Hospital Survey on Patient Safety Culture (HSPSC): a systematic review of the psychometric properties of 62 international studies

Patrick Waterson,1 Eva-Maria Carman,2 Tanja Manser,3 Antje Hammer4

ABSTRACT
Objectives To carry out a systematic review of the psychometric properties of international studies that have used the Hospital Survey on Patient Safety Culture (HSPSC).

Design Literature review and an analysis framework to review studies.

Setting Hospitals and other healthcare settings in North and South America, Europe, the Near East, the Middle East and the Far East.

Data sources A total of 62 studies and 67 datasets made up of journal papers, book chapters and PhD theses were included in the review.

Primary and secondary outcome measures Psychometric properties (eg, internal consistency) and sample characteristics (eg, country of use, participant job roles and changes made to the original version of the HSPSC).

Results Just over half (52%) of the studies in our sample reported internal reliabilities lower than 0.7 for at least six HSPSC dimensions. The dimensions ‘staffing’, ‘communication openness’, ‘non-punitive response to error’, ‘organisational learning’ and ‘overall perceptions of safety’ resulted in low internal consistencies in a majority of studies. The outcomes from assessing construct validity were reported in 60% of the studies. Most studies took place in a hospital setting (84%); the majority of survey participants (62%) were drawn from nursing and technical staff. Forty-two per cent of the studies did not state what modifications, if any, were made to the original US version of the instrument.

Conclusions While there is evidence of a growing worldwide trend in the use of the HSPSC, particularly within Europe and the Near/Middle East, our review underlines the need for caution in using the instrument. Future use of the HSPSC needs to be sensitive to the demands of care settings, the target population and other aspects of the national and local healthcare contexts. There is a need to develop guidelines covering procedures for using, adapting and translating the HSPSC, as well as reporting findings based on its use.

INTRODUCTION
Over the course of the last few decades, the field of patient safety has expanded and evolved in a number of directions. Early work concentrated on understanding the extent and the types of human error, which were contributing to patient harm.12 More recently, the focus has shifted towards understanding the role played by human and organisational issues such as leadership, teamwork and communication in contributing to, as well as preventing, adverse incidents.3 4 One prominent organisational factor, safety culture, has in particular generated much interest and discussion.5–8 Use of the term ‘safety culture’ first came about as a result of the Chernobyl nuclear accident9 and has since been used as a way of understanding accidents in a wide variety of industries including aviation, oil and gas and most recently healthcare.10 11 Safety culture is typically taken to refer to ‘the attitudes, beliefs, perceptions and values that employees share in relation to safety in the workplace’.12 Safety culture as a construct is related to ‘safety climate’ and is typically associated with ‘the underlying assumptions and values that guide behaviour in organisations’, whereas safety climate focuses on...
‘the direct perceptions of individuals’ of the underlying culture.13

Patient safety culture
Within healthcare, the design of survey instruments and other methods for measuring patient safety culture (PSC) has expanded considerably in the past few years.6 10 14 In Europe, for example, the European Network for Patient Safety Project15 identified 19 different survey instruments and other methods in use throughout the European Union (EU) member states until 2010. At that time, the most frequently used were the Hospital Survey on Patient Safety Culture (HSPSC,16 used in 12 EU member states), the Manchester Patient Safety Framework17 (used in 3 EU member states) and the Safety Attitudes Questionnaire18 (used in 4 EU member states). A similar review identified 26 studies within Europe, which had used the HSPSC instrument between January 2003 and February 2012.19 In the USA, the Middle East and the Far East, there is likewise evidence of a widening interest in measuring safety culture.20–24

Psychometric properties of PSC instruments
While increased efforts to measure PSC have been welcomed,16 others have suggested the need for a degree of caution. More than a decade ago, Flin et al argued in a series of articles25 26 that the psychometric properties of PSC instruments demonstrated poor levels of reliability and validity. More recently, Pronovost and Sexton27 warned that ‘the enthusiasm for measuring culture may be outpacing the science’. One explanation for low levels of reliability and validity may be the different interpretations of what constitutes a positive PSC among the diverse occupational groupings and subcultures involved in healthcare delivery.28–30 Previous work, for example, has shown that physicians report higher levels of PSC when there is good teamwork across units and a unit-level management that promotes PSC.31 Nursing staff, by contrast, are more likely to report higher levels of PSC safety when they feel that there is enough staff in the unit.31–33 In addition to these concerns, the psychometric properties of PSC instruments may be affected by other factors such as national culture.34 The extent to which healthcare workers are likely to report incidents and errors, for example, has been shown to be related to cultural norms such as fear of ‘losing face’ and the desire to avoid uncertainty and ambiguity,35 characteristics that are often associated with Eastern as compared with Western national cultures.36–38

Motivation for the review and aims
The use of a PSC instrument requires significant investment of time and other resources by healthcare providers. In addition, the outcomes from using the instrument often form the basis with which to identify areas for improvement and intervention within hospitals and other healthcare settings. In other cases, healthcare organisations sometimes use the findings to benchmark levels of PSC across units, departments and hospitals.18 It is therefore very important that these instruments demonstrate acceptable levels of reliability and validity when they are used.

Currently, there are no examples of systematic reviews of the psychometrics of PSC instruments. Hammer et al published a review of the psychometrics of the HSPSC, but this was limited to Europe. The aim of this review was therefore to fill this gap in knowledge and to carry out a review of the international use of psychometric properties of one of the most well-known and widely used surveys on PSC, the HSPSC instrument19 39 In addition, the review examined how the psychometric properties of the HSPSC varied across a range of different types of sample characteristics (eg, country of use, healthcare setting and participant job roles) in order to identify any trends or patterns in the data.

METHODS

Literature search
A literature search for publications that had used the HSPSC was conducted using the Scopus, Web of Science, PubMed and PsycINFO electronic databases. The literature search was initially carried out by the lead author and then checked and updated by the second author. The outcomes from subsequent updates of the literature search were checked by the other authors. The search terms used were ‘HSPSC’ and the following acronyms and terms that have been used to describe the survey, that is, ‘HSPSC’, ‘HSOPSC’, ‘HSOPS’, ‘Survey on Patient Safety Culture’ and ‘SOPS’. All of these search terms were used in conjunction with the Boolean operator ‘OR’. As the original report on the psychometric properties of the HSPSC was published in 2004, the search was limited to studies published between 2004 and July 2018. In addition, citations and references to other publications describing use of the HSPSC were consulted, as well as a list of national contacts provided by the statistical services company (Westat–Agency for Healthcare Research and Quality (AHRQ)40), which provides support to the AHRQ and their work with the HSPSC. Online supplementary file 1 provides details of the full search strategy for the Scopus database.

Inclusion and exclusion criteria
We included all international studies that published psychometric properties on healthcare professionals’ perception of PSC measured with the HSPSC. To be eligible for inclusion in the review, studies had to be published as journal papers, book chapters or reports that are in the public domain (eg, PhD theses) and written in the English language. A wide variety of study designs were included in the review (eg, studies comparing datasets from two or more countries or regions, studies conducted in one country). The original report9 outlining the details of the HSPSC and the process used to compile the survey was also included in the review.
FINDINGS

Search results

The initial search yielded a total of 8604 studies of potential relevance from the four databases. Sixty-two studies met the inclusion criteria and reported psychometric properties on the HSPSC for a total of 67 datasets (one study used the survey in two different countries, 25 two studies used the survey in two different languages in the same country, 42 43 and two studies used the same survey over two different time periods 44 45). The screening and filtering process of the database results is shown in Figure 2.

Psychometric properties

Internal reliabilities

Table 1 shows the values of internal reliabilities using Cronbach’s $\alpha$ coefficient for 67 datasets from the 62 studies. An acceptable value of Cronbach’s $\alpha$ is typically taken as 0.70–0.90, 46–48; however, the first psychometric testing of the HSPSC 16 considered levels of $\alpha \geq 0.60$ as acceptable. Eleven studies did not report values of Cronbach’s $\alpha$ for one or more of the dimensions in the HSPSC. Of the remaining 769 values of Cronbach’s $\alpha$ given in Table 1, there were 355 instances of HSPSC subscales (46.16% of all values) where the internal reliability reported was lower than a Cronbach’s $\alpha$ value of 0.70. The internal reliability for the dimension ‘staffing’ obtained an acceptable level (e.g., Cronbach’s $\alpha$ score >0.70) for only 8 of the 62 datasets that reported this dimension. The dimension ‘frequency of events’ was above the level of acceptability for 61 of the 65 datasets that reported the internal reliability for this item. Thirty-five datasets reported Cronbach’s $\alpha$ values of $\alpha < 0.70$ for 6 or more of the 12 dimensions in the HSPSC. Two datasets, from Lebanon 21 and Turkey, 49 reported values of $\alpha < 0.70$ for 10 dimensions, and another used in Oman 50 reported values of $\alpha < 0.70$ for 11 dimensions.

Five dimensions demonstrated consistently weak internal reliability (i.e., mean $\alpha < 0.70$) across the 67 HSPSC datasets. The dimension staffing was the weakest (mean $\alpha = 0.56$). This was followed by ‘overall perceptions of safety’ (43 datasets, where $\alpha < 0.7$, mean $\alpha = 0.60$), ‘organisational learning’ (47 datasets, where $\alpha < 0.7$, mean $\alpha = 0.63$), ‘communication openness’ (46 datasets where $\alpha < 0.70$, mean $\alpha = 0.64$) and ‘non-punitive response to error’ (37 datasets, where $\alpha < 0.7$, mean $\alpha = 0.65$).

Construct validity using factor analysis

Of the 62 articles, 37 (59.68%) tested construct validity using factor analysis. For a total of 16 articles, both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) had been conducted. For another 3 datasets, EFA had been carried out, and in 18, CFA had been conducted. EFA provides information about the optimal number of factors, whereas CFA determines how well the original 12 factorial structures developed by Sorra and Nieva 16 or alternatives fit the dataset. The appropriateness of the CFA model was assessed by measures of global and local fit. 46 47 51 The report by Sorra and Nieva 16 is
### Table 1  Internal consistencies of the 67 datasets in which the HSPSC was used, from the 62 articles

<table>
<thead>
<tr>
<th>Authors</th>
<th>Country</th>
<th>Language</th>
<th>Hospital management support for patient safety</th>
<th>Supervisor /manager expectations /actions</th>
<th>Teamwork across hospital units</th>
<th>Teamwork within units</th>
<th>Communication openness</th>
<th>Hospital handoffs and transitions</th>
<th>Non-punitive response to error</th>
<th>Feedback and communication about error</th>
<th>Staffing</th>
<th>Organisational learning</th>
<th>Overall perceptions of safety</th>
<th>Frequency of event reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorra and Nieva</td>
<td>USA</td>
<td>English</td>
<td>0.83</td>
<td>0.75</td>
<td>0.80</td>
<td>0.83</td>
<td>0.72</td>
<td>0.80</td>
<td>0.79</td>
<td>0.78</td>
<td>0.63</td>
<td>0.76</td>
<td>0.74</td>
<td>0.84</td>
</tr>
<tr>
<td>Sorra and Dyer</td>
<td>USA</td>
<td>English</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
<td>0.83</td>
<td>0.73</td>
<td>0.81</td>
<td>0.78</td>
<td>0.78</td>
<td>0.62</td>
<td>0.71</td>
<td>0.74</td>
<td>0.85</td>
</tr>
<tr>
<td>Blegen et al</td>
<td>USA</td>
<td>English</td>
<td>0.70</td>
<td>0.65</td>
<td>0.68</td>
<td>0.77</td>
<td>0.64</td>
<td>0.76</td>
<td>0.72</td>
<td>0.82</td>
<td>0.44</td>
<td>0.64</td>
<td>0.53</td>
<td>0.83</td>
</tr>
<tr>
<td>Castle</td>
<td>USA</td>
<td>English</td>
<td>0.81</td>
<td>0.67</td>
<td>0.74</td>
<td>0.75</td>
<td>0.72</td>
<td>0.77</td>
<td>0.70</td>
<td>0.79</td>
<td>0.57</td>
<td>0.67</td>
<td>0.61</td>
<td>0.81</td>
</tr>
<tr>
<td>Castle and Sonon</td>
<td>USA</td>
<td>English</td>
<td>0.66</td>
<td>0.62</td>
<td>0.74</td>
<td>0.80</td>
<td>0.71</td>
<td>0.75</td>
<td>0.72</td>
<td>0.70</td>
<td>0.81</td>
<td>0.77</td>
<td>0.78</td>
<td>0.71</td>
</tr>
<tr>
<td>Eriert et al</td>
<td>USA</td>
<td>English</td>
<td>–</td>
<td>0.83</td>
<td>–</td>
<td>0.80</td>
<td>0.68</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.84</td>
<td>0.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marsteller et al</td>
<td>USA</td>
<td>English</td>
<td>0.85</td>
<td>0.71</td>
<td>0.81</td>
<td>0.78</td>
<td>0.68</td>
<td>0.82</td>
<td>0.68</td>
<td>0.73</td>
<td>0.80</td>
<td>0.76</td>
<td>0.75</td>
<td>0.86</td>
</tr>
<tr>
<td>Wagner et al</td>
<td>USA and</td>
<td>English</td>
<td>0.84</td>
<td>0.81</td>
<td>0.84</td>
<td>0.79</td>
<td>0.74</td>
<td>0.83</td>
<td>0.79</td>
<td>0.78</td>
<td>0.70</td>
<td>0.69</td>
<td>0.76</td>
<td>0.80</td>
</tr>
<tr>
<td>Tomazoni et al</td>
<td>Brazil</td>
<td>Portuguese</td>
<td>0.60</td>
<td>0.74</td>
<td>0.60</td>
<td>0.61</td>
<td>0.64</td>
<td>0.64</td>
<td>0.47</td>
<td>0.72</td>
<td>0.46</td>
<td>0.74</td>
<td>0.43</td>
<td>0.88</td>
</tr>
<tr>
<td>Silva-Batalha and Melleiro</td>
<td>Brazil</td>
<td>Portuguese</td>
<td>0.72</td>
<td>0.77</td>
<td>0.56</td>
<td>0.68</td>
<td>0.63</td>
<td>0.70</td>
<td>0.40</td>
<td>0.67</td>
<td>0.66</td>
<td>0.60</td>
<td>0.47</td>
<td>0.87</td>
</tr>
<tr>
<td>Andrade et al</td>
<td>Brazil</td>
<td>Portuguese</td>
<td>0.74</td>
<td>0.73</td>
<td>0.67</td>
<td>0.70</td>
<td>0.59</td>
<td>0.73</td>
<td>0.53</td>
<td>0.74</td>
<td>0.59</td>
<td>0.61</td>
<td>0.63</td>
<td>0.87</td>
</tr>
<tr>
<td>Vlayen et al</td>
<td>Belgium</td>
<td>Dutch</td>
<td>0.72</td>
<td>0.77</td>
<td>0.66</td>
<td>0.66</td>
<td>0.65</td>
<td>0.71</td>
<td>0.68</td>
<td>0.78</td>
<td>0.57</td>
<td>0.59</td>
<td>0.58</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>French</td>
<td>0.77</td>
<td>0.75</td>
<td>0.68</td>
<td>0.82</td>
<td>0.72</td>
<td>0.72</td>
<td>0.64</td>
<td>0.76</td>
<td>0.52</td>
<td>0.59</td>
<td>0.63</td>
<td>0.87</td>
</tr>
<tr>
<td>Vlayen et al</td>
<td>Belgium</td>
<td>Dutch</td>
<td>0.72</td>
<td>0.77</td>
<td>0.69</td>
<td>0.65</td>
<td>0.66</td>
<td>0.70</td>
<td>0.70</td>
<td>0.76</td>
<td>0.55</td>
<td>0.50</td>
<td>0.54</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>French</td>
<td>0.79</td>
<td>0.74</td>
<td>0.66</td>
<td>0.84</td>
<td>0.71</td>
<td>0.70</td>
<td>0.68</td>
<td>0.70</td>
<td>0.50</td>
<td>0.58</td>
<td>0.58</td>
<td>0.84</td>
</tr>
<tr>
<td>Stoyanova et al</td>
<td>Bulgaria</td>
<td>Bulgarian</td>
<td>0.86</td>
<td>0.81</td>
<td>0.65</td>
<td>0.63</td>
<td>0.77</td>
<td>0.92</td>
<td>0.75</td>
<td>0.63</td>
<td>0.33</td>
<td>0.61</td>
<td>0.34</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bulgarian</td>
<td>0.81</td>
<td>0.81</td>
<td>0.65</td>
<td>0.63</td>
<td>0.77</td>
<td>0.92</td>
<td>0.75</td>
<td>0.63</td>
<td>0.33</td>
<td>0.61</td>
<td>0.34</td>
<td>0.86</td>
</tr>
<tr>
<td>Br borovic et al</td>
<td>Croatia</td>
<td>Croatian</td>
<td>0.73</td>
<td>0.79</td>
<td>0.79</td>
<td>0.76</td>
<td>0.64</td>
<td>0.64</td>
<td>0.62</td>
<td>0.74</td>
<td>0.35</td>
<td>0.53</td>
<td>0.57</td>
<td>0.91</td>
</tr>
<tr>
<td>Boussat et al</td>
<td>France</td>
<td>French</td>
<td>0.66</td>
<td>0.81</td>
<td>0.56</td>
<td>0.71</td>
<td>0.63</td>
<td>0.60</td>
<td>0.60</td>
<td>0.66</td>
<td>0.60</td>
<td>0.62</td>
<td>0.65</td>
<td>0.84</td>
</tr>
<tr>
<td>Occelli et al</td>
<td>France</td>
<td>French</td>
<td>0.73</td>
<td>0.83</td>
<td>0.59</td>
<td>0.63</td>
<td>0.62</td>
<td>0.66</td>
<td>0.57</td>
<td>0.64</td>
<td>0.46</td>
<td>0.59</td>
<td>0.67</td>
<td>0.84</td>
</tr>
<tr>
<td>Kuormanen et al</td>
<td>Finland</td>
<td>Finnish</td>
<td>0.79</td>
<td>0.78</td>
<td>0.72</td>
<td>0.81</td>
<td>0.52</td>
<td>0.71</td>
<td>0.72</td>
<td>0.67</td>
<td>0.62</td>
<td>0.60</td>
<td>0.66</td>
<td>0.67</td>
</tr>
<tr>
<td>Turunen et al</td>
<td>Finland</td>
<td>Finnish</td>
<td>0.83</td>
<td>0.75</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.79</td>
<td>0.78</td>
<td>–</td>
<td>–</td>
<td>0.74</td>
<td>0.84</td>
</tr>
<tr>
<td>Hammer et al</td>
<td>Germany</td>
<td>German</td>
<td>0.87</td>
<td>0.69</td>
<td>0.78</td>
<td>0.77</td>
<td>0.64</td>
<td>0.83</td>
<td>0.73</td>
<td>0.79</td>
<td>0.73</td>
<td>0.61</td>
<td>0.73</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Continued
### Table 1 Continued

<table>
<thead>
<tr>
<th>Authors</th>
<th>Country</th>
<th>Language</th>
<th>Hospital management support for patient safety</th>
<th>Supervisor/manager expectations/actions</th>
<th>Teamwork across hospital units</th>
<th>Teamwork within units</th>
<th>Communication openness</th>
<th>Hospital handoffs and transitions</th>
<th>Non-punitive response to error</th>
<th>Feedback and communication about error</th>
<th>Staffing</th>
<th>Organisational learning</th>
<th>Overall perceptions of safety</th>
<th>Frequency of event reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gambashidze et al 84</td>
<td>Germany</td>
<td>German</td>
<td>0.83</td>
<td>0.75</td>
<td>0.79</td>
<td>0.78</td>
<td>−*</td>
<td>−†</td>
<td>0.73</td>
<td>0.83</td>
<td>0.79</td>
<td>−†</td>
<td>−†</td>
<td>0.87</td>
</tr>
<tr>
<td>Kapaki and Souliotis 85</td>
<td>Greece</td>
<td>Greek</td>
<td>0.79</td>
<td>0.84</td>
<td>0.88</td>
<td>0.80</td>
<td>0.77</td>
<td>−††</td>
<td>0.71</td>
<td>−‡‡</td>
<td>0.60</td>
<td>0.60</td>
<td>−††</td>
<td>0.82</td>
</tr>
<tr>
<td>Tereanu et al 86</td>
<td>Italy</td>
<td>Italian</td>
<td>0.82</td>
<td>0.75</td>
<td>0.70</td>
<td>0.85</td>
<td>0.79</td>
<td>0.77</td>
<td>0.71</td>
<td>0.76</td>
<td>−‡‡</td>
<td>−†††</td>
<td>−††</td>
<td>0.93</td>
</tr>
<tr>
<td>Smits et al 87</td>
<td>The Netherlands</td>
<td>Dutch</td>
<td>0.68</td>
<td>0.70</td>
<td>0.72</td>
<td>0.66</td>
<td>0.72</td>
<td>0.65</td>
<td>0.69</td>
<td>0.78</td>
<td>0.58</td>
<td>−§§</td>
<td>0.64</td>
<td>0.79</td>
</tr>
<tr>
<td>Verbeek-Van Noord et al 88</td>
<td>The Netherlands</td>
<td>Dutch</td>
<td>0.72</td>
<td>0.78</td>
<td>0.78</td>
<td>0.60</td>
<td>0.67</td>
<td>0.66</td>
<td>0.55</td>
<td>0.75</td>
<td>0.47</td>
<td>−†††</td>
<td>−†</td>
<td>0.67</td>
</tr>
<tr>
<td>Olsen et al 89</td>
<td>Norway</td>
<td>Norwegian</td>
<td>0.79</td>
<td>0.77</td>
<td>0.65</td>
<td>0.77</td>
<td>0.68</td>
<td>0.65</td>
<td>0.64</td>
<td>0.70</td>
<td>0.65</td>
<td>0.51</td>
<td>0.76</td>
<td>0.82</td>
</tr>
<tr>
<td>Olsen and Aase 90</td>
<td>Norway</td>
<td>Norwegian</td>
<td>0.78</td>
<td>0.76</td>
<td>0.68</td>
<td>0.74</td>
<td>0.65</td>
<td>0.68</td>
<td>0.67</td>
<td>0.69</td>
<td>0.68</td>
<td>0.53</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>Haugen et al 91</td>
<td>Norway</td>
<td>Norwegian</td>
<td>0.80</td>
<td>0.85</td>
<td>0.85</td>
<td>0.73</td>
<td>0.75</td>
<td>0.67</td>
<td>0.68</td>
<td>0.73</td>
<td>0.59</td>
<td>0.64</td>
<td>0.78</td>
<td>0.82</td>
</tr>
<tr>
<td>Storm et al 92</td>
<td>Norway</td>
<td>Norwegian</td>
<td>0.74</td>
<td>0.79</td>
<td>0.82</td>
<td>0.83</td>
<td>0.46</td>
<td>0.69</td>
<td>0.73</td>
<td>0.63</td>
<td>0.63</td>
<td>0.67</td>
<td>0.68</td>
<td>0.70</td>
</tr>
<tr>
<td>Eiras et al 93</td>
<td>Portugal</td>
<td>Portuguese</td>
<td>0.77</td>
<td>0.72</td>
<td>0.69</td>
<td>0.73</td>
<td>0.67</td>
<td>0.71</td>
<td>0.57</td>
<td>0.76</td>
<td>0.48</td>
<td>0.71</td>
<td>0.62</td>
<td>0.90</td>
</tr>
<tr>
<td>Tereanu et al 94</td>
<td>Romania</td>
<td>Romanian</td>
<td>0.73</td>
<td>0.78</td>
<td>0.74</td>
<td>0.75</td>
<td>0.83</td>
<td>−***</td>
<td>0.68</td>
<td>−†††</td>
<td>−†††</td>
<td>0.56</td>
<td>−†††</td>
<td>−†††</td>
</tr>
<tr>
<td>Sarac et al 95</td>
<td>UK</td>
<td>English</td>
<td>0.79</td>
<td>0.79</td>
<td>0.70</td>
<td>0.80</td>
<td>0.73</td>
<td>0.74</td>
<td>0.77</td>
<td>0.78</td>
<td>0.60</td>
<td>0.64</td>
<td>0.71</td>
<td>0.84</td>
</tr>
<tr>
<td>Waterson et al 96</td>
<td>UK</td>
<td>English</td>
<td>0.69</td>
<td>0.68</td>
<td>0.70</td>
<td>0.73</td>
<td>0.67</td>
<td>0.77</td>
<td>0.65</td>
<td>0.80</td>
<td>0.58</td>
<td>0.66</td>
<td>0.67</td>
<td>0.83</td>
</tr>
<tr>
<td>Robida 97</td>
<td>Slovenia</td>
<td>Slovene</td>
<td>0.82</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.66</td>
<td>0.81</td>
<td>0.72</td>
<td>0.65</td>
<td>0.36</td>
<td>0.65</td>
<td>0.88</td>
</tr>
<tr>
<td>Robida 98</td>
<td>Slovenia</td>
<td>Slovene</td>
<td>0.81</td>
<td>0.75</td>
<td>0.72</td>
<td>0.76</td>
<td>0.73</td>
<td>0.69</td>
<td>0.60</td>
<td>0.75</td>
<td>0.63</td>
<td>0.65</td>
<td>0.62</td>
<td>0.89</td>
</tr>
<tr>
<td>Hedsköld et al 99</td>
<td>Sweden</td>
<td>Swedish</td>
<td>0.81</td>
<td>0.78</td>
<td>0.72</td>
<td>0.76</td>
<td>0.68</td>
<td>0.74</td>
<td>0.74</td>
<td>0.76</td>
<td>0.67</td>
<td>0.66</td>
<td>0.71</td>
<td>0.87</td>
</tr>
<tr>
<td>Nordin et al 100</td>
<td>Sweden</td>
<td>Swedish</td>
<td>0.82</td>
<td>0.75</td>
<td>0.68</td>
<td>0.67</td>
<td>0.62</td>
<td>0.72</td>
<td>0.74</td>
<td>0.72</td>
<td>0.60</td>
<td>0.61</td>
<td>0.74</td>
<td>0.87</td>
</tr>
<tr>
<td>Danielsson et al 101</td>
<td>Sweden</td>
<td>Swedish</td>
<td>0.81</td>
<td>0.82</td>
<td>0.74</td>
<td>0.79</td>
<td>0.68</td>
<td>0.77</td>
<td>0.76</td>
<td>0.78</td>
<td>0.73</td>
<td>0.73</td>
<td>0.79</td>
<td>0.86</td>
</tr>
<tr>
<td>Poussette et al 102</td>
<td>Sweden</td>
<td>English</td>
<td>0.78</td>
<td>0.76</td>
<td>0.65</td>
<td>0.74</td>
<td>0.67</td>
<td>0.69</td>
<td>0.70</td>
<td>0.68</td>
<td>−§§§</td>
<td>0.62</td>
<td>−§§§</td>
<td>−§§§</td>
</tr>
<tr>
<td>Perneger et al 103</td>
<td>Switzerland</td>
<td>French</td>
<td>0.79</td>
<td>0.75</td>
<td>0.71</td>
<td>0.80</td>
<td>0.67</td>
<td>0.78</td>
<td>0.60</td>
<td>0.77</td>
<td>0.61</td>
<td>0.57</td>
<td>0.68</td>
<td>0.86</td>
</tr>
<tr>
<td>Pfeiffer and Manser 104</td>
<td>Switzerland</td>
<td>German</td>
<td>0.83</td>
<td>0.78</td>
<td>0.76</td>
<td>0.73</td>
<td>0.64</td>
<td>0.71</td>
<td>0.71</td>
<td>0.79</td>
<td>0.61</td>
<td>0.68</td>
<td>0.75</td>
<td>0.88</td>
</tr>
</tbody>
</table>
| Near and Middle East
| Moghri et al 105 | Iran         | Farsi    | 0.79                                          | 0.80                                   | 0.77                          | 0.68                  | 0.68                  | 0.76                          | 0.77                            | 0.74                                      | 0.64                  | 0.70                    | 0.57                            | 0.78                             |
| Khater et al 106 | Jordan       | Arabic   | 0.60                                          | 0.54                                   | 0.61                          | 0.73                  | 0.46                  | 0.71                          | 0.61                            | 0.70                                      | 0.41                  | 0.52                    | 0.50                            | 0.78                             |
| Saleh et al 107 | Jordan       | English  | 0.77                                          | 0.67                                   | 0.66                          | 0.82                  | 0.64                  | 0.71                          | 0.63                            | 0.73                                      | 0.77                  | 0.60                    | 0.65                            | 0.73                             |

Continued
<table>
<thead>
<tr>
<th>Authors</th>
<th>Country</th>
<th>Language</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
<th>Percentage &lt;0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suliman et al</td>
<td>Jordan</td>
<td>Arabic</td>
<td>0.72</td>
<td>0.63</td>
<td>0.73</td>
<td>0.75</td>
<td>0.69</td>
</tr>
<tr>
<td>El-Jardali et al</td>
<td>Lebanon</td>
<td>Arabic</td>
<td>0.63</td>
<td>0.57</td>
<td>0.68</td>
<td>0.46</td>
<td>0.74</td>
</tr>
<tr>
<td>Al-Mandhari et al</td>
<td>Oman</td>
<td>English</td>
<td>0.59</td>
<td>0.53</td>
<td>0.53</td>
<td>0.59</td>
<td>0.63</td>
</tr>
<tr>
<td>Najjar et al</td>
<td>Palestine</td>
<td>Arabic</td>
<td>0.66</td>
<td>0.75</td>
<td>0.77</td>
<td>0.41</td>
<td>0.73</td>
</tr>
<tr>
<td>Hamdan and Saleem</td>
<td>Palestine</td>
<td>Arabic</td>
<td>0.76</td>
<td>0.76</td>
<td>0.78</td>
<td>0.38</td>
<td>0.72</td>
</tr>
<tr>
<td>Abohashaikah</td>
<td>Saudi Arabia</td>
<td>English</td>
<td>0.59</td>
<td>0.61</td>
<td>0.70</td>
<td>0.53</td>
<td>0.70</td>
</tr>
<tr>
<td>Suliman et al</td>
<td>Jordan</td>
<td>Arabic</td>
<td>0.58</td>
<td>0.44</td>
<td>0.64</td>
<td>0.54</td>
<td>0.78</td>
</tr>
<tr>
<td>Alonazi et al</td>
<td>Saudi Arabia</td>
<td>Arabic</td>
<td>0.65</td>
<td>0.75</td>
<td>0.75</td>
<td>0.67</td>
<td>0.77</td>
</tr>
<tr>
<td>Alswat et al</td>
<td>Saudi Arabia</td>
<td>Arabic</td>
<td>0.52</td>
<td>0.40</td>
<td>0.60</td>
<td>0.53</td>
<td>0.78</td>
</tr>
<tr>
<td>Günes et al</td>
<td>Turkey</td>
<td>Turkish</td>
<td>0.65</td>
<td>0.68</td>
<td>0.74</td>
<td>0.54</td>
<td>0.63</td>
</tr>
<tr>
<td>Bodur and Filiz</td>
<td>Turkey</td>
<td>Turkish</td>
<td>0.59</td>
<td>0.67</td>
<td>0.84</td>
<td>0.67</td>
<td>0.70</td>
</tr>
<tr>
<td>Bodur and Filiz</td>
<td>Turkey</td>
<td>Turkish</td>
<td>0.59</td>
<td>0.67</td>
<td>0.77</td>
<td>0.54</td>
<td>0.72</td>
</tr>
<tr>
<td>Nie et al</td>
<td>China</td>
<td>Chinese</td>
<td>0.67</td>
<td>0.51</td>
<td>0.63</td>
<td>0.72</td>
<td>0.64</td>
</tr>
<tr>
<td>Shu et al</td>
<td>China</td>
<td>Chinese</td>
<td>0.54</td>
<td>0.63</td>
<td>0.76</td>
<td>0.48</td>
<td>0.44</td>
</tr>
<tr>
<td>Wang and Tao</td>
<td>China</td>
<td>Chinese</td>
<td>0.61</td>
<td>0.64</td>
<td>0.79</td>
<td>0.68</td>
<td>0.44</td>
</tr>
<tr>
<td>Ito et al</td>
<td>Japan</td>
<td>Japanese</td>
<td>0.61</td>
<td>0.70</td>
<td>0.83</td>
<td>0.62</td>
<td>0.73</td>
</tr>
<tr>
<td>Fujita et al</td>
<td>Japan</td>
<td>Japanese</td>
<td>0.62</td>
<td>0.70</td>
<td>0.82</td>
<td>0.64</td>
<td>0.74</td>
</tr>
<tr>
<td>Ito et al</td>
<td>Japan</td>
<td>Japanese</td>
<td>0.65</td>
<td>0.72</td>
<td>0.77</td>
<td>0.54</td>
<td>0.72</td>
</tr>
<tr>
<td>Chen and Li</td>
<td>Taiwan</td>
<td>Chinese</td>
<td>0.70</td>
<td>0.73</td>
<td>0.78</td>
<td>0.51</td>
<td>0.76</td>
</tr>
</tbody>
</table>

**Descriptive Statistics**

- **Minimum**: 0.52, 0.40, 0.56, 0.53, 0.38, 0.44, 0.13, 0.36, 0.02, 0.36, 0.15, 0.53
- **Maximum**: 0.87, 0.85, 0.88, 0.85, 0.83, 0.92, 0.79, 0.83, 0.81, 0.79, 0.87, 0.93
- **Mean**: 0.73, 0.72, 0.70, 0.75, 0.64, 0.72, 0.65, 0.72, 0.56, 0.63, 0.60, 0.83
- **SD**: 0.09, 0.09, 0.07, 0.07, 0.10, 0.07, 0.11, 0.09, 0.15, 0.09, 0.16, 0.07
- **Percentage <0.7**: 32.31, 29.85, 49.23, 19.70, 71.88, 31.15, 56.06, 29.69, 87.10, 75.81, 69.35, 6.15
Authors | Country | Language | Sample characteristics
--- | --- | --- | ---

Table 1 Continued

| Authors | Country | Language | Sample characteristics |
| --- | --- | --- | ---
| <0.6 | 7 | 6 | 3 |
| 0.6 | 17 | 2 | 14 |
| Percentage <0.6 | 10.77 | 8.96 | 4.62 |

Bold values are lower than a value of 0.7 Cronbach alpha.

*The dimensions ‘communication openness’ and ‘feedback and communication about error’ were combined into one factor.
†The dimensions ‘hospital handoffs and transitions’ and ‘teamwork across hospital units’ were combined into one factor.
‡The dimension ‘organisational learning’ was omitted.
§The dimensions ‘staffing’ and ‘overall perceptions of safety’ were combined into one factor.
¶The dimensions ‘hospital handoffs and transitions’ and ‘teamwork across hospital units’ were combined into one factor.
**The dimensions ‘communication openness’ and ‘feedback and communication about error’ were combined into one factor.
††The dimension ‘overall perceptions of safety’ was omitted.
‡‡The dimensions ‘staffing’ and ‘overall perceptions of safety’ were combined into one factor.
§§The dimensions ‘organisational learning’ and ‘feedback and communication about error’ were combined into one factor.
***As the dimensions ‘handoffs and transitions’ had only two items, the internal reliability was not determined for this dimension.
****The dimensions ‘feedback and communication about error’ and ‘communication openness’ were combined into one factor named ‘communication’.
*****The dimensions ‘staffing’ and ‘overall perceptions of patient safety’ were omitted.
******The dimensions ‘staffing’, ‘overall perceptions of safety’ and ‘frequency of events reporting’ were omitted.
†††The dimension ‘hospital management support for patient safety’ was grouped with the dimension ‘teamwork across hospital units’.
‡‡‡The dimension ‘communication openness’ was grouped with the dimension ‘feedback and communication about error’.

| Authors | Country | Language | Sample characteristics |
| --- | --- | --- | ---
| <0.6 | 7 | 6 | 3 |
| 0.6 | 17 | 2 | 14 |

The study sample consisted of a total of 30 different countries. Eleven studies reported the use of the survey in North America, 15 studies in Europe, 15 in Europe, 13 in the Near East and the Middle East, and 7 in the Far East. On the growth of the use of the HSPSC from 2004 to 2018 across the world is shown in Figure 3.

Setting and participant job roles

The setting within which the survey was used was predominantly hospitals, with 47 of the 62 published studies (75.81%) using the survey in acute general, public, private, psychiatric, community, and secondary and tertiary hospitals. Of the remaining 15 studies, 3 used the survey in long-term care settings. Four studies sampled specific job roles (e.g., nursing staff and medical directors of hospitals). The survey was used in intensive care (72.81%), emergency, critical, and transport and neonatal intensive care. The majority of participants across the studies (table 2) indicated that the survey was used in different aspects of the care process for each patient (e.g., emergency, critical, and transport and neonatal intensive care). The remaining eight studies used the survey in different aspects of the care process for each patient (e.g., emergency, critical, and transport and neonatal intensive care). The remaining eight studies used the survey in different aspects of the care process for each patient (e.g., emergency, critical, and transport and neonatal intensive care).

Sample size, response rate and adaptations to the original version of the HSPSC

Sample sizes varied across the studies. In total, data were reported from 408 563 respondents from 1119 healthcare organisations across the world.

Open access

Figure 3 Growth of use of the HSPSC from 2004 to 2018 across world regions of the studies included in our sample.
organisations. Site sample size ranged from 331 hospitals to studies carried out within a single site. Most studies (85.48%) did not identify the number of hospital units in which respondents worked. Sample sizes ranged from 51 to 111,478 individuals, with a median of 1026. Response rates ranged from 17% to 98% (mean=38.94±29.54%). Of the 62 studies, the HSPSC instrument was applied in 20 different languages. Vlayen et al used a French and a Dutch translation. One study used the HSPSC in Japan and Taiwan, translating the survey into Japanese and Mandarin. From the 67 datasets, 16 obtained results through the use of an English version of the survey. Another 27 datasets required the HSPSC to be translated into a language other than English using several different translation procedures, and 24 datasets used the survey translated in previous studies. Two datasets used the original AHRQ translation guidelines. Fourteen datasets reported using forward–backward translation. Twelve datasets reviewed and verified the translation using an independent expert group. Five datasets made no modifications to the original version of the instrument, and 28 datasets (41.79%) did not state whether modifications were made. Online supplementary files 3 and 4 provide details of the sample sizes, response rates and adaptations to the HSPSC in the 67 datasets.

DISCUSSION
Implications of the review for users of the HSPSC
A clear implication of our review is that researchers and healthcare practitioners should adopt a degree of caution when using the original, unmodified version of the HSPSC to measure PSC. Nearly 50% of items from the 12 HSPSC dimensions in our sample demonstrate weak levels of reliability. In some cases, the internal reliabilities of specific dimensions (eg, staffing and overall perceptions of safety) were very weak. By contrast, the internal reliabilities for other dimensions (eg, ‘teamwork across/within units’) were stronger. While most studies (59.68%) that used CFA found a fit with the original 12-factor model, inconsistencies in carrying out assessments of construct validity (eg, testing EFA models using CFA and carrying out cross validation) make it very difficult to draw firm conclusions about the validity of the HSPSC. This difficulty is compounded by the fact that only 43% of the datasets in our sample report the outcomes from using factor analysis to assess construct validity. It has been 15 years since the original report describing the HSPSC instrument was released by AHRQ, and one recommendation from our review is that many of the instrument’s dimensions and items need to be revised. Further development and revision of the HSPSC might result in improved psychometric properties and increase its suitability for use across a range of international settings and contexts.

The need for guidance covering adaptations and other changes to the HSPSC
Just under a quarter of the studies in our sample made changes to the original survey items and/or dimensions. Over a third of the sample did not state what modifications, if any, were made to the original US version of the survey. It is well known that changing items and dimensions, alongside other variations in procedure (eg, procedures used to translate the HSPSC and reverse scoring questions), are likely to impact on the psychometric properties of the instrument. Alterations to the HSPSC also make it difficult to obtain valid measurements and to carry out cross-national comparisons. We would recommend scientific support and guidance for users of the HSPSC covering necessary adaptations in survey items to meet language requirements of different national, regional and healthcare contexts. Some guidelines already exist but are limited in coverage (eg, focusing solely on translation) and could be expanded to consider, among other things, advice regarding changes to items and dimensions and the reporting of findings. In addition, the development of guidelines covering the reporting of data from PSC surveys might make it easier for healthcare managers and decision makers to benchmark data within their organisations in order to promote organisational learning processes.

Educational support, resources and the development of standards
We also found a lot of variation across the studies with regard to the reporting of psychometric data, particularly the outcomes of using factor analysis. While recognising that this is something of a ‘grey area’ (eg, agreement with regard to CFA fit indices) and that the analysis and interpretation of psychometric properties require some skills and knowledge, there is room for further improvement in the future, particularly in terms of the provision of educational resources and the development of standards and guidance covering criteria and levels of acceptability for psychometric data. One way forward may be to work towards a set of networks for sharing, comparing and benchmarking the results from using the HSPSC and other PSC instruments and tools. Some of these networks would have to be coupled with educational materials to ensure that such networks are able to meet the needs of healthcare managers and decision makers.
already exist (eg, the web resources provided for the HSPSC by AHRQ) but may be expanded in the same way that other types of collaborative networks and ‘observatories’ work in other domains of safety (eg, road safety) to include protocols and standards for data collection, analysis and reporting, as well as other educational and training resources.

Implications of the review for researchers

The field of PSC is relatively new in comparison to the range of other scientific disciplines that have examined safety and organisational culture over the last 30 or so years (eg, safety science, human factors and organisational psychology). Likewise, the use of safety culture instruments is well established in a variety of industrial sectors outside of healthcare (eg, nuclear energy, aviation and construction). There is much that could be learnt both in terms of theory and methodology from this large body of knowledge and experience. We would point to three specific areas for the future. First, there is a need to understand the influence of a range of other influences on PSC, including the role played by national culture and professional subculture. Wagner et al, for example, found clear differences between the Netherlands, the USA and Taiwan in terms of responses to some HSPSC dimensions (eg, communication openness and non-punitive response to error). The large variability in terms of the reliability and validity of the HSPSC across our sample might be due to the influence of professional groups and other subcultures within healthcare. Many HSPSC items may be very differently interpreted and may elicit very different opinions among different healthcare groups (eg, nursing staff, physicians and managers) .

Second, there is a need to expand the range of theoretical constructs that influence and shape PSC. Research on, for example, the relationship between ‘speaking up’ and ‘voice’, levels of management control and PSC is starting to appear; however, there is potential for expansion. Similarly, research on the extent of agreement on levels of PSC among healthcare units and groups (‘climate strength’) is starting to appear. A range of other constructs influencing PSC, which are well established in other disciplines (eg, organisational psychology), including leadership styles and organisational justice, might also form part of an agenda for future research. Finally, we know very little about the levels of change in PSC over time, how these levels may decline, improve or stagnate, and what triggers these changes.

Thirdly, we would point to the value of using mixed methods to measure PSC as a means of addressing and taking into account the important role played by diverse social, cultural and subcultural contexts within different healthcare settings. Combining survey data with data drawn from interviews or focus groups, for example, is currently under-represented in the literature on PSC, despite being relatively common in studies of PSC in other safety-critical industries (eg, aviation, where interviews and focus groups are often used as a means to check and validate other data, including those from survey instruments).

Limitations of the review

Our review is the first to systematically examine the psychometric properties of a PSC instrument across a broad range of international contexts; however, it focuses on only one of several instruments currently in use (eg, refs 18, 67). We would point to the need for similar systematic reviews of the psychometric properties of these instruments to be carried out in the future. Finally, the diverse ways in which studies in our sample reported details of the procedure, which was used to administer the instrument, as well as how they reported findings from its use, made it difficult to derive firm conclusions about the reliability and validity of the HSPSC.

Acknowledgements

We thank Professor Paula Griffiths (School of Sport, Exercise and Health Sciences, Loughborough University) for her comments on an earlier draft of the paper. We also thank the reviewers for BMJ Open for their helpful comments on an earlier version of this paper.

Contributors

PW was responsible for the conception and design of the study. The paper was primarily written by PW. Data collection and analysis were initially carried out by PW and subsequently completed by E-MC. TM and AH commented on successive drafts of the paper and suggested areas for revision and improvement.

Funding

The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests

None declared.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data availability statement

All data relevant to the study are included in the article or uploaded as supplementary information.

Open access

This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES

Culture: Theory, Methods and Application Farnham, UK: CRC Press, 2014: 349–70.


68. Sorra JS, Dyer N. Multilevel psychometric properties of the AHRQ Hospital survey on patient safety culture. BMC Health Serv Res 2010;10:199.


75. Silva-Batalha EMS, Melleiro MM. Patient safety culture in a teaching hospital: differences in perception existing in the different scenarios of this institution. Texto contexto - enferm. 2015;24:432–41.


