The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0952-6862.htm

The hospital survey on patient safety culture in Portuguese hospitals

Instrument validity and reliability

Margarida Eiras
School of Health Technologies of Lisbon, Polytechnic Institute of Lisbon, Lisbon, Portugal

Ana Escoval
National School of Public Health, Lisbon, Portugal

Isabel Monteiro Grillo
Unit of Nutrition and Metabolism, Institute of Molecular Medicine, Faculty of Medicine of Lisbon, Lisbon, Portugal and Santa Maria Hospital, Lisbon, Portugal, and

Carina Silva-Fortes
Center of Statistics and Applications, University of Lisbon, Lisbon, Portugal and School of Health Technologies of Lisbon, Polytechnic Institute of Lisbon, Lisbon, Portugal

Received 28 December 2011
Revised 11 November 2012
Accepted 15 February 2013

Abstract

Purpose – Quantitative instruments to assess patient safety culture have been developed recently and a few review articles have been published. Measuring safety culture enables healthcare managers and staff to improve safety behaviours and outcomes for patients and staff. The study aims to determine the AHRQ Hospital Survey on Patient Safety Culture (HSPSC) Portuguese version's validity and reliability.

Design/methodology/approach – A missing-value analysis and item analysis was performed to identify problematic items. Reliability analysis, inter-item correlations and inter-scale correlations were done to check internal consistency, composite scores. Inter-correlations were examined to assess construct validity. A confirmatory factor analysis was performed to investigate the observed data's fit to the dimensional structure proposed in the AHRQ HSPSC Portuguese version. To analyse differences between hospitals concerning composites scores, an ANOVA analysis and multiple comparisons were done.

Findings – Eight of 12 dimensions had Cronbach's alphas higher than 0.7. The instrument as a whole achieved a high Cronbach's alpha (0.91). Inter-correlations showed that there is no dimension with redundant items, however dimension 10 increased its internal consistency when one item is removed.

Originality/value – This study is the first to evaluate an American patient safety culture survey using Portuguese data. The survey has satisfactory reliability and construct validity.

Keywords Decision making, Assessment, Patient safety

Paper type Research paper

International Journal of Health Care Quality Assurance
Vol. 27 No. 2, 2014
pp. 111-122

0952-6862
DOI 10.1108/IJHCQA-07-2012-0072

Introduction

Like other highly complex organisations, healthcare is vulnerable to error and so are all healthcare environments and professionals involved in complex care processes. Recently and especially since the IOM report (Kohn *et al.*, 1999) almost all countries and healthcare organisation staff are attending to patient-safety issues. Before implementing patient safety programmes, healthcare staff must understand their safety culture (NPSA, 2009). Quantitative instruments designed to assess safety culture have been developed recently and a few review articles have been published (Flin *et al.*, 2006; Singla *et al.*, 2006; Scott *et al.*, 2003). Measuring healthcare safety-culture enables us to identify improvements, safety behaviours and outcomes for both patients and staff. These instruments should also serve as decision making tools, especially for managers.

Safety culture has been defined as "the extent to which individuals and groups will commit to personal responsibility for safety, act to preserve, enhance and communicate safety concerns, strive to actively learn, adapt and modify behaviour based on lessons learned from mistakes, and be rewarded in a manner consistent with these values" (Wiegmann *et al.*, 2004), and Nieva and Sorra, 2003; Wagner, 2007; Vincent, 2006 have used it. When defining safety culture, one tries to include shared values among all organisation members; formal safety issues; contribution; willingness to learn from errors and accidents; attitudes and behaviour towards an organisation's on-going health and safety performance and assuming active roles in error prevention either from the staff or from the leaders' perspective.

Safety culture, therefore, is seen as a continuum, which means that by measuring it, healthcare staff can improve their safety culture by introducing culture change. Our aim, therefore, was to determine the validity and reliability of the Agency for Healthcare Research and Quality (AHRQ) Hospital Survey on Patient Safety Culture, Portuguese version.

Methods

Hospital selection

Northern, central and southern Portuguese hospitals were chosen based on manager commitment to quality and safety and motivation to engage in a patient-safety culture assessment project.

Survey

We followed Nieva and Sorra's (2003) recommendations and selected the AHRQ survey based on:

- the domains it assesses;
- staff groups included in the sample;
- the hospital setting as the survey's focus; and
- the questionnaire's reliability and validity.

Based on the healthcare safety-culture literature (Flin *et al.*, 2006; Flin, 2007; Pronovost and Sexton, 2005; Fleming, 2005; Colla *et al.*, 2005), we selected the Hospital Survey on Patient Safety Culture (HSOPSC). The HSOPSC was developed by AHRQ for hospital settings (Sorra and Nieva, 2004). The questionnaire emphasises patient-safety issues, errors and event reporting. It has 42 items grouped into 12 patient safety-culture dimensions or composites; two outcome dimensions and ten safety dimensions, seven unit-level and three hospital-level (Table I). The questionnaire also includes two questions that ask respondents to provide an overall grade on patient safety for their work area and to indicate the total events they reported over the past 12 months. All items were assessed with a five-point Likert scale reflecting the agreement rate, ranging from (1) "strongly disagree" or "never", to (5) "strongly agree" or "always". There was also a middle category "neither" or "sometimes".

Preparing the Portuguese HSOPSC version

An English language expert, whose native language is Portuguese, translated the questionnaire before an independent translator, who had not seen the original survey translated it back to English. To validate the translated version, eight healthcare quality and safety experts (mangers, physicians, nurses and researchers) were selected. All experts were contacted by e-mail. They assessed questionnaire ambiguity, cultural differences and taxonomy, especially error and patient-safety issues. A chart was adapted from Wagner (2007) for identifying survey dimensions to guarantee interpretation robustness. All experts were asked to complete the Portuguese version and to send it back to the researchers. As a Spanish HSOPSC version (MSPS, 2009) already existed and all experts were fluent in Spanish, they were also asked to complete this version and send it to the researchers by e-mail. Answers were compared to see if they matched. Differences were discussed with the experts to help us understand the main difficulties and problems detected during completing both questionnaires. Again, interpreting some Portuguese words like "error" or "patient safety" were discussed and evaluated.

Distribution

An open session on patient-safety issues took place at all hospitals included in the sample. All hospital staff were invited to attend and questionnaires were distributed to all individuals with direct or indirect clinical contact with patients. Response rates were maximised as follows:

Culture dimensions pertaining to patient safety	Unit-level	Supervisor/manager expectations and actions promoting patient safety (four items)
		Organizational learning – continuous improvement (three items)
		Teamwork within units (three items)
		Communication openness (three items)
		Feedback and communications about errors (three items)
		Non-punitive response to error (three items)
		Staffing (four items)
	Hospital-level	Hospital management (three items)
		Teamwork in the hospital units (four items)
		Hospital handoffs and transitions (four items)
Patient safety outcomes	Overall safety-perceptions (four items)	
	Events reporting (three items)	
Total events reported		
Overall patient-safety grade		

Table I. Patient safety composites and items

- Hospital A (HA) – a Department of Quality member personally distributed 1,063 questionnaire and 231 were returned, representing a 22 per cent response rate.
- Hospital B (HB) – the survey was distributed together with hospital administrative information (paper-based); 411 questionnaires were distributed and 146 were returned, a 36 per cent response rate.
- Hospital C (HC) – 23 units were selected for the sample and two staff members (one nurse and one physician) from each unit were the key focal point for the distribution process; 620 questionnaires were returned, a 24 per cent response rate.

Analysis

Our goal was to assess the HSOPSC validity and reliability in Portuguese hospital contexts, by verifying if the 12 patient-safety culture dimensions or composites were appropriate for the Portuguese population. The SPSS 17.0 software package was used for statistical analysis and the items negatively worded were reverse-scored for further analysis.

Data screening and pre-analysis

Descriptive statistics were used to examine response variability and missing data. To identify and eliminate those items with most missing data, an individual descriptive item analysis was performed. A missing-value analysis was performed to verify if it was necessary to remove questionnaires from the dataset. Composite response-frequencies where also calculated by averaging negative, neutral and positive responses.

Reliability analysis

Cronbach's alpha is a popular reliability testing method. It indicates the extent to which questionnaire items can be treated as a single latent construct. A 0.7 reliability is considered adequate for a survey instrument (Bland and Altman, 1997), although some authors consider 0.6 and higher adequate (Field, 2000). For the entire survey, Cronbach's should be at least 0.9 (Bland and Altman, 1997).

Item analysis

Item analysis, using inter-item and inter-scale correlations, tell us what items should be removed. Low correlations among all items that compose the dimension or items with low correlation with the composite total score should be removed. Spearman's correlations were used.

Validity analysis

Validity refers to how well the instrument measures what it is intended to quantify. Construct validity is considered the most valuable indicator (Sorra and Nieva, 2004). Composite scores and inter-correlations allow us to analyse construct validity. The construct validity of each safety culture dimension would be reflected in composite scores moderately related to one another. Spearman's correlation coefficient was used because normality assumptions were not verified (Sorra and Nieva, 2004).

Confirmatory factor analysis

We used confirmatory factor analysis (CFA) to compare the Portuguese sample factor structure to the factor structure reported for the original HSOPSC (Sorra and Nieva, 2004). We used chi-square divided by degrees of freedom (χ^2/df), where the model fit is considered good if the quotient is less than 2. Less than 5 is acceptable and values greater than 5 are unacceptable. We also used the goodness-of-fit index (GFI), which accounts for the proportion of observed covariance between the manifest variables (items), explained by the fitted model (a concept similar to the coefficient of determination in linear regression). Generally GFI values between 0.9 and 0.95 indicate good fit and GFI values above 0.95 indicate a very good fit. Bentler's Comparative Fit Index (CFI) was used to correct the under-estimation that can occur when samples are small. CFI is independent from the sample size. Values between 0.9 and 0.95 indicate good fit and values equal to or above 0.95 indicate a very good fit. The Tucker-Lewis index (TLI) varies between 0 and 1; values close to 1 indicate a good fit. Parsimony CFI (PCFI) and parsimony GPI (PGFI) are obtained to compensate for the "artificial" improvement in the model, which is achieved simply by adding more parameters; i.e. a more complex model may have better fit than a simpler model (parsimonious). Values between 0.6 and 0.8 indicate a reasonable fit and values above 0.8 a good fit. The index Root Mean Square Error of Approximation (RMSEA) was used to adjust the model simply by adding more parameters. Empirical studies suggest that the model fit is considered good for [0.05, 0.08] and very good for values less than 0.05.

Analysis of variance

A one-way analysis of variance is used when data are divided into groups according to only one factor, in this case hospitals. Our main questions were: is there a significant difference between hospital composite scores? If so, which hospitals are significantly different from the others? If it is assumed that group variances are statistically equal, an F statistic is used. If this assumption is not valid then the resulting F statistic is invalid. The Brown-Forsythe test is a statistical test for group-variances resulting from an ordinary one-way analysis of variance on the absolute deviations from the median. If there are significant differences between hospitals, multiple comparisons should be performed to analyse which hospital pairs are different. The most popular multiple comparisons procedure is Tukey's HSD, when equal variation across observations is verified, or the Games-Howell statistic when equal variances assumption is not verified (Bland and Altman, 1997).

Results

Data screening and pre-analysis

We performed a pre-test with 12 staff members. Their feedback on item comprehension was analysed and reformulated when appropriate. All questionnaires were disseminated on paper that used barcode reading. When a respondent chose two or more options in one item, this item response was considered missing. We did a missing value analysis and only four from 1,113 respondents had completed less than half the items; however, only respondents who responded to all the items were considered. The final dataset totalled 884 questionnaires. Most missing values were found in composite 8 (frequency of events reported), 4.2 to 8.6 per cent of the responses to these items were missing. No items were excluded based on missing values percentage. There were no items with an average negative responses or average positive responses higher than 80 per cent. Average composite response-frequencies were obtained (Table II), namely

Dimensions/composites	Average positive response (%)
1. Teamwork within units	70
2. Supervisor/manager expectations and actions promoting patient safety	63
3. Management support for patient safety	37
4. Organizational learning – continuous improvement	65
5. Overall patient-safety perceptions	54
6. Feedback and communication about error	54
7. Communication openness	53
8. Frequency of events reported	40
9. Teamwork across units	44
10. Staffing	39
11. Handoffs and transitions	54
12. Non-punitive response to errors	25

Table II.
Patient safety composite responses

average negative, neutral and positive responses. Almost all composite responses reflected positive opinions. The lowest positive scores were found in:

- non-punitive response to error, with an overall 25 per cent positive average response, ranging from 13-31 per cent;
- management support for patient safety, with an overall 37 per cent positive average response, ranging from 27-43 per cent; and
- staffing, with an overall 39 per cent positive average response, ranging from 27 to 56 per cent.

The values in Table II were obtained before removing all questionnaires with missing responses, in this case these values were obtained using the leastwise procedure. The dimensions with highest scores were:

- teamwork within units, with an overall 70 per cent positive average response, ranging from 54-78 per cent;
- organizational learning – continuous improvement, with an overall 65 per cent positive average response, ranging from 48-74 per cent; and
- supervisor/manager expectations and actions promoting patient safety, with an overall 63 per cent positive average response, ranging from 49-79 per cent.

Reliability analysis

Reliability analyses, using Cronbach's alpha (α), were performed on 12 composites to ensure that individuals were responding consistently to items (Table III).

There are five dimensions with Cronbach's alphas lower than 0.7; however dimension nine (teamwork across units) has a Cronbach's alpha very close to 0.7. Dimension 10 (staffing) had the lowest value, but removing the item: "we use more agency/temporary staff than is best for patient care", from this dimension increases this value up to 0.57. Dimension eight (frequency of events reported) achieved the highest Cronbach's alpha.

Composite	Items	Cronbach's alpha	Average inter-item correlation	Average inter-scale correlation
1. Teamwork within units	4	0.73	0.42	0.53
2. Supervisor/manager expectations and actions promoting patient safety	4	0.72	0.39	0.51
3. Management support for patient safety	3	0.77	0.52	0.60
4. Organizational learning – continuous improvement	3	0.71	0.45	0.53
5. Overall patient-safety perceptions	4	0.62	0.30	0.41
6. Feedback and communication about error	3	0.76	0.51	0.60
7. Communication openness	3	0.67	0.40	0.48
8. Frequency of events reported	3	0.90	0.76	0.80
9. Teamwork across units	4	0.69	0.36	0.48
10. Staffing	4	0.48	0.19	0.29
11. Handoffs and transitions	4	0.71	0.38	0.50
12. Non-punitive response to errors	3	0.57	0.30	0.38
Entire scale	42	0.91	0.20	0.42

Table III. Internal consistency statistics

Item analysis

Other internal-consistency measures are inter-item correlations and inter-scale correlations. Inter-item correlation analysis measures the internal consistency – how items within a subscale correlate. If the inter-item correlations are high then there is evidence that the items are measuring the same underlying construct. A rule-of-thumb is that the inter-item correlations should be at least 0.3. Items belonging to the same composite will correlate as they measure the same patient-safety culture aspect. Another internal consistency indicator is the item-total correlation or the specific item's correlation with the total items comprising the composite (Table III). Dimension ten (staffing) revealed the lowest values for inter-item and item-total statistics. Removing "we use more agency/temporary staff than is best for patient care" from this dimension increases these values to 0.309 and 0.383, respectively. We also checked whether higher correlations between items occurred. Ideally, every patient-safety culture aspect uniquely contributes towards patient safety culture. Additionally, Bartlet's test demonstrated that the inter-item correlations were sufficient, $\chi^2 = 12983.6$, df $= 861$, $p < 0.001$.

Validity analysis

Composite scores and inter-correlations were examined to assess construct validity (Table IV). A composite score was calculated by averaging all item responses within a dimension. Each safety culture dimension's construct validity is reflected by composite scores moderately related to one another, indicated by correlations between 0.2 and 0.4. Correlations less than 0.2 indicate that two safety-culture dimensions were weakly related. Correlations equalling 0.85 or higher indicate that the dimensions are essentially the same concept and these dimensions could be combined, and some items eliminated.

Dimension 8 (frequency of events reported) correlates with values lower than 0.2, reflecting a weak relation between them. The remaining dimensions showed moderate to high inter-correlations (0.20-0.56). None was exceptionally high, indicating that no two safety culture dimensions appeared to measure the same construct. The highest inter-correlation was calculated between the composites feedback and communication

Table IV.
Composite score means, standard deviations and inter-correlations

	Mean	SD**	1	2	3	4	5	6	7	8	9	10	11	12
1. Teamwork within units	3.7	0.65	1	0.37	0.26	0.47	0.36	0.45	0.38	0.19***	0.34	0.39	0.32	0.30
2. Supervisor/manager expectations and actions promoting patient safety	3.6	0.68		1	0.36	0.39	0.34	0.46	0.41	0.23	0.30	0.24	0.24	0.30
3. Management support for patient safety	3.1	0.75			1	0.45	0.40	0.40	0.26	0.24	0.47	0.24	0.27	0.29
4. Organizational learning – continuous improvement	3.6	0.68				1	0.44	0.53	0.33	0.32	0.40	0.25	0.29	0.29
5. Overall patient-safety perceptions	3.4	0.69					1	0.33	0.27	0.24	0.38	0.37	0.32	0.28
6. Feedback and communication about error	3.4	0.81						1	0.56	0.43	0.34	0.25	0.32	0.26
7. Communication openness	3.4	0.73							1	0.19***	0.25	0.29	0.29	0.29
8. Frequency of events reported	3.1	1.03								1	0.15***	0.09*	0.19***	0.10***
9. Teamwork across units	3.3	0.6									1	0.20	0.46	0.23
10. Staffing	3.0	0.67										1	0.28	0.36
11. Handoffs and transitions	3.5	0.64											1	0.21
12. Non-punitive response to errors	2.8	0.69												1

Notes: *Correlation is significant at the 0.01 level; all other correlations are significant at the 0.05 level; **Empirical standard deviation; ***Correlations less than 0.2 indicate that two safety culture dimensions were weakly related about error and communication openness (r = 0.56). The second highest inter-correlation was between organisational learning and continuous improvement and feedback and communication about errors (r = 0.53). The lowest relationship was between Frequency of events reported and Staffing (r = 0.09).

Confirmatory factor analysis

The model needs to be validated in a different sample from that where the model was adjusted. A common model-validation is the cross-validation strategy when the sample size is large. In this case, two thirds of the whole sample, randomly selected, is used to adjust the model and the remainder is used to evaluate the model's invariance. If the fitted model in the first sample provides a good fit in the second, then we can assume that the model is invariant in the two samples and, if both samples represent the population, it can be concluded that the model is valid for the population. The sample ($n = 884$) was split into approximately a two-thirds for evaluating the quality adjustment ($n1 = 573$) and the remainder ($n2 = 311$) to validate the model. The analysis indicated that Sorra and Nieva's (2004) model is a very good overall fit considering the RMSEA index, and analysing the other indexes all showed a good overall fit (χ^2/df = 2.323, CFI = 0.875, GFI = 0.869; TLI = 0.857; PGFI = 0.724, PCFI = 0.766; RMSEA = 0.048).

The model was also applied to a validation sample ($n2 = 337$) and the values showed a very good fit considering the RMSEA and the qui-square divided by degrees of freedom. Analysing the other indexes all showed a good overall fit (χ^2/df = 1.692, CFI = 0.885, GFI = 0.839; TLI = 0.869; PGFI = 0.7, PCFI = 0.774, RMSEA = 0.047).

Analysis of variance: difference between hospitals

A one-way analysis of variance (ANOVA) was conducted on all 12 safety culture dimensions to determine the extent to which composites scores on these culture scales differed between three hospitals. The F-Snedecor statistic was used if variances were homogeneous, otherwise the Brown-Forsythe F statistic was used (Table V).

Dimension	Test statistic value and p-value
1. Teamwork within units	$F = 0.031^{*}$, $p = 0.969$
2. Supervisor/manager expectations and actions promoting patient safety	$F = 1.465$, $p = 0.232$
3. Management support for patient safety	$F = 22.144$, $p = 0^{**}$
4. Organizational learning – continuous improvement	$F = 1.428$, $p = 0.240$
5. Overall patient-safety perceptions	$F = 2.525$, $p = 0.081$
6. Feedback and communication about error	$F = 3.317^{*}$, $p = 0.037^{**}$
7. Communication openness	$F = 1.437$, $p = 0.238$
8. Frequency of events reported	$F = 0.272$, $p = 0.762$
9. Teamwork across units	$F = 6.803$, $p = 0.001^{**}$
10. Staffing	$F = 3.824$, $p = 0.022^{**}$
11. Handoffs and transitions	$F = 0.906$, $p = 0.404$
12. Non-punitive response to errors	$F = 4.544$, $p = 0.011^{**}$

Notes: *Brown-Forsythe F-statistics, all others are F-statistics; **significant at 0.01 level; ***significant at 0.05 level

Table V.
ANOVA results

There were significant differences between hospitals on the following composite scores: management support for patient safety; feedback and communication about error; teamwork across units; staffing and non-punitive response to error. Among hospitals with significant differences, we undertook pair-wise comparisons using Tukey multiple comparisons and Games-Howell when the variances were not homogeneous (Table VI).

Management support for patient safety is different in all three hospitals; the other dimensions are different mostly between hospital B and C.

Discussion and conclusions

We conclude that the questionnaire has satisfactory reliability as seven of 12 dimensions had $\alpha > 0.7$ and a high global Cronbach's alpha (0.9). Removing an item from the staffing dimension increased internal consistency. The construct validity was satisfactory for all composites. The composite moderate correlations show that there are no two dimensions measuring the same construct. The highest inter-correlation was obtained between the composites feedback and communication about error and communication openness. Considering the CFA analysis, the dimensional structure proposed in the AHRQ Hospital Survey on Patient Safety Culture Portuguese version, we concluded that Sorra and Nieva's (2004) original model needs to be adjusted to the Portuguese scenario.

The dimension management support for patient safety was answered differently by respondents in all hospitals. There are large differences between hospital characteristics (type and size) and management commitment to quality and safety issues. We also compared the Cronbach's alphas results with other countries (Figure 1).

Measuring Portuguese healthcare safety culture is still at a relatively immature development stage. Our study has limitations; hospitals were not randomly selected so our findings may not represent Portuguese hospitals as a whole. We also note that

Dimension	p-value obtained in all pair-wise comparisons within each dimension
3. Management support for patient safety	C vs B $p = 0^{*}$
	C vs A $p = 0^{*}$
	B vs A $p = 0.022^{**}$
6[a]. Feedback and communication about error	C vs B $p = 0.021^{**}$
	C vs A $p = 0.892$
	B vs A $p = 0.139$
9. Teamwork across units	C vs B $p = 0.299$
	C vs A $p = 0.001^{*}$
	B vs A $p = 0.432$
10. Staffing	C vs B $p = 0.041^{**}$
	C vs A $p = 0.172$
	B vs A $p = 0.669$
12. Non-punitive response to errors	C vs B $p = 0.034^{**}$
	C vs A $p = 0.209$
	B vs A $p = 0.472$

Notes: [a]Games-Howell pairwise comparisons, all others are Tukey's pairwise comparisons; *significant at 0.01 level; **significant at 0.05 level

Table VI.
Pairwise comparisons of Hospital A, B, and C

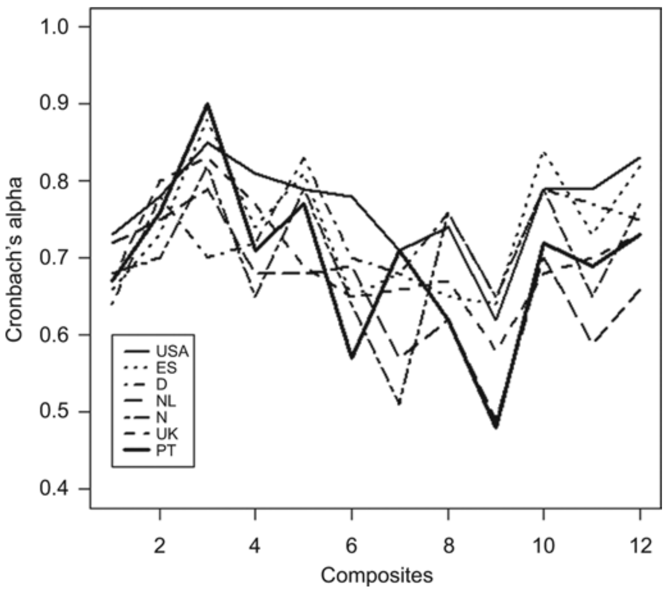

Figure 1.
Cronbach's alphas comparison: USA (Sorra and Dyer, 2010); ES (MSPS, 2009); D (Pfeiffer and Manser, 2010); NL (Smits *et al.*, 2008); N (Haugen *et al.*, 2010); UK (Waterson *et al.*, 2010)

response rates were low; if response rates fall below 60 per cent then the results should be used cautiously (Pronovost and Sexton, 2005). It is unclear, therefore, if these findings can be replicated across Portuguese Hospitals nationwide.

We plan to explore this measurement tool to see how well the data fit the original model and then compare the results with other European publications. Meanwhile, we will conduct a nationwide study by distributing the HSOPSC to all Portuguese hospital staff via a project involving Directorate-General for Health, a research university team and the Portuguese Association for Hospital Development.

Safety culture assessments can empower hospital staff and help them to work on quality and safety improvement strategies to achieve safer environments. This project, which started with communicating survey results to staff and managers, encompasses setting priorities for action, making changes to improve healthcare delivery and measuring their effect on patient safety. We believe that the information we acquire from hospital culture assessment tools is vital to developing effective patient safety strategies and projects in specific healthcare systems.

References

Bland, M. and Altman, G. (1997), "Statistics notes: Cronbach's alpha", *British Medical Journal*, Vol. 314, p. 572.

Colla, B., Bracken, C., Kinney, M. and Weeks, B. (2005), "Measuring patient safety climate: a review of surveys", *Quality and Safety in Health Care*, Vol. 14, pp. 364-366.

Field, A. (2000), *Discovering Statistics Using SPSS for Windows*, SAGE Publications, London.

Fleming, M. (2005), "Patient safety culture measurement and improvement: a 'how to' guide", *Healthcare Quarterly*, No. 8, pp. 14-19.

Flin, R. (2007), "Measuring safety culture in healthcare: a case for accurate diagnosis", *Safety Science*, Vol. 45, pp. 653-667.

Flin, R., Bums, C., Mearns, K., Yule, S. and Robertson, E. (2006), "Measuring safety climate in health care", *Quality and Safety in Health Care*, Vol. 15 No. 2, pp. 109-115.

Haugen, A., Softeland, E., Eide, G., Nortvedt, M., Aase, K. and Harthug, S. (2010), "Patient safety in surgical environments: cross-countries comparison of psychometric properties and results of the Norwegian version of the Hospital Survey on Patient Safety", *BMC Health Services Research*, Vol. 10, p. 279.

Kohn, T., Corrigan, M. and Donaldson, M. (1999), *Is Human: Building a Safer Health System*, 1st ed, National Academy Press, Washington, DC.

Ministerio de Sanidad y Política Social (2009), "Análisis de la cultura sobre seguridad del paciente en el ámbito hospitalario del sistema nacional de salud español", available at: www.msc.es/organizacion/sns/planCalidadSNS/docs/Analisis_cultura_SP_ambito_hospitalario.pdf (accessed 14 June 2011).

Nieva, V. and Sorra, J. (2003), "Safety culture assessment: a tool for improving patient safety in healthcare organizations", *Journal of Quality and Safety in Healthcare*, Vol. 12, pp. ii17-ii23.

NPSA (2009), "National Patient Safety Agency: seven steps to patient safety: the full reference guide", available at: www.npsa.nhs.uk/sevensteps (accessed 25 September 2011).

Pfeiffer, I. and Manser, T. (2010), "Development of the German version of the hospital survey on patient safety culture: dimensionality and psychometric properties", *Safety Science*, Vol. 48, pp. 1452-1462.

Pronovost, P. and Sexton, B. (2005), "Assessing safety culture: guidelines and recommendations", *Quality and Safety in Health Care*, Vol. 14, pp. 231-233.

Scott, T., Mannion, R., Davis, H. and Marshall, M. (2003), "The quantitative measurement of organizational culture in health care: a review of the available instruments", *HSR: Health Services Research*, Vol. 38, p. 923.

Singla, K., Kitch, T., Weissman, S. and Campbell, G. (2006), "Assessing patient safety culture: a review and synthesis of the measurement tools", *Journal of Patient Safety*, Vol. 2 No. 3, pp. 105-115.

Smits, M., Christiaans-Dingelhoff, I., Wagner, C., Van der Wal, G. and Groenewegen, P. (2008), "The psychometric properties of the 'hospital survey on patient safety culture' in Dutch hospitals", *BMC Health Services Research*, Vol. 8, p. 230.

Sorra, J. and Dyer, N. (2010), "Multilevel psychometric properties of the AHRQ hospital survey on patient safety culture", *BMC Health Services Research*, Vol. 10, p. 199.

Sorra, J. and Nieva, F. (2004), "Hospital survey on patient safety culture: 2004 report", AHRQ Publication no. 04-0041, April 2004, Agency for Healthcare Research and Quality, Rockville, MD, available at: www.ahrq.gov/qual/hospsurveydb/

Vincent, C. (2006), *Patient Safety*, 2nd ed., Churchill Livingstone, London.

Wagner, C. (2007), "What role for safety cultures?", working paper, WHO World Alliance for Patient Safety, Patient Safety Research: shaping the European agenda, Oporto, 24-26 September.

Waterson, P., Griffiths, P., Stride, C., Murphy, J. and Hignett, S. (2010), "Psychometric properties of the hospital survey on patient safety culture: findings from the UK", *Quality and Safety in Health Care*, Vol. 19, p. e2.

Wiegmann, A., Zhang, H., Von Thaden, L., Sharma, G. and Gibbons, M. (2004), "Safety culture: an integrative review", *The International Jounal of Aviation Psychology*, Vol. 14 No. 2, pp. 117-134.

Corresponding author
Margarida Eiras can be contacted at: margarida.eiras@estesl.ipl.pt