Effect of a two year national quality improvement program on surgical checklist implementation

Effekt eines 2-jährigen nationalen Verbesserungsprogramms zur Einführung der chirurgischen Checkliste.

Anna C Mascherek[a *], Paula Bezzola[a], Katrin Gehring[a], David LB Schwappach[a, b *, †]

[a] Swiss Patient Safety Foundation, Asylstrasse 77, 8032 Zurich, Switzerland

[b] Institute of Social and Preventive Medicine (ISPM), University of Bern, Finkenhubelweg 11, 3012 Bern, Switzerland

* Authors contributed equally; † Corresponding author;

Email addresses:

AM: mascherek@patientensicherheit.ch

KG: katrin.gehring@puk.zh.ch

PB: paula.bezzola@bluewin.ch

DS: schwappach@patientensicherheit.ch

Abstract

Use of the surgical checklist in Switzerland is still incomplete and unsatisfactory. A national improvement program was developed and conducted in Switzerland to implement and improve the use of the surgical safety checklists. The aims of the implementation program were to implement comprehensive and correct checklist use in participating hospitals in every patient and in every surgical procedure; and to improve safety climate and teamwork as important cultural context variables.

10 hospitals were selected for participation in the implementation program. A questionnaire assessing use, knowledge, and attitudes towards the checklist and the Safety Climate Survey were conducted at two measurement occasions each in October/November 2013 and January/February 2015. Significant increases emerged for frequency of checklist use ($F_{(1,1001)}$ = 340.9, p <0.001), satisfaction ($F_{(1,1232)}$ = 25.6, p<0.001), and knowledge($F_{(1,1294)}$ = 184.5, p<0.001). While significant differences in norms ($F_{(1,1284)}$ = 17.9, p<0.001) and intentions ($F_{(1,1284)}$ = 7.8, p<0.01) were observed, this was not the case for attitudes ($F_{(1,1283)}$ = .8, *n.s.*) and acceptances ($F_{(1,1284)}$ = 0.1, *n.s.*). Significant differences for safety climate and teamwork emerged in the present study ($F_{(1,3555)}$ = 11.8, p<0.001 and $F_{(1,3554)}$ = 24.6, p<0.001, respectively). However, although statistical significance was reached, effects are very small and practical relevance is thus questionable. The results of the present study suggest that the quality improvement program conducted by the Swiss Patient Safety Foundation in 10 hospitals led to successful checklist implementation. The strongest effects were seen in aspects concerning behaviour and knowledge specifically related to checklist use. Less impact was achieved on general cultural variables safety climate and teamwork. However, as a trend was observable, these variables may simply need more time in order to change substantially.

Keywords: Surgical safety checklist, implementation program Switzerland, Safety Climate, Attitudes, Survey data.

Abstract

Die Anwendung der chirurgischen Checkliste ist in der Schweiz heute noch immer nicht zufriedenstellend. Ein nationales Verbesserungsprogramm wurde entwickelt und umgesetzt, um die chirurgische Checkliste einzuführen und ihre Anwendung zu verbessern. Ziele des Verbesserungsprogramms waren erstens die umfassende und richtige Anwendung der Checkliste in den teilnehmenden Spitälern bei jedem Patienten und jedem chirurgischen Eingriff und zweitens die Verbesserung des Sicherheitsklimas und der Teamzusammenarbeit als wichtige Aspekte der Betriebskultur. 10 Spitäler wurden für die Teilnahme am Programm ausgewählt. Zu zwei Erhebungszeitpunkten (Oktober/November 2013 und Januar/Februar 2015) wurden mit einem Fragebogen Anwendung, Wissen und Einstellungen gegenüber der chirurgischen Checkliste sowie Sicherheitsklima und Teamzusammenarbeit gemessen. Es zeigten sich signifikante Veränderungen für die Häufigkeit der Anwendung ($F_{(1,1001)} = 340.9$, $p < 0.001$), Zufriedenheit mit der Anwendung ($F_{(1,1232)} = 25.6$, $p<0.001$) und Wissen ($F_{(1,1294)} = 184.5$, $p<0.001$). Signifikante Unterschiede zeigten sich auch für Normen ($F_{(1,1284)} = 17.9$, $p<0.001$) und Intentionen ($F_{(1,1284)} = 7.8$, $p<0.01$), allerdings nicht für Einstellungen ($F_{(1,1283)} = .8$, *n.s.*) und Akzeptanz ($F_{(1,1284)} = 0.1$, *n.s.*). Auch Sicherheitsklima und Teamzusammenarbeit veränderten sich signifikant ($F_{(1,3555)} = 11.8$, $p<0.001$ und $F_{(1,3554)} = 24.6$, $p<0.001$), die geringen Effektstärken deuten aber auf eine geringe praktische Relevanz dieser Unterschiede hin. Die Ergebnisse der vorliegenden Studie zeigen, dass das von der Stiftung Patientensicherheit Schweiz durchgeführte Verbesserungsprogramm eine erfolgreiche Einführung der chirurgischen Checkliste bewirkt hat. Die stärksten Effekte zeigten sich bei konkretem Verhalten und Wissen in direkter Verbindung zur Anwendung der chirurgischen Checkliste. Geringe Unterschiede zeigten sich in den allgemeineren Kulturvariablen Sicherheitsklima und Teamzusammenarbeit. Der Trend deutet darauf hin, dass eine Veränderung in diesen Variablen mehr Zeit braucht.

Schlüsselwörter: chirurgische Checkliste, Implementierungsprogramm in der Schweiz, Sicherheitsklima, Einstellungen, Fragebogenerhebung.

1. Introduction

The WHO estimated that more than 234 million major surgical procedures are undertaken every year worldwide [1]. In high-expenditure countries, the mean surgical rate is 11’110 per 100’000 population. Research suggests that 3-16% of patients undergoing surgery will suffer adverse events, with a considerable fraction of these events deemed preventable [2-4]. Surgery has also been identified as one major predictor of patient reported hospital-acquired infections [5]. Given the high volume of surgery, these figures make improvements in safety of surgery a global public health issue. During the last years the WHO surgical safety checklist has emerged as a powerful intervention to increase safety in surgery. The checklist has become one of the most strongly recommended “single-intervention” safeguards worldwide [6]. The checklist has been proven to be an effective intervention to reduce morbidity and mortality in surgical procedures [7]. However, recent evidence also suggests that simply making the surgical checklist mandatory by policy-makers does not necessarily improve surgical outcomes [8-10]. Research revealed that in order to be effective, the surgical safety checklist must be used and applied correctly, completely and in all patients [10-12]. While the checklist itself rather serves as a cognitive aid, the introduction of the checklist is a complex team intervention making its sustainable implementation challenging. The checklist is more than a ticking box exercise and requires change and the willingness to change on different levels within an organization. For example, team communication behaviours and information sharing among team members during intraoperative phases have been associated with decreased mortality and major complications in surgery [13]. But interaction and communication within surgical teams require acceptance, acknowledgement and a joint

understanding of the different roles of co-workers, in particular, across the professions. A recent systematic review confirmed that operating room teamwork and communication can be improved by checklist use.This may be one mechanism through which patient outcomes are improved with checklist implementation [14,15].

Checklist use in Switzerland is still incomplete and unsatisfactory. In a recent study, Mascherek et al. [16] found that about 80% of the study participants use a surgical checklist on an everyday level. Only 61% of the nursing staff and 72% percent of the doctoral staff were satisfied with the use of the checklist, indicating that the quality of use was less than perfect. Several European countries already work together in the High 5s project to improve checklist use in hospitals [17]. However, because Switzerland is not part of the European program, a national improvement program was developed and conducted in Switzerland to implement and improve the use of the surgical safety checklists. The program was developed based on implementation research, international quality improvement programs, and expert opinion. Leadership, inter-professional project teams, local champions, training, and education were substantial parts of the implementation process.

The aims of the implementation program were twofold: first, to implement comprehensive and correct checklist use in participating hospitals in every patient and in every surgical procedure; and second, to improve safety climate and teamwork as important cultural context variables. The program was accompanied by extensive evaluation on different levels. In a pre-post intervention design, safety climate and staff attitudes were assessed at two measurement occasions.

In the present paper we provide an overview of the results of the two-year national improvement program. We describe changes in checklist use, knowledge and attitudes towards checklists, and changes in safety climate and teamwork between measurement points on group level

2. Method

2.1. Design

The national quality improvement program was developed and conducted by the Swiss Patient Safety Foundation between 2012 and 2015 (for a detailed study protocol see Kobler et al., 2014 [18]).

The program was designed to compromise 2 axes to mutually affect one another. On axis 1, a national campaign was conducted between 2012 and 2015 to promote patient safety in general and the use of the WHO surgical checklist in particular. Axis 2 consisted of a specific improvement intervention with 10 participating hospitals. Hospitals were recruited via a national open tender. All Swiss acute-care hospitals could apply for participation. Participation was promoted as prestigious. Application was connected to the evaluation program from which the data for the present study originates. No financial compensation was offered, but participating hospitals had to pay an attendance fee of CHF 6000. In total, 32 hospitals applied for participation from all areas of Switzerland. 10 out of the 32 applicants were selected for participation. Criteria for hospital selection were hospital size, hospital speciality, geographical region in Switzerland, explicit commitment from senior surgeons and anaesthetists, and whether or not they had already implemented any surgical checklist in their operating room–routine (OR).

The improvement program was scheduled between summer 2013 and summer 2015. Hospitals were contractually bound to implement the checklist and execute mandatory activities during the program such as training of checklist use, education of staff, and local adaptation of checklist. Also, explicit support by leadership, support by local champions, and establishing a cross-professional team were mandatory components of the program. Additionally, to facilitate exchange and learning between hospitals, 4 mandatory workshops were held during the 2 years of program execution.

Data in the present study were collected with two separate questionnaire instruments during the program at two time points each. The time between data collections was 15 months. Print versions of the surveys were sent to hospitals and locally distributed. Surveys were provided in German and French. The survey sample consisted of all members of the Operating Room (OR) teams of the participating hospitals (doctors, nurses, scrub nurses, surgical technicians, and attendants for surgical positioning). The sample for the assessment of safety climate and teamwork additionally included ward staff (doctors, nurses, nursing assistants, and further professionals who were subsumed under "others" involved in the pre- and postoperative care of surgical patients). Subjects were invited to participate by the hospitals' project teams and repeatedly reminded throughout the data-collecting period. As the study was conducted as part of a quality improvement project, study design and data collection did not require approval of an ethical committee in Switzerland referring to Article 1 and 2 of the Federal Act on Research involving Human Beings (Human Research Act, HRA) [19].

2.2. Survey instruments

2.2.1. Use, knowledge, and attitudes towards the checklist ("Attitudes survey")

The first survey was developed to assess use of, knowledge of and attitudes towards the surgical checklist. It was developed on the basis of extensive review of the literature and own pilot studies. The survey consisted of three conceptual parts. The first part referred to the general use of surgical checklists in the OR (e.g., "which of the following checklists do you use?") and relative frequency of use (rated on a 0-100% scale subdivided into 6 categories). Satisfaction with using the checklist was also assessed in part one and rated on a 5-point-Likert-scale ranging from "very unsatisfied" to "very satisfied". General attitudes towards the use of the checklist in clinical practice were assessed in a semantic differential. The semantic differential acts as "barometer" and addresses the perceived atmosphere of checklist implementation. The differential used six pairs of adjectives with opposing meaning to

describe checklist use. Paired adjectives were “easy-difficult”, “pleasant-unpleasant”, “familiar-unfamiliar”, “important-unimportant”, and “good for staff-bad for staff”, “good for patients-bad for patients”. The continuum consisted of a 7-point-scale with “7” anchoring a positive attitude and “1” a negative attitude.

In the second part, knowledge of the WHO-checklist was assessed. Subjects rated their subjective knowledge of the checklist on a 5-point-Likert-scale from “very bad” to “very good”. In the third part, attitudes towards, acceptance of, norms, and intention to use the checklist were assessed. Norms (e.g. “Surgeons look down upon checklist use”) were measured with 4 items, acceptance (e.g. “The checklist is evidence-based.”) with 5 items, attitudes (e.g. “Checklist use enhances paying attention to patient safety”) with 7 items, and intentions with (e.g. “It is my plan to carefully mind the use of the checklist”) with 6 items. All items were rated on a 7-point-Likert-scale ranging from “do not agree at all” to “completely agree” (for the wording of all survey-items see Appendix A). High scores indicate a positive evaluation of the checklist. Four items had to be reverse coded for data analyses. The survey was pre-tested with individuals from all professions and languages.

2.2.2. Safety Climate Survey

As second questionnaire the Safety Climate Survey was applied [20,21]. The original version of the Safety Climate Survey was translated by a professional translator from English to German and back-translated to English by a second translator. Differences in translation and back-translation were discussed and resolved by the research team. The survey was also translated to French and proofread by bilingual researchers in French (for the wording of all survey-items see Appendix B).Two versions (“OR” and “ward”) were developed differing in the wording of single items referring to the specific working area. Details on survey development are provided elsewhere [22]. The questionnaire consisted of the 19 items of the

Safety Climate Survey and was rated on a 5-point Likert-scale from 1 = "disagree strongly" to 5 = "agree strongly". At the end of the survey, socio-demographic variables were assessed.

2.3. Data analyses

Individuals who answered less than 60% of the items were dropped from analysis. ANOVA was conducted to test for differences between time points on different variables. The Scheffé correction was applied to significance tests to account for multiple comparisons and correct for cumulative type I error. Effect sizes were calculated with Cohen's *d*. According to convention, 0.2 is considered a small, 0.5 a medium, and 0.8 a large effect [23]. Although for some of the measures the assumption of homoscedasticity was violated, we conducted one-way ANOVA as the sample was large enough and balanced. ANOVA is a robust estimator against violations of homoscedasticity in large, balanced samples [24]. Cluster effects were not considered when analysing the data. All analyses were conducted using STATA v13.1 [25].

3. Results

3.1. Sample

742 individuals completed the attitudes-questionnaire at time-point 1 (T1) and 660 at time-point 2 (T2). 1966 individuals completed the safety-climate-questionnaire at T1 and 1604 at T2. 217 individuals only answered the attitudes -questionnaire, 1441 individuals only answered the safety-climate-questionnaire, and 525 individuals returned both questionnaires at T1. The corresponding figures for T2 were 142, 1086, and 518, respectively. This equals response rates of 37.9% at T1 and 31.9% at T2 for the attitudes -questionnaire and 42.7% at T1 and 33.3% at T2 for the safety-climate-questionnaire.

Sample characteristics separated by time-points are presented in table 1. Although some of the characteristics differ significantly between time-points, significance is due to the large sample

size rather than to practically relevant differences. Samples are hence to be considered as being comparable between time points with respect to socio-demographic variables.

3.2. Means at T1

Means of the variables under study at T1 and T2 are presented in table 2. The mean for frequency of use was *M* = 3.5 (*SD* = 2.2), for satisfaction with use *M* = 3.4 (*SD* = 1.1), and for subjective knowledge *M* = 3.2 (*SD* = 1.2), hence, in the mid-range of the scale. Means for safety climate (*M* = 3.8, *SD* = 0.5) and teamwork (*M* = 3.9, *SD* = 0.7) were in the mid-range of the scale as well. For norms, attitudes, acceptance, and intentions means were *M* = 5.9 (*SD* = 1.0), *M* = 6.0 (*SD* = 1.0), *M* = 6.1 (*SD* = 0.9), and *M* = 6.2 (*SD* = 1.0), respectively. On a 7-point scale, these values are already very high, hence, imposing the possibility of ceiling effects at T2. For norms, the highest score on the scale accounted for 31% of the answers. This applied to 34% in attitudes, 40% in acceptance, and even 50% in intentions, emphasizing the possibility of ceiling effects at T2. Means for the paired adjectives “easy-difficult” (*M* = 5.9, *SD* = 1.3), “pleasant-unpleasant” (*M* = 5.4, *SD* = 1.4), “familiar-unfamiliar” (*M* = 5.3, *SD* = 1.8), “important-unimportant” (*M* = 6.4, *SD* = 1.2), “good for staff-bad for staff” (*M* = 6.2, *SD* = 1.1), “good for patients-bad for patients” (*M* = 6.5, *SD* = 0.9) were also already high at T1. High scores accounted for a considerable amount of answers in all of the paired adjectives, with the least of 26% in “pleasant-unpleasant” (easy-difficult” 43%, “pleasant-unpleasant” 26%, “familiar-unfamiliar” 32%, “important-unimportant” 67%, “good for staff-bad for staff” 57%, “good for patients-bad for patients” 73%), hence, again, imposing the possibility of ceiling effects at T2.

3.3. Mean Differences

Checklist use and knowledge

Significant differences between time points concerning frequency of checklist use emerged, indicating that the checklist was used more often at the end of the program ($F_{(1,1001)} = 340.9$, $p < 0.001$). The difference in use denotes a large effect; hence, the increase can also be interpreted as being of practical significance. Satisfaction with use was also significantly higher at T2 ($F_{(1,1232)} = 25.6$, $p<0.001$). However, effect size for this difference was small, indicating that the difference in satisfaction was less pronounced. Results for knowledge also show higher values at T2 ($F_{(1,1294)} = 184.5$, $p<0.001$). Effect size for this difference is large, indicating that differences in subjective knowledge are of practical significance. Items of the semantic differential describe the perceived usability and practical significance of the checklist for staff and patients. Results for these items were mixed. Significant differences emerged only on the dimensions "easy-difficult" and "familiar-unfamiliar" with $F_{(1,1220)} = 46$, $p<0.001$ and $F_{(1,1215)} = 100.5$, $p<0.001$, respectively. On the remaining dimensions ("pleasant-unpleasant", "important-unimportant", "good for staff-bad for staff", and "good for patients-bad for patients") no significant differences emerged. Data are presented in table 2.

3.4. Attitudes towards checklist use

Effects of the program were also assessed for personal attitudes, norms and intention to use the checklist. The scales reflect an individual's personal attitude towards the checklist, his/her intention to use the checklist and his/her perception of checklist use as being normative. Results were mixed. While significant differences in norms ($F_{(1,1284)} = 17.9$, $p<0.001$) and intentions ($F_{(1,1284)} = 7.8$, $p<0.01$) were observed, this was not the case for attitudes ($F_{(1,1283)} = .8$, *n.s.*) and acceptances ($F_{(1,1284)} = 0.1$, *n.s.*). However, effect sizes for differences in norms and intentions were small, indicating limited practical relevance. Data are presented in table 2.

3.5. Safety climate and teamwork

Significant differences for safety climate and teamwork emerged in the present study ($F_{(1,3555)}$ = 11.8, p<0.001 and $F_{(1,3554)}$ = 24.6, p<0.001, respectively). However, although statistical significance was reached, effects are very small and practical relevance is thus questionable. Means are presented in table 2.

4. Discussion

The goal of the present study was to provide a broad overview of differences on various dimensions related to surgical checklist implementation in Switzerland. Different aspects relevant for implementation programs were addressed. Differences were most pronounced in a group of variables that mainly reflect practical aspects of checklist use. The largest differences emerged for frequency of checklist use, familiarity, and knowledge. These variables directly reflect a changed routine of checklist use. Staff of participating hospitals was explicitly educated about objectives and aims of the checklist. Also, checklist use was made mandatory in the treatment of all patients undergoing surgery during the study period. Increase in usage may be reflected in greater familiarity. This is supported by an increase in the "ease of use" rating of the checklist. Although no causal inferences are warranted, it seems reasonable that differences in those variables are a result of the activities during the implementation program which mainly focused on correct checklist use. Staff, for example, had to participate in trainings outside the OR before checklist implementation. Results indicate that this aim was achieved including side-effects on familiarity and ease of use.

A second set of variables assessed attitudes, norms, intentions, and acceptance of checklist use. While significant effects for norms and intentions emerged, no significant differences were observed in acceptance and attitude. Effects that were found were of small to medium size. It seems reasonable that every-day-checklist-use, education of staff and changed hospital policies strengthened norms perceived by staff. At least superficially, not using the checklist in surgery is more and more perceived as unprofessional behaviour. However, no differences

for acceptance of the checklist as a relevant tool to improve patient safety were observed. Two explanations are possible. First, no intervention was tailored to explicitly affect attitudes, as the program explicitly focused on the practical implementation of the checklist. Hence, attitudes simply remained unaffected by the program... Second, as scores on attitudes were already high at T1 with a considerably proportion of the sample scoring at the upper boundary of the scale, the results may reflect ceiling effects. High levels of acceptance have also been reported by Rothmund et al in Germany [26]. In their study, checklists were acknowledged as useful tool, however, leaving room for improvement in daily practice. This result resembles the pattern found in Switzerland before the national improvement program, underlining the interpretation of the checklist being an accepted patient safety tool even before the intervention. The results suggest that the interventions mainly worked on closing the gap between what is considered a relevant tool in theory and what is actually applied in the field. Impact of the program mainly affected behavioural aspects and knowledge of checklist use on daily basis. One may conclude that the main difference after the program was the shift from "theory to practice". This interpretation is supported by several personal experiences and feedback from project managers during the program.

A second aim of the quality improvement study was to improve overall safety climate and teamwork in the course of checklist implementation and use. Although significant differences emerged, effect sizes were very small. As sample size was large, the differences found in the present study must be interpreted as being of little practical relevance. However, a trend emerged, pointing into the targeted direction. Stronger effects might be observable on item level. The more detailed analysis of the data from T1 revealed stronger mean difference on item level than on scale level [27]. However, as hospital organizational culture needs time to change, the interval between measurements in our study might simply have been too short to exhibit substantial differences on scale level.

This study has several limitations. First, no clinical outcome measures were assessed, hence, direct effects on adverse events could not be observed. However, as compliance with checklist use was high, the inference seems justified that the program not only led to comprehensive checklist use but, as a consequence, also improved patient safety in surgery. A second limitation lies in the sampling itself. Although study requirements specifically defined sample composition, this was only controlled for by the local project manager. Hence, sample composition might have varied between sites. Finally, only 10 hospitals participated in the program. Although hospital selection took various aspects into account, generalizability of the results is limited. This is especially the case as no causal inferences can be made from the results of the present study. If, and if so which aspect of the intervention actually caused the differences on the measures between time points remains unknown. Strong experimental or RCT-studies are needed in order to extract causal factors underlying achieved differences in large quality improvement programs.

5. Conclusion

The results of the present study suggest that the quality improvement program conducted by the Swiss Patient Safety Foundation in 10 hospitals led to successful checklist implementation. The strongest effects were seen in aspects concerning behaviour and knowledge specifically related to checklist use. Less impact was achieved on the general cultural variables safety climate and teamwork. However, as a trend was observable, these variables may simply need more time in order to change substantially. Taken together, the results of the present study show that the program was suitable for successful implementation of comprehensive checklist use in Swiss hospitals.

Table 1. Sample characteristics (N_{total}=3929) *N* for subsamples are***:*** $\boldsymbol{n}_{t1attitude}$ = 742; $\boldsymbol{n}_{t2attitude}$ = 660; $\boldsymbol{n}_{t1safetyclimate}$ = 1966; $\boldsymbol{n}_{t2safetyclimate}$ = 1604. Because some individuals answered both questionnaires the sum of the subsamples is larger than the total *n* per time-point. Data not adding up to 100% are due to missing values

Characteristic†	**T1**		**T2**	
	$\boldsymbol{n}_{t1}$ **= 2183**	**%**	$\boldsymbol{n}_{t2}$ **= 1746**	**%**
Survey language				
German	1460	66.9	1162	66.6
French	723	33.1	584	33.4
Gender				
Male	622	29.1	505	29.6
Female	1515	70.9	1203	70.4
Mean age in years*	38.9 (SD: 11.3)		38.1 (SD: 11.1)	
Profession				
Physician	573	27.3	502	30.1
Nurse	1176	56.1	913	54.8
Other	347	16.6	252	15.1
Managerial function				
Yes	511	24.3	385	22.9
No	1587	75.6	1295	77
Years of professional experience*				
0 -2	286	13.4	296	17.2
2 - 5	387	18.1	317	18.5
5 - 10	427	20	317	18.5
10 - 20	541	25.3	426	24.8
More than 20	490	23	361	21
Hours of direct patient care per week*				
0	295	14	214	12.7
0-8	328	15.5	237	14.1
8-16	259	12.3	208	12.3
16-24	295	14	230	13.7
24-32	299	14.1	240	14.2
32-40	355	16.8	298	17.7
More than 40	276	13.1	250	14.8

† *Note:* Characteristics marked with "*" differ significantly between time-points on $p < .05$.

Table 2. Differences in mean scores of assessed constructs between time points

	$Mean_{T1}$ (SD)	n_{T1}	$Mean_{T2}$ (SD)	n_{T2}	Effect sizes*	p-value
Frequency of use	3.5 (2.2)	491	5.8 (.80)	512	1.2	<0.001
Satisfaction with use	3.4 (1.1)	419	3.8 (1.1)	507	0.3	<0.001
Semantic differential						
easy-difficult†	5.9 (1.3)	584	6.3 (1.0)	638	0.4	<0.001
pleasant-unpleasant†	5.4 (1.4)	579	5.8 (1.4)	635		n.s.
familiar-unfamiliar†	5.3 (1.8)	580	6.1 (1.1)	637	0.6	<0.001
important-unimportant†	6.4 (1.2)	594	6.5 (1.0)	643		n.s
good for staff-bad for staff†	6.2 (1.1)	595	6.3 (1.1)	641		n.s.
good for patients-bad for patients†	6.5 (.9)	591	6.6 (.9)	641		n.s.
Subjective knowledge	3.2 (1.2)	647	4.1 (1.1)	649	0.8	<0.001
Norms	5.9 (1.0)	637	6.1 (.8)	649	0.3	<0.001
Attitude	6.0 (1.0)	636	6.0 (1.0)	649		n.s.
Acceptance	6.1 (.9)	637	6.2 (.9)	649		n.s.
Intentions	6.2 (1.0)	637	6.4 (.9)	649	0.2	<0.001
Safety Climate	3.8 (.5)	1957	3.9 (.6)	1600	0.12	0.001
Teamwork	3.9 (.7)	1956	4.0 (.7)	1600	0.12	<0.001

Note: T1 = time point 1, T2 = time point 2, *we report Cohen's *d* for differences between time points. †higher values indicate more positive evaluation

Competing interests

The authors declare that they have no competing interests.

Authors' contribution

DS and PB designed the study. AM and KG substantially contributed to design and conception. AM, KG, DS developed the survey. AM conducted statistical analyses. AM and DS drafted the manuscript. KG and PB revised the manuscript critically for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements

We thank Denis Hemmes from the Hôpital de l'enfance, Fiona Schreve Bloem from the Hôpital de la Tour, Dominique Thorin from the hôpital fribourgeois, Monica Wyss Baur from the Kantonsspital Baden, Marianna Friedli-Braun from the Kantonsspital Graubünden, Markus Röthlin from the Kantonsspital Münsterlingen, Tobias Haefliger from the Kantonsspital Uri, Antoine Oesch from the Spitalzentrum Biel, Henry Hoffmann from the Unversitätsspital Basel, and Ursula Schäfer from the UniversitätsSpital Zürich as project leader in charge at their respective institution for their support with data collection.

Funding

The program was mainly funded by the Federal Office of Public Health, Switzerland (BAG).

References

[1] Weiser TG, Regenbogen S, Thompson K, Haynes A, Lipsitz S, Berry W *et al*.: An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet* 2008, 372: 139-144.

[2] Anderson O, Davis R, Hanna GB, Vincent CA: Surgical adverse events: a systematic review. *The American Journal of Surgery* 2013, 206: 253-262.

[3] de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA: The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care* 2008, 17: 216-223.

[4] Seiden SC, Barach P: Wrong-side/wrong-site, wrong-procedure, and wrong-patient adverse events: Are they preventable? *Arch Surg* 2006, 141: 931-939.

[5] Schwappach DLB: Frequency of patient-reported infections among sicker adults in high-income countries: An international perspective. *Am J Infect Control* 2012, doi: 10.1016/j.ajic.2012.02.011.

[6] Shekelle PG, Pronovost PJ, Wachter RM, McDonald KM, Schoelles K, Dy SM *et al*.: The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now. *Ann Intern Med* 2013, 158: 365-368.

[7] Borchard A, Schwappach DLB, Barbir A, Bezzola P: A Systematic Review of the Effectiveness, Compliance, and Critical Factors for Implementation of Safety Checklists in Surgery. *Ann Surg* 2012, 256: 925-933.

[8] Urbach DR, Govindarajan A, Saskin R, Wilton AS, Baxter NN: Introduction of Surgical Safety Checklists in Ontario, Canada. *N Engl J Med* 2014, 370: 1029-1038.

[9] Reames BN, Scally CP, Thumma JR, Dimick JB: Evaluation of the Effectiveness of a Surgical Checklist in Medicare Patients. *Med Care* 2015, 53.

[10] Levy SM, Senter CE, Hawkins RB, Zhao JY, Doody K, Kao LS *et al.*: Implementing a surgical checklist: More than checking a box. *Surgery* 2012, 152: 331-336.

[11] van Klei WA, Hoff RG, van Aarnhem EEHL, Simmermacher RKJ, Regli LPE, Kappen TH *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.

[12] Russ SJ, Sevdalis N, Moorthy K, Mayer EK, Rout S, Caris J *et al.*: A Qualitative Evaluation of the Barriers and Facilitators Toward Implementation of the WHO Surgical Safety Checklist Across Hospitals in England: Lessons From the "Surgical Checklist Implementation Project". *Ann Surg* 9000, Publish Ahead of Print.

[13] Mazzocco K, Petitti DB, Fong KT, Bonacum D, Brookey J, Graham S *et al.*: Surgical team behaviors and patient outcomes. *The American Journal of Surgery* 2008, 197: 678-685.

[14] Russ S, Rout S, Sevdalis N, Moorthy K, Darzi A, Vincent C: Do Safety Checklists Improve Teamwork and Communication in the Operating Room? A Systematic Review. *Ann Surg* 2013, 258: 856-871.

[15] Neily J, Mills PD, Young-Xu Y, Carney BT, West P, Berger DH *et al.*: Association Between Implementation of a Medical Team Training Program and Surgical Mortality. *JAMA* 2010, 304: 1693-1700.

442 [16] Mascherek A, Schwappach D, Bezzola P: Frequency of use and knowledge of the WHO-
443 surgical checklist in Swiss hospitals: a cross-sectional online survey. *Patient safety in*
444 *surgery* 2013, 7: 36.

445 [17] Renner D, Fishman L, Berning D, Thomeczek C: Zwischenstand des High 5s-Projekts.
446 *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 2014, 108: 56-
447 58.

448 [18] Kobler I, Mascherek A, Bezzola P: Wie kann die Implementierung von
449 Patientensicherheitsprogrammen gelingen? Ein Beispiel aus der Schweiz. *Zeitschrift*
450 *für Evidenz, Fortbildung und Qualität im Gesundheitswesen*.

451 [19] The Federal Authorities of the Swiss Confederation. Federal Act on Research involving
452 Human Beings. http://www.admin.ch/opc/en/classified-
453 compilation/20061313/index.html . 2011. 9-7-2014.

454 [20] Shteynberg G, Sexton B.J., Thomas E. Test retest reliability of the safety climate scale.
455 Edited by The University of Texas Center of Excellence for Patient Safety Research
456 and Practice. Technical Report 01-05. 2013. 16-4-2013.

457 [21] Sexton J, Helmreich R, Neilands T, Rowan K, Vella K, Boyden J *et al*.: The Safety
458 Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging
459 research. *BMC Health Serv Res* 2006, 6: 44.

460 [22] Gehring K, Mascherek AC, Bezzola P, Schwappach DLB: Safety climate in Swiss
461 hospital units: Swiss version of the Safety Climate Survey. *J Eval Clin Pract* 2015, 21:
462 332-338.

[23] Cohen J: *Statistical power analysis for the behavioral sciences*, Rev. edn. edn. Hillsdale: Erlbaum; 1987.

[24] Alexander RA, Govern DM: A New and Simpler Approximation for ANOVA under Variance Heterogeneity. *Journal of Educational Statistics* 1994, 19: 91-101.

[25] StataCorp: *Stata Statistical Software: Release 12.* College Station, TX: StataCorp LP.; 2011.

[26] Rothmund M, Kohlmann T, Heidecke CD, Siebert H, Ansorg J: Einführung und Beurteilung von Maßnahmen zur Fehlerprävention in chirurgischen Kliniken: Ergebnisse einer aktuellen online-Befragung. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 2015, 109: 384-393.

[27] Gehring K, Mascherek AC, Bezzola P, Schwappach DLB: Safety climate in Swiss hospital units: Swiss version of the Safety Climate Survey. *J Eval Clin Pract* 2015, doi: 10.1111/jep.12326.

Appendix A. Wording items of Attitude-survey (items translated from the German original)

Item No.	Item
V04	Checklist use facilitates speaking up in the OR. (attitude)
V06	Checklist use is far too time-consuming. (attitude)
V09	Checklist use decreases adverse events. (attitude)
V13	Checklist use hinders the flow of information among OR-team members. (attitude)
V19	Checklist use enhances paying attention to patient safety. (attitude)
V20*	Checklist use interferes with my tasks. (attitude)
V22*	A well-functioning OR-team does not need a checklist. (attitude)
V15	Surgeons look down upon checklist use. (norms)
V18	My colleagues take checklist use serious. (norms)
V23	My supervisor promotes checklist use. (norms)
V24	I am expected to use the checklist seriously. (norms)
V07	I will promote checklist use. (intentions)
V08	I will support my colleagues with using the checklist. (intentions)
V10	Next time I am up to decide, I will apply the checklist. (intentions)
V11	It is my plan to carefully mind the use of the checklist. (intentions)
V16	I want the checklist to be used with every patient. (intentions)
V25†	It is my duty to correctly use the checklist. (intentions)
V26	The checklist is not necessary. (acceptance)
V14	The checklist is a well-designed tool to enhance patient safety (acceptance)
V12	If I had to have an operation, I would want the checklist to be applied. (acceptance)
V02	The checklist is not more than a ticking-exercise. (acceptance)
V01	The checklist is evidence-based. (acceptance)

491 **Appendix B.** Items of the Safety Climate Survey (SCS) and Teamwork (TW)

Item No.	Item
V1	The culture of this clinical area makes it easy to learn from the mistakes of others. (SCS)
V2	Medical errors** are handled appropriately in this clinical area. (SCS) *** Medical error is defined as any mistake in the delivery of care, by any healthcare professional, regardless of outcome.*
V3	The senior leaders in my hospital listen to me and care about my concerns. (SCS)
V4	The physician and nurse leaders in my area listen to me and care about my concerns. (SCS)
V5	Leadership is driving us to be a safety- centred institution. (SCS)
V6	My suggestions about safety would be acted upon if I expressed them to management. (SCS)
V7	Management/leadership does not knowingly compromise safety concerns for productivity. (SCS)
V8	I am encouraged by my colleagues to report any safety concerns I may have. (SCS)
V9	I know the proper channels to direct questions regarding patient safety. (SCS)
V10	I receive appropriate feedback about my performance. (SCS)
V11	I would feel safe being treated here as a patient. (SCS)
V12	Briefing personnel before the start of a shift (i.e., to plan for possible contingencies) is an important part of patient safety. (SCS)
V13	Briefings are common here. (SCS)
V14	I am satisfied with the availability of clinical leadership.
V14a	Physician(SCS)
V14b	Nursing
V14c	Pharmacy
V15	This institution is doing more for patient safety now than it did one year ago. (SCS)
V16	I believe that most adverse events occur as a result of multiple system failures and are not attributable to one individual's actions. (SCS)
V17	The personnel in this clinical area take responsibility for patient safety. (SCS)
V18	Personnel frequently disregard rules or guidelines that are established for this clinical area. (SCS)
V19	Patient safety is constantly reinforced as the priority in this clinical area. (SCS)
V20	Input from nursing staff is well-received in my area of work. (TW)
V21	Conflicts in my area of work are solved appropriately. (that is the focus is not on *who* is right but on *what* is best for the patient) (TW)
V22	I receive the support I need from my colleagues to adequately take care of the

	patients. (TW)
V23	It is easy for staff to ask questions in case of lack of knowledge. (TW)
V24	It is difficult in my area of work to address problems concerning patient care. (TW)
V25	Doctors and Nurses work together well as an experienced team. (TW)

492