ISQua®

International Journal for Quality in Health Care, 2017, 29(5), 685–692
doi: 10.1093/intqhc/mzx086
Advance Access Publication Date: 18 July 2017
Article

OXFORD

Article

Hospital survey on patient safety culture (HSOPS): variability of scoring strategies

JORIS GIAI[1], BASTIEN BOUSSAT[2,3], PAULINE OCCELLI[4,5], GERALD GANDON[2], ARNAUD SEIGNEURIN[2,3], PHILIPPE MICHEL[4], and PATRICE FRANÇOIS[2,3]

[1]Service de Biostatistique, Hospices Civils de Lyon, Laboratoire de biométrie et biologie évolutive, UMR 5558 CNRS, F-69003 Lyon, France, [2]Quality of Care Unit, Grenoble University Hospital, F-38043 Grenoble, France, [3]TIMC UMR 5525 CNRS, Université Grenoble Alpes, France, [4]Quality and Patient Safety Department, Pôle IMER, University Hospital of Lyon, F-69003 Lyon, France, and [5]EA 7425 Health Services and performance Research (HESPER), Université Claude Bernard Lyon 1, France

Address reprint requests to: Bastien Boussat, Quality of Care Unit, Pavillon Taillefer, CHU CS 10-217, F-38043, Grenoble Cedex, France. Tel: +33-4-76-76-69 -96; Fax: +33-4-76-76-88-31; E-mail: bboussat@chu-grenoble.fr

Editorial Decision 1 July 2017; Accepted 3 July 2017

Abstract

Objective: To assess the variability of safety culture dimension scores and their associated rankings depending on three different scoring strategies using the Hospital Survey On Patient Safety Culture (HSOPS).
Design: Cross-sectional study using a self-administered questionnaire.
Setting: The study was conducted in an 1836-bed acute-care French university hospital with an annual volume of 135 999 stays, between April 2013 and November 2014.
Participants: All caregivers and technical-administrative staff with at least 6 months of employment, spending at least half of their working time in the hospital, were asked to participate.
Intervention: None.
Main outcome measure: The variability of the HSOPS results using three different scoring methods: the percentage of positive responses recommended by the Agency for Healthcare Research and Quality, the averaged individual means and the averaged individual sums.
Results: The response rate was 78.6% ($n = 3978$). The percentage of positive responses resulted in lower scores compared to averaged individual means and averaged individual sums in the six least developed dimensions, and gave more widely spread scores and greater 95CIs in the six most developed dimensions. Department rankings also varied greatly depending on the scoring methods.
Conclusion: The values of the HSOPS scores and their corresponding rankings greatly depended on the computation method. This finding shows how important it is to agree on the use of the same scoring strategies, before broadly comparing results within and across organizations.

Key words: patient safety, healthcare benchmarking, safety management, survey

Introduction

Patient safety culture, as defined by the Agency for Healthcare Research and Quality (AHRQ), is 'the product of individual and group values, attitudes, perceptions, competencies and patterns of behavior that determine the commitment to, and the style and proficiency of, an organization's health and safety management' [1].

© The Author 2017. Published by Oxford University Press in association with the International Society for Quality in Health Care. All rights reserved.
For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/intqhc/article-abstract/29/5/685/3978047 by Ministry of Health Iran TRIAL ACCESS user on 13 December 2019

The need to measure patient safety in healthcare appears central and has been the subject of numerous studies [2–6].

Numerous tools have been designed to measure safety culture in various designs and settings. Among them, the most widely used in hospital-wide designs is the Hospital Survey on Patient Safety Culture (HSOPS), set up in 2004 by the US Agency for Healthcare Research and Quality (AHRQ) [1]. The HSOPS is a self-administered questionnaire including 42 items used to calculate composite scores for 12 dimensions of safety culture. Although the main purpose of this questionnaire is the assessment of quality improvement, the results of the HSOPS could have substantial policy implications when used to conduct comparisons within and across organizations. In 2014, 653 US hospitals participated in the sixth nationwide HSOPS [7]. The managers received a feedback report comparing their results to the comparative database, helping them identify strengths and opportunities for improvement in their hospital's patient safety culture. In 2010, the French National Health Authority (Haute Autorité de Santé, HAS) introduced a new criterion relative to the patient safety culture in the national accreditation process of all French hospitals, demonstrating the growing importance of the survey results for healthcare decision-makers [8].

Using the HSOPS, many studies have reported culture safety variations among healthcare facilities, departments or occupational categories of healthcare workers in North America [9–12], Europe [13–22], Asia [23–27] and the Middle East [28–30]. Surprisingly, three different aggregation methods were used to compound dimension-level scores in these studies. Whatever the context or the purpose of the studies, the three methods were used without specific explanation Briefly, dimension scores were calculated (i) as recommended by AHRQ guidelines by dividing the total number of positive responses in the dimension by the total number of responses, (ii) by averaging individual means of the items within a dimension, or (iii) by averaging individual sums of the items within a dimension. See the Methods section for further details.

Several studies found substantial variations for HSOPS results between hospitals and between countries [7, 22, 27]. While HSOPS score differences undoubtedly reflected true variations in patient safety culture, the use of different aggregation rules for the composite scores could bias the comparisons. Although never considered for the HSOPS questionnaire, inconsistencies relating to the scoring methods have already been reported for the use of other composite scores in the field of health service performance studies [31, 32]. Knowing that comparison of HSOPS results could impact safety culture enhancement policies, the possible variability between the three scoring methods raises questions.

The main goal of this study was to assess the variability of safety culture dimension scores and their associated rankings depending on three different scoring strategies. We expect that our findings will guide researchers who plan to compare patient safety culture levels in their units or hospitals based on the most appropriate scoring method.

Methods

Study design

We conducted a cross-sectional study using a self-administered questionnaire. The project was submitted for advisory purposes to an independent protection committee (IRB00006705), and received its approval from the French National Commission for Information Technology and Civil Liberties (CNIL).

Setting

The study was conducted at a single university-affiliated hospital with a capacity of 1836 beds. The study site reported 135 999 stays in 2014. The hospital staff comprised 4422 registered paramedical staff and 642 board-certified physicians, the vast majority of them hospital specialists.

Participants

Eligible participants were full-time or part-time (half-time or more) employees with at least 6 months of employment in clinical, laboratory/pathology, radiology or pharmacy departments.

Data collection

The data collection was conducted anonymously on a volunteer basis, department by department, between April 2013 and September 2014.

Questionnaire and variables

The French version of the HSOPS was used to assess the level of safety culture in our setting. The present analysis takes into account the 42 items covering the 12 dimensions that were included in the original HSOPS questionnaire. Eighteen items were negatively worded

Each item was answered on a five-point Likert scale, from 'Strongly disagree' (1 point) to 'Strongly agree' (5 points), or from 'Never' to 'Always' when relevant. A global safety grade between 'poor' and 'excellent' and the numbers of reported incidents in the past 12 months were also assessed but were not used to compute any of the 12 dimension scores.

Scoring methods

Before analysis, the coding of negatively worded items were reversed so that high-value answers would always be associated with a more highly developed safety culture. Then three methods were compared: (1) the percentage of positive responses, (2) the averaged individual means and (3) the averaged individual sums.

1) Percentage of positive responses (M1)

The first dimension score computation method was the percentage of positive answers defined as values of 4 ('Agree') or 5 ('Strongly agree'). Following AHRQ guidelines, the first step was to obtain, for each item, the percentage of respondents who answered it positively among non-missing answers [1]. Then unweighted averages of those percentages were computed for each dimension. By definition this method was a mean of percentages, hence resulting in dimension scores ranging from zero to 100.

2) Averaged individual means (M2)

The second method first required the case-wise computation of individual means across the three or four items in a dimension, followed by an averaging within dimensions. Given that each individual item response ranged from one to five, M2 dimension scores varied to the same extent.

3) Averaged individual sums (M3)

The last method we found was only present in three studies [23, 24, 27]. First, individual sums were assessed over dimensions with missing

Downloaded from https://academic.oup.com/intqhc/article-abstract/29/5/685/3978047 by Ministry of Health Iran TRIAL ACCESS user on 13 December 2019

responses accounting for zero. Those were then averaged within dimensions. Thus, three-item dimensions scores ranged from 1 to 15, whereas four-items dimension scores ranged from 1 to 20.

Statistical analysis

The usual descriptive statistics were used to describe the sample characteristics. Qualitative variables were reported as numbers and percentages, while quantitative variables were reported as means with standard deviations or medians with the range.

Dimension scores were assessed with M1, M2 and M3 within each of the 14 departments. Rankings were then obtained for each dimension and each scoring method by simply ordering the computed scores, with the first rank attributed to the highest score. Absolute rank differences were calculated for each department in each dimension as the absolute value of the difference between its rank using a method and its rank using another method. The sum of such absolute rank differences between two methods x and y across all departments is noted $SARD_{x,y}$. These SARDs can range from zero (no difference in ranks between the two methods) to $\text{trunc}\left(\frac{c^2}{2}\right)$ (the two methods give opposite rankings) for a variable with c categories. Dimension scores, rankings and SARDs were also computed for the other respondent characteristics included, for illustrative purposes.

For comparability, dimension scores obtained with the M2 and M3 methods were rescaled in order to vary from zero to 100. Since this was done using linear transformations only, they did not impair the ranking process.

Seeking a single way to calculate 95% confidence intervals (95CIs) for dimension scores of all three scoring strategies, we estimated them using the percentile bootstrap method from 5000 samples. Bootstrap estimators of M2 and M3 95CIs where compared with those obtained from normal distribution approximation.

Analyses were performed using Stata software release 11 (StataCorp. 2009. College Station, TX, USA: StataCorp LP).

Results

Respondent characteristics

The overall response rate was 78.6% ($n = 3978$). After the exclusion of 90 questionnaires due to the discovery of exclusion criteria (Fig. 1), the analytical sample consisted of 3888 survey questionnaires. Females accounted for 80.6% of the responses (Table 1).

Scores variability

M2 and M3 strategies globally attributed higher dimension scores than M1. According to M1, the two most developed dimensions were 'teamwork within hospital units' (M1 = 62.7%; M2 = 63.4% and M3 = 65.2%) and 'communication openness' (M1 = 59.4%; M2 = 62.4% and M3 = 62.9%), while the two least developed dimensions were 'hospital handoffs and transitions' (M1 = 25.4%; M2 = 46.9% and M3 = 51.7%) and 'hospital management support' (M1 = 20.2%; M2 = 41.7% and M3 = 46.1%).

In all 12 dimensions, score values varied across scoring methods for a given department. For instance, in the most developed dimension, the score of the worst department (F) increased by more than 10% from M1 to M3, increasing from 46.9% to 57.7% (Fig. 2). In contrast, for the best ranked departments in the same dimension, scores tended to decrease from M1 to M3 on a smaller scale: department G for example decreased its score from 72.4% to 70.2%. This pattern of variability reduction was also found in five other dimensions (see additional figures in Supplemental Digital Content A, which show caterpillar-type plots with M1, M2 and M3 department score point estimates and bootstrap 95CIs for all dimensions). Notably, these six dimensions were the most developed ones in the

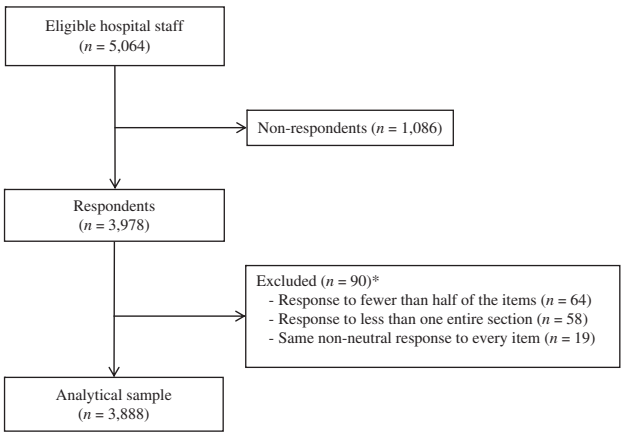


Figure 1 Survey participation.

Table 1 Characteristics of the participants ($n = 3888$)

Characteristics	*N* (%)
Female	3016 (80.6)
Age classes (years)	
Up to 35	1406 (37.5)
36–45	966 (25.7)
46–55	967 (25.7)
56 or older	415 (11.1)
Seniority in current unit (years)	
Up to 1	460 (12.1)
1–2	652 (17.2)
3–5	982 (25.9)
6 or more	1698 (44.8)
Healthcare occupation	2665 (69.9)
Nurse	1323 (34.7)
Nursing assistant	706 (18.5)
Physician	457 (12.0)
Other healthcare	179 (4.7)
Non-healthcare occupation	1148 (30.1)
Technical	775 (20.3)
Administrative	347 (9.1)
Other non-healthcare	26 (0.7)
Department	
Couple–children	591 (15.2)
Orthopedics, traumatology	466 (12.0)
Multidisciplinary medicine	415 (10.7)
Anesthesia, reanimation	337 (8.7)
Thorax, vessels	310 (8.0)
Biology	303 (7.8)
Emergency, acute medicine	302 (7.8)
Gastroenterology, uronephrology	300 (7.7)
Psychiatry, neurology	266 (6.9)
Oncology	158 (4.1)
Operating rooms	153 (3.9)
Imaging	150 (3.9)
Pharmacy	97 (2.5)
Management, home care	32 (0.8)

Downloaded from https://academic.oup.com/intqhc/article-abstract/29/5/685/3978047 by Ministry of Health Iran TRIAL ACCESS user on 13 December 2019

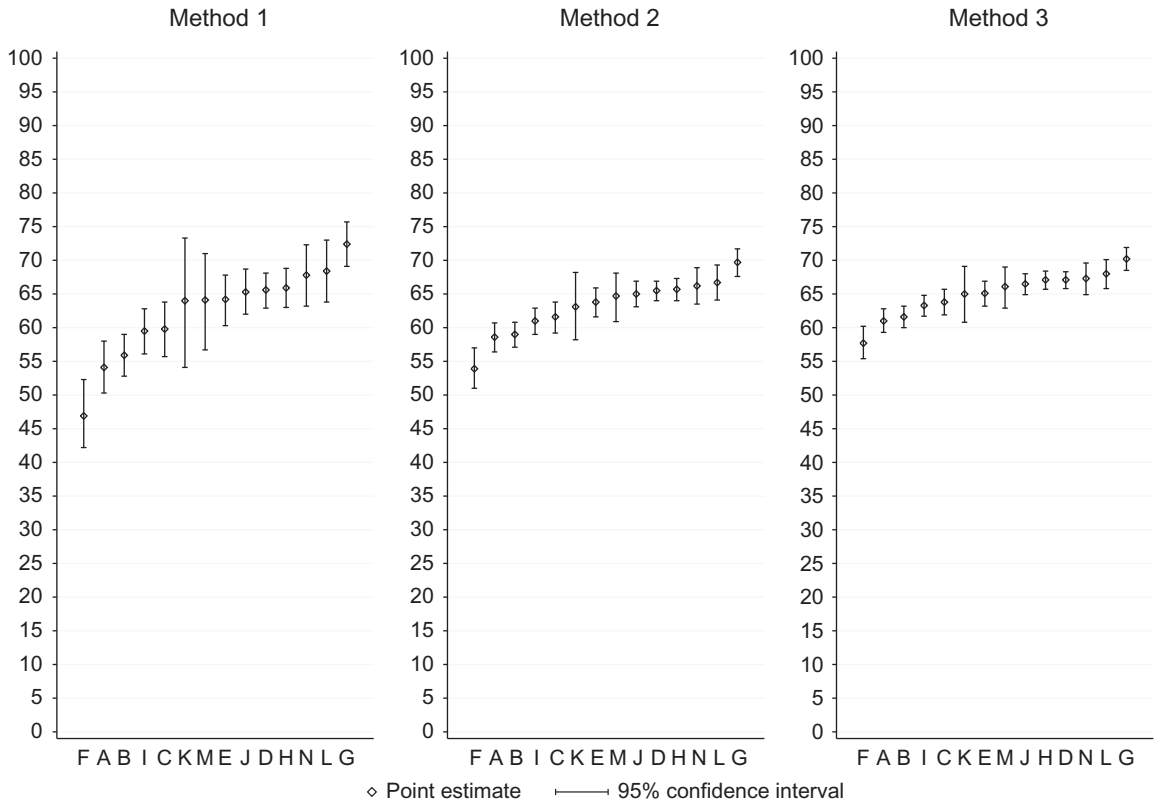


Figure 2 'Teamwork within hospital units' dimension scores. *X*-axis: ordered de-identified departments for each method. *Y*-axis: rescaled score.

study according to M1 scores from the AHRQ, all above the median of 42%.

In the other six dimensions, M1 scores were always lower than their M2 and M3 counterparts. For example, in the least developed 'hospital management support' dimension, the last department according to M1 more than tripled its score, increasing from 14.6% with M1 to 47.5% with M3 (Fig. 3).

Ranking variability

The greatest department ranking discordances between methods in terms of SARD were found in the 'hospital handoffs and transitions' dimension (Table 2). As depicted, SARDs could reach moderately high values with $SARD_{1,2} = 48$, $SARD_{1,3} = 48$ and $SARD_{2,3} = 4$. The closeness shown between M2 and M3 rankings, compared to M1 and M2 or M1 and M3 was observed in all dimensions (see additional tables in Supplemental Digital Content B, which show subgroup scores using M1, M2 and M3 with their associated ranks and SARD for all dimensions).

As for score values, variability patterns were different for the six most developed dimensions and the six least developed dimensions. In the six most developed dimensions cited above, $SARD_{1,2}$ for departments ranged from 2 in 'teamwork within hospital units' to 20 in 'frequency of event reporting,' with a median of 12 (mean = 11.0). Although for the six least developed dimensions these sums were much higher and increased from 12 in 'staffing' to 48 in 'hospital management support,' with a median of 14 (mean = 24.7).

Finally, bootstrap estimations of 95CIs with the percentile method were consistent with normal approximation estimations for M2 and M3 (data not shown). In all dimensions, M1 always had the widest 95CIs while they were quite similar in width between M2 and M3. Moreover, the 95CIs of the scores were almost systematically overlapping within each scoring method. Thus, few departments could be confidently placed in the top or bottom quartiles of the league table.

Discussion

This study highlights the heterogeneity of the results obtained by the three scoring methods used to assess levels of perceived safety culture from the HSOPS, showing that dimensional score values as well as their associated rankings could vary substantially across these methods. In particular, the recommended method, M1, gave lower scores in the six least developed dimensions and overall varied more. In conjunction with the overlapping 95CIs, this led to substantial ranking variations across methods for the 14 departments.

We found seven studies conducted in various settings (three in Europe, three in the Middle East, one in the USA) that computed scores using M1 and M2 [12–14, 18, 28–30]. Many discrepancies between M1 and M2 could be observed: for instance, in the study reported by El-Jardali *et al.* [29], who analyzed a comparable number of questionnaires ($n = 2572$) in a similar setting, the 'overall perception of safety' dimension, which ranked fourth with M1, dropped to the eighth place with M2, while the 'frequency of event reporting' dimension, which ranked eighth, rose to fifth place.

Unweighted means of M1, M2 and M3 dimensional scores along with several summary statistics at a continental level are presented in the Supplemental Digital Content C. Interestingly, the same pattern as in our study appeared. In all studies, greater variability was noted in the six most developed dimensions according to M1, while in the six others it attributed lower dimensional scores [13, 14, 18,

Downloaded from https://academic.oup.com/intqhc/article-abstract/29/5/685/3978047 by Ministry of Health Iran TRIAL ACCESS user on 13 December 2019

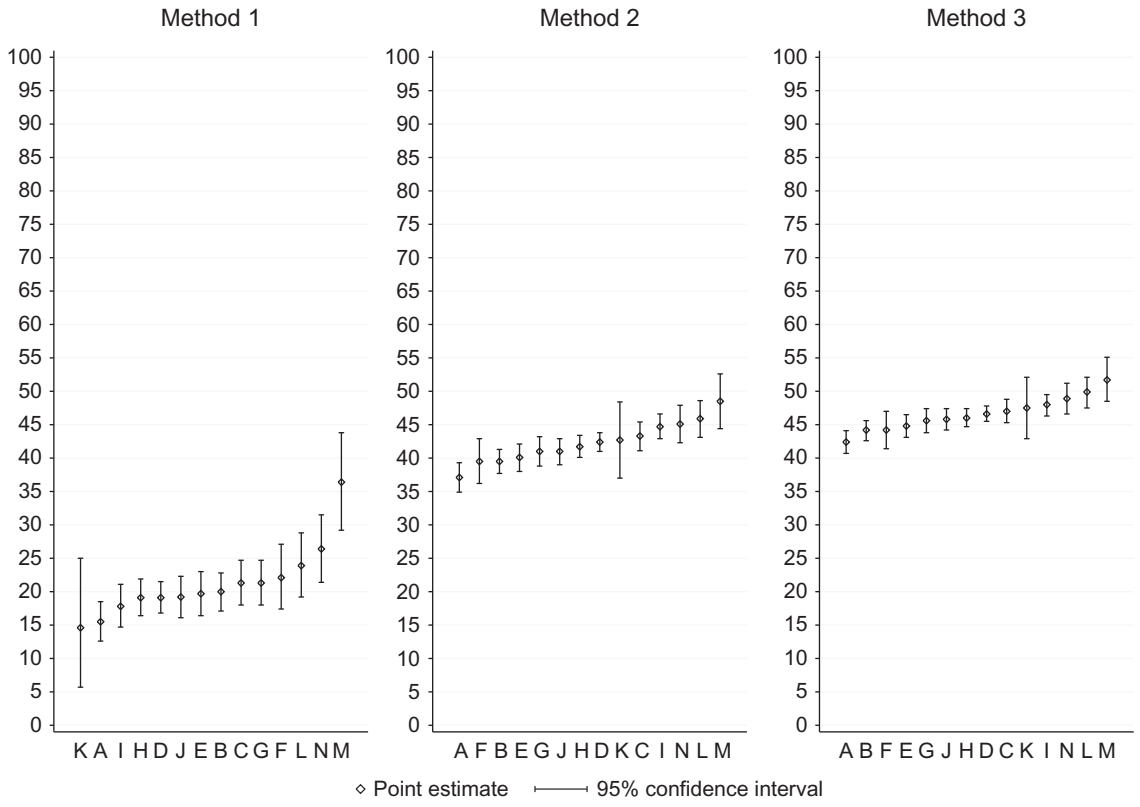


Figure 3 'Hospital management support' dimension scores for each department. *X*-axis: ordered de-identified departments for each method. *Y*-axis: rescaled score.

28, 29]. As an example from the study of Ballangrud *et al.* [14], even though the 'frequency of event reporting' dimension was the least developed one according to both M1 and M2, its associated scores were 18% and 42.5%, respectively. As shown in the Supplemental Digital Content C, this remained true for continental scores as well, despite the diversity of study designs and settings.

Variations in score values and rankings can be explained by the specific properties of the three scoring methods. The main discrepancies were found when comparing M1 with M2 and M1 with M3. M1 is a percentage of positive responses obtained by a prior dichotomization of item responses. This process undoubtedly leads to a loss of information and consequently an increase in variance, which could help explain the lower accuracy illustrated by the widest 95CI. Neutral responses are considered as non-positive, which could provide lower scores for the least developed dimensions. Both M2 and M3 do not transform item responses before computing the scores. Indeed, M2 provides an averaged individual mean and M3 an averaged individual sum of all the items in each dimension. Those two methods preserve all the information from the items and consequently yield more accurate dimensions scores, with narrower 95CIs.

Even if close results between M2 and M3 were found, several arguments tip the scale in favor of M2. M3 is found in only three studies, whereas individual means are already widely used. This method is recommended and validated for most other self-administered questionnaires, in the same field [33] or in other similar depression scales [34]. Finally, the psychometric properties of the original HSOPS questionnaire and most of their translated versions were validated using M2 [1, 16, 17, 21, 24, 35].

Although the HSOPS questionnaire can be used for various purposes, several studies compared HSOPS results between hospitals and between countries. In 2014, 405 281 hospital staff respondents from 653 US hospitals participated in the sixth nationwide HSOPS questionnaire-based survey [7]. This database identifies areas for internal improvement for each hospital and also provides hospital rankings using M1. In 2013, Fujita *et al.* compared HSOPS results between Japan, Taiwan and the USA using M1 and M3 [23]. The development of such benchmarking studies demonstrates the need to agree on the use of the same scoring strategies. These comparisons may impact the policy-making process, including external evaluation such as accreditation, purchasing or incentive policies. Indeed, the ranking of hospitals based on performance composite indicators has become common worldwide. In Great Britain, the National Health Service (NHS) disseminated league tables, rating hospitals according a composite score calculated with 47 indicators, to the public. In 2016, the NHS published the 'Learning from mistakes league' based on a nationwide hospital staff survey, and produced a ranking for 230 hospitals [36]. In this survey, the 30 last hospitals were publicly noted as having a 'Poor reporting culture.' If similar studies used the HSOPS as a methodology, the impact of the variability of the scoring method might have a huge impact for the hospitals standing at the bottom of the league. In the USA, the National Committee for Quality Assurance (NCQA) recently developed comparable hospital rankings based on composite scores up to 100, which reflect member satisfaction and success in preventing and treating illness compared with other Medicare plans [37]. Parallel to this, the Centers for Medicare & Medicaid Services (CMS) and the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO)

Downloaded from https://academic.oup.com/intqhc/article-abstract/29/5/685/3978047 by Ministry of Health Iran TRIAL ACCESS user on 13 December 2019

Table 2 'Hospital handoffs and transitions' subgroup scores using M1[a], M2[b] and M3[c] with their associated ranks and sums of absolute rank differences (SARD)

Variable	*N* (%)	Scores			Ranks			Sums of absolute rank differences (SARD)		
		M1[a]	M2[b]	M3[c]	M1[a]	M2[b]	M3[c]	$SARD_{1,2}$	$SARD_{1,3}$	$SARD_{2,3}$
Age classes (years)								2	2	4
45–54	967 (25.8)	27.0	47.7	52.1	1	1	2			
<35	1406 (37.5)	26.3	47.1	52.2	2	2	1			
35–44	966 (25.7)	24.0	46.1	51.2	3	4	3			
≥55	415 (11.1)	22.2	46.6	50.8	4	3	4			
Occupation								14	12	4
Nursing assistant	706 (18.5)	34.7	48.9	53.6	1	1	1			
Nurse	1323 (34.7)	29.0	46.6	51.7	2	3	3			
Other non-healthcare	26 (0.7)	24.1	45.1	48.2	3	6	7			
Other healthcare	179 (4.7)	23.7	45.7	51.3	4	5	4			
Physician	457 (12.0)	23.2	45.0	49.9	5	7	6			
Technical	775 (20.3)	19.2	47.6	51.8	6	2	2			
Administrative	347 (9.1)	11.4	46.2	51.2	7	4	5			
Department								36	34	14
G	310 (8.0)	31.4	47.1	52.4	1	6	4			
D	591 (15.2)	29.0	48.5	53.0	2	1	2			
F	153 (3.9)	28.7	47.6	52.3	3	4	6			
B	415 (10.7)	28.6	46.8	51.9	4	8	8			
H	466 (12.0)	28.5	47.1	52.2	5	7	7			
J	337 (8.7)	28.4	48.3	52.8	6	2	3			
E	300 (7.7)	26.7	46.0	51.3	7	9	9			
N	158 (4.1)	25.9	47.9	53.2	8	3	1			
A	302 (7.8)	24.8	47.5	52.3	9	5	5			
C	266 (6.9)	21.9	45.3	50.4	10	11	10			
K	32 (0.8)	21.1	41.2	47.3	11	14	14			
M	97 (2.5)	14.4	44.6	48.7	12	12	12			
L	150 (3.9)	13.6	43.7	48.8	13	13	11			
I	303 (7.8)	9.4	45.5	48.5	14	10	13			
Gender								2	2	0
Female	3016 (80.6)	25.6	46.9	51.7	1	2	2			
Male	725 (19.4)	24.6	47.2	52.1	2	1	1			
Speciality								2	0	2
Home care, nursing home, follow-up care	378 (9.7)	32.4	48.5	53.5	1	2	1			
Anesthesia, reanimation, emergency	771 (19.8)	28.8	48.8	53.3	2	1	2			
Surgery, gynecology, obstetrics	990 (25.5)	28.6	47.1	52.1	3	3	3			
Medicine, psychatry	1163 (29.9)	24.9	45.8	51.1	4	4	4			
Biology, imaging, pharmacy	586 (15.1)	11.5	45.2	49.0	5	5	5			

Ordered according to the AHRQ M1 score, $n = 3888$.

[a]M1: Percent positive scores.

[b]M2: Average individual mean.

[c]M3: Average individual sum.

produced several rankings based on composite indicators publicly disseminated on their website as 'Hospital Compare' [38] and 'Quality Check' [39]. One part of the CMS ranking is produced with the results of the composite score of the Survey of patients' experiences (HCAHPS). The use of similar ranking based on the HSOPS may also be considered by similar organizations. In France, the HAS has produced several patient safety indicators for many years and published the hospitals' ranking based on the composite scores on a public website. Moreover, the results of these indicators were recently included in the financial incentive program for quality improvement in healthcare, implemented in 2016 [40]. For the moment, the HSOPS results have not been included in this program, but we can presume a future use, given the promotion of the HSOPS and its use in the French hospitals' national accreditation process [8]. Overall, given our findings on the huge variability according to the scoring methods, healthcare decision-makers should consider development of similar ranking based on the HSOPS with great caution.

This study has several limitations. First, participants were included on a volunteer basis and even with high answer rates this may have created a selection bias. However, it is very unlikely that the extent of this bias could have been heterogeneous among departments. Moreover, this inclusion method has been used by every HSOPS-related study so far. Second, this survey was monocentric. Although the results might apply less broadly, they are still valuable at least for similar settings. As we estimated the dimensional scores for the three strategies on the same sample, this should not challenge the corresponding findings to a significant extent. Third, our study was cross-sectional and therefore could not assess the impact of scoring methods on the trending in patient safety culture over time.

Downloaded from https://academic.oup.com/intqhc/article-abstract/29/5/685/3978047 by Ministry of Health Iran TRIAL ACCESS user on 13 December 2019

The percentage of positive responses method might be best suited for this purpose, considering the widest range of M1 scores for the 12 dimensions observed in the survey. However, only another survey with a longitudinal design can address this question adequately.

Conclusion

The values of the HSOPS scores and their corresponding rankings greatly depended on the computation method used. This finding shows that agreement must be reached on the use of the same scoring strategies, before broadly comparing results within and across organizations. Overall, the variability of the three scoring methods raises questions about the usefulness of the HSOPS results.

Supplementary material

Supplementary material is available at *International Journal for Quality in Health Care* online.

Acknowledgements

This work was developed within the framework of the Grenoble Alpes Data Institute, which is supported by the French National Research Agency under the 'Investissements d'avenir' program (ANR-15-IDEX-02). We acknowledge Linda Northrup for her assistance in editing the manuscript in English.

Funding

This work was supported by a grant from the French Ministry of Health (Programme de recherche sur la performance du système de soins-PREPS 12-026-0001)

References

1. Sorra J, Nieva VF. *Hospital Survey on Patient Safety Culture*. Rockville, MD: Agency for Healthcare Research and Quality, 2004. (Prepared by Westat, under Contract No. 290-96-0004). AHRQ Publication No. 04-0041.
2. An J, Kim SJ, Park S *et al*. The effects of patient education on patient safety: can we change patient perceptions and attitudes? Lessons from the Armed Forces Capital Hospital in Korea. *Int J Qual Health Care* 2017. doi:10.1093/intqhc/mzx037.
3. Mallouli M, Tlili MA, Aouicha W *et al*. Assessing patient safety culture in Tunisian operating rooms: a multicenter study. *Int J Qual Health Care* 2017;**29**:176–82. doi:10.1093/intqhc/mzw157.
4. Tereanu C, Smith SA, Sampietro G *et al*. Experimenting the hospital survey on patient safety culture in prevention facilities in Italy: psychometric properties. *Int J Qual Health Care* 2017;**29**:269–75. doi:10.1093/intqhc/mzx014.
5. Tricarico P, Castriotta L, Battistella C *et al*. Professional attitudes toward incident reporting: can we measure and compare improvements in patient safety culture? *Int J Qual Health Care* 2017;**29**:243–9. doi:10.1093/intqhc/mzx004.
6. Zecevic AA, Li AH, Ngo C *et al*. Improving safety culture in hospitals: facilitators and barriers to implementation of Systemic Falls Investigative Method (SFIM). *Int J Qual Health Care* 2017. doi:10.1093/intqhc/mzx034.
7. Sorra J, Famolaro T, Dyer N *et al*. *Hospital Survey on Patient Safety Culture: 2014 User Comparative Database Report* http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/quality-patient-safety/patientsafetyculture/hospital/2014/hsops14pt1.pdf (2014, date last accessed).
8. Occelli P. La culture de sécurité des soins: du concept à la pratique https://www.has-sante.fr/portail/upload/docs/application/pdf/2011-02/culture_de_securite_des_soins__du_concept_a_la_pratique.pdf (2010, date last accessed).
9. Bowman C, Neeman N, Sehgal NL. Enculturation of unsafe attitudes and behaviors: student perceptions of safety culture. *Acad Med* 2013;**88**:802–10. doi:10.1097/ACM.0b013e31828fd4f4.
10. Jasti H, Sheth H, Verrico M *et al*. Assessing patient safety culture of internal medicine house staff in an academic teaching hospital. *J Grad Med Educ* 2009;**1**:139–45. doi:10.4300/01.01.0023.
11. Jones KJ, Skinner AM, High R *et al*. A theory-driven, longitudinal evaluation of the impact of team training on safety culture in 24 hospitals. *BMJ Qual Saf* 2013;**22**:394–404. doi:10.1136/bmjqs-2012-000939.
12. Sorra JS, Dyer N. Multilevel psychometric properties of the AHRQ hospital survey on patient safety culture. *BMC Health Serv Res* 2010;**10**:199. doi:10.1186/1472-6963-10-199.
13. Agnew C, Flin R, Mearns K. Patient safety climate and worker safety behaviours in acute hospitals in Scotland. *J Safety Res* 2013;**45**:95–101. doi:10.1016/j.jsr.2013.01.008.
14. Ballangrud R, Hedelin B, Hall-Lord ML. Nurses' perceptions of patient safety climate in intensive care units: a cross-sectional study. *Intensive Crit Care Nurs* 2012;**28**:344–54. doi:10.1016/j.iccn.2012.01.001.
15. Bodur S, Filiz E. A survey on patient safety culture in primary healthcare services in Turkey. *Int J Qual Health Care* 2009;**21**:348–55. doi:10.1093/intqhc/mzp035.
16. Hammer A, Ernstmann N, Ommen O *et al*. Psychometric properties of the Hospital Survey on Patient Safety Culture for hospital management (HSOPS_M). *BMC Health Serv Res* 2011;**11**:165. doi:10.1186/1472-6963-11-165.
17. Occelli P, Quenon JL, Kret M *et al*. Validation of the French version of the Hospital Survey on Patient Safety Culture questionnaire. *Int J Qual Health Care* 2013;**25**:459–68. doi:10.1093/intqhc/mzt047.
18. Robida A. Hospital Survey on Patient Safety Culture in Slovenia: a psychometric evaluation. *Int J Qual Health Care* 2013;**25**:469–75. doi:10.1093/intqhc/mzt040.
19. van Noord I, de Bruijne MC, Twisk JW. The relationship between patient safety culture and the implementation of organizational patient safety defences at emergency departments. *Int J Qual Health Care* 2010;**22**:162–9. doi:10.1093/intqhc/mzq013.
20. Verbeek-Van Noord I, Wagner C, Van Dyck C *et al*. Is culture associated with patient safety in the emergency department? A study of staff perspectives. *Int J Qual Health Care* 2014;**26**:64–70. doi:10.1093/intqhc/mzt087.
21. Vlayen A, Hellings J, Claes N *et al*. Measuring safety culture in belgian psychiatric hospitals: validation of the dutch and French translations of the hospital survey on patient safety culture. *J Psychiatr Pract* 2015;**21**:124–39. doi:10.1097/01.pra.0000462605.17725.48.
22. Wagner C, Smits M, Sorra J *et al*. Assessing patient safety culture in hospitals across countries. *Int J Qual Health Care* 2013;**25**:213–21. doi:10.1093/intqhc/mzt024.
23. Fujita S, Seto K, Ito S *et al*. The characteristics of patient safety culture in Japan, Taiwan and the United States. *BMC Health Serv Res* 2013;**13**:20. doi:10.1186/1472-6963-13-20.
24. Ito S, Seto K, Kigawa M *et al*. Development and applicability of Hospital Survey on Patient Safety Culture (HSOPS) in Japan. *BMC Health Serv Res* 2011;**11**:28. doi:10.1186/1472-6963-11-28.
25. Liu C, Liu W, Wang Y *et al*. Patient safety culture in China: a case study in an outpatient setting in Beijing. *BMJ Qual Saf* 2014;**23**:556–64. doi:10.1136/bmjqs-2013-002172.
26. Nie Y, Mao X, Cui H *et al*. Hospital survey on patient safety culture in China. *BMC Health Serv Res* 2013;**13**:228. doi:10.1186/1472-6963-13-228.
27. Wu Y, Fujita S, Seto K *et al*. The impact of nurse working hours on patient safety culture: a cross-national survey including Japan, the United States and Chinese Taiwan using the Hospital Survey on Patient Safety Culture. *BMC Health Serv Res* 2013;**13**:394. doi:10.1186/1472-6963-13-394.
28. El-Jardali F, Jaafar M, Dimassi H *et al*. The current state of patient safety culture in Lebanese hospitals: a study at baseline. *Int J Qual Health Care* 2010;**22**:386–95. doi:10.1093/intqhc/mzq047.

Downloaded from https://academic.oup.com/intqhc/article-abstract/29/5/685/3978047 by Ministry of Health Iran TRIAL ACCESS user on 13 December 2019

29. El-Jardali F, Sheikh F, Garcia NA *et al*. Patient safety culture in a large teaching hospital in Riyadh: baseline assessment, comparative analysis and opportunities for improvement. *BMC Health Serv Res* 2014;**14**:122. doi:10.1186/1472-6963-14-122.
30. Hamdan M, Saleem AA. Assessment of patient safety culture in Palestinian public hospitals. *Int J Qual Health Care* 2013;**25**:167–75. doi:10.1093/intqhc/mzt007.
31. Minvielle E. Analyse critique du développement d'indicateurs composites: le cas de l'infarctus du myocarde après la phase aiguë In: Haute Autorité de Santé; 2010.
32. Jacobs R, Goddard M, Smith PC. How robust are hospital ranks based on composite performance measures? *Med Care* 2005;**43**:1177–84.
33. Pronovost PJ, Weast B, Holzmueller CG *et al*. Evaluation of the culture of safety: survey of clinicians and managers in an academic medical center. *Qual Saf Health Care* 2003;**12**:405–10.
34. Zung WW. A Self-Rating Depression Scale. *Arch Gen Psychiatry* 1965; **12**:63–70.
35. Perneger TV, Staines A, Kundig F. Internal consistency, factor structure and construct validity of the French version of the Hospital Survey on Patient Safety Culture. *BMJ Qual Saf* 2014;**23**:389–97. doi:10.1136/bmjqs-2013-002024.
36. Transparency data. Learning from mistakes league https://www.gov.uk/government/publications/learning-from-mistakes-league (2016, date last accessed).
37. Shetty P, Rivas R, Hristidis V. Correlating ratings of health insurance plans to their providers' attributes. *J Med Internet Res* 2016;**18**:e279. doi:10.2196/jmir.6475.
38. Hospital Compare. https://www.medicare.gov/hospitalcompare/about/what-is-HOS.html (2017, date last accessed).
39. Quality Check. https://www.jointcommission.org/facts_about_quality_check_and_quality_reports/ (2017, date last accessed).
40. Incitation financière à l'amélioration de la qualité (IFAQ) https://www.has-sante.fr/portail/jcms/c_493937/fr/incitation-financiere-a-l-amelioration-de-la-qualite-ifaq-generalisation (2016, date last accessed).

Downloaded from https://academic.oup.com/intqhc/article-abstract/29/5/685/3978047 by Ministry of Health Iran TRIAL ACCESS user on 13 December 2019