World J Surg (2014) 38:287–295
DOI 10.1007/s00268-013-2300-6

Understanding WHO Surgical Checklist Implementation: Tricks and Pitfalls. An Observational Study

Pedro J. Saturno · Víctor Soria-Aledo · Zenewton A. Da Silva Gama · Felipe Lorca-Parra · Marina Grau-Polan

Published online: 19 October 2013

Abstract
Background The purpose of the present study was to assess the reliability of implementation data regarding the surgical safety checklist (SSC) and to identify which factors influence actual implementation.
Methods The study was a retrospective record-based evaluation in a regional network of nine Spanish hospitals, combined with a complementary direct-observation study that included a survey of the surgical teams' attitudes. SSC compliance and associated factors were assessed and compared in a retrospective sample of 280 operations and a concurrent sample of another 85 surgical interventions.
Results In the retrospective evaluation the SSC was present in 83.1 % of cases, fully completed in 28.4 %, with 69.3 % of all possible items checked. The concurrent direct-observation study showed that recorded compliance was unreliable ($\kappa < 0.13$ for all items) and significantly higher ($p < 0.001$) than actual compliance. Over-registration occurred across hospitals and surgical specialties. Factors associated with recorded compliance included hospital size, surgical specialty, and the use of an electronic format. In actual (direct-observation) compliance, a positive attitude on the part of the surgeon is an overriding significant factor (OR 12.8), along with using the electronic format, which is consistently and positively associated with recorded compliance but negatively related to actual compliance.
Conclusions Recorded SSC compliance may be widely unreliable and higher than actual compliance, particularly when recording is facilitated by using an electronic format. A positive attitude on the part of the surgical team, particularly surgeons, is associated with actual compliance. Effective use of the SSC is a far more complex adaptive process than the usual mandatory strategy.

Funding for this study was provided by an Agreement between the Spanish Ministry of Health and the Murcia Region to Foster Safe Practices in Health Services, project FR05IPS119.

P. J. Saturno (✉) · Z. A. Da Silva Gama · F. Lorca-Parra · M. Grau-Polan
Unidad docente de Medicina Preventiva y Salud Pública, Facultad de Medicina, Universidad de Murcia, 30100 Murcia, Spain
e-mail: psaturno@um.es

V. Soria-Aledo
Universidad de Hospital General Universitario Morales Meseguer, Avenida Marqués de los Vélez, 3008 Murcia, Spain

Introduction

Surgical care is ubiquitous in health systems worldwide, with an estimated 234 million major surgical procedures performed annually [1]. Although undeniably preventing loss of life and limb, it requires complex coordination of different health professionals and technologies and is associated with nearly two-thirds of in-hospital adverse events [2]. In addition, at least half of surgical complications seem to be avoidable [3]. Surgical checklists have been proposed to improve this situation and increase the safety of surgical care. A number of these have been developed [4–7], but the most widely known and implemented is the WHO surgical safety checklist (SSC) [8]. This intervention was developed within the safe surgery saves lives challenge, along with an implementation manual [9], aiming to improve surgical care safety around the world by ensuring adherence to proven standards of care. In studies conducted to test the SSC [3, 10], compliance with standards improved and complications from surgery decreased, with significant results overall and in three of the eight pilot hospitals where it was evaluated. Accordingly, the SSC has since been adopted by thousands of hospitals in several countries around the world [11], and made mandatory in some. However, its usefulness for developed countries has been questioned [12], given that other studies have shown mixed outcomes [13–15]. Moreover, it has been suggested that improved outcomes are related to actual compliance with the SSC [14–16], thereby promoting the gathering of evidence on the characteristics and factors influencing successful implementation strategies [16, 17]. However, published studies may be difficult to compare because the type of data sources used to assess compliance (recorded, observation, mixed) may influence results and professional behavior [12, 15].

The present study describes the effectiveness of the SSC implementation strategy adopted by a regional network of nine public hospitals in Spain, and analyzes the influence of structural and socio-professional variables, comparing results from record-based and direct-observation data sources. The main objective of the study was to better understand which factors should be taken into account in order to reliably assess and improve the implementation process.

Materials and methods

SSC implementation strategy

The Murcia Region Health Authority implemented the SSC in all nine of its public hospitals, and discussed the main components of the implementation strategy with management and surgical services representatives. The eventual strategy closely followed WHO recommendations [9], and included information and training sessions in all hospitals (first with middle management, nursing directors, medical-surgery directors, chiefs of surgery services, and operating room coordinators, and then within each service and surgical unit), where the objective, contents, roles, responsibilities, and practicalities of the SSC were explained. Suggestions for the implementation process were gathered and discussed. Hospitals and services were allowed to adapt the SSC, provided changes did not affect the main contents. A number of hospitals introduced the SSC in electronic format. An indicator of SSC compliance was included in the annual contract between the Regional Authority and the hospitals in order to assess accountability.

Study design

Two complementary studies were designed. First, SSC compliance and associated structural and socio-professional factors were assessed in a retrospective random sample of 280 surgical interventions obtained from the previous month's interventions. The sample was non-proportionally stratified by hospital and common surgical specialties (general, gynecology, ophthalmology, and traumatology), selecting a minimum of 10 interventions per specialty within each center. Minor surgery, and cardiac surgery were excluded, as in the SSC pilot study [3]. Gynecology was present in two of the nine hospitals; ophthalmology, in eight. Therefore total cases for these two specialties were 20 and 80, respectively, and 90 for each of the other specialties. Data were abstracted from medical and operating room records. Second, a concurrent direct-observation study of 85 interventions was performed immediately after the retrospective study to confirm the correctness of SSC application and the reliability of recorded data, in addition to reassessing compliance-related factors. This direct-observation sample was also non-proportionally stratified by hospital and surgical specialty, focusing on the "time out" portion of the SSC, and including an interview with the surgical team to survey their attitudes towards the SSC. Direct-observation and recorded data were independently gathered for the same interventions. "Time-out" was chosen for observation both because of feasibility issues and because most critical safety items are checked here along with confirmation of most "sign-in" items. The attitude survey consisted of two questions summarizing the main aspects that could influence implementation, as follows: "Do you think using the SSC is a waste of time? and "Do you think using the SSC improves safety? Both questions had three response options (yes, no, and neutral).

Data gathering in all studies was performed by two trained nurses from another hospital. Inter-observer reliability was assessed using the kappa coefficient for both recorded and observational data. The κ coefficient was >0.85 for all SSC items for the recorded data, and 1 for all items in direct-observation data after disagreements in the first sample were discussed. In the direct-observation study, auditors were introduced to the surgical team and the objectives of the study were explained before performance was observed.

Study variables

SSC compliance was assessed overall and by item. Overall compliance was measured by (1) presence of the SSC in the medical record, (2) percentage of cases in compliance with all SSC items (all explicitly checked), and (3) percentage of items checked, a less restrictive composite indicator that may more accurately demonstrate existing variability, calculated by dividing the sum of items in compliance in a sample of cases by the total number of items assessed in that sample. Compliance with SSC items was measured individually and grouped by SSC parts ("Sign-in," "Time-out," and "Sign-out"). The same approach, but limited to the "Time-out" phase, was used for the concurrent study. The variables that explained the variation in compliance were the hospital (H1–H9); hospital size (small: <150 beds; medium: 150–400 beds; large: >400 beds); surgical specialty (general, gynecology, ophthalmology, traumatology); type of anesthesia (local, regional, general); shift (regular: morning; complementary: afternoon/evening); electronic format (yes, no) and the age and gender of the patient. In the direct-observation study, the age and gender of health personnel (nurses, surgeons, and anesthesiologists) and their attitude toward the SSC were included in data analysis. The variable "attitude" for each type of professional (surgeon, anesthesiologist, nurse) and the team as a whole, was transformed for analysis into a composite indicator jointly considering both items of the questionnaire as either "positive" (positive answer to the question "It improves safety" and negative to "It is a waste of time") or "neutral or negative."

Statistical analysis

Point estimates of SSC compliance and 95 % confidence intervals were calculated using the appropriate formula for non-proportional stratified samples in both the retrospective and direct-observation studies. Estimates from the retrospective and concurrent record-based data were compared to determine the effect of direct observation on data recording. Record-based and direct-observation compliance data in the concurrent study were compared with the κ coefficient and the z-value for the differences between both data sources. The factors associated with compliance were analyzed for both the retrospective and direct-observation studies by estimating adjusted odds ratios with stepwise multivariable logistic regression models, where the composite indicators "Compliance with all items" (total and for a group of items) were the dependent variables. Two models were tested, because data from the retrospective study did not include the professionals' attitudes. To achieve a parsimonious model for the concurrent study, the age and gender of the patient (both nonsignificant in the retrospective study and in bivariate analysis in the concurrent study) and of the surgical team (nonsignificant in bivariate analysis in the concurrent study) were not included. The association between health professionals' attitude and the measures of compliance in direct-observation data was analyzed with the χ^2 test (for the measure "% of all time-out items checked") and the t test (for the averages of "Proportion of items checked").

Results

Recorded SSC compliance in the retrospective study: the apparent reality

The SSC was included in medical records in 83.1 % of cases, fully completed in 28.4 and 69.3 % of all possible items were checked (Table 1). There were no significant differences in full completion by group of items, although the "Time-out" group exhibited a higher percentage than the other two groups (51.8 % full completion vs 49.3 for "Sign-in" and 43.1 % for "Sign-out"). Compliance varied somewhat by single items, but no clear priorities for improvement were identified. The lowest compliance figures corresponded to items in the "Sign-out" group ("confirmation of specimen labeling": 55.7 %; and of "equipment problems": 58.4 %), followed by the display of essential imaging (60.6 %) in the "Time-out" group, and marking of the intervention site (63.1 %) in the "Sign-in" group. Compliance, however, was not homogeneous among hospitals. Four institutions included the SSC in the medical records of all interventions, and in four others it was missing in over 20 % of cases. Even greater variation was observed in relation to full compliance (all items), ranging from 5.4 to 81.8 %, and percentage of items completed (from 33.3 to 97.4 %). The two hospitals using the SSC in electronic format ranked among the highest in full compliance, one of which ranked first in both composite measures.

Surgical safety checklist compliance in the direct observation study: a different reality

Data from the concurrent study, both record-based and direct-observation, differ significantly from the record-based retrospective study (Table 2). Record-based SSC compliance is higher in the concurrent study than the retrospective study, and direct-observation compliance is lower than recorded in both studies. Recorded estimates of full "Time-out" compliance was 51.8 % (95 % CI 45.9–57.6) in the retrospective study, and 68.2 % (95 % CI 58.3–78.1) in the concurrent study, more than twice as high as that obtained in the direct-observation study (24.7 %, 95 % CI 15–34.4). Recorded data are not reliable when compared to those obtained by direct observation (κ coefficient is <0.13 for all items). Recorded compliance was significantly higher ($p < 0.001$) (Table 2), the only exception being the display of essential imaging (less recorded than performed) and the introduction of team members (equal estimates according to recorded and observed data, but without case-by-case agreement). The apparent (recorded) high compliance in checking anticipated critical events and correct antibiotic prophylaxis does not correspond to observed reality. These items are actually checked in <50 % of cases. Over-registration of compliance in the concurrent study occurred across all hospitals (with only one exception) and surgical specialties (Table 3). Hospitals using electronic format are among the lowest in actual compliance (direct-observation data), yet one ranks first (with 100 % compliance) according to

Table 1 Estimates of checklist use based on clinical record data ($n = 280$)

Measures	Estimate (95 % CI)
% Checklist present in the clinical record	83.1 (78.7–87.5)
% Checklist fully checked	28.4 (23.1–33.7)
Percentage of items checked[a]	69.3 (68.1–70.6)
% Sign in fully checked	49.3 (43.5–55.2)
% Item 1 checked: Patient has confirmed identity, site, procedure, and consent	77.2 (72.2–82.1)
% Item 2 checked: site marked/not applicable	63.1 (57.5–68.7)
% Item 3 checked : anesthesia safety check completed	69.6 (64.2–75.0)
% Item 4 checked: pulse oximeter on patient and functioning	78.8 (74.0–83.5)
% Item 5 checked: known allergy?	78.7 (73.9–83.6)
% Item 6 checked: difficult airway/aspiration risk?	67.8 (62.3–73.3)
% Item 7 checked: risk of hemorrhage and IVs planned	66.6 (61.1–72.1)
% Time out fully checked	51.8 (45.9–57.6)
% Item 8 checked: confirm all team members have introduced themselves by name and role	76.3 (71.3–81.3)
%Item 9 checked: surgeon, anesthesia professional, and nurse verbally confirm patient's name, site, and procedure.	74.0 (68.9–79.1)
% Item 10 checked: antibiotic prophylaxis	73.3 (68.1–78.5)
% Item 11 checked: surgeon reviews Anticipated Critical Events	69.3 (63.9–74.6)
% Item 12 checked: anesthetist reviews Anticipated Critical Events	66.9 (61.5–72.4)
% Item 13 checked: nursing reviews Anticipated Critical Events	78.6 (73.9–83.4)
% Item 14 checked: essential imaging displayed	60.6 (55.2–66.1)
% Sign out fully checked	43.1 (37.4–48.9)
% Item 15 checked: nurse confirms the name of the procedure recorded	77.7 (72.8–82.5)
% Item 16 checked: instrument, sponge, and needle counts	66.2 (60.7–71.6)
% Item 17 checked: specimen is labeled and includes patient's name	55.7 (50.0–61.4)
% Item 18 checked: equipment problems to be addressed	58.4 (52.7–64.0)
% Item 19 checked: key concerns for the patient's recovery and management	67.2 (61.8–72.6)

CI confidence interval

[a] $n = 5{,}320$ items

record-based data. Direct-observation data illustrate priority items for intervention, clearly demonstrating the checking of anticipated critical events, followed by the correct administration of antibiotic prophylaxis.

Factors associated with SSC compliance

The variable most consistently associated with SSC compliance in recorded data is the use of electronic format (Table 4), the only significant variable in both the retrospective and concurrent study for full completion of the SSC (adjusted OR: 11.8 and 10.6, $p < 0.001$ and < 0.01, R^2 of the models 0.29 and 0.35, respectively). The "Sign-out" group of items tends to be less often recorded in interventions with regional anesthesia (adjusted OR: 0.4, 95 % CI 0.2–0.7, $p < 0.01$), and significantly higher when using electronic format (adjusted OR:13.0, 95 % CI 6.1–27.8), once again the only significant factor explaining the full recording of "Sign-in" items (adjusted OR: 5.9, 95 % CI 3.0–11.8). Recorded completion of "Time-out" items is significantly associated with hospital size (more so in small hospitals), shift (more in the morning shift), and—only in the concurrent study—surgical specialty (traumatology and gynecology interventions had estimated adjusted OR of 0.2 and 0.1, respectively, in relation to general surgery). However, according to analysis of direct-observation data, where health professionals' attitude is included in the model, completion is more frequent (adjusted OR: 12.8, 95 % CI 1.4–118.6) when the surgeon's attitude is positive, and 53 % of completion variability is explained by this variable and two other factors (electronic format and hospital size), with no variability for electronic format or cases from the large hospitals (no cases using electronic format or large hospitals completing all "Time-out" items). The influence of health professionals' attitude was consistently positive for both compliance measures ("% of time-out items checked" and "proportion of items checked" Table 5), albeit statistically significant only for surgeons, physicians, and the team as a whole. Surgeons' attitude was more strongly associated with SSC compliance than the attitudes of the rest of the team, and it was the only attitudinal variable included in the final model of logistic multivariate analysis.

Table 2 Clinical record-based and direct observation estimates of "time-out" compliance differences and concordance

measures	n[a]	(1) Clinical Record (95 % CI)	(2) Direct Observation (95 % CI)	Difference (1)–(2) (p value)	Concordance κ
% of Items checked	595	88.7 (87.4–90.0)	59.5 (57.5–61.5)	29.2 (< *0.001*)	–
% Full time-out checked	85	68.2 (58.3–78.1)	24.7 (15.5–33.9)	43.5 (< *0.001*)	0.11
% Team members introduced	85	95.3 (90.8–99.8)	95.3 (90.8–99.8)	0	–0.05
% Patient's name, procedure, and incision site confirmed	85	88.2 (82.0–94.4)	55.3 (44.7–65.9)	32.9 (< *0.001*)	0.13
% Correct timing (or not applicable) antibiotic prophylaxis confirmed	85	91.8 (86.0–97.6)	50.6 (40.0–61.2)	41.2 (< *0.001*)	0.07
% Surgeon anticipated critical events checked	85	89.4 (82.9–95.9)	40.0 (29.6–50.4)	49.4 (< *0.001*)	0.11
% Anesthetist anticipated critical events checked	85	87.1 (80.0–94.2)	40.0 (29.6–50.4)	47.1 (< *0.001*)	0.10
% Nursing anticipated critical events checked eventcríticos enfermería	85	95.3 (90.8–99.8)	42.4 (31.9–52.9)	52.9 (< *0.001*)	0.03
% Essential imaging displayed (or not) applicable)	85	74.1 (64.8–83.4)	92.9 (87.5–98.3)	–18.8 (< *0.001*)	0.04

[a] *n* represents the denominator for the measure, which is the total number of surgical safety checklist (SSC) items (595) or total interventions observed (85)

Table 3 Percentage of "time-out" items checked by hospitals and surgical specialty according to clinical records and direct observation data ($n = 85$)

	Total items	Clinical record (1) (95 % CI)	Observation (2)(95 % CI)	Difference (1)−(2)	p value
Hospital[a]					
H5[b]	63	100.0 (100.0–100.0)	38.1 (32.0–44.2)	61.9	<0.001
H3	63	96.8 (94.6–99.0)	39.7 (33.5–45.8)	57.1	<0.001
H9[b]	63	63.5 (57.4–69.6)	44.4 (38.2–50.7)	19.1	0.016
H7	84	91.7 (88.7–94.7)	45.2 (39.8–50.7)	46.5	< 0.001
H1	63	92.1 (88.7–95.5)	55.6 (49.3–61.8)	36.5	<0.001
H2	63	85.7 (81.3–90.1)	65.1 (59.1–71.1)	20.6	0.004
H4	63	90.5 (86.8–94.2)	71.4 (65.7–77.1)	19.1	0.003
H6	70	92.9 (89.8–95.9)	80.0 (75.2–84.8)	12.9	0.013
H8	63	84.1 (79.5–88.7)	98.4 (96.8–100.0)	−14.3	0.002
Surgical specialty					
Traumatology	189	77.2 (71.3–83.2)	52.9 (45.8–60.0)	24.3	<0.001
General surgery	189	93.7 (90.2–97.1)	59.3 (52.3–66.3)	34.4	<0.001
Gynecology	42	83.3 (72.1–94.6)	64.3 (49.8–78.8)	19.0	0.024
Ophthalmology	150	97.1 (94.7–99.6)	65.7 (58.7–72.7)	31.4	<0.001
Total	595	88.7 (87.4–90.0)	59.5 (57.5–61.5)	29.2	<0.001

CI confidence interval

[a] Ranked by ascendant % items checked according to direct observation

[b] Hospital with electronic format for SSC

Discussion

All hospitals in our study have included the SSC in their routine practices, following a regional strategy containing all the main ingredients recommended in the WHO Implementation Manual [9]. These include training, middle management commitment, meetings with health professionals where roles and responsibilities were specified, and adaptation of the SSC for ownership by each hospital. Eventually, the SSC was made mandatory, as in many

Table 4 Factors associated with compliance with the safe surgery checklist

Dependent variable	Initial model	Final model/significant independent variables	Recorded data (retrospective study $n = 280$) or (95 % CI)	Recorded data (concurrent study $n = 85$) or (95 % CI)	Direct observation (concurrent study, $n = 85$) OR (95 % CI)
Checklist fully checked	M1	Electronic format	11.8 (6.1–22.7)***	10.6 (2.8–41.7)**	NA
		R^2	0.29	0.35	
"Sign in" fully checked	M1	Electronic format	5.9 (3.0–11.8)*** 0.14	NA	
"Time out" fully checked	M1	Electronic format	–	–	0.0[a]
		Shift (morning)	2.3 (1.2–4.3)*	13.7 (1.2–142.9)*	–
		Hospital size	0.4(0.2–0.6)***	–	0.15 (0.04–0.5)
		Medium	0.2(0.1–0.5)***		0.0[a]
		Big			
		Surgical specialty	–	0.2 (0.1–0.9)*	–
		Traumatology		0.1 (0.01–0.7)*	
		Gynecology			
		R^2	0.12	0.25	0.44
	M2	Electronic format	NA	–	0.0[a]
		Shift (morning)		13.7(1.2–142.9)*	
		Hospital size		–	0.14 (0.04–
		Medium			0.54)**
		Big			0.0[a]
		Attitude of surgeon (positive)			12.8 (1.4–118.6)*
		Surgical specialty		0.2(0.1–0.9)*	
		Traumatology		0.1(0.01–0.7)*	
		Gynecology			
		R^2		0.25	0.53
"Sign out" fully checked	M1	Electronic format	13.0 (6.1–27.8)***	NA	NA
		Type of anesthesia (regional)	0.4 (0.2–0.7)**		
		R^2	0.29		

M1 initial model with the following independent variables: hospital size, electronic format, surgical specialty, shift, type of anesthesia, and gender, and age of patient, *M2* initial model with the independent variables: hospital size, electronic format, surgical specialty, shift, type of anesthesia, and health professionals' attitude, *NA* not applicable; – indicates the datum was nonsignificant ($p < 0.05$)

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

[a] Variables in the model with no variability: no case using electronic format, or in big hospital had all the "time out" items completed in direct observation data

other countries and regions, and annual compliance assessment was established. A number of hospitals use the electronic format in their system, while others use a paper-based operating room and medical records. However, in spite of this uniform region-wide strategy, the resulting recorded compliance was deficient and uneven in the hospitals. Recorded compliance was clearly influenced by structural factors such as use of electronic format, and real compliance was much lower across hospitals and surgery specialties when assessed by direct observation, which revealed the unreliability of recorded data and the importance of health professionals' attitudes.

Uneven recorded compliance in spite of an even implementation strategy

The SSC was present in 83.1 % of the records reviewed in the retrospective study, fully completed in 28.4 %, and not checked in around 30 % of the items. These figures are somewhat lower that those reported by other hospitals in Spain and countries such as Canada and the UK [13, 18], suggesting varying effectiveness of the implementation strategy. However, while this strategy was similar for the whole region, the differences in recorded compliance among hospitals are significant (estimated % of fully

Table 5 Health team positive attitudes toward SSC and associated compliance with "time-out" items (direct observation, $n = 85$ interventions)

	(a): Is completing the SSC a waste of time? (% "No")	(b): Does the SSC improve safety? (% "Yes")	Positive attitude (% answering jointly "No" to [a] and "Yes" to [b])	% all time out items checked[a]	Difference in proportion of items checked[b]
Team (altogether)	35.3	35.3	24.7	57.1*	0.15*
Physicians	70.6	55.3	51.8	71.4*	0.07
Surgeons	84.7	77.6	75.3	95.2**	0.09
Anesthesiologists	78.8	72.9	68.2	76.2	0.03
Nurses	44.7	54.1	37.6	52.4	0.05
Circulating nurse	58.8	61.2	45.7	61.9	0.06
Scrub nurse	63.5	65.9	51.8	66.7	0.04

*$p < 0.05$

**$p = 0.01$

[a] Cases with positive attitude over total cases with all time out items checked

[b] Average in cases with positive attitude—average in cases with negative or mixed attitude

completed SSC ranging from 5.4 to 81.8). The main factor explaining this variability is the use of electronic format. This system (structural factor) seems to facilitate apparent compliance. Thus, comparing recorded compliance data between hospitals, both nationally and internationally, without taking this factor into account, may be misleading.

Data source does matter when assessing SSC compliance

One of the relevant findings of our study is the lack of reliability of recorded SSC compliance, along with confirmation of the Hawthorne effect (also known as the observer effect), suggested by greater recorded compliance in the concurrent observational study in relation to retrospective data from the previous month. The differences are so remarkable (recorded estimated "Time-out" fully completed: 51.8 % in the retrospective study, 68.2 % in the concurrent study, and 24.7 % by direct-observation) that they cast reasonable doubt on the validity of most published SSC compliance studies, indicating that results comparing hospitals and estimating the effect of SSC compliance on outcomes may also be seriously flawed if they are not adjusted according to data source. A number of inconsistencies described in articles relating SSC compliance to outcomes, including the pioneering study by Haynes et al. [3], as well as other types of checklists [19], may be partially due to the unreliability of recorded data, along with a possible Hawthorne effect. This was also demonstrated in a UK pilot study, where SSC compliance declined when the researchers were not present [15]. In this respect, our results refer only to the "Time out" portion of the SSC, but are consistent and highly significant, not only in general, but also across hospitals and surgical specialties, and for practically all "Time out" items. Our study also suggests that while the Hawthorne effect may be present in direct-observation data, it seems to have more influence on the recorded data, probably because compliance is usually assessed using recorded data. Overall, the use of direct-observation data may be a better way of assessing SSC compliance, perhaps using analysis of routine or random videotaping in order to lessen the possible Hawthorne effect and also the potential ethical concerns when avoiding obtrusive intervention, as we did in the concurrent direct observation.

The problem of relying on recorded data to assess SSC compliance worsens when an electronic format is used. It seems that electronic recording facilitates the use of the SSC as a "tick box" exercise. It is noteworthy that hospitals using the electronic format are among the highest in record-based compliance, but rank among the lowest according to direct-observation data. One institution is the highest in (apparent) compliance both in the retrospective and concurrent study, but the lowest according to direct-observation data. The design of the checklist itself may have contributed to this apparent adoption of the SSC as a "tick and flick" routine, where some elements on the list may be acknowledged but not really considered carefully. In the absence of literature reports regarding methods for building medical checklists, the SSC was compiled by drawing on lessons from the airline industry [20]; however, some of the seemingly relevant guidelines proposed for cockpit checklist design and use [21] are not applied in the SSC. For instance, it is recommended that checklist responses portray the desired status or value of the item being considered, not just "checked." This design feature may be important for items such as "Anticipated Critical Events" or the "Completion of instrument, sponge and needle counts," which should be made more explicit.

Deficient compliance with important items

Analysis of recorded data in the present study does not show priorities for improvement. However, direct-observation data reveal a different situation, one in which items more potentially related to complications are overlooked too frequently and should be prioritized for improvement. This is the case of "Anticipated Critical Events" (actually checked explicitly in only around 40 % of cases) and Antibiotic prophylaxis administration within the previous 60 min (checked in around 50 % of the cases). It seems that the role of the SSC in improving communication among the operating room team and anticipating safety problems is not being fully achieved. As mentioned above, some SSC design features may have contributed to this situation, particularly in relation to the actual checking of "Anticipated Critical Events." The items "Introduction of the team" and "Display of essential imaging" have the highest actual compliance according to direct-observation data (95.3 and 92.9 %, respectively). A possible explanation is that these two items may be very visible and new (particularly introduction of the team members) to the operating room routine, whereas other items form part of usual performance before SSC implementation, whether implicitly or in other phases of the intervention. Combining previous tasks and routines with SSC requirements has proved to be one of the barriers to its implementation [15].

Overall, the SSC seems to be viewed as just more paperwork, perhaps because it was made compulsory and subject to evaluation by the Regional Health Authority to assess surgical services accountability within a strategy that did not adequately consider the required cultural change and adaptive process [22]. Greater efforts are needed to recruit true innovators who would lead by example [23], as well as demonstrating the SSC effect in outcomes, a very difficult and costly endeavor without which any strategy may be compromised [12]. Our data suggest that mandating the use of the SSC checklist may, by itself, be a weak intervention for improving safety. Other actions grounded on an understanding of how organizations and people work may be needed, as some critical voices have pointed out [16, 22]. It is also reasonable to think that SSC compliance would improve with overall safety culture improvement, an objective that should be explicitly considered in all training programs for residents and health professions curricula.

Importance of structural factors and health professionals' attitude

Direct-observation data have permitted some clarification of the factors associated with actual SSC implementation. Not surprisingly, the use of electronic format is stipulated in the model; however, in spite of being recorded, in none of the cases where SSC was used in electronic format were all "Time out" items actually checked. By contrast, a positive attitude on the part of health professionals, particularly surgeons, is significantly associated with actual compliance. These two factors and hospital size (more frequent compliance in small and medium-sized hospitals) explained more than 50 % of the variability in actual, directly observed SSC compliance. It could be argued that the responses to attitudinal questions may have been biased because team members were interviewed shortly after the intervention. In theory, negative attitudes are expressed to justify misuse of the SSC, but the variability found both between and among professionals and for both compliance and noncompliance cases, together with the relatively low agreement in positive attitudes for the team as a whole (24.7 %) and the high level of recorded compliance, cast serious doubt on this argument. The joint positive attitude of the team is correlated with SSC implementation, but when adjusted for other structural variables and the attitudes of particular members, does not show a significant relationship. It seems that the surgeon's attitude is of paramount importance. Consequently, while training and convincing the entire surgical team may be relevant [9], the outlook and leadership of surgeons appears to be particularly important.

Limitations of the study

Sample size for the retrospective study, though representative at the regional level, was relatively small at the hospital level. For the concurrent study, we used a smaller convenience sample and a different data-gathering method; therefore, comparisons between both studies should be made with caution. However, most of the lessons learned are drawn from the concurrent study, particularly when recorded and direct observation data are compared for the same cases in the sample, surveying health professionals' attitude and assessing the success of a standard mandatory implementation strategy. The significant factors maintained in the final model for direct-observation data exhibited little or no variability. As such, their odds ratios are either extreme or show wide confidence intervals, despite explaining a considerable amount of compliance variability. Professionals' attitude could be measured with a longer and previously validated questionnaire, but the simple two-question instrument addressing the main concerns about the SSC has proved to be useful enough to discriminate positive and negative attitudes. We did not consider other potentially influential variables such as the previous occurrence of preventable complications or the duration of the surgical intervention.

Conclusions

Recorded SSC compliance may be unreliable and differ significantly from actual compliance, suggesting that studies comparing compliance using different data sources or assessing SSC effectiveness on the basis of recorded data may be flawed. Additionally, using an electronic recording format within the standard mandatory strategy facilitates apparent compliance and the use of the SSC as a tick box exercise. It seems that the main trick to improving safety is a strategy leading to positive attitudes on the part of the health professionals involved, involving a far more complex adaptive process than merely mandating the use of a checklist [22].

References

1. Weiser TG, Regenbogen SE, Thompson KD et al (2008) An estimation of the global volume of surgery: a modeling strategy based on available data. Lancet 372:139–144
2. De Vries EN, Ramrattan MA, Smorenburg SM et al (2008) The incidence and nature of in-hospital adverse events: a systematic review. Qual Saf Health Care 17:216–223
3. Haynes AB, Weiser TG, Berry WR et al (2009) A surgical safety checklist to reduce morbidity and mortality in a global population. N Engl J Med 360:491–499
4. De Vries EN, Hubert AP, Crolla R et al (2010) Effect of a comprehensive surgical safety system on patient outcomes. N Engl J Med 363:1928–1937
5. Lyons MK (2010) Eight-year experience with a neurosurgical checklist. Am J Med Qual 25:285–288
6. Lindgard L, Espin S, Rubin B et al (2005) Getting teams to talk: development and pilot implementation of a checklist to promote interprofessional communication in the OR. Qual Saf Health Care 14:340–346
7. Messael FM, Al-Qahtani AS (2009) Benchmarking of World Health Organization surgical safety checklist. Saudi Med J 30:422–425
8. Surgical safety checklist. http://whqlibdoc.who.int/publications/2009/9789241598590_eng_Checklist.pdf (Accessed May 2013)
9. World Health Organization (2009) Implementation manual WHO surgical safety checklist 2009. WHO, Geneva
10. Weiser TG, Haynes AB, Berry WR et al (2010) Effect of a 19-item surgical safety checklist during urgent operations in a global patient population. Ann Surg 251:976–980
11. Surgical safety web map. http://maps.cga.harvard.edu:8080/Hospital/. Accessed May 2013
12. Vijayasekar C, Steele RJC (2009) The World Health Organization's surgical safety checklist. Surgeon 7:260–262
13. Sewell M, Adebibe M, Jayakumar P et al (2011) Use of the WHO surgical safety checklist in trauma and orthopaedic patients. Int Orthop 35:897–901
14. Van Klei WA, Hoff RG, Van Aarnhem E et al (2012) Effects of the introduction of the WHO "Surgical Safety Checklist" on in-hospital mortality: a cohort study. Ann Surg 255:44–49
15. Vats A, Vincent CA, Nagpal K et al (2010) Practical challenges of introducing WHO surgical checklist: UK pilot experience. Br Med J 340:b5433
16. Conley DM, Singer SJ, Edmondson BA et al (2011) Effective surgical safety checklist implementation. J Am Coll Surg 212:873–879
17. Majan RP (2011) The WHO surgical checklist. Best Prac Res Clin Anaesthesiol 25:161–168
18. Taylor B, Slater A, Reznick R (2010) The surgical safety checklist effects are sustained, and team culture is strengthened (editorial). Surgeon 8:1–4
19. Ko HC, Turner TJ, Finnigan MA (2011) Systematic review of safety check list for use by medical care teams in acute hospital settings—limited evidence of effectiveness. BMC Health Services Res 11:211
20. Weiser TG, Haynes AB, Lashoher A et al (2010) Perspectives in quality: designing the WHO surgical safety checklist. Int J Qual Health Care 22:365–370
21. Degani A (1993) Cockpit checklists: concepts, design and use. Hum Factors 35:28–43
22. Bosk CL, Dixon-Woods M, Goeschel CA et al (2009) Reality check for checklists. Lancet 374:444–445
23. Berwick DM (2003) Disseminating innovations in health care. JAMA 289:1969–1975