%)			0.272
ASA I	21.8	24.0	
ASA II	43.3	43.5	
ASA III	31.3	29.6	
ASA IV	3.4	2.9	
ASA V	0.1	0.0	
No ASA score (n)	87	73	
Surgical specialty (%)			**< 0.001**
Orthopedic	32.7	55.3	
Thoracic	13.4	12.5	
Neuro	17.6	9.3	
General*	27.0	16.9	
Urologic	9.3	5.9	
Surgery (%)			**0.001**
Elective	59.0	54.2	
Emergency	41.0	45.8	
Anesthesia			**< 0.001**
Regional	31.8	45.3	
General	68.2	54.7	
Patients by hospitals (%)			**< 0.001**
Tertiary hospital	63.7	77.2	
Central community hospital	36.3	22.8	
Mean length of in-hospital stay (d)	7.8	7.0	**0.022**

P value indicates Pearson exact χ^2 test with Bonferroni correction and independent samples *t* test for continuous variables. Significant values are in bold.
*Includes procedures with gastrointestinal, endocrine, and plastic surgery.
ASA indicates American Society of Anesthesiologists comorbidity/risk score.

RESULTS

A total of 5295 surgical procedures were carried out throughout the stepped wedge cluster RCT, that is, 2212 in control and 3083 (of which 2263 had the SSC performed) after implementation of the SSC. Patients (14.9%; 667/4475) underwent more than 1 procedure. The control and SSC study steps included 1778 and 2033 unique patients, respectively. Characteristics of patients and their distribution across study steps are reported in Table 1. Patients did not differ in sex, age, or comorbidity between the control/SSC stages. However, patients were more likely to undergo orthopedic and emergency surgery and regional anesthesia in the SSC than the control cohort ($P < 0.001$). Compliance with the SSC was 73.4%.

The overall complication rate during the study period was 15.7% and decreased from 19.9% in the control steps to 11.5% in the SSC steps ($P < 0.001$) (Table 2). Absolute risk reduction was 7.5 (95% CI, 5.5–9.5) post-Checklist implementation when all procedures (3083) were included. Absolute risk reduction was 8.4 (95% CI, 6.3–10.5) when all 3 parts of the SSC were used. Relative risk reduction was 0.42 (95% CI, 0.33–0.50). A large effect size by parametric η^2 at 0.14 was found. Number needed to treat (overall Checklist use) in order to prevent one of these complications was 12 (95% CI, 9–16). Using logistic regression, we adjusted the Checklist effect on complications for possible confounding factors including sex, age, comorbidity (American Society of Anesthesiologists score), surgical specialty, urgency of surgery, type of anesthesia, and times (this refers to study time points from August 2009 to June 2010). Even with these adjustments, the Checklist remained significantly related to complication outcome, with OR of 1.95 (95% CI, 1.59–2.40) in the final regression model (Table 3). To control for possible differences in complication rates and complexity within surgical specialties from pre- to postintervention, we adjusted for interactions with time (study time points). In the final step, the Checklist effect remained with OR 1.84 (95% CI, 1.27–2.65). Checklist introduction resulted in significant decreases of complications in 4 of the 5 surgical specialties included but not in general surgery.

The overall in-hospital mortality rate during the whole study period was 1.3% and decreased from 1.6% in the control steps to 1.0% in the steps after SSC implementation ($P = 0.151$). The result did not change after controlling for possible confounders including sex, age, comorbidity (American Society of Anesthesiologists score), surgical specialty, urgency of surgery, type of anesthesia, and times (study time points from August 2009 to June 2010) (Supplemental Digital Content 2, available at http://links.lww.com/BRS/A868). Analysis of mortality by hospital revealed a significant decrease from 1.9% to 0.2% ($P = 0.020$) post-SSC implementation in the smaller community hospital of the study.

Patients' LOS was compared at control and SSC intervention stages of the study. The total in-hospital LOS for both study hospitals was significantly reduced from 7.8 days to 7.0 days after introduction

TABLE 2. Morbidity and Mortality Outcome of the Stepped Wedge Cluster Randomized Controlled Trial in 2 Hospitals in Western Norway in 2009–2010

Procedures (n = 4475) With 1 or More Complications	Control (%)	SSC (%)	ARR Percent Points	ARR 95% CI	*P*
Respiratory complication	**6.4**	**3.2**			**<0.001**
Pneumonia	**3.7**	**1.9**			**<0.001**
Respiratory failure	1.0	0.5			0.062
Other*	**1.8**	**0.8**			**0.008**
Cardiac complication	**6.4**	**4.3**			**0.004**
Cardiac arrest	0.5	0.4			0.644
Arrhythmia	3.3	2.7			0.188
Congestive heart failure	0.7	0.3			0.061
Acute myocardial infarction	1.0	0.5			0.062
Angina pectoris	0.9	0.4			0.058
Infections	**6.0**	**3.4**			**<0.001**
Sepsis	0.6	0.3			0.075
Surgical site	2.2	1.5			0.149
Urinary tract	**2.8**	**1.4**			**0.001**
Other†	0.7	0.3			0.089
Surgical wound rupture	**1.2**	**0.3**			**<0.001**
Nervous system complication‡	0.5	0.3			0.232
Bleeding§	**2.3**	**1.2**			**0.008**
Embolism¶	0.5	0.2			0.092
Mechanical implant complication	**1.1**	**0.4**			**0.005**
Anesthesia complication	0.3	0.2			0.772
All other complications‖	**2.0**	**0.7**			**<0.001**
Unplanned return to operating theatre	**1.7**	**0.6**			**<0.001**
Complications (SSC)	**19.9**	**11.5**	**8.4**	**6.3–10.5**	**<0.001**
Complications in all (n = 2212/3083)	**19.9**	**12.4**	**7.5**	**5.5–9.5**	**<0.001**
In-hospital death (n = 3811)					
Tertiary hospital (n = 2715)	1.4	1.3	0.1	−0.7–1.1	0.865
Central hospital (n = 1083)	**1.9**	**0.2**	**1.7**	**0.4–3.0**	**0.020**
Total deaths	1.6	1.0	0.5	−0.2–1.3	0.151

P value indicates Pearson exact $\chi 2$ test. Significant values are in bold. No interactions between SSC and the other variables in the final model were significant.
*Including asthma, pleura-effusion, and dyspnea.
†Including meningitis, peri- and endocarditis, and gastroenteritis.
‡Including delirium and somnolence.
§Bleeding: A complication to surgical or medical procedures and valid for major or severe acute bleedings associated with the surgical procedure that required erythrocyte transfusions unplanned for and noted in the medical record by the surgeon.
¶Including arterial-, venous-, lung-, and air emboli.
‖Including circulatory collapse, unintended punctures or lacerations, kidney failure, complications after surgical and medical procedures, and complications to surgery not classified.
ARR indicates absolute risk reduction.

of the Checklist, with a mean difference of −0.8 days, $t = 2.30$ (95% CI, 0.11–1.43). Furthermore, there were no significant changes in length of surgery or in total time spent in the operating room.

DISCUSSION

Comparison With Other Studies

To our knowledge, this is the first stepped wedge cluster RCT on the clinical effectiveness of the WHO SSC. The study showed substantial improvements in surgical outcomes. Across 2 hospitals of a well-developed and funded health care system (Norway) including 5 surgical specialties, complication rates fell by 42% on average when the SSC was introduced. The effect was largest when all 3 parts of the SSC were conducted. The effect was significant even when surgical procedures included "intention to treat" with the SSC (in all 3083 surgical procedures postintervention, Table 2). The findings support our hypothesis and are consistent with previous pre-/poststudies having found similar effects of the WHO Checklist use.[14–16] Our results of reduction in morbidity also correspond to findings on use of the comprehensive "surgical patient safety checklist system" (SURPASS) in The Netherlands.[13]

The in-hospital stay decreased significantly in this study by almost a day. This is the first time the WHO SCC is shown to reduce LOS. The finding is consistent with a reduction in LOS by 0.6 days previously obtained after introducing the SURPASS checklist, which, however, did not reach statistical significance.[13] Furthermore, our study reflects similar findings in intensive care units, where LOS has been significantly reduced after use of a daily checklist (goal sheet).[38] LOS reduction provides a potential of significant cost savings in surgical care by improved patient outcome, as costs of complications and unplanned returns to operating room are reduced.[39] Although the WHO SSC was designed for quality improvement, a secondary effect—cost savings—should further encourage health care leaders to adopt and support its use.

After implementation of the WHO SSC, the overall study mortality deceased from 1.6% to 1.0% but did not improve significantly

TABLE 3. Results From Logistic Regression Analyses of Complications on Patient and Treatment Variables in the Stepped Wedge Cluster Randomized Controlled Trial of the WHO SSC in 2 Hospitals in Norway in 2009–2010

		Unadjusted			Fully Adjusted			Final Model*		
	n	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
Absence of the WHO SSC (after = reference)	2212 (2263)	1.91	1.62–2.26	<0.001	2.01	1.40–2.88	<0.001	1.95	1.59–2.40	<0.001
Age	4475	1.03	1.02–1.03	<0.001	1.01	1.01–1.02	<0.001	1.01	1.01–1.02	<0.001
Male sex (female = reference)	2516 (1959)	1.32	1.12–1.55	0.001	1.14	0.93–1.40	0.205	—	—	
Comorbidity by ASA score	4266			<0.001			<0.001			<0.001
Healthy (reference)	977	1			1			1		
Mild systemic disease	1852	3.04	2.06–4.50		2.42	1.59–3.69		2.41	1.59–3.68	
Moderate systemic disease	1298	14.09	9.67–20.53		5.74	3.69–8.94		5.78	3.70–8.93	
Severe systemic disease	135	25.92	15.84–42.41		8.14	4.60–14.40		8.15	4.61–14.41	
Morbid	4	91.55	9.26–905.19		23.71	1.54–365.39		23.98	1.54–373.27	
Surgical specialty	4475			<0.001			<0.001			<0.001
Orthopedic (reference)	1975	1			1			1		
Cardiothoracic	580	8.40	6.75–10.45		8.15	5.56–11.96		8.49	5.82–12.40	
Neuro	600	1.27	0.96–1.68		1.61	1.12–2.33		1.66	1.16–2.38	
General	981	1.24	0.97–1.57		1.61	1.16–2.24		1.64	1.19–2.26	
Urology	339	0.71	0.46–1.09		1.02	0.62–1.68		1.09	0.67–1.77	
Surgery emergency (elective = reference)	2532 (1943)	1.18	1.00–1.39	0.045	3.18	2.45–4.12	<0.001	3.19	2.46–4.13	<0.001
Anesthesia† general (regional = reference)	2528 (1588)	1.78	1.48–2.15	<0.001	1.56	1.19–2.04	0.001	1.55	1.18–2.03	0.002
Time: Study time points	4475	0.90	0.88–0.93	<0.001	1.01	0.94–1.07	0.884	—	—	

P values in the regression model are based on the likelihood ratio test.
*Backward stepwise selection from the fully adjusted model at $P < 0.05$, with SSC entered into all models. No interactions between SSC and the other variables in the final model were significant.
†General anesthesia included patients induced with anesthesia requiring respirator support through laryngeal mask or endotracheal tubes. Regional anesthesia included patients anesthetized through epidural-, spinal-, or plexus anesthesia. Combinations of regional and general anesthesia were classified as general anesthesia.
ASA indicates American Society of Anesthesiologists; OR, odds ratio; WHO SSC, World Health Organization Surgical Safety Checklist.

($P = 0.151$). However, we observed a highly significant reduction of mortality from 1.9% to 0.2% ($P = 0.02$) in the smaller community hospital (albeit on fewer cases due to small hospital size), with a relative risk reduction of 91%. The Checklist effect on mortality was thus present but weaker in our RCT than in previous reports from pre-/postintervention studies.[13–16]

Strengths and Limitations of This Study

In our view, this study's major contribution to our better understanding of Checklist effects lies in its stepped wedge cluster RCT design. Such designs have been considered unfeasible because in countries such as the United Kingdom, the WHO SSC is now national policy (and hence a control arm is not available) and also due to contamination and biases resulting from "control" operating room teams treating control patients as patients assigned to the checklist arm.[13] However, such contaminations and biases were minimized by randomization of the study clusters in "stepped wedges."[33] Each cluster acted as its own control and hence provided data in both the control and SSC stages, comparable with a crossover design, with all data being compared between the control and SSC stages. To reduce uncertainty of variation in surgical procedure complication rates and complexity within each cluster from pre- to postintervention, we adjusted for possible risk factors as age, sex, comorbidity, surgical specialty, emergency or planned surgery, type of anesthesia, and time (study time points). The stepped wedges provided the possibility to control complication and morbidity for time effects during the study period. Complications rates varied at different study time points but when controlled for, time was not a confounding factor for the Checklist effect on complications (Table 3). The stepped wedge cluster RCT design is considered particularly appropriate for studying patient safety interventions.[30,32] To control for leakage and possible contamination of surgeons between the 2 hospitals and the 5 surgical specialties, we did not include the same surgical specialty in both hospitals. The SSC was first introduced to the intervention groups. Hence, any possible contamination would have leaked from the intervention group to improve care in the controls, eventually. The results do not suggest that this was apparent.

The degree of blinding is important for the validity of RCTs, and in our study, operating room staff were not informed of the study outcomes, as they routinely registered the patient data on the electronic data system of their operating rooms. To further prevent information bias, the outcome assessors were masked to which cohort (control and SSC stages) patients were enrolled. Furthermore, to reduce the risk of performance and information bias, all recovery and ward staff carried out care as usual and were blinded to the study cohorts and outcomes, following the extended CONSORT statement for nonpharmacological randomized trials.[40]

Our study has several limitations. First, the clusters that had not yet received the intervention could have been contaminated by possible enthusiasm for the SSC from colleagues in other specialties that were in the SSC study stage. Such bias would have likely minimized any positive effects of the Checklist. The substantial and robust decrease of complications that we found suggests that such bias did not affect the study significantly. A second limitation is the way in which the data were registered. A selection of *ICD-10* codes was used to identify complications. It is possible that surgeons and ward doctors reported the *ICD-10* codes variably. As far as we could account for, there were no changes in the *ICD-10* code implemented during the study period. Furthermore, variable recording would equally affect the control and the SSC stages of the study. A final limitation is that recording of complications was confined to the in-hospital admission period. Data on complications after discharge were not recorded or obtained. The total number of postoperative complications could, therefore, be higher. A more extensive follow-up of the patients after discharge would be beneficial in future studies, though costly.

Further Research

Further research should investigate how use of the SSC and other checklists achieves its positive impact on patient outcomes. Improved outcomes post–checklist implementation have been explained by improvements in communication and teamwork in the operating room[27] and a wider improvement in safety attitudes.[20,22,24–26] In a concurrent with this study evaluation of the impact of the introduction of the WHO SSC on patient safety climate in operating rooms, we did not find the hypothesized improvement in culture—although we did find that operating room teams reported being better able to handle a complex situation when the Checklist is used.[34] We also anecdotally observed that the introduction of the WHO SSC drove behavior change, as the team members paused, introduced themselves, and carried out team briefings prior to the operative list. Such behavioral changes may precede deeper changes in organizational safety culture—which may in turn underline the sustainability of long-term appropriate implementation of a checklist and improved patient outcomes. These questions require longitudinal controlled research designs to be addressed.

CONCLUSIONS

This stepped wedge cluster RCT adds to this growing body of evidence on the positive effects on patient outcomes driven by the WHO SSC. We conclude that the use of the WHO Checklist prevents complications and reduces in-hospital length of stay and potentially also mortality across a wide range of patients undergoing simple or complex surgical procedures in hospitals within a well-developed and funded health care system.

ACKNOWLEDGMENTS

The authors thank the Norwegian Knowledge Center for Patient Safety for collaboration with translating the WHO Surgical Safety Checklist. The study was endorsed by the Norwegian National Knowledge Center for Patient Safety and the Patient Safety Office of the World Health Organization.

REFERENCES

1. Weiser TG, Regenbogen SE, Thompson KD, et al. An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet.* 2008;372:139–144.
2. Finks JF, Osborne NH, Birkmeyer JD. Trends in hospital volume and operative mortality for high-risk surgery. *N Engl J Med.* 2011;364:2128–2137.
3. Gawande AA, Thomas EJ, Zinner MJ, et al. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery.* 1999;126:66–75.
4. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl J Med.* 2009;360:1418–1428.
5. Pearse RM, Moreno RP, Bauer P, et al. Mortality after surgery in Europe: a 7-day cohort study. *Lancet.* 2013;380:1059–1065.
6. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med.* 2009;361:1368–1375.
7. Davies P, Lay-Yee R, Briant R, et al. Adverse events in New Zealand public hospitals I: occurrence and impact. *N Z Med J.* 2002;15:U271.
8. de Vries EN, Ramrattan MA, Smorenburg SM, et al. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care.* 2008;17:216–223.
9. Kable AK, Gibberd RW, Spigelman AD. Adverse events in surgical patients in Australia. *Int J Qual Health Care.* 2002;14:269–276.
10. Arriaga AF, Bader AM, Wong JM, et al. Simulation-based trial of surgical-crisis checklists. *N Engl J Med.* 2013;368:246–253.
11. Bliss LA, Ross-Richardson CB, Sanzari LJ, et al. Thirty-day outcomes support implementation of a surgical safety checklist. *J Am Coll Surg.* 2012;215:766–776.
12. de Vries EN, Prins HA, Bennink MC, et al. Nature and timing of incidents intercepted by the SURPASS checklist in surgical patients. *BMJ Qual Saf.* 2012;21:503–508.
13. de Vries EN, Prins HA, Crolla RMPH, et al. Effect of a comprehensive surgical safety system on patient outcomes. *N Engl J Med.* 2010;363:1928–1937.
14. Haynes A, Weiser T, Berry W, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med.* 2009;360:491–499.
15. van Klei WA, Hoff RG, van Aarnhem EEHL, et al. Effects of the introduction of the WHO "Surgical Safety Checklist" on in-hospital mortality: a cohort study. *Ann Surg.* 2012;255:44–49.
16. Weiser TG, Haynes AB, Dziekan G, et al. Effect of a 19-item surgical safety checklist during urgent operations in a global patient population. *Ann Surg.* 2010;251:976–980.
17. Weiser TG, Haynes AB, Lashoher A, et al. Perspectives in quality: designing the WHO Surgical Safety Checklist. *Int J Qual Health Care.* 2010;22:365–370.
18. Borchard A, Schwappach DLB, Barbir A, et al. A systematic review of the effectiveness, compliance, and critical factors for implementation of safety checklists in surgery. *Ann Surg.* 2012;256:925–933.
19. Fudickar A HK, Wiltfang J, Bein B. The effect of the WHO Surgical Safety Checklist on complication rate and communication. *Dtsch Arztebl Int.* 2012;109:6.
20. Kearns RJ, Uppal V, Bonner J, et al. The introduction of a surgical safety checklist in a tertiary referral obstetric centre. *BMJ Qual Saf.* 2011;20:818–822.
21. Nilsson L, Lindberget O, Gupta A, et al. Implementing a pre-operative checklist to increase patient safety: a 1-year follow-up of personnel attitudes. *Acta Anaesthesiol Scand.* 2010;54:176–182.
22. Takala RSK, Pauniaho SL, Kotkansalo A, et al. A pilot study of the implementation of WHO Surgical Checklist in Finland: improvements in activities and communication. *Acta Anaesthesiol Scand.* 2011;55:1206–1214.
23. Böhmer AB, Kindermann P, Schwanke U, et al. Long-term effects of a perioperative safety checklist from the viewpoint of personnel. *Acta Anaesthesiol Scand.* 2013;57:150–157.
24. Böhmer AB, Wappler F, Tinschmann T, et al. The implementation of a perioperative checklist increases patients' perioperative safety and staff satisfaction. *Acta Anaesthesiol Scand.* 2012;56:332–338.
25. Helmiö P, Blomgren K, Takala A, et al. Towards better patient safety: WHO Surgical Safety Checklist in otorhinolaryngology. *Clin Otolaryngol.* 2011;36:242–247.
26. Haynes AB, Weiser TG, Berry WR, et al. Changes in safety attitude and relationship to decreased postoperative morbidity and mortality following implementation of a checklist-based surgical safety intervention. *BMJ Qual Saf.* 2011;20:102–107.
27. Russ S, Rout S, Sevdalis N, et al. Do safety checklists improve teamwork and communication in the operating room? A systematic review. *Ann of Surg.* 2013;258:856–871.
28. Birkmeyer JD. Strategies for improving surgical quality—checklists and beyond. *N Engl J Med.* 2010;363:1963–1965.
29. Lauvrak V, Jeppesen E, Krogstad U. Method alert for the WHO Surgical Safety Checklist. Norwegian Knowledge Centre for the Health Services Web site. Available at: http://www.kunnskapssenteret.no/publikasjoner/who-sjekkliste-for-trygg-kirurgi. Published 2010. Accessed February 4, 2014.
30. Brown C, Lilford R. The stepped wedge trial design: a systematic review. *BMC Med Res Methodol.* 2006;6:54.
31. World Health Organization. WHO Safe Surgery Saves Lives Web site. Available at: http://www.who.int/patientsafety/safesurgery/en. Published 2009. Accessed June 17, 2013.
32. Mdege ND, Man MS, Taylor CA, et al. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *J Clin Epidemiol.* 2011;64:936–948.
33. Brown C, Hofer T, Johal A, et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 2. Study design. *Qual Saf Health Care.* 2008;17:163–169.
34. Haugen AS, Søfteland E, Eide GE, et al. Impact of the World Health Organization's Surgical Safety Checklist on safety culture in the operating theatre: a controlled intervention study. *Br J Anaesth.* 2013;110:807–815.
35. World Health Organization. International Classification of Diseases (ICD) Web site. Available at: http://www.who.int/classifications/icd/en/. Published 2010. Accessed June 17, 2013.

36. Cohen J. *Statistical Power Analysis for the Behavioural Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
37. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med.* 1988;318:1728–1733.
38. Pronovost P, Berenholz S, Dorman T, et al. Improving communication in the ICU using daily goals. *J Crit Care.* 2003;18:71–75.
39. Semel ME, Resch S, Haynes AB, et al. Adopting a surgical safety checklist could save money and improve the quality of care in U.S. hospitals. *Health Aff (Millwood).* 2010;29:1593–1599.
40. Boutron I, Moher D, Altman DG, et al. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med.* 2008;148:295–309.