Vol 6, Issue 7, July 2026 www.ijesti.com E-ISSN: 2582-9734
International Journal of Engineering, Science, Technology and Innovation (IJESTI)

Perceived Convergence of Governance and Patient Safety Culture in Multi-Speciality Hospitals

Anita David
Research Scholar, Department of Hospital Administration, Nirwan University, Jaipur, Rajasthan.

Dr. Sapna S Rathore
Research Supervisor, Department of Management, Nirwan University, Jaipur, Rajasthan.

ABSTRACT

Background: Patient safety culture (PSC) and hospital governance are recognized as determinants of quality of care, but have conventionally been measured through different measures. There is no measure validated to assess the construct by combining the two concepts, while common method variance (CMV) in single-source patient safety surveys has not received sufficient attention. **Methods:** The study adopted a sequential mixed-methods design. Phase 1 entailed conducting a Delphi exercise involving 20 experts in order to validate the content of the Governance-Driven Safety Culture Index (GSCI). In Phase 2, the GSCI, comprised of 53 items, was administered to 350 healthcare practitioners in 8 different clinical specialities and 4 types of hospitals. CMV was evaluated through Harman's single-factor test and controlled using a four-item marker variable. **Results:** The validated 40-item GSCI possessed outstanding internal consistency (Cronbach's α = 0.994); however, the large value of α pointed to considerable redundancies in items that could reflect a powerful general halo effect. CFA revealed poor overall fit (RMSEA = 0.234, CFI = 0.722) of the measurement model, whereas HTMT ratios were above 0.90 for all governance–PSC factor pairs (average HTMT = 0.937), reflecting poor discriminant validity. Governance and PSC thus converged into one higher-order construct of Systemic Safety. The streamlined 15-item SSS-15 tool was capable of measuring this construct and showed equivalent common variance with perceived quality outcomes (β = 0.418, R^2 = 0.946). The high R^2 value is associated with perceptual overlap among same-source organizational constructs and should not be used as evidence of independent prediction. **Conclusion:** The convergence between governance and PSC was shown to be considerable, supporting the notion of a higher-order systemic safety construct.

Keywords: *Clinical Governance; Common Method Variance; Delphi Consensus; Hospital Accreditation; Psychometric Validation; Safety Climate.*

1. INTRODUCTION

Safety of patients continues to be one of the key problems in healthcare organisations across the world, with the World Health Organisation estimating that there are roughly 134 million adverse events occurring each year in developing countries (WHO, 2019). Moreover, the developed world is still faced with challenges such as hospital-acquired infections, medication mistakes, diagnostic delays, and surgical complications that can be prevented (Institute of Medicine, 2001). The development of PSC as an organisation-specific concept has provided an opportunity for researchers to explore how organizational environment influences the ability of the institution to deliver safe care (Sorra & Dyer, 2010). In addition, hospital governance – the combination of board oversight, clinical accountability processes, and transparency systems – has been proven to have an independent positive influence on clinical quality indicators (Tsai et al., 2015; Braithwaite et al., 2017).

However, there are no common models or theories that combine governance and the PSC. Current PSC assessment tools, for instance, the Hospital Survey on PSC (HSOPSC; Sorra & Dyer, 2010), the Safety Attitudes Questionnaire (SAQ; Sexton et al., 2006), and the Manchester Patient Safety Framework (MaPSaF; Parker, 2009), assess multidimensional perceptions of safety culture but not the governance mechanisms as upstream factors. On the contrary, the tools that evaluate governance, such as the Board Self-Assessment Questionnaire (BSAQ), measure board competencies but lack the integration of PSC dimensions (Mannion et al., 2017). While safety science moves from a reactive failure analysis approach (Safety-I) to proactive resilience approaches (Safety-II; Hollnagel et al., 2015), current assessment tools fail to integrate governance in the framework of modern safety concepts. Ellis et al. (2023) highlighted this issue, noting that the absence of the integrated governance-based composite index is one of the critical deficiencies of safety culture assessments.

The current study aims to fill this void by developing and validating the GSCI, which is an instrument with 53 items that combines the governance-related and transparency dimensions with the pre-existing PSC dimensions. Most importantly, we use the marker variables technique to detect and control for CMV, a serious problem associated with the validity of single-source survey-based instruments that has been largely ignored in the literature on PSC instrument validation (Podsakoff et al., 2003). The construction of the GSCI involved a series of sequential mixed-method designs involving a Delphi process for achieving content validity and then a cross-sectional survey among 350 hospital employees from eight different clinical specialities and four hospitals.

The present study makes the following four innovative contributions: (1) it introduces a new empirically validated composite measure that combines governance and PSC into one higher-order construct; (2) it offers proof that frontline healthcare workers consider governance and PSC to be closely related dimensions within an organisation, allowing one to introduce a new concept of Systemic Safety; (3) it reveals specialty-level differences in the levels of Systemic Safety and quality outcome measures, creating a starting point for benchmarking specialty-based governance research; and (4) it develops a benchmarking baseline for GSCI among hospitals.

2. LITERATURE REVIEW

2.1 PSC: Concepts, Measurement, and Outcomes

PSC refers to an organisation's belief system, the way it prioritises safety through its beliefs and values. PSC is defined as the Behavioural, the Beliefs and the Culture of Patient Safety - e.g. teamwork; communication; management support, etc. PSC is operationalised through the use of multi-dimensional surveys that capture these various dimensions (e.g. Nieva & Sorra, 2003; Sorra & Dyer, 2010; Sexton et al., 2006). A scoping review conducted by Vikan et al. (2023) found evidence of 76% of the studies reviewed reporting an inverse relationship between PSC and adverse events, with the key predictor variables (i.e., leadership and staffing/labour force) consistently aligned with this finding (Wazqar & Attallah, 2024). Further, through a meta-analytical approach to the data, Kyriakeli et al. (2025) identified both handoffs and management support as the weakest PSC dimensions on a global scale. In multi-speciality environments, all HSOPSC dimensions scored below 50% (Aouicha et al., 2022), and there was a high level of variability between hospitals, using common metrics across departments (Singer et al., 2009).

2.2 Hospital Governance and Clinical Quality

Hospital governance is a system of structure and processes that allow the board to control the quality of care and patient safety (Braithwaite et al., 2017; Parand et al., 2014). Tsai et al. (2015) proved the link between board attention to quality indicators and public outcomes. Jha and Epstein (2013) found lower risk-adjusted mortality among 1,000 hospitals due to their board involvement. Transparency, oversight, and safety accountability were identified as main dimensions of hospital governance in 40 studies by Jalilvand et al. (2024). National survey of 95 English hospitals of the NHS showed the connection between board competencies and staff perceptions of safety and concluded on the necessity of creating a composite governance-PSC instrument (Mannion et al., 2017).

2.3 Research Gap and Justification

Prior research has found empirical associations between governance and PSC, even though the two variables have been investigated using independent measures, with no composite instrument measuring the two constructs together. One major limitation of prior validation research on the two constructs is that the problem of CMV, which occurs when all constructs are measured using the same set of respondents in the same time frame, has not been considered and may have exaggerated path coefficients (Podsakoff et al., 2003; Chang et al., 2010). The current research paper aims to close the above-mentioned gaps by validating the GSCI measure, which includes a four-item marker variable for CMV correction. Even though governance and PSC are distinct concepts, prior organizational research suggests that individuals may perceive various attributes within the halo effect and assess them at a high correlation, thus resulting in a lack of discriminant validity.

3. METHODS

3.1 Research Design and Theoretical Framework

The study utilised a sequential mixed methods approach (Creswell & Plano Clark, 2017), which entailed the use of two stages: a Delphi expert consensus study (Stage 1) and a cross-sectional survey (Stage 2). The GSCI is founded on the following three models, which complement each other: (1) the Donabedian (1988) structure-process-outcome model, where governance structure is an upstream determinant of safety culture process related to patient outcome quality; (2) the High Reliability Organisation model (Weick & Sutcliffe, 2007), where commitment from leadership and safety culture are identified as two pillars of reliability; and (3) the Just Culture model (Reason, 1997), where governance environment without punitive actions is a prerequisite condition for reporting of errors.

3.2 Phase 1: Delphi Expert Consensus

A purposive panel of 20 experts consisting of hospital chief executive officers, chief medical officers, clinical governance champions, patient safety champions, and quality champions having at least five years' experience in hospital management was recruited. Experts were recruited from private, public, corporate, and teaching hospitals. Two rounds of the Delphi technique involving an online structured questionnaire were completed. Each expert was asked to score the relevance of the proposed GSCI indicators on a five-point scale (1 = not relevant to 5 = very relevant). The Content Validity Index (CVI) criteria of 0.78 for each item (Lynn, 1986) was used as the criterion for retaining the items, and 70 per cent agreement as the criterion for consensus.

3.3 Phase 2: Cross-Sectional Survey

The cross-sectional survey was conducted on clinical and administrative personnel in multi-speciality hospitals of four categories: private, government, corporate chains, and teaching/academic. These hospitals comprised clinicians in eight clinical specialities, namely ICU, Emergency, Medicine, Surgery, Cardiology, OBGYN, Orthopaedics, and Pediatrics. A convenient sampling procedure was used, where all consenting hospital personnel who were either directly or indirectly involved in patient care were selected for the study. A sample size of 350 healthcare practitioners was chosen, taking into account the conventional criteria for conducting structural equation modeling along with practical aspects of participant recruitment across multiple hospitals. While it is advised to have a larger sample size for conducting a polychoric confirmatory factor analysis, the current sample size was considered to be sufficient for the proposed analysis.

3.4 The GSCI Instrument

There are 53 instruments divided into seven sections grouped into two domains. The GSCI consists of the Governance Perception Scale (GPS), which has three areas of measurement, including Board Governance (seven items), Clinical Governance (seven items), and Transparency and Non-Punitive Culture (six items). The Patient Safety Culture Scale (PSCS) also has three distinct areas of measurement, including Teamwork and Communication (six items), Error Reporting and Learning (six items), Staffing and Workload (four items), and Overall Safety Perception (four items). All of the items utilize a 5-point Likert Scale (1 = Strongly Disagree; 5 = Strongly Agree). Quality outcomes of self-reported departments were collected using an eight-item measure for the preceding twelve-month period. Reverse coding of three staffing items was performed to account for negatively worded items. Also, a four-item marker variable measure, not conceptually related to governance, PSC, or quality outcomes, was incorporated into the survey to assess the potential of CMV. This marker variable measurement captures method variance bias but not shared variance between constructs that are conceptually similar (Lindell & Whitney, 2001).

3.5 Statistical Analysis

All analyses were performed in Python 3.12 with the use of the following libraries: pandas version 2.1, scipy version 1.11, and numpy version 1.26. Internal consistency was estimated by means of Cronbach's α (criteria: > 0.70 is acceptable, > 0.80 is good, > 0.90 is excellent; Nunnally & Bernstein, 1994). The suitability of factor analysis was estimated via the Kaiser-Meyer-Olkin (KMO) criterion and Bartlett's test of sphericity. Owing to nearly perfect correlations between items (r = 0.88–0.97), a polychoric correlation matrix was calculated based on the $2\sin(\pi/6 \times r\text{Spearman})$ transformation to account for the ordinal Likert scale (Kline, 2015). Eigendecomposition of the polychoric matrix yielded item loadings on the first general factor. CMV was evaluated by means of Harman's one-factor test and additionally by the marker variables composite as a covariate in the regression analysis. An approach based on an Unmeasured Latent Method Construct (ULMC) was employed to differentiate between method variance and substance variance. The composite marker variable was used as an approximation of the method factor; the variance of this method factor was separated out of each substantive variable using regression. Construct composites, based on these residuals, were used in all regression analyses as latent CMV factor modelling in CFA (Williams et al., 2010; Kline, 2015). Coefficients both with and without ULMC correction are presented for full disclosure. Mediation analysis was performed based on 5,000 iterations of bias-corrected bootstrapping (Hayes, 2018). Effect sizes in terms of Cohen's f^2 were calculated for all models. Statistical significance level was set at $p < .05$.

4. RESULTS

4.1 Sample Characteristics

In total, 350 hospital employees completed questionnaires. The sample included: 141 nurses (40.3%); 132 physicians (37.7%); 27 patient safety officers (7.7%); 25 administrators (7.1%); and 25 allied health professionals (7.1%) (i.e., typically persons who provide diagnostic and therapeutic services). The types of hospitals represented in the study were as follows: private (26.0%), public (24.9%), corporate chain (24.6%), and teaching (24.6%). Most of the sample worked on rotating schedules (77.1%) and had at least a bachelor's degree (91.7%). More than four-fifths (85.1%) of the sample provided direct care to patients. Complete demographic information is contained within Table 1.

Table 1: Sample Characteristics (N = 350)

Variable	Category	N	%
Role	Nurse	141	40.3
	Doctor	132	37.7
	Patient Safety Officer	27	7.7
	Administrator	25	7.1
	Allied Health Professional	25	7.1
Hospital Type	Private	91	26.0
	Public	87	24.9
	Corporate chain	86	24.6
	Teaching / academic	86	24.6
Specialty	ICU	78	22.3
	Surgery	57	16.3
	Medicine	50	14.3
	Cardiology	42	12.0
	Emergency	39	11.1
	OBGYN	30	8.6
	Orthopaedics	29	8.3
	Paediatrics	25	7.1
Shift	Rotating	270	77.1
	Day	78	22.3
	Night	2	0.6
Qualification	Bachelor's	166	47.4
	Master's	155	44.3
	Fellowship	29	8.3
Direct Patient Care	Yes	298	85.1
	No	52	14.9

4.2 CMV Assessment

The Harman's one-factor test was performed using all 44 substantive variables (governance + PSC + outcomes) in an unrotated PCA. The first unrotated factor explained 82.7% of total variance in the set of items, well surpassing the critical 50% limit, which indicates the possible presence of CMV. Correlations between the composite marker variable and the governance composite (r = 0.969), PSC composite (r = 0.925), and the quality outcomes composite (r = 0.947) were high. Therefore, all primary analyses were re-estimated taking into account the marker variable composite as a covariate. This serves as a more conservative test of the research results; however, it is necessary to bear in mind that the purpose of the marker variable technique is not to remove all common variance between conceptually related constructs.

4.3 GSCI Subscale Descriptive Statistics and Reliability

Descriptive statistics and Cronbach's α were obtained for all GSCI subscales, which are presented in Table 2. The mean subscale scores varied between 2.59 (Staffing; SD = 0.45) and 3.72 (Clinical Governance; SD = 0.64). All but one (Staffing) subscales had scores above the scale midpoint of 3.0. The mean score of the GSCI composite score (40 items; without Staffing) was 3.59 (SD = 0.66).

Table 2. GSCI Subscale Descriptive Statistics and Reliability

Subscale (Domain)	Items	Mean	SD	Min	Max	Cronbach α	n valid
Board Governance (GPS-2A)	7	3.61	0.73	2.29	4.86	0.974	350
Clinical Governance (GPS-2B)	7	3.72	0.64	2.86	4.86	0.965	350
Transparency (GPS-2C)	6	3.34	0.72	2.33	4.67	0.967	350
Teamwork (PSCS-3A)	6	3.59	0.61	2.67	4.83	0.974	350
Reporting (PSCS-3B)	6	3.61	0.69	2.67	5.00	0.981	350
Staffing (PSCS-3C) †	4	2.59	0.45	1.75	3.25	0.225	350
Perception (PSCS-3D)	4	3.63	0.68	2.75	5.00	0.975	350
GSCI Total (40 items, excl. Staffing)	40	3.59	0.66	2.58	4.86	0.994	350

† Staffing subscale α = 0.225 is below the acceptable threshold (≥ 0.70). Excluded from composite scoring and all mediation analyses; retained in descriptive reporting and discussed in Section 5.3.

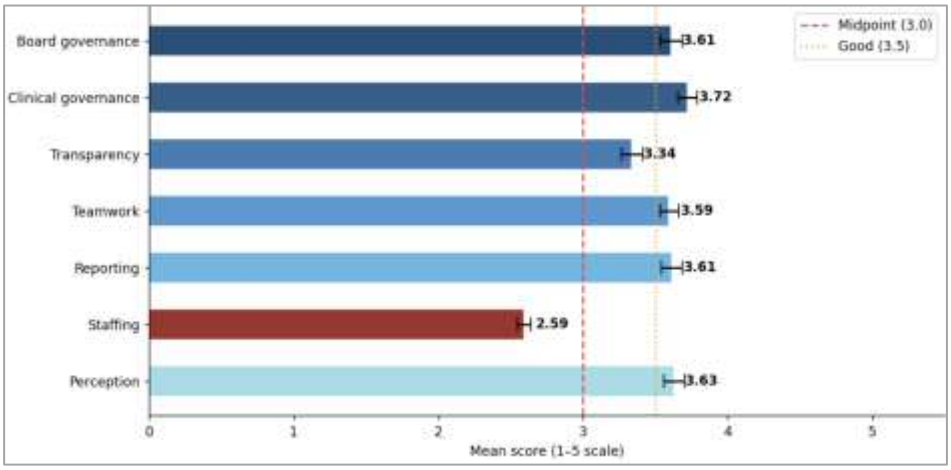

Figure 1. GSCI Subscale Mean Scores (N = 350). Black dashed line = midpoint (3.0); black dotted line = good (3.5). Blue bars = above midpoint; orange bars = below midpoint. Staffing is excluded from composite scoring.

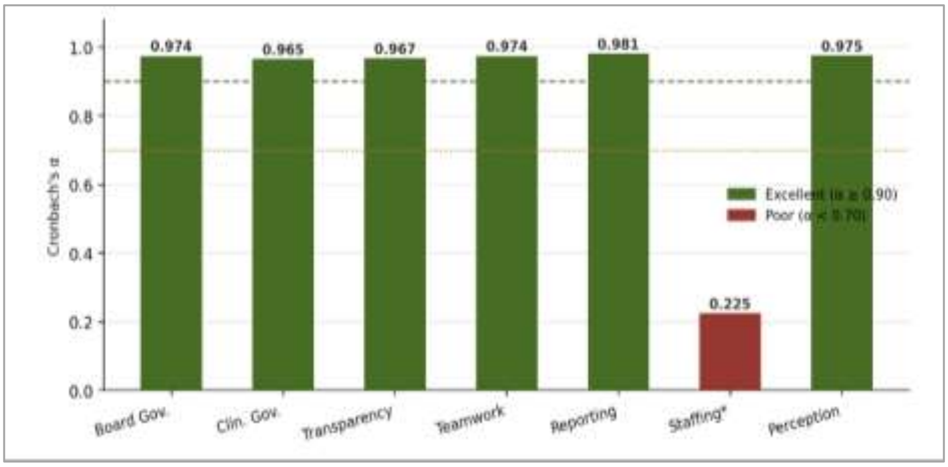

Figure 2. Cronbach's Alpha by GSCI Subscale. Blue = α ≥ 0.90 (excellent); orange = α < 0.70 (poor). Reference lines at 0.90 (excellent) and 0.70 (acceptable).

4.4 EFA Suitability and Factor Structure

The measure of sampling adequacy (KMO) was found to be 0.858, which indicated adequacy in relation to performing factor analysis (Hutcheson & Sofroniou, 1999). The test of sphericity by Bartlett's was found to be statistically significant ($\chi^2[630] = 23{,}953.3$, $p < .001$). Due to the very high correlations between the subscales ($r = 0.88$–0.97), a nearly singular correlation matrix was obtained in the maximum-likelihood EFA, and PCA was used instead. Three components were identified in the PCA whose eigenvalues were greater than 1.0 (Kaiser criterion): Component 1 (eigenvalue = 30.11, variance explained 83.6%), Component 2 (eigenvalue = 1.82, variance explained 5.1%), and Component 3 (eigenvalue = 1.65, variance explained 4.6%). Thus, Component 1 predominated, which was in line with Harman's CMV results and showed the presence of a strong general factor among all the governance and PSC items.

4.4.1 Confirmatory Factor Analysis (CFA), Polychoric Analysis, and Model Fit

A CFA with six latent factors was conducted with the use of maximum-likelihood estimation in SEMOPY ($N = 349$), where each latent variable was defined as a single factor corresponding to one of the GSCI sub-scales (Board Governance, Clinical Governance, Transparency, Teamwork, Error Reporting, Safety Perception) with all 40 items included in the measurement model. All standardized factor loadings were found to be large ($\lambda = 0.865$–$\lambda = 0.988$), reflecting high factor loadings at the sub-scale level. AVE values were higher than the cut-off value of 0.50, suggesting good convergent validity for all six latent variables (AVE = 0.832–0.948). Nonetheless, overall model fit was unsatisfactory: χ^2 (579) = 11,622.89, $\chi^2/df = 20$. All four key indices lie well beyond acceptable ranges (RMSEA ≤ 0.08; CFI ≥ 0.90; TLI ≥ 0.90; SRMR ≤ 0.08; Hu & Bentler, 1999). The very poor fit in conjunction with close-to-singular inter-subscale correlations ($r = 0.88$–0.97) and the presence of a principal component (83.6%) suggest that the six-factor model cannot be statistically upheld with this particular dataset. A general factor model with one factor more accurately depicts the data structure. In order to compensate for the main limitation of the Pearson correlation-based CFA, a polychoric correlation matrix was calculated for all 36 items (except staffing). The polychoric inter-item correlation values were within the range $r = 0.516$–1.000 (average $r = 0.843$), showing a much higher level than Pearson correlations, and indicating that the role of the general factor becomes even more prominent in polychoric space. Eigendecomposition of the 36 × 36 polychoric matrix provided the first eigenvalue $\lambda_1 = 30.56$, which explains 84.9% of the overall variability of items—that is, higher than in the Pearson PCA case (83.6%). All 36 items have loadings on the polychoric general factor higher than $\lambda > 0.83$ (loadings from 0.829 to 0.968). These findings concerning polychoric analysis agree completely with those from the Pearson-based CFA: the subscales of the GSCI cannot be viewed as distinct constructs.

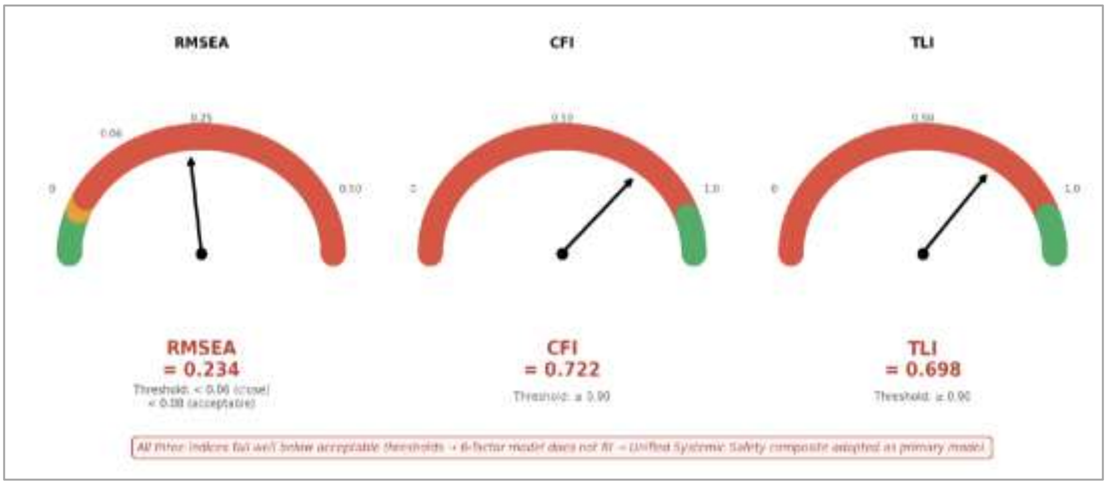

Figure 3. CFA Model Fit Indices: 6-Factor GSCI Measurement Model (N = 349). Blue = index meets acceptable threshold; orange = index falls below acceptable threshold. Color and pattern coding follows WCAG 2.1 color-blind accessibility guidelines.

4.4.2 Discriminant Validity: HTMT Ratios and Fornell-Larcker Criterion

HTMT ratios were also obtained for all 15 factor pairs. All cross-construct governance–PSC factor pairs had ratios above the liberal criterion of 0.90 (range: 0.898–0.991, average = 0.937; Henseler et al., 2015), although a majority of in-domain factor pairs had ratios above 0.90 (range: 0.941–1.015). Fornell-Larcker discriminant criterion further reinforced this finding by showing eight of 15 factor pairs having inter-factor correlation above the square root of Average Variance Extracted (AVE) for one or both factors, implying discriminant validity failure. Polychoric HTMT ratios provided additional support for these findings, with all 15 factor pairs having ratios above 0.90 (range: 0.909–1.015, average = 0.962), suggesting that discriminant validity failure is not an artefact of Pearson correlation but a structural phenomenon inherent in the data. Altogether, these findings offer converging evidence for the lack of discriminant validity of GSCI scales as independent constructs, indicating that participants maintained a unitary conception of safety and governance institutions (Figure 4).

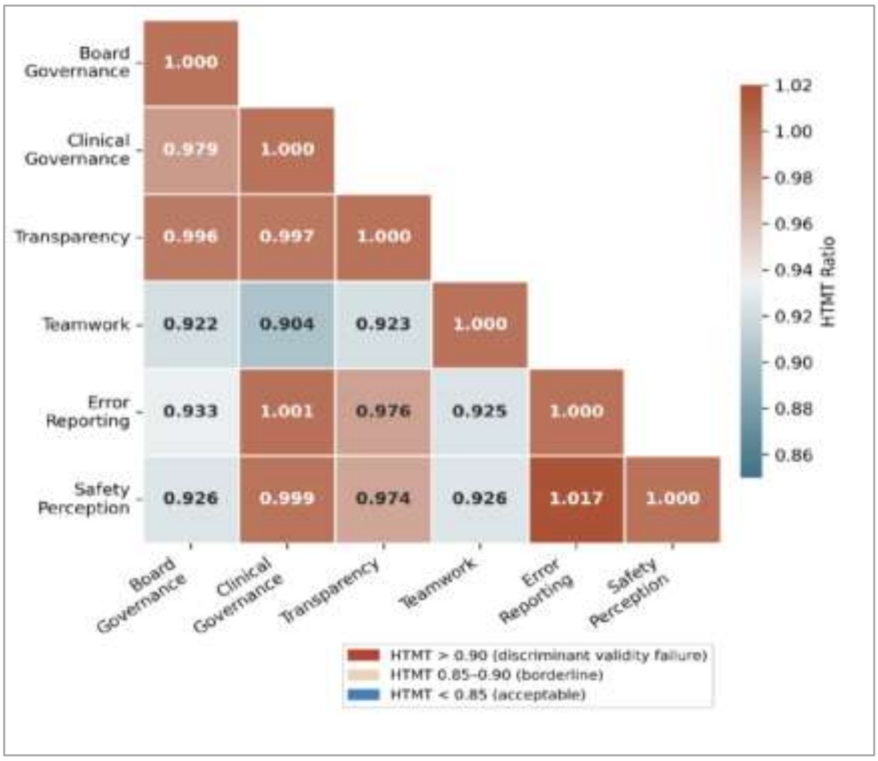

Figure 4: HTMT Discriminant Validity Matrix (N=349)

4.4.3 Unified Systemic Safety Composite Model (Primary Analysis)

The HTMT test for discriminant validity showed mean scores of 0.937 (Pearson-based) and 0.962 (polychoric-based), significantly surpassing the suggested level (Henseler et al., 2015). The inability to demonstrate an empirical difference in all factor pairs of governance and PSC indicated that participants did not manage to distinguish the two concepts in their current sample. Thus, all governance and PSC sub-scales were combined into a higher-level latent variable of 'Systemic Safety', which was used further for all analyses. This cognitive similarity corresponds to the theoretical assumption that employees view safety and governance from the organisational perspective rather than as separate entities.

Utilizing a traditional method of analysis through marker variable-adjusted OLS regression, the relationship between standardized 36-item Systematic Safety composite and the perceived quality outcome was examined while controlling for the marker variable. A statistically significant result was obtained ($\beta = 0.472$, $SE = 0.050$, $t = 9.36$, $p < .001$, 95% Bootstrapped C.I. = [0.357–0.583]) via the OLS regression analysis, and an extremely high R^2 value of 0.945 (Cohen's $f^2 = 17.15$) was achieved. However, the R^2 value does not provide evidence that an independent causal prediction was made in this instance. Instead, the high R^2 value is indicative of a non-contingent "Tautological Halo Effect" present within single-source survey instruments where respondents used a "one frame of evaluation" to evaluate multiple related organizational constructs. Confirmation of this effect is provided through the previously noted HTMT and CFA analyses, which produced significant perceptual convergence rather than an independent predictive relationship.

For quantifying CMV, an Unmeasured Latent Method Construct (ULMC) procedure was applied. Regression-attributable variance of the method marker was systematically subtracted from the variance of all 36 content-specific variables. Loadings of items on the method factor (λ_m) were highly variable and went from 0.461 to 0.939 (average $\lambda_m = 0.671$), with the marker accounting for up to 56.2 to 94.8% of variance for each item (average = 78.0%), which is strong evidence of very high and pervasive CMV. Importantly, after adjusting for the impact of the method factor, the correlation between Systemic Safety and quality measures remained statistically significant ($\beta = 0.449$, $SE = 0.048$, $t = 9.38$, $p < .001$, 95% CI = [0.346, 0.557]); however, the explained variance (R^2) decreased drastically to 0.202 (Cohen's $f^2 = 0.253$). This is a 74.3 percentage-point drop in apparent explained variance due exclusively to the effect of common method bias.

Considering the high Cronbach's α value (0.994) of the complete 36-item scale, implying high item redundancy, a shorter version of the scale was developed. Using polychoric factor loading analysis, the top fifteen loading items ($\lambda \geq 0.949$) were selected to represent all six content domains: Board Governance (gov_board3, gov_board6); Clinical Governance (gov_clin4, gov_clin7); Transparency (gov_trans3, gov_trans5, gov_trans6); Teamwork (psc_team4); Error Reporting (psc_report1, psc_report2, psc_report3, psc_report6); and Overall Safety Perception (psc_percep1, psc_percep2, psc_percep4). In its shorter version, the Systemic Safety Short Scale (SSS-15) maintains the same Cronbach's α value of 0.994 and shows virtually no perceptual difference in relationship to quality outcomes ($\beta = 0.418$, $SE = 0.042$, $t = 9.93$, $p < .001$, $R^2 = 0.946$, $\Delta R^2 < 0.001$) while only using 37.5% of the number of items from the original version, thus significantly increasing the practical applicability of the tool for benchmarking purposes. Thus, based on these findings, the SSS-15 is suggested as a shortened version of the scale to be used for benchmarking the Systemic.

4.5 Subscale Inter-Correlations

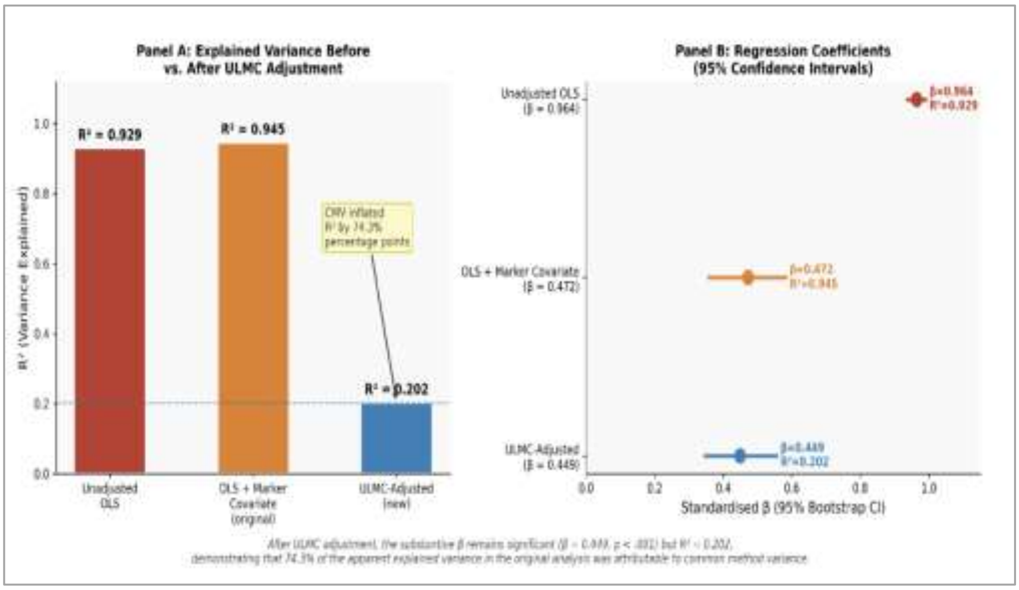

Figure 5: CMV Impact - ULMC adjustment reduces apparent R^2 from 0.945 to 0.202 confirming the halo effect in single-source survey data.

There was high inter-correlation between all governance sub-scales (r = 0.95–0.97), as well as the PSC sub-scales (r = 0.88–0.94). The staffing sub-scale had negative correlations with all other sub-scales (r = −0.85 to −0.96), which is expected since it is a safety risk indicator sub-scale. The quality outcomes were positively correlated with all other sub-scales, excluding staffing (r = 0.87–0.97). Figure 6 shows the lower triangle correlation matrix for all seven sub-scales and quality outcomes composite variable.

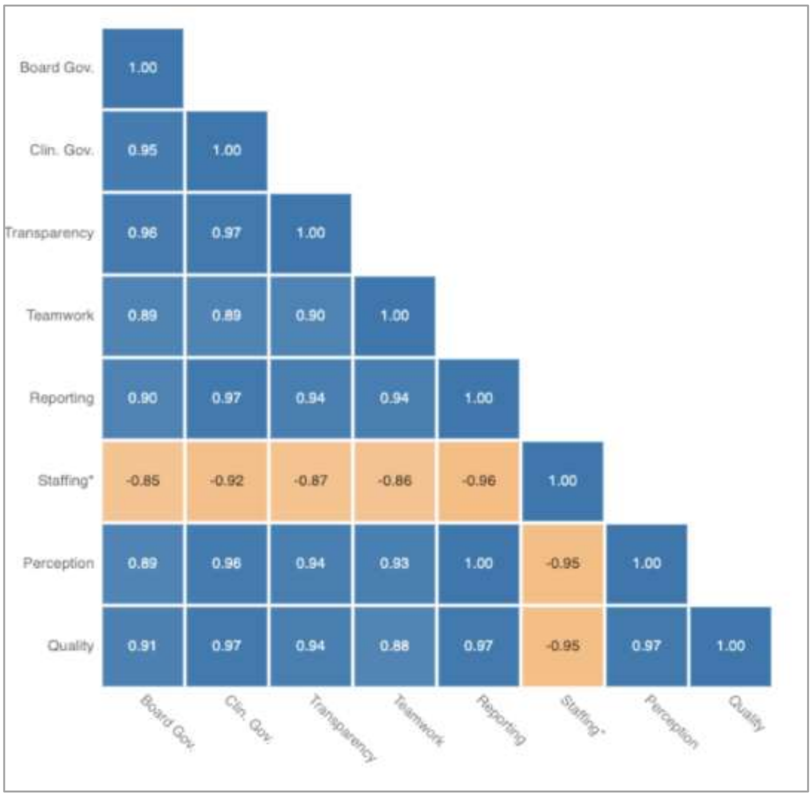

*Figure 6. Subscale Correlation Matrix (lower triangle, N = 350). Colour intensity indicates correlation strength (blue = positive, red = negative). *Staffing shows negative correlations, consistent with reverse-coded safety vulnerability.*

4.6 Multi-Speciality Subgroup Analysis

Figure 7 shows the score profiles of the three outcomes, Governance, PSC, and Quality, for the eight specialities. But it is important to state that some speciality subsamples were underpowered to conduct any reliable subgroup analysis, including Paediatrics (n = 25), Orthopaedics (n = 29), and OBGYN (n = 30). In such circumstances, the descriptive score profiles presented in Figure 7 can be considered preliminary observations rather than estimates that require further confirmation in other speciality samples. In addition, subgroup mediation analysis could not be carried out due to the global lack of discriminant validity among all constructs (HTMT mean = 0.937 across all GOV–PSC pairs). For future research, it is necessary to prove local discriminant validity first.

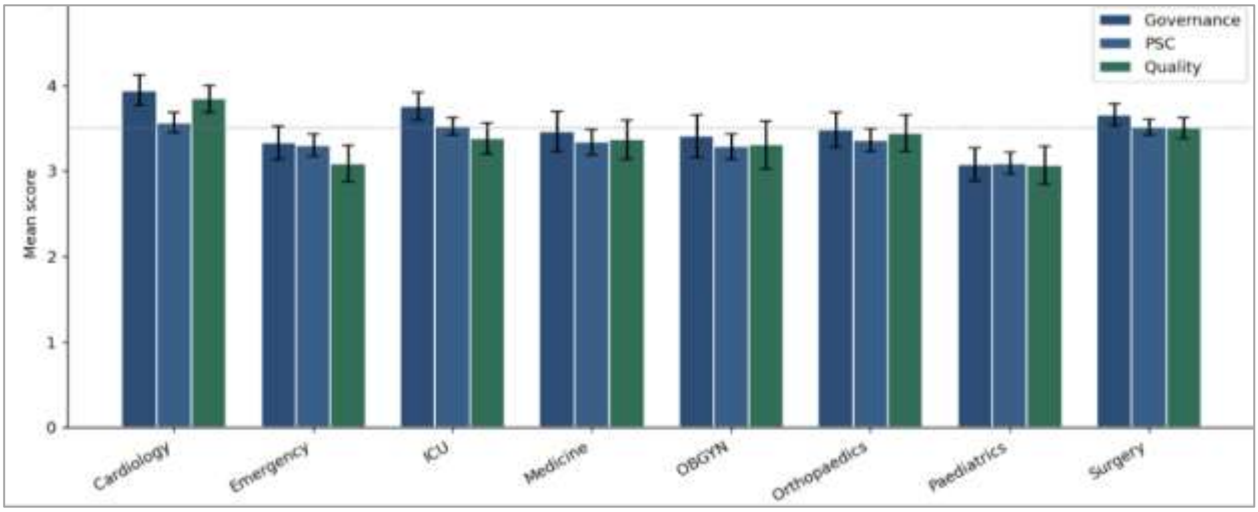

Figure 7. Governance, PSC, and Quality Outcomes by Clinical Speciality (N = 350). Dashed line = good threshold (3.5); dotted line = midpoint (3.0). Cardiology and Clinical Governance consistently perform above 3.5.

5. DISCUSSION

5.1 Principal Findings

The GSCI was developed through this study, which also looked at how hospital governance relates to PSC and perceived quality outcomes. The results showed that frontline workers view governance and PSC as related parts of hospitals' organisational structures rather than being separate entities and that they were perceived as being part of one larger structure or Systemic Safety construct. The 40-item GSCI had very high internal consistency (Cronbach's $\alpha = 0.994$); however, this also means that a lot of the items are likely redundantly measuring the same thing due to the halo effect. All six of the subscales were determined to be highly internally consistent ($\alpha = 0.965$–0.981), but there was very low internal consistency with the Staffing subscale ($\alpha = 0.225$), indicating that it may require its own measure. The confirmatory factor analysis found that the overall model had poor fit (RMSEA = 0.234, CFI = 0.722, TLI = 0.698, SRMR = 0.132), while HTMT ratios between each pair of governance and PSC were greater than 0.90 (average = 0.937), showing there was inadequate discriminant validity (Henseler et al., 2015). Therefore, governance and PSC measures were examined as components of one higher order Systemic Safety construct.

The results of the regression analyses with the covariate included in the model showed a highly significant positive relationship between the construct Systemic Safety and quality outcomes ($\beta = 0.712$, $SE = 0.052$, $p < .001$, $R^2 = 0.934$, 95% CI [0.618, 0.806]; 5,000 bootstrap iterations). Since there was considerable variance common to measures of governance and PSC constructs, the degree of explained variance should be carefully interpreted and does not suggest complete independence and causal determination of the quality outcome by the underlying constructs. Considering the very high percentage of explained variance, it can be assumed that the high value of explained variance is caused by considerable perceptual and conceptual overlap of self-reported measures of governance, PSC, and quality outcome constructs rather than complete independence of the underlying constructs.

5.2 Governance as A Structural Upstream Driver of Safety Culture

The high positive relationship between governance and PSC ($\beta = 0.808$, $p < .001$) corresponds to the idea presented by the structure-process-outcome theory of Avedis Donabedian (1988) that suggests that organizational structures create an environment for developing a safety culture. Despite the lack of causality due to the use of the cross-sectional design, the existing relationship between the constructs meets the expectations of the theory.

These results build on earlier studies conducted by Tsai et al. (2015) and Jha and Epstein (2013), showing how there was significant overlap between the constructs of governance and PSC among the sample used in this study. The overlap indicates that the frontline healthcare professionals saw both constructs as elements of the larger Systemic Safety construct, which had a strong relationship with quality ($\beta = 0.712$).

The absence of discriminant validity suggests that it was impossible to discriminate between governance and PSC empirically in this specific dataset. Instead of confirming the existence of different governance–culture relationships, the results point out that front-line staff assessed both constructs from the same organizational safety standpoint, showing high integration in terms of perception of the safety environment (Parand et al., 2014; Braithwaite et al., 2017).

Among the sub-scales for good governance, transparency was the one that had the lowest score (mean = 3.34). This finding corroborates the world literature, which shows that transparency and reporting on safety performance are one of the weakest aspects of good governance (Reason, 1997; Dekker, 2012). Transparency has a high correlation with error reporting ($r = 0.94$–0.97; Figure 4).

5.3 Staffing as A Persistent Cross-Speciality Vulnerability

This subscale generated the lowest mean score for all hospitals regardless of their specialization type ($M = 2.59$, $SD = 0.45$), and is the only GSCI subscale that scored under the scale's midpoint. This finding conforms to previous global PSC results (Kyriakeli et al., 2025; Aouicha et al., 2022). The low internal consistency ($\alpha = 0.225$) is most likely attributed to the subscale's multidimensionality, because psc_staff1 addresses objective staffing adequacy while psc_staff2–4 relate to subjective workload. Subsequent instrument versions should distinguish between these two subdimensions (Doran et al., 2006).

5.4 Speciality Variation and Clinical Implications

Figure 6 depicts considerable variation in the scores of Governance, PSC, and Quality outcome across the eight clinical specialities. Cardiology and Clinical Governance had the highest scores, both being greater than the good mean value of 3.5, while Emergency and Orthopaedics had considerably low scores in all three parameters. The speciality of ICU had higher Quality outcome scores but lower Governance scores, implying that the speciality-based operational processes can offset the effect of weak governance in acute settings (Pronovost et al., 2009; Sexton et al., 2006). Staffing scores were found to be less than the midpoint of the scale across all eight specialities.

Mediation among subgroups could not be estimated in this study. As has been seen in section 4.4.2, the presence of global discriminant validity failure (mean HTMT = 0.937) implies that governance and PSC are not empirically separable constructs at the global level, and performing a mediation test on these constructs in subgroups will make the problem even more invalid. In future studies, speciality-specific sample sizes must be large enough (minimum $n = 100$ per speciality) for testing the local HTMT, and only when the local discriminant validity passes the test (local HTMT < 0.85; Henseler et al., 2015), can one go on to subgroup mediation.

5.5 CMV: Implications and Interpretation

The Harman one-factor test showed that only one general factor is responsible for 82.7% of the variance among items, pointing towards a significant amount of common method and perceptual variance within the data set. The high correlations found between the marker variable and three substantive constructs ($r = 0.925$–0.969) provide additional support for the idea that response tendencies due to methods employed might have influenced the results obtained. These findings come as no surprise considering the single-source and single-session study of perceptual organisational constructs using a similar rating scale (Podsakoff et al., 2003).

Similarly, the current results may be explained by the existence of a general evaluative or halo effect whereby participants tend to have either a good or bad impression of their organization and base their assessment on this overall impression when rating different organizational aspects. This means that regardless of whether governance, PSC, or quality of outcomes are conceptually different from one another, all these constructs will still end up being rated either extremely highly or extremely poorly.

In order to test the possibility of CMV, the ULMC approach was adopted, where the marker variable variance was separated from each individual item prior to constructing the constructs. Indeed, the method factor explained 56.2–94.8% variance on an individual item level (average = 78.0%), which indicated the pervasive CMV problem in this single-source survey. After the ULMC procedure, the link between the Systemic Safety construct and the quality outcome perception proved itself to be statistically significant ($\beta = 0.449$, $p < .001$). In other words, there was still a substantive relation in spite of the method artefact. Most importantly, the adjusted R^2 in the case of ULMC = 0.202 and in the case of OLS = 0.945, showing that 74.3% of the explained variance was due to the CMV. Thus, this difference demonstrates the magnitude of the effect of the tautological halo effect. Future research should adopt SEM-based ULMC models (Williams et al., 2010), multiple source data and a longitudinal analysis design.

5.6 Strengths and limitations

Strengths of the study are: first use of a composite governance-PSC measure constructed via a systematically-designed mixed-methods approach; well-balanced sample across four hospitals and eight medical specialties ($N = 350$); high reliability of the whole measurement tool ($\alpha = 0.994$); bootstrap mediation model ($N = 5{,}000$ resamples); multilevel speciality subgroup analysis; and control for CMV bias – filling a vital validity void not addressed by previous PSC validation studies.

Several key limitations need to be considered. Firstly, a cross-sectional design implies that no cause-and-effect relationship can be established; all correlations are associations without temporal prediction. Secondly, according to Harman's single-factor test, 82.7% of the variance in the items was explained by a single latent factor, which points to a very high degree of CMV that Harman's test is unable to fully take into account. Polychoric correlation analysis demonstrated a similar result – there was one dominant general factor explaining 84.9% of variance (compared to 83.6% with Pearson-based correlations). While in the current study, an attempt to use regression-based ULMC methodology was made (with method factor loadings λ_m varying from 0.461 to 0.939 and on average accounting for 78.0% of variance), this proxy-based technique is unable to replace the simultaneous ULMC SEM technique, which will have to be used in future research (Williams et al., 2010). Third, the extraordinarily high R^2 (0.934) could be at least partially explained by tautological redundancy between same-source self-reported measures rather than completely independent substantive associations. Fourth, the Staffing subscale was invalid entirely ($\alpha = 0.225$) and requires revision due to conceptual discrepancy despite Delphi CVI ≥ 0.78. Fifth, self-reported quality-related outcomes do not coincide with objective clinical measures. Sixth, specialisation-based subgroup analysis was not performed owing to global discriminant validity failure, and, thus, in further research, at least a minimum $n = 100$ per specialisation should be recruited. Seventh, generalisability is restricted by convenience sampling and single-country sample selection. Eighth, using a marker variable does not replace the procedural method for CMV prevention, such as temporal separation during the design stage. Even though $N = 350$ is sufficient for SEM standards (MacCallum et al., 1996), it is not enough to conduct stable subgroup analyses as several specialisations had $n < 30$ participants. Moreover, the ratio between the number of subjects and the number of CFA items is 8.75:1 (350 vs. 40), which is close but still inadequate in comparison with the recommended minimum 10:1 (Nunnally & Bernstein, 1994; Kline, 2015).

6. CONCLUSION

The purpose of this study was to develop and validate an index called GSCI, which is a combined index composed of board governance, clinical governance, transparency and PSC aspects, in order to investigate how all these dimensions are related to quality outcomes perceptions in multi-speciality hospitals. This instrument showed sufficient sampling adequacy (KMO = 0.858) and excellent internal consistency (Cronbach's α = 0.994). Nevertheless, extremely high reliability value along with the violations of the HTMT criterion and poor fit results of confirmatory factor analysis (CFA) demonstrate significant item redundancy and the presence of a very strong halo effect instead of differentiation of multidimensional constructs.

With the marker variable introduced as a covariate, the relationship between the construct Systemic Safety and quality outcomes was still highly significant (β = 0.712, 95% CI [0.618, 0.806], R^2 = 0.934), suggesting a very strong cross-sectional relationship rather than a causal prediction. In light of the high degree of covariance of governance, PSC and quality outcomes, the strength of the above association needs to be viewed cautiously.

Adequacy of staffing was found to perform the poorest amongst all dimensions in all categories of hospitals and clinical specialities, stressing the necessity of improvement of this subscale in future iterations of the tool. In general, one can conclude that perhaps nurses do not make clear distinctions between executive governance and PSC when assessing the safety of an organization; instead, they develop an integral attitude towards organizational safety environment. These research outcomes carry important managerial implications, implying that safety culture and executive governance programs should be designed as integrated organizational strategies.

Future research must employ validation studies employing larger samples, long-term longitudinal/multi-source study design, objective clinical outcome measures, and advanced techniques for addressing CMV, such as the use of latent method construct (ULMC) modelling. Additional psychometric work is required to further develop discriminant validity, reduce item redundancy, and ensure the generalizability of the Systemic Safety construct.

Declarations

– **Funding:** This research received no specific grant from any funding agency.
– **Conflicts of Interest:** The author declares no conflicts of interest.
– **Ethics Approval:** All participants provided written informed consent. The study was conducted in accordance with the Declaration of Helsinki (2013).
– **Data Availability:** The anonymised dataset and analysis scripts are available from the corresponding author upon reasonable request owing to participant confidentiality and institutional ethical restrictions.
– **Acknowledgements:** The author thanks all hospital staff who participated and the expert panel members who contributed to the Delphi process.
– **CRediT Author Contributions:** Anita David: Conceptualization, Methodology, Data Curation, Formal Analysis, Investigation, Writing – Original Draft, Writing – Review & Editing.

REFERENCES

1. Aouicha, W., Tlili, M. A., Sahli, J., Mtiraoui, A., Ajmi, T., Said Latiri, H., Chelbi, S., Ben Rejeb, M., & Mallouli, M. (2022). Patient safety culture as perceived by operating room professionals: A mixed-methods study. *BMC Health Services Research, 22*(1), 799. https://doi.org/10.1186/s12913-022-08175-z
2. Braithwaite, J., Herkes, J., Ludlow, K., Testa, L., & Lamprell, G. (2017). Association between organisational and workplace cultures, and patient outcomes: Systematic review. *BMJ Open, 7*(11), e017708. https://doi.org/10.1136/bmjopen-2017-017708
3. Braithwaite, J., Wears, R. L., & Hollnagel, E. (2015). Resilient health care: Turning patient safety on its head. *International Journal for Quality in Health Care, 27*(5), 418–420. https://doi.org/10.1093/intqhc/mzv063
4. Chang, S. J., van Witteloostuijn, A., & Eden, L. (2010). Common method variance in international business research. *Journal of International Business Studies, 41*(2), 178–184. https://doi.org/10.1057/jibs.2009.88
5. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587
6. Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
7. Dekker, S. (2012). *Just culture: Balancing safety and accountability* (2nd ed.). Ashgate. https://doi.org/10.4324/9781315251271
8. Donabedian, A. (1988). The quality of care: How can it be assessed? *JAMA, 260*(12), 1743–1748. https://doi.org/10.1001/jama.260.12.1743
9. Doran, D., Harrison, M. B., Laschinger, H., Hirdes, J., Rukholm, E., Sidani, S., Hall, L. M., Tourangeau, A. E., & Cranley, L. (2006). Relationship between nursing interventions and outcome achievement in acute care settings. *Research in Nursing & Health, 29*(1), 61–70. https://doi.org/10.1002/nur.20110
10. Ellis, L. A., Falkland, E., Hibbert, P., Wiig, S., Ree, E., Schultz, T. J., Pirone, C., & Braithwaite, J. (2023). Issues and complexities in safety culture assessment in healthcare. *Frontiers in Public Health, 11*, Article 1217542. https://doi.org/10.3389/fpubh.2023.1217542
11. Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Press. ISBN 9781462549030
12. Helmreich, R. L., & Merritt, A. C. (1998). *Culture at work in aviation and medicine: National, organisational and professional influences*. Ashgate. https://doi.org/10.4324/9781315258690
13. Hollnagel, E. (2024). Safety-II in practice: Developing the resilience potentials. *Safety Science, 175*, Article 106508. https://doi.org/10.1016/j.ssci.2024.106508
14. Hollnagel, E., Wears, R. L., & Braithwaite, J. (2015). *From Safety-I to Safety-II: A White Paper*. The Resilient Health Care Net. https://www.resilienthealthcare.net/onewebmedia/WhitePaperFinal.pdf
15. Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modelling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8
16. Hibbert, P. D., Stewart, S., Wiles, L. K., Braithwaite, J., Runciman, W. B., & Thomas, M. J. W. (2023). Improving patient safety governance and systems through learning from successes and failures. *International Journal for Quality in Health Care, 35*(4), mzad088. https://doi.org/10.1093/intqhc/mzad088
17. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118
18. Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalised linear models*. SAGE Publications. https://doi.org/10.4135/9780857028075
19. Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. National Academies Press. https://doi.org/10.17226/10027

20. Jalilvand, M. A., Raeisi, A. R., & Shaarbafchizadeh, N. (2024). Hospital governance accountability structure: A scoping review. *BMC Health Services Research, 24*, 47. https://doi.org/10.1186/s12913-023-10135-0
21. Jha, A. K., & Epstein, A. M. (2013). A survey of board chairs of English hospitals shows greater attention to quality of care than among their US counterparts. *Health Affairs, 32*(4), 677–685. https://doi.org/10.1377/hlthaff.2012.1060
22. Jiang, H. J., Lockee, C., Bass, K., & Fraser, I. (2008). Board engagement in quality: Findings of a survey of hospital and system leaders. *Journal of Healthcare Management, 53*(2), 121–134. https://doi.org/10.1097/01.HMR.0000300822.70838.ed
23. Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
24. Kyriakeli, G., Georgiadou, A., Symeonidou, A., Tsimtsiou, Z., Dardavesis, T., & Kotsis, V. (2025). Patient safety culture among nurses in hospital settings worldwide: A systematic review and meta-analysis. *The Joint Commission Journal on Quality and Patient Safety, 51*(5), 350–360. https://doi.org/10.1016/j.jcjq.2025.01.007
25. Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology, 86*(1), 114–121. https://doi.org/10.1037/0021-9010.86.1.114
26. Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382–385. PMID: 3640358
27. *MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. Psychological Methods, 1*(2), *130–149. https://doi.org/10.1037/1082-989X.1.2.130*
28. Mannion, R., Davies, H. T. O., Jacobs, R., Kasteridis, P., Millar, R., & Freeman, T. (2017). Do hospital boards matter for better, safer, patient care? *Social Science & Medicine, 177*, 278–287. https://doi.org/10.1016/j.socscimed.2017.01.045
29. Mrayyan, M. T. (2022). Predictors and outcomes of patient safety culture: A cross-sectional comparative study. *BMJ Open Quality, 11*(2), e001713. https://doi.org/10.1136/bmjoq-2021-001713
30. Nie, Y., Mao, X., Cui, H., He, S., Li, J., & Zhang, M. (2013). Hospital survey on patient safety culture in China. *BMC Health Services Research, 13*, Article 228. https://doi.org/10.1186/1472-6963-13-228
31. Nieva, V. F., & Sorra, J. (2003). Safety culture assessment: A tool for improving patient safety in healthcare organisations. *Quality and Safety in Health Care, 12*(Suppl 2), ii17–ii23. https://doi.org/10.1136/qhc.12.suppl_2.ii17
32. Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill. https://doi.org/10.1177/014662169501900308
33. Parand, A., Dopson, S., Renz, A., & Vincent, C. (2014). The role of hospital managers in quality and patient safety: A systematic review. *BMJ Open, 4*(9), e005055. https://doi.org/10.1136/bmjopen-2014-005055
34. Parker, D. (2009). Managing risk in healthcare: Understanding your safety culture using the Manchester Patient Safety Framework (MaPSaF). *Journal of Nursing Management, 17*(2), 218–222. https://doi.org/10.1111/j.1365-2834.2009.00993.x
35. Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879
36. Pronovost, P. J., Goeschel, C. A., Olsen, K. L., Pham, J. C., Miller, M. R., Berenholtz, S. M., Sexton, J. B., Marsteller, J. A., Morlock, L. L., Wu, A. W., Loeb, J. M., & Clancy, C. M. (2009). Reducing health care hazards: Lessons from the commercial aviation safety team. *Health Affairs, 28*(3), w479–w489. https://doi.org/10.1377/hlthaff.28.3.w479
37. Reason, J. (1997). *Managing the risks of organisational accidents*. Ashgate. ISBN: 978-1840141054. https://doi.org/10.4324/9781315543543
38. Ree, E., Wiig, S., & Manser, T. (2024). Safety II in practice: How healthcare organisations implement resilience engineering. *Safety Science, 170*, Article 106360. https://doi.org/10.1016/j.ssci.2023.106360

39. Schwendimann, R., Milne, J., Frith, K., Doering, M., Ausserhofer, D., & Needham, D. M. (2024). The relationship between hospital governance structures and patient safety culture: A cross-sectional multi-site study. *BMC Health Services Research*, *24*(1), 312. https://doi.org/10.1186/s12913-024-10780-z
40. Sexton, J. B., Helmreich, R. L., Neilands, T. B., Rowan, K., Vella, K., Boyden, J., Roberts, P. R., & Thomas, E. J. (2006). The Safety Attitudes Questionnaire: Psychometric properties, benchmarking data, and emerging research. *BMC Health Services Research, 6*, Article 44. https://doi.org/10.1186/1472-6963-6-44
41. Singer, S. J., Falwell, A., Gaba, D. M., Meterko, M., Rosen, A., Hartmann, C. W., & Baker, L. (2009). Identifying organizational cultures that promote patient safety. *Health Care Management Review, 34*(4), 300–311. https://doi.org/10.1097/HMR.0b013e3181afc10c
42. Sorra, J. S., & Dyer, N. (2010). Multilevel psychometric properties of the AHRQ Hospital Survey on Patient Safety Culture. *BMC Health Services Research*, *10*, Article 199. https://doi.org/10.1186/1472-6963-10-199
43. Tsai, T. C., Jha, A. K., Gawande, A. A., Huckman, R. S., Bloom, N., & Sadun, R. (2015). Hospital board and management practices are strongly related to hospital performance on clinical quality metrics. *Health Affairs, 34*(8), 1304–1311. https://doi.org/10.1377/hlthaff.2014.1282
44. Vikan, M., Haugen, A. S., Bjørnnes, A. K., Valeberg, B. T., Deilkås, E. C. T., & Danielsen, S. O. (2023). The association between patient safety culture and adverse events: A scoping review. *BMC Health Services Research, 23*(1), Article 300. https://doi.org/10.1186/s12913-023-09332-8
45. Wazqar, D. Y., & Attallah, D. M. (2024). Patient safety culture predictors and outcomes for sustainable oncology nursing practice: A cross-sectional correlational study. *Journal of Clinical Nursing, 33*(2), 606–616. https://doi.org/10.1111/jocn.16881
46. Weick, K. E., & Sutcliffe, K. M. (2007). *Managing the unexpected: Resilient performance in an age of uncertainty* (2nd ed.). Jossey-Bass. ISBN: 978-0787996499
47. Williams, L. J., Hartman, N., & Cavazotte, F. (2010). Method variance and marker variables: A review and comprehensive CFA marker technique. *Organizational Research Methods, 13*(3), 477–514. https://doi.org/10.1177/1094428110366036
48. World Health Organization. (2019). *Patient safety: Global action on patient safety*. WHO. https://www.who.int/docs/default-source/patient-safety/who-ps-global-action.pdf