

Development of the safety attitudes questionnaire - Korean version (SAQ-K) and its novel analysis methods for safety managers

Abstract

Patient safety has emerged as the key quality indicator of health care, and safety culture among health care workers (HCWs) is also highlighted as it serves a foundational role in safety improvement endeavors. This study developed and validated the Korean version of the Safety Attitudes Questionnaire (SAQ-K) and administered it to HCWs in a large hospital. Analysis results showed a satisfactory level of internal consistency for all six domains of SAQ-K and a good fit based on confirmatory factor analysis. We have also provided novel analysis methods that can be applied to analyze SAQ data even from a large hospital with multiple clinical areas. First, variance components models were developed to examine whether SAQ-K scores differed by clinical area. Likelihood-ratio tests were conducted, and a statistically significant difference was observed for each of the six domains. Then, we applied the empirical Bayes (EB) method to obtain more accurate estimates of SAQ scores for clinical areas with too much variance. Confidence intervals were significantly reduced for those areas, offering improved accuracy, in return for EB means shrinking toward the overall mean. In sum, the current study provides a valuable instrument with which to measure safety culture in Korea, as well as solid methodologies that can be readily applied in health care organizations not only in Korea but also in other countries.

Keywords: safety culture, safety attitude questionnaire, patient safety, variance components model, random effect, empirical bayes method

Volume 2 Issue 1 - 2015

Heon Jae Jeong,¹ Su Mi Jung,² Eun Ae An,³ So Yeon Kim,⁴ HyeYeon Yoon,⁵ Minji Kim,⁶ Byung Joo Song⁴

¹Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, USA

²Data Analytics center, Kangwon National University, Korea

³Leadership Development Division, The Catholic Education Foundation, Korea

⁴Performance Improvement Team, Seoul St. Mary's Hospital, The Catholic University of Korea, Korea

⁵Center for Patient Education, The Care Quality Research Group, USA

⁶Annenberg School for Communication, University of Pennsylvania, USA

Correspondence: Heon-Jae Jeong, Department of Health Policy and Management, Bloomberg School of Public Health, Johns Hopkins University, 624 North Broadway, Rm. 455, Baltimore, 21205, USA, Tel 410-955-5315, Fax 410-955-6959 Email hjeong1@jhu.edu

Received: January 20, 2015 | **Published:** February 06, 2015

Introduction

As the importance of patient safety has been widely acknowledged, numerous improvement programs have been designed and implemented internationally. ¹ These programs aimed not only to solve specific problems such as central line-associated bloodstream infection and patient falls, but also to nurture organizational and individual health care workers' (HCWs') safety culture. ² Although safety culture has been defined in various ways, the core idea involves "the collective beliefs and perceptions of workers regarding the organization and safety of their workplace operations". ³ Safety culture has served as the foundation of safety improvement endeavors. ⁴ Thus, there is high demand for instruments to measure safety culture among HCWs in hospitals.

In response to the demand, various tools have been developed and tested, ⁵⁻⁷ and one of the most embraced and thoroughly validated instruments is the Safety Attitudes Questionnaire (SAQ), which is composed of six domains⁸ (Table 1).

Since its inception, SAQ and its variants have been adopted in diverse health care settings, including intensive care units, ambulatory care, emergency departments, operating rooms, and pharmacies. ^{3,8-10} Also, SAQ has been translated into various languages and validated; therefore, it has provided benchmarks not only for healthcare organizations but also for countries. ^{10,11-15} In addition, SAQ has been used to examine the effectiveness of patient safety programs. ^{2,16,17} Unfortunately, however, despite huge improvements in patient safety, Korea does not have a validated safety culture measurement tool that can be applied nationally, although some healthcare organizations have developed their own instrument. The lack of such tools might have slowed down improvement in patient safety in hospitals across the country.

Table 1 SAQ Domain definitions

SAQ Domain	Definition
Teamwork Climate (TC)	Perceived quality of collaboration between personnel
Safety Climate (SC)	Perception of a strong and proactive organizational commitment to safety
Job Satisfaction (JS)	Positivity about the work experience
Stress Recognition (SR)	Acknowledgment of how performance is influenced by stressors
Perception of Management (PM)	Approval of managerial action
Working Conditions (WC)	Perceived quality of the work environment and logistical support

In addition to the dire need for a validated nationwide tool to measure safety culture, more advanced analytical methods of the collected safety culture data are also required. To date, most analyses of SAQ surveys have been conducted by simply calculating the mean and standard deviation of SAQ domain scores^{14,18,19} or the percentage of respondents reporting positive perceptions of SAQ domains for each cluster, most frequently a clinical area,^{15,20} and then examining patterns across clinical areas. ¹⁹ From a statistical standpoint, this traditional approach leaves huge room for improvement, although its simplicity has attracted hospital safety managers with limited knowledge of statistics.

The most frequently omitted step is testing whether SAQ scores are statistically different across clusters (i.e., testing the effect of clusters in explaining the variance in SAQ before calculating summary statistics like the mean for each cluster). Only if there is

a difference in SAQ scores across clusters should cluster-specific statistics be estimated; otherwise, the overall SAQ score (pooled estimate) for the entire hospital is sufficient. For ordinary cases, analysis of variance (ANOVA) can be used, but for a highly unbalanced dataset with unequal variances, such as SAQ scores over clinical areas, the assumptions of ANOVA are easily violated,²¹ which requires seeking a different approach.

Another issue from the above-mentioned traditional approach arises from many work areas having too small a number of HCWs to provide meaningful information about the precision of clinical area-specific statistics. Compared to hospitals in the US, hospitals in Korea are understaffed, and a lot of clinical areas even have fewer than five HCWs (e.g., outpatient clinics in a hospital). For these small areas, most dispersion measures, such as standard deviation, inter quartile range, or median absolute deviation, cannot be easily calculated, and even if calculated they do not provide much information.

Therefore, we examined methodologies that can be applied to evade the aforementioned issues. First, to test for differences in SAQ scores by cluster, we used the variance components model, which is designed to estimate such between-cluster variances (i.e., within-cluster correlations).²² Meanwhile, the empirical Bayesian (EB) method was used to obtain a more accurate estimate for each cluster by allowing small clusters to borrow information from large clusters.²²⁻²⁵

This study was conducted with three goals:

- To develop and validate the Korean version of SAQ (SAQ-K).
- To apply the variance components model to test whether SAQ scores vary significantly over clinical areas (i.e., whether SAQ scores differ by clinical area), which is one of the most commonly used clustering categories.²⁰
- To obtain precise estimates of SAQ scores of each cluster with the empirical Bayes method.

Methods

Development and validation of SAQ-K

With permission of the original developer of SAQ, a team composed of physicians, nurses, patient safety managers, and communication experts translated SAQ into the Korean language. Not only the language difference but also structural and cultural differences in the health care sector were addressed based on the cross-cultural survey adaptation guidelines from.²⁶ Another researcher who did not participate in the translation, back-translated the Korean version into English to check for unintended distortions in meaning. The Korean version was then pilot tested with various health care professionals, including but not limited to physicians, nurses, pharmacists, paramedics, and managerial personnel. Two negatively worded items (item 2: “In your clinical area, it is difficult to speak up if you perceive a problem with patient care” and item 11: “In this clinical area, it is difficult to discuss errors”) of the original survey were pointed out as difficult to understand and answer properly. This might be attributed to the different language structures of English and Korean; Koreans are not accustomed to answering to negatively worded questionnaire items. Since there was no other way to properly translate these items, they were excluded, and a questionnaire composed of 34

items was developed (TC: 5 items, SC: 6 items, JS: 5 items, SR: 4 items, PM, 10 items, and WC: 4 items).

The questionnaire was administered anonymously to health care professionals working at a large metropolitan hospital in Seoul from October 2013 through November 2013. Participants were allowed to choose the modality of survey between the intranet or paper version. SAQ-K responses were measured on a 5-point Likert scale (1 = Disagree Strongly, 2 = Disagree Slightly, 3 = Neutral, 4 = Agree Slightly, 5 = Agree Strongly) and then converted into a scale of 0 to 100 with the interval of 25 as the original SAQ rubric. For each respondent, six domain scores were obtained by calculating the mean of the items in each domain. Cronbach’s alpha was calculated to examine internal consistency, and confirmatory factor analysis (CFA) was conducted to test construct validity. These analyses were performed with statistical software packages, IBM SPSS Statistics 21 and AMOS 21 (SPSS Inc., Chicago Illinois).

Testing differences in SAQ-K domain scores over clinical areas

The following steps apply to each of the six SAQ-K domains. We began by modeling the SAQ-K score of a certain domain for a person (model 0). This model did not include random effects and is denoted as

$$y_j = \beta + E_j$$

Where β is the overall hospital mean of the SAQ-K scores of all survey participants from a certain domain and E_j is the random deviation of the SAQ-K score of a respondent j from the overall mean β . E_j is assumed to follow a normal distribution with a mean of 0 and variance ϕ^2 . Then we built a variance components model addressing the variance introduced by clinical area (model 1), the model that describes the score of a person j in clinical area i of a certain domain as

$$y_{ij} = \beta + \zeta_i + E_{ij}$$

Where β is the overall hospital mean of SAQ scores for a certain domain and ζ_i is a random deviation of the mean SAQ score for clinical area i from β . ζ_i , the random effect of clinical areas, is assumed to follow a normal distribution with a mean of 0 and variance τ^2 and to be independent over clinical areas. Thus, E_{ij} is the residual within clinical area and random deviation of y_{ij} from the mean of clinical area i . E_{ij} is assumed to be normally distributed with a mean of 0 and variance σ^2 and independent from both clinical areas and survey participants.

Then we tested the null hypothesis, $H_0 : \tau^2 = 0$, with the likelihood-ratio (LR) test to compare the above variance components model (model 1) to the model without the random effects (model 0). Because variance, τ^2 , cannot be negative, the p -value from the LR test should be divided by 2,²² which is usually done automatically by statistical software; henceforth, p -values in this article denote the divided p -value. If the result of the LR test rejects the null hypothesis, $H_0 : \tau^2 = 0$, then we can conclude that clinical area explains the variance in SAQ-K score with statistical significance, and therefore it is worth calculating clinical area-specific SAQ-K scores. If the LR test does not reject the null hypothesis, then the overall mean of SAQ-K for the entire organization is sufficient. For this and the rest of the sections, Stata 13.1 (Stata Corp, College Station, Texas) was used.

Obtaining more accurate estimates of cluster-level SAQ-K scores with the empirical bayes method

The EB method was used in this study to obtain more precise estimates of the SAQ-K domain score for each clinical area. Because this method calculates the cluster estimates by pulling the ordinary mean of unit scores to the overall mean, it is also called empirical Bayes shrinkage. The amount of shrinkage for clinical area is determined by the shrinkage factor B_i .

$$B_i = \frac{\sigma_i^2}{\tau^2 + \sigma_i^2} \tag{3}$$

Where σ_i^2 is the within-cluster variance of clinical area i . Then, the EB mean of clinical area i , $\hat{\theta}_i$ is denoted as

$$\hat{\theta}_i = (1 - B_i)\bar{y}_i + B_i\hat{\mu} \tag{4}$$

Where \bar{y}_i is the mean of clinical area i and $\hat{\mu}$ is the overall mean score of a certain domain of SAQ-K. Therefore, if clinical area i has a wide variance and the shrinkage factor is large, then the EB estimate of the area moves closer to the overall mean accordingly. This shrinkage is prominent usually for the clusters with a small sample size.

Results

Characteristics of respondents

A total of 1,381 questionnaires were returned. After excluding those with too many missing values, especially in clinical area variables, the main interest of this study, 1,142 questionnaires were analyzed. As depicted in (Table 2), most of the survey respondents (73.7%) were female, which is understandable in that most nurses in Korea are female and the number of nurses is much larger than the number of physicians. Health care professionals with 5 to 10 years of experience (25.4%) formed the largest group and were followed by those with 3 to 4 years (21.8%). The tails on both sides of work years accounted for the smallest proportion at 6.7% for respondents with less than 6 months of experience and 5.3% for those with 21 years or more. For job type, nurses constituted the majority at 53.3%, followed by physicians at 33.1%. Supporting staff was 11.6% of the respondents and there was a small number of pharmacists, administrators, and uncategorized HCWs.

Table 2 Characteristics of respondents

Characteristics	N	%
Gender		
Male	300	26.3
Female	842	73.7
Work years		
Less than 6 months	77	6.7
7 ~ 11 months	122	10.7
1 ~ 2 years	193	16.9
3 ~ 4 years	249	21.8
5 ~ 10 years	290	25.4
11 ~ 20 years	150	13.1
More than 21 years	61	5.3
Job type		
Physician	378	33.1
Nurse	609	53.3

Table Continued

Characteristics	N	%
Pharmacist	10	0.9
Supporting Staff	132	11.6
Administration	9	0.8
Other	4	0.4
Total	1142	100.0
0*	222234	
0.	5566666667778888899999	
1*	001222222344	
1.	5788889999	
2*	01233	
2.	55567888	
3*	133	
3.	7	
4*	03	
4.	8	
5*	3	

To examine the dispersion of HCWs by clinical area, we summarized the data in a stem-and-leaf plot, as seen in Figure 1. The numbers of HCWs were heavily dispersed over 72 clinical areas, ranging from 2 to 53 HCWs; many of the areas had fewer than 10 HCWs working in them. Specific names of clinical areas were not of interest in this study and are therefore not listed in this article.

Psychometric properties of SAQ-K

Internal consistency was evaluated with Cronbach's alpha,²⁷ a coefficient of equivalence among multiple items measuring the same construct. The alphas for SAQ-K were 0.836 for TC, 0.841 for SC, 0.907 for JS, 0.734 for SR, 0.928 for PM, and 0.758 for WC. Since alphas greater than .70 is generally considered to be large enough,^{27,28} this suggests that SAQ-K shows high reliability.

The results and fit indices of CFA are provided in (Table 3 & 4). To help readers better understand the CFA results, items in (Table 3) are shown in English as the original SAQ, although actual survey administration and analysis used SAQ-K. GFI, NFI, and CFI were higher than 0.9, AGFI was 0.894, and RMSEA was less than 0.05, indicating a good model fit based on commonly used criteria.^{29,30}

Effects of clinical area on the score of SAQ-K domains

(Table 5) shows the overall means β and standard deviations ϕ of the six SAQ-K domain scores from model 0; the overall means β and standard deviations τ of the clinical area random effect around the overall mean, and standard deviations σ of residuals of SAQ-K scores around the clinical area-specific mean for each of the six domains from (model 1).

The means β s of both models were similar, ranging from approximately 55 (WC) to 69 (SR). Among the six domains, JS showed the largest standard deviation of random effects, τ , at 7.84, and SR showed the smallest τ , at 3.29. Though τ , appeared smaller than the standard deviation σ of the residuals within clinical areas

for all six domains, the LR test results showed that random effects were statistically significant for all six domains, suggesting that SAQ-K scores differed by clinical area and thus clinical area-specific SAQ-K scores are worth estimating.

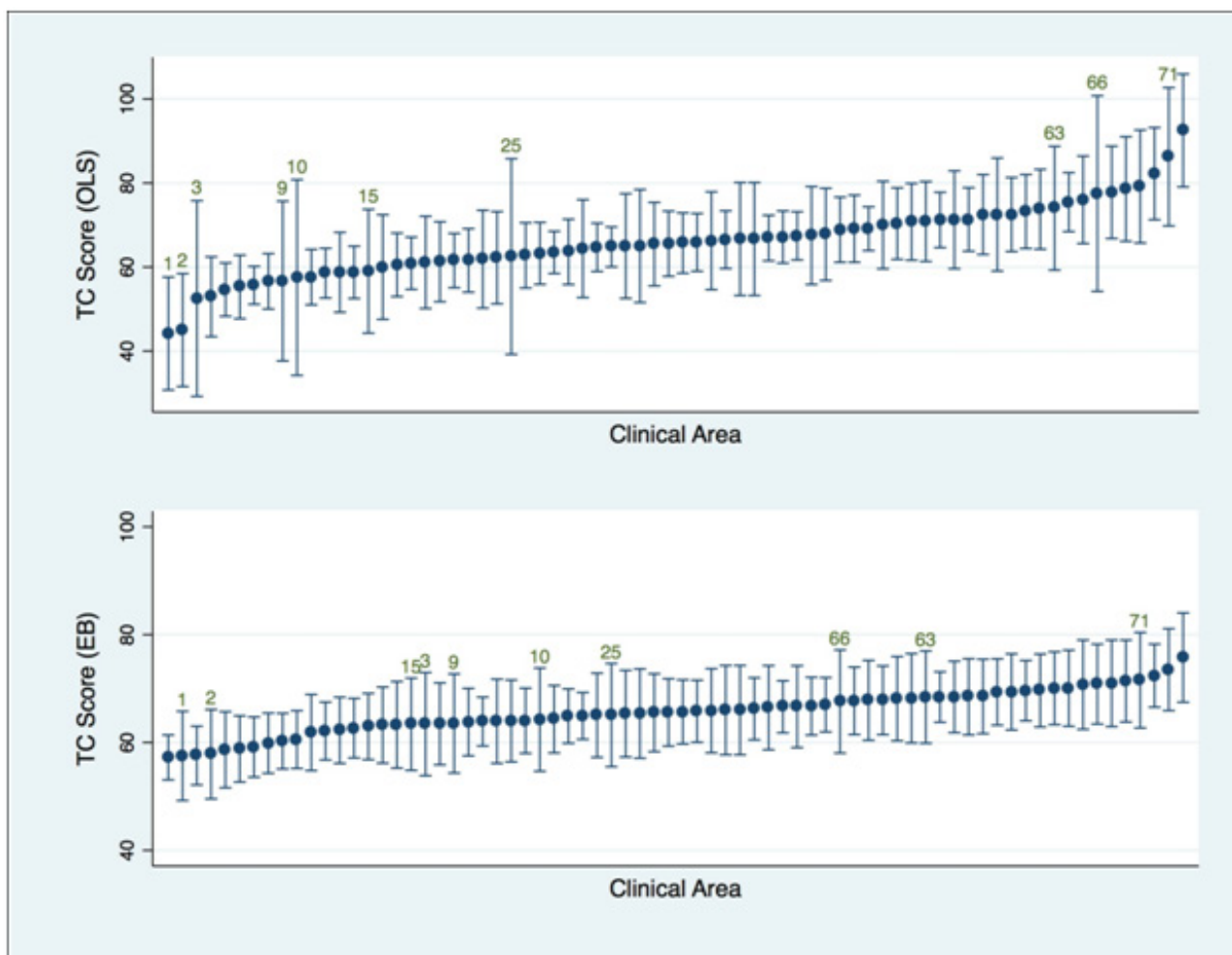


Figure 1 Stem-and-Leaf Plot of the number of respondents by clinical area.

Table 3 Standardized and un standardized coefficients for CFA

Domain	Items	β	B	SE
TC	Nurse input is well received in this clinical area	.68	1.00	
	Disagreements in this clinical area are resolved appropriately (i.e., not who is right, but what is best for the patient)	.76	1.20	.05
	I have the support I need from other personnel to care for patients	.74	1.17	.05
	It is easy for personnel here to ask questions when there is something that they do not understand	.70	1.09	.05
	The physicians and nurses here work together as a well-coordinated team	.68	1.13	.06
	I would feel safe being treated here as a patient	.71	1.00	
SC	Medical errors are handled appropriately in this clinical area	.73	.99	.04
	I know the proper channels to direct questions regarding patient safety in this clinical area	.66	1.04	.05
	I receive appropriate feedback about my performance	.72	.99	.04
	I am encouraged by my colleagues to report any patient safety concerns I may have	.68	1.03	.05
	The culture in this clinical area makes it easy to learn from the errors of others	.62	.84	.04

Table Continued

Domain	Items	β	B	SE
JS	I like my job	.78	1.00	
	Working here is like being part of a family	.82	1.09	.04
	This is a good place to work	.87	1.09	.03
	I am proud to work in this clinical area	.84	1.01	.03
	Morale in this clinical area is high	.79	.93	.04
SR	When my workload becomes excessive, my performance is impaired	.56	1.00	
	I am less effective at work when fatigued	.57	1.10	.08
	I am more likely to make errors in tense or hostile situations	.71	1.41	.09
	Fatigue impairs my performance during emergency situations (e.g., emergency resuscitation, seizure)	.70	1.39	.09
	Unit management supports my daily efforts	.79	1.00	
PM	Hospital management supports my daily efforts	.79	1.00	.04
	Unit management doesn't knowingly compromise patient safety	.58	.73	.04
	Hospital management doesn't knowingly compromise patient safety	.63	.77	.04
	Unit management is doing a good job	.80	.99	.03
	Hospital management is doing a good job	.77	.97	.04
	Problem personnel are dealt with constructively by our unit management	.79	.99	.03
	Problem personnel are dealt with constructively by our hospital management	.75	.94	.03
	I get adequate, timely info about events that might affect my work from unit management	.80	.94	.03
	I get adequate, timely info about events that might affect my work from hospital management	.74	.92	.03
	The levels of staffing in this clinical area are sufficient to handle the number of patients	.59	1.00	
WC	This hospital does a good job of training new personnel	.73	1.28	.08
	All the necessary information for diagnostic and therapeutic decisions is routinely available to me	.73	1.13	.07
	Trainees in my discipline are adequately supervised	.78	1.18	.07

β : Standardized coefficients; B: Standardized coefficient; SE: Standard error of B

Table 4 Confirmatory factor analysis results

Goodness-of-Fit Index	Value
Chi-square Test of Model Fit (df, p-value)	1782.13 (497, $p < 0.01$)
Goodness-of-Fit Index (GFI)	0.911
Adjusted Goodness-of-Fit Index (AGFI)	0.894
Normed Fit Index (NFI)	0.924
Comparative Fit Index (CFI)	0.944
Root Mean Square Error of Approximation (RMSEA)	0.044

Empirical bayesian estimation of SAQ-K domain scores by clinical area

To obtain more accurate estimates of clinical area-specific means, we applied the empirical Bayesian method. (Figure 2) shows the means estimated for the TC domain by ordinary least square (OLS) regression (upper panel) and the EB method (lower panel). Each bar stands for the mean and 95% confidence interval (CI) of a clinical area. The numbers over the CI bars are the clinical area ID numbered by the rank of OLS means; ten clinical areas with the widest CIs by OLS were marked, and these area IDs were marked on the EB means and CIs to clearly show the changes.

For those ten clinical areas, CIs hugely shrunk and their ranks changed, indicating that area mean estimates from the EB method differed from OLS results. Note that any changes in rank were toward the center, the overall TC score mean that was around 65, as denoted in (Table 5). These findings exemplify the shrinkage idea of the EB method, where clusters with large variance borrow information from other clusters to yield more accurate estimates; in other words, CIs become narrower. In return, the means from the EB method are pulled toward the overall mean even to the level at which the rank orders change significantly.

Figure 3 shows the estimates from OLS and the EB method side by side for all six domains, helping the reader understand the effects of the EB method in obtaining more accurate estimates. Due to space limits, IDs of only seven clinical areas are displayed. Although the TC domain was shown in (Figure 2), we show it again to better compare the results.

Discussion

The primary aim of this study was to develop and validate SAQ-K. The results showed that SAQ-K had acceptable internal consistency and construct validity, suggesting that it is ready for use in health care organizations across Korea. Note that two items from the original SAQ, one in the TC domain and the other in PM, were excluded in the final version of SAQ-K. These two items

were negatively phrased and their Korean translation was hard for respondents to answer, which emphasizes the difficulty of cross-cultural adaptation of an instrument and the need for exceptional caution.²⁶

One reason we want to measure safety culture is to predict future safety incidents. Indeed, many studies have shown that safety improvement programs with a component that nurtures

safety culture have significantly reduced the incidence rates of harmful events.^{2,31} Therefore, hospital safety managers naturally ask, “Which areas of my hospital have a low safety culture?” The answer to this question guides resource allocation for safety improvement efforts across various areas of the hospital. In this sense, the more accurate the SAQ scores we obtain, the better we can allocate resources, which will eventually enable a hospital to get the most out of its safety programs.

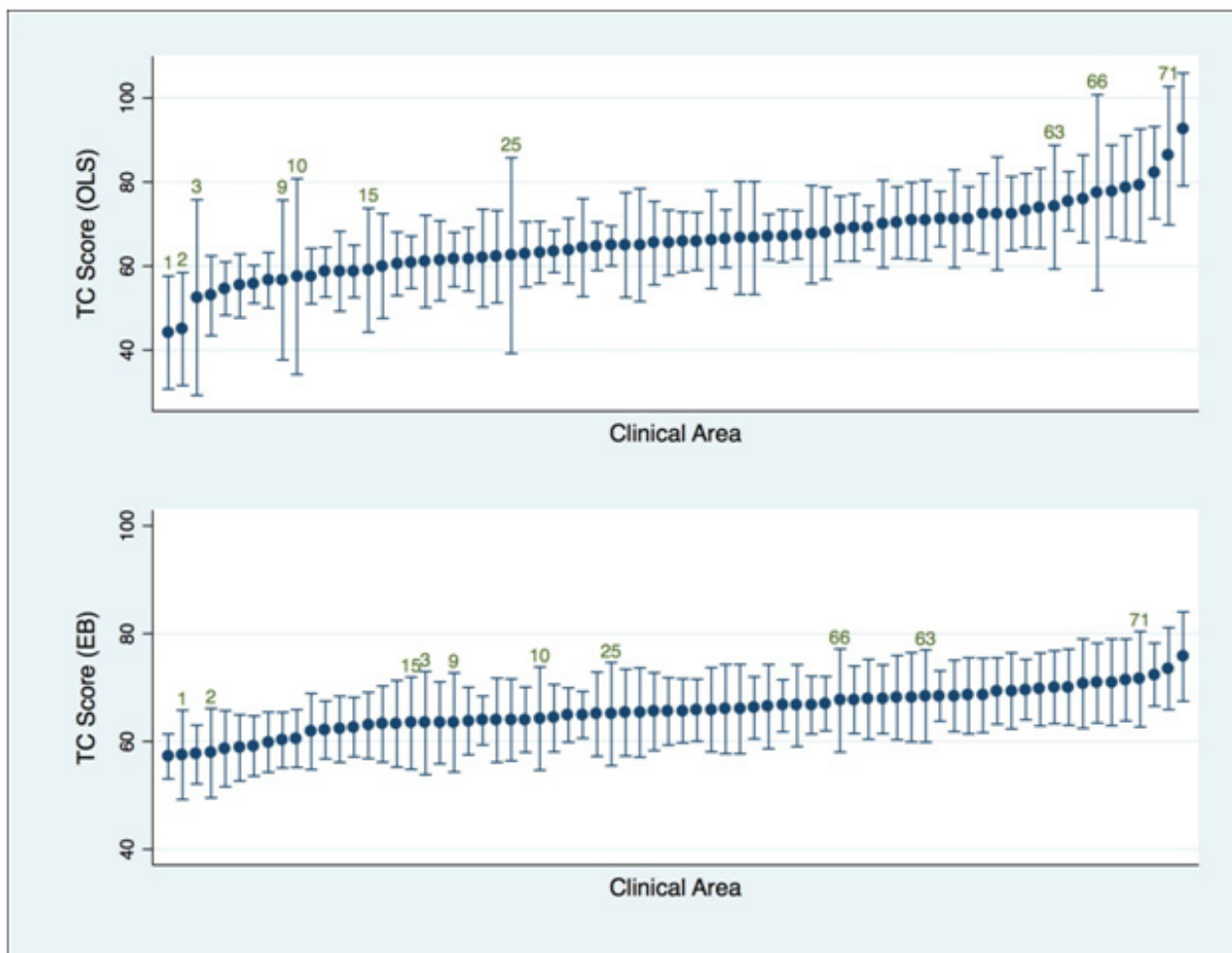


Figure 2 Work area means and 95% confidence intervals of the TC domain of SAQ-K.

Traditional approaches have simply calculated the mean and standard deviation of SAQ domain scores by clinical area, and more likely they only calculated the percentage of respondents with positive perceptions. Both of these approaches are limited in use by overly small sample sizes in many clinical areas. As shown in Figure 1, several clinical areas have only a few HCWs working in them and, therefore, the clinical area mean of such units cannot be estimated accurately due to large standard deviations, as shown in (Figures 2 & 3). However, in the most cases, those simple traditional approaches have prevailed, probably because safety managers more often than not have limited statistical background, especially in advanced statistics like the Bayesian approach.

This study, therefore, applied the empirical Bayes method to provide more accurate mean SAQ domain scores for each clinical

area. The empirical Bayes predictor is the best linear unbiased predictor (BLUP), and in this study the EB method improves the accuracy of area-specific estimates by borrowing strengths from other clinical areas.²² In this process, the mean of each area shrinks toward the overall mean, yielding results as seen in the lower panel of (Figure 2). The advantage of the EB method is that it can be applied even when there are clusters contain very small sample sizes, like one or two,^{22-24,32} which is frequently the case in hospitals in Korea.

Despite the above-mentioned statistical strengths of the EB method, one might point out overlaps in the CIs of clinical areas and therefore this EB approach might not be necessary. Though seemingly logical, this doubt is easily resolved if one understands how SAQ scores are used in hospitals. Due to the scarcity of

resources, a hospital can run safety improvement programs like teamwork training only for a certain number of clinical areas (e.g., ten units in a given year). Selection of those ten units is based on SAQ scores (e.g., bottom ten areas), regardless of whether CIs overlap or not. Therefore, estimating mean SAQ scores of clinical areas as accurately as possible is essential to guaranteeing the just allocation of resources. In addition, SAQ is usually measured annually, leading safety managers to examine changes in SAQ

scores by clinical area over time. One might consider building a longitudinal model with the SAQ score of each HCW as the unit of analysis. However, in most cases, this can be challenging because new HCWs are joining and also many HCWs are leaving the area. This naturally leads us to use the mean SAQ score of a clinical area as the unit of analysis, where a sound longitudinal model can only be built with accurate mean SAQ scores of clinical areas.

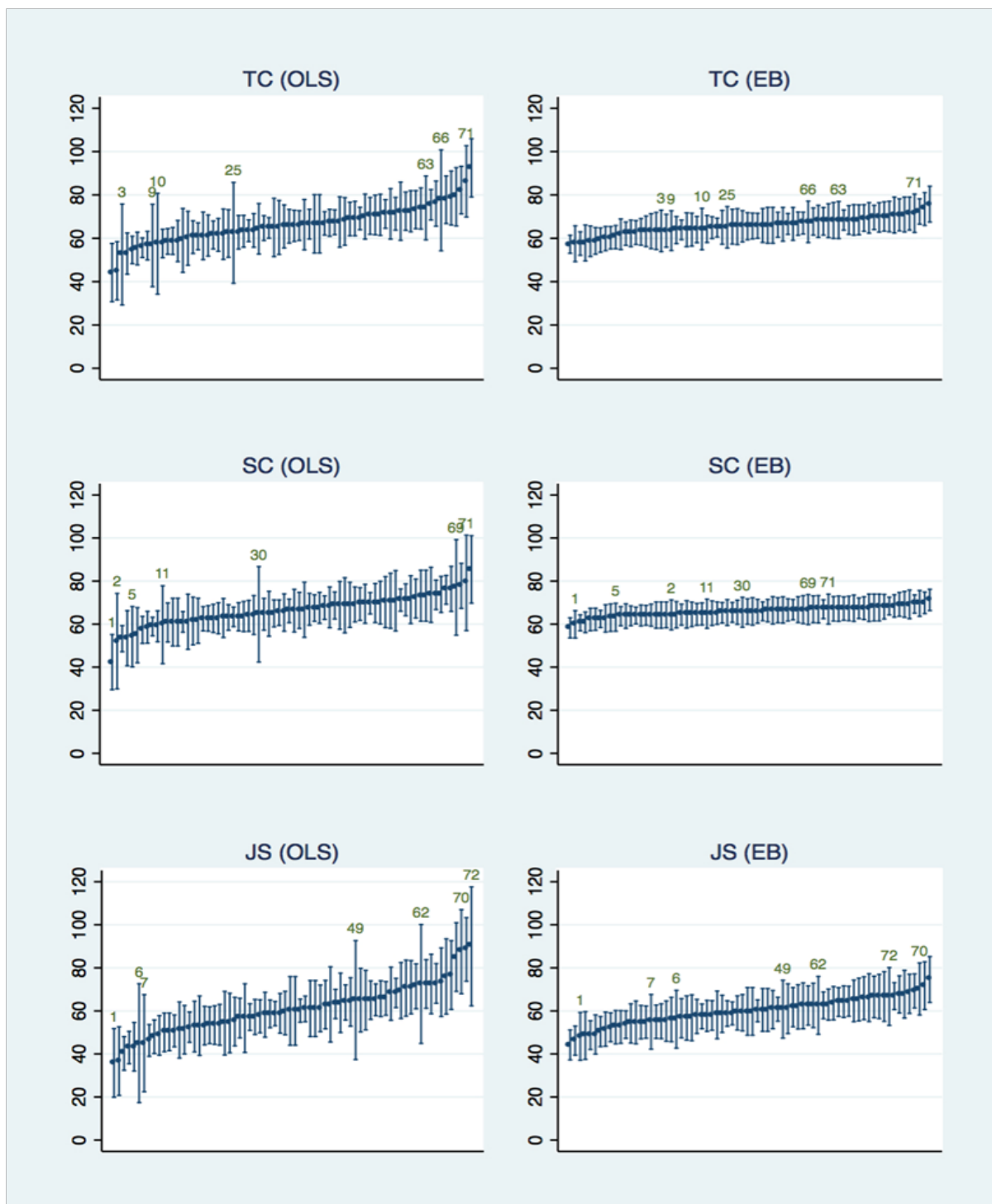


Figure 3 (a) Work area means and 95% confidence intervals of all six domains of SAQ-K.

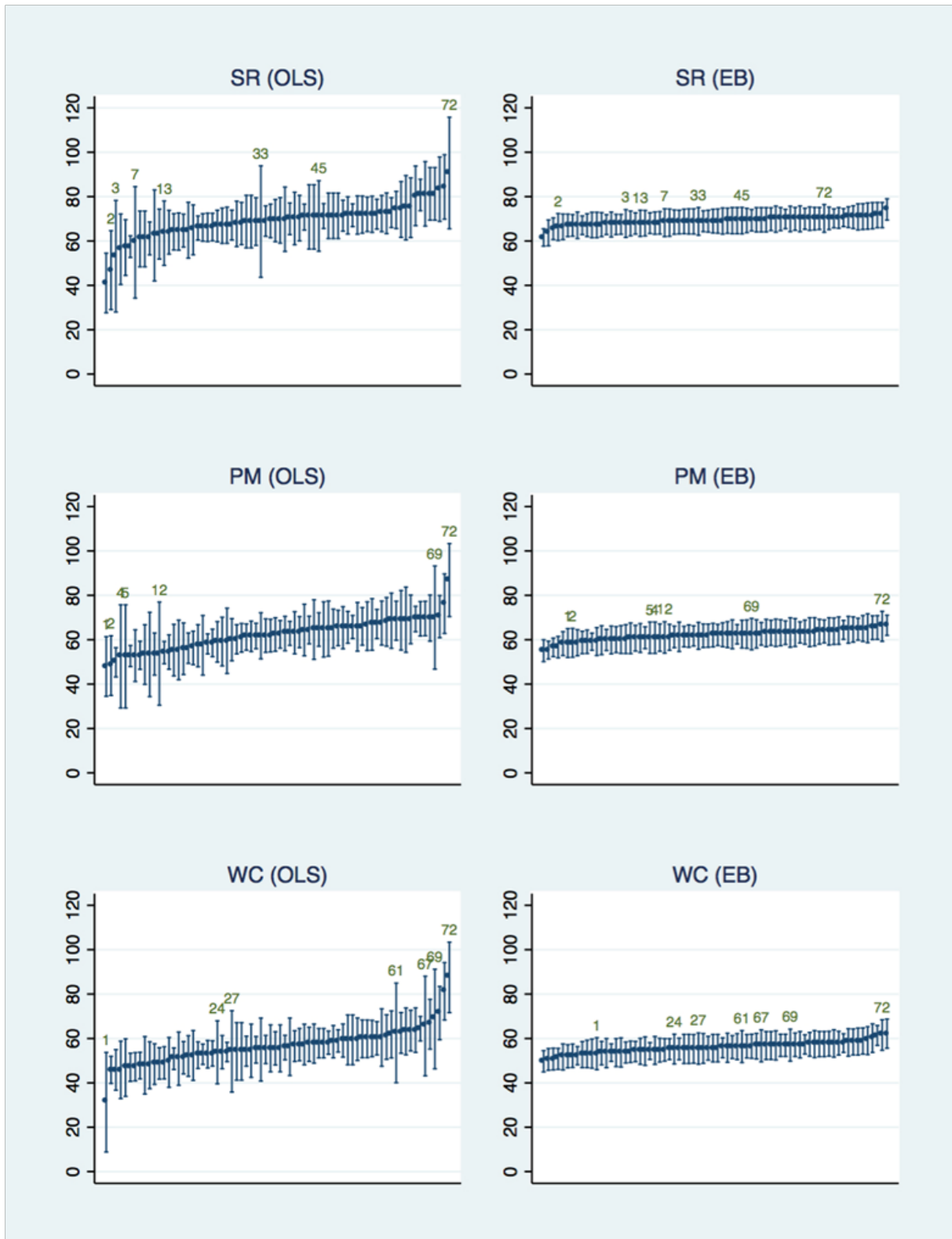


Figure 3 (b) Work area means and 95% confidence intervals of all six domains of SAQ-K.

Table 5 SAQ-K domain scores and variance components model parameters

	TC	SC	JS	SR	PM	WC
Model 0						
β	64.86	65.32	58.55	68.66	61.39	55.23
ϕ	17.60	16.47	21.26	18.51	17.23	16.70
Model 1						
β	65.60	65.65	59.52	68.84	61.78	55.65
τ	5.35*	3.74*	7.84*	3.29*	3.79*	3.88*
σ	16.85	16.03	20.00	18.19	16.78	16.27

*: $p < 0.01$ from the LR test comparing model 1 to model 0 ($H_0: \tau^2 = 0$)

Before we applied the EB method, we tested whether SAQ scores were significantly different among clinical areas. This step was particularly important in this study in that it guided us in judging whether EB estimates were worth applying. For a dataset with not many clusters, one-way ANOVA can be used, but this study covered 72 clinical areas which were heavily unbalanced in sample sizes and had unequal variance, as seen in (Figures 1 & 2). Therefore, we built variance components models with clinical area-level random effect and tested the difference by observing whether the random effect was statistically significant. In this study, we used the likelihood-ratio test because we were only interested in whether the random effect model was necessary, as seen in (Table 5). If the estimates of CIs for random effect are of interest, score tests (Lagrange multiplier tests) can be used instead.^{22,33} Employing variance components models is useful, especially when multiple clustering factors are of interest, such as clinical areas and job type of HCW. In an earlier section, how resources can be allocated over clinical areas was introduced, and the same logic can be applied to job type instead of clinical area. The problem arises when a safety manager needs to decide which clustering factors between clinical area and job type should be used for resource allocation. To illustrate, if variance in TC scores over job type is much bigger than that over clinical area, then the manager can develop a program for the job type with the lowest TC score to match other job types with higher TC scores.³⁴

The variance components model allows us to perform such comparison between variances. For example, by expanding the above variance component model, a crossed random effects model can be built on the assumption that the SAQ score of an HCW is correlated with both clinical area and job type, without assuming hierarchy between clinical area and job type. We can test whether each of the clustering criteria is significant and compare their variances. Also, we can test whether the effect of job type is nested within a clinical area, or the other way around, with hierarchical models. If there are only a few different job types in a hospital and just marginal differences in SAQ scores among job types are of interest, a generalized estimating equation (GEE) with clinical area as the clustering variable³⁵ can be used. Though building a GEE model is relatively simple, it is not discussed in this study because it is not within the scope of the study.

Although we developed and validated SAQ-K successfully and provided methodology for its analysis, some issues remain to be addressed. First, many HCWs work in multiple areas of the hospital. For example, a physician in the Department of Anesthesiology and

Clinical Care Medicine (ACCM) works in both operating rooms and intensive care units. Which clinical area should the physician choose? A more fundamental question might be which area one perceives as the culture in which one is embedded. SAQ defines one's clinical area as the area where one 'typically' spends time, but this is not clear for many HCWs. Also, sufficient time is required for an HCW to be assimilated into an area's culture. If the period from when an HCW begins to work in an area to SAQ administration is only one week, it might not be wise to think that the HCW's responses accurately reflect the area. In hospitals where HCWs frequently change their working areas or where new trainees join regularly, this issue is amplified. These issues probably cannot be solved perfectly. Rather, what we can do is maintain consistency across the hospital. For example, a SAQ participant inclusion criterion in the Johns Hopkins Hospital is "HCWs with a 50% commitment to one clinical area for at least 4 consecutive weeks before the survey administration".¹⁷ Though still arbitrary, this offers better internal validity when a comparison of SAQ scores within a hospital is of interest.

Conclusion

Culture is complex issue. Measuring culture, therefore, is rightly difficult, and safety culture in hospitals is no exception. However, such difficulty cannot justify our negligence in measuring it. Indeed, without measuring safety culture, we will never be able to improve it. Thus, instruments to measure safety culture are crucial. Through this study, we developed the Korean version of such measurement tool, and we also developed methodologies to better analyze it; these methodologies can be used in other countries as well. We hope that this study helps Korean hospitals provide patients in Korea with safer care, the care that patients well deserve.

Contact the corresponding author at hjeong1@jhu.edu for SAQ-K and Stata commands for the analyses in this article.

Acknowledgement

None.

Conflict of interests

Authors declare that there is no conflict of interest.

References

1. Jeong HJ, Pham JC, Kim M, et al. Major cultural-compatibility complex: considerations on cross-cultural dissemination of patient

- safety programmes. *BMJ Qual Saf.* 2012;21(7):612–615.
2. Timmel J, Kent PS, Holzmueller CG, et al. Impact of the Comprehensive Unit-based Safety Program (CUSP) on safety culture in a surgical inpatient unit. *Jt Comm J Qual Patient Saf.* 2010;36(6):252–260.
 3. Patterson PD, Huang DT, Fairbanks RJ, et al. The emergency medical services safety attitudes questionnaire. *Am J Med Qual.* 2010;25(2):109–115.
 4. Dekker S. *Just culture: Balancing safety and accountability.* 2nd ed. Ashgate Publishing Ltd: UK; 2012. 200 p.
 5. Colla J, Bracken A, Kinney L, et al. Measuring patient safety climate: a review of surveys. *Qual Safety in Health Care.* 2005;14(5):364–366.
 6. Etchegaray JM, Thomas EJ. Comparing two safety culture surveys: safety attitudes questionnaire and hospital survey on patient safety. *BMJ Qual Saf.* 2012;21(6):490–498.
 7. Morello RT, Lowthian JA, Barker AL, et al. Strategies for improving patient safety culture in hospitals: a systematic review. *BMJ Qual Saf.* 2013; 22(1): 11–18.
 8. Makary MA, Sexton JB, Freischlag JA, et al. Operating room teamwork among physicians and nurses: teamwork in the eye of the beholder. *J Am Coll Surg.* 2006;202(5):746–752.
 9. Modak I, Sexton JB, Lux TR, et al. Measuring safety culture in the ambulatory setting: the safety attitudes questionnaire-ambulatory version. *J Gen Intern Med.* 2007;22(1):1–5.
 10. Norden-Hagg A, Sexton JB, Kalvemark-Sporrong S, et al. Assessing Safety Culture in Pharmacies: The psychometric validation of the Safety Attitudes Questionnaire (SAQ) in a national sample of community pharmacies in Sweden. *BMC Clin Pharmacol.* 2010;10:8.
 11. Carvalho REFLD, Cassiani SHDB. Cross-cultural adaptation of the Safety Attitudes Questionnaire-Short Form 2006 for Brazil. *Rev Lat Am Enfermagem.* 2012;20(3):575–582.
 12. Deilkas ET, Hofoss D. Psychometric properties of the Norwegian version of the Safety Attitudes Questionnaire (SAQ), generic version (short form 2006). *BMC Health Serv Res.* 2008;8:191.
 13. Lee WC, Wung HY, Liao HH, et al. Hospital safety culture in Taiwan: a nationwide survey using Chinese version safety attitude questionnaire. *BMC Health Serv Res.* 2010;10:234.
 14. Raftopoulos V, Pavlakis A. Safety climate in 5 intensive care units: A nationwide hospital survey using the Greek-Cypriot version of the Safety Attitudes Questionnaire. *J Crit Care.* 2013;28(1):51–61.
 15. Zimmermann N, Kung K, Sereika SM, et al. Assessing the safety attitudes questionnaire (SAQ), German language version in Swiss university hospitals—a validation study. *BMC Health Serv Res.* 2013;13:347.
 16. Berenholtz SM, Lubomski LH, Weeks K, et al. Eliminating central line-associated bloodstream infections: a national patient safety imperative. *Infect Control Hosp Epidemiol.* 2014;35(1):56–62.
 17. Pronovost PJ, Goeschel CA, Marsteller JA, et al. Framework for patient safety research and improvement. *Circulation.* 2009;119(2):330–337.
 18. Relihan E, Glynn S, Daly D, et al. Measuring and benchmarking safety culture: application of the safety attitudes questionnaire to an acute medical admissions unit. *Ir J Med Sci.* 2009;178(4):433–439.
 19. Watts BV, Percarpio K, West P, et al. Use of the safety attitudes questionnaire as a measure in patient safety improvement. *J Patient Saf.* 2010;6(4):206–209.
 20. Singer SJ, Gaba DM, Falwell A, et al. Patient safety climate in 92 US hospitals: differences by work area and discipline. *Med Care.* 2009;47(1):23–31.
 21. Shaw RG, Mitchell-Olds T. ANOVA for unbalanced data: an overview. *Ecology.* 1993;74:1638–1645.
 22. Rabe-Hesketh S, Skrondal A. *Multilevel and Longitudinal Modeling Using Stata.* 2nd ed. Stata Press, Texas: USA; 2002. 562 p.
 23. Berger J. Parametric Empirical Bayes Inference: Theory and Applications: Comment. *Journal of the American Statistical Association.* 1983;78(381):55–57.
 24. Carlin BP, Louis TA. Bayes and empirical Bayes methods for data analysis. *Statistics and Computing.* 1997;7(2):153–154.
 25. Militino A, Ugarte M, Goicoa T. A BLUP synthetic versus an EBLUP estimator: an empirical study of a small area estimation problem. *Journal of Applied Statistics.* 2007;34(2):153–165.
 26. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine.* 2000;25(24):3186–3191.
 27. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16(3):297–334.
 28. Nunnally J. *Psychometric Theory.* McGraw-Hill, New York: USA; 1978. 640 p.
 29. Hair JF, Tatham RL, Anderson RE, et al. *Multivariate data analysis.* 6th ed. Upper Saddle River, Pearson Prentice Hall, New Jersey: USA; 2006. 899 p.
 30. Sexton JB, Helmreich R, Neilands T, et al. The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res.* 2006;6:44.
 31. Pronovost PJ, Needham D, Berenholtz S, et al. An Intervention to Decrease Catheter-Related Bloodstream Infections in the ICU. *N Engl J Med.* 2006;355(26): 2725–2732.
 32. Morris CN. Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association.* 1983;78(381):47–55.
 33. Breusch TS, Pagan AR. The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies.* 1980;47(1):239–253.