Published in final edited form as:

Arch Dis Child Fetal Neonatal Ed. 2012 March ; 97(2): F127–F132. doi:10.1136/archdischild-2011-300612.

The Safety Attitudes Questionnaire as a Tool for Benchmarking Safety Culture in the NICU

Jochen Profit, MD, MPH^{a,b,c}, Jason Etchegaray, PhD^d, Laura A Petersen, MD, MPH^{b,c}, J Bryan Sexton, PhD^e, Sylvia J Hysong, PhD^c, Minghua Mei, MS^{b,c}, and Eric J Thomas, MD, MPH^d

^aDepartment of Pediatrics, Baylor College of Medicine, Texas Children's Hospital, Houston, TX, USA

^bSection of Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, TX, USA

^cHouston Veterans Affairs (VA) Health Services Research and Development Center of Excellence, Health Policy and Quality Program, Michael E. DeBakey VA Medical Center; Houston, TX, USA

^dUniversity of Texas – Memorial Hermann Center for Healthcare Quality and Safety, University of Texas Medical School, Houston, TX, USA

^eDepartment of Psychiatry, Duke University School of Medicine; Duke University Health System, Durham, NC, USA

Abstract

background—NICU safety culture, as measured by the Safety Attitudes Questionnaire (SAQ), varies widely. Associations with clinical outcomes in the adult ICU setting make the SAQ an attractive tool for comparing clinical performance between hospitals. Little information is available on the use of the SAQ for this purpose in the NICU setting.

objectives—To determine whether the dimensions of safety culture measured by the SAQ give consistent results when used as a NICU performance measure.

Correspondence to: Jochen Profit, MD, MPH, Assistant Professor, Baylor College of Medicine, Section of Neonatology, Houston Center for Quality of Care & Utilization Studies, VA HSR&D (152), 2002 Holcombe Boulevard, Houston, TX 77030, Tel: 713.794.8609, Fax: 713.748.7359, profit@bcm.edu.

Address for reprint requests: Jochen Profit, MD, MPH, Houston VA HSR&D Center of Excellence (152), 2002 Holcombe Boulevard, Houston, TX 77030, phone: 713.794.8609, fax: 713.748.7359, email: profit@bcm.edu

The authors have no financial relationships relevant to this article to disclose.

AUTHOR CONTRIBUTIONS

JP - Designed the study questions, analyses, and drafted the manuscript. JE - Contributed to the design and framing of the study questions and the analyses. He also edited the manuscript. LP - Helped to frame the study question in a variation and performance measurement context. Helped interpret study findings, reviewed and edited the manuscript. BS - Conducted the original data collection, transmitted a de-identified data set to Dr. Profit, helped frame and design study questions, interpreted study findings, and reviewed and edited the manuscript. SH - Contributed to interpretation of analysis, framing of results in an industrial/organizational context, and edited the manuscript. MM - Developed an analysis plan, conducted analyses, reviewed results with study members, and edited the manuscript. ET - Acted as senior advisor to the project. Helped with framing of study questions, interpretation of results, contextual background and review/editing of manuscript.

methods—Cross-sectional survey of caregivers in twelve NICUs, using the six scales of the SAQ: teamwork climate, safety climate, job satisfaction, stress recognition, perceptions of management, and working conditions. NICUs were ranked by quantifying their contribution to overall risk-adjusted variation across the scales. Spearman Rank Correlation coefficients were used to test for consistency in scale performance. We then examined whether performance in the top four NICUs in one scale predicted top four performance in others.

results—There were 547 respondents in twelve NICUs. Of fifteen NICU-level correlations in performance ranking, two were greater than 0.7, seven were between 0.4 and 0.69, the six remaining were less than 0.4. We found a trend towards significance in comparing the distribution of performance in the top four NICUs across domains with a binomial distribution $p = .051$, indicating generally consistent performance across dimensions of safety culture.

conclusion—A culture of safety permeates many aspects of patient care and organizational functioning. The SAQ may be a useful tool for comparative performance assessments among NICUs.

Keywords

Infant; newborn; quality of care; benchmarking; safety culture

INTRODUCTION

The National Initiative for Children's Healthcare Quality has highlighted that "promoting safety requires changing the culture of medicine to recognize that the potential for errors exists and that teamwork and communication are the basis to guarantee change."¹

A culture of safety is the shared values, attitudes, perceptions, and patterns of behavior that determine the observable degree of effort with which organizational members direct their attention and actions towards minimizing patient harm.² Of the several safety culture survey instruments in the literature, the Safety Attitudes Questionnaire (SAQ) is widely used, has good psychometric properties³, and is associated with clinical outcomes.^{4–8}

The SAQ measures clinician assessments of "the way we do things around here," providing a snapshot of the unit-level care delivery context. Given that safety culture is associated with clinical outcomes, SAQ scores themselves might be used as a unit level clinical outcome tool for use in comparative performance measurement. Whether the SAQ would be valuable for this purpose has not been studied, however it does meet normative criteria in that a) significant variation in quality of care among providers exists, that b) this variation is not random, and that c) the measurement of provider performance will provide an impetus and path to improvement.

When a performance measurement instrument measures multiple aspects of quality, it is important to know whether performance across these aspects is consistent.^{9,10} High performance consistency suggests that care quality can be classified with a high degree of confidence, therefore supporting the use of the SAQ for purposes of comparative performance assessment. Performance consistency across the SAQ's domains would suggest that the instrument reflects caregiver perceptions of a unified systems-based construct that

permeates the care delivery system. Information from ongoing comparative measurement of safety culture in the NICU setting would offer important complementary information to current measurements based solely on clinical outcomes.

This study examines the extent to which the SAQ detects consistency of performance across NICUs.

METHODS

sample and procedure

The SAQ (ICU Version) was administered to all caregivers in 12 NICUs in a faith-based non-profit health system in July and August of 2004. All staff with a 50% commitment to the NICU for at least the four consecutive weeks prior to survey administration was invited to participate. This included critical care and other staff physicians, fellows/residents, critical care RNs, charge nurses, nurse managers, pharmacists, respiratory therapists, and nursing assistant/aides. In two NICUs there were no physician respondents because those physicians were assigned to complete surveys for other pediatric units where they met inclusion criteria more fully (i.e., they spent significantly more time in units other than the NICU). Surveys were administered during pre-existing departmental and staff meetings, together with a pencil and return sealable envelope to maintain confidentiality. Individuals not captured in pre-existing meetings, were hand delivered a survey, pencil and return envelope. This administration technique has generated high response rates.¹¹ The original study was approved by the Johns Hopkins University Institutional Review Board, and the analysis of a de-identified data set was approved by the Institutional Review Board at Baylor College of Medicine.

measures

The ICU version of the SAQ contains 65 items with response scales ranging from 1 (disagree strongly) to 5 (agree strongly). In previous SAQ development work 30 items loaded on six domains: teamwork climate, safety climate, job satisfaction, perceptions of management, stress recognition, and working conditions.

The SAQ also captures respondent characteristics including job position, time at institution, gender, race/ethnicity, and predominant work shift. To facilitate analysis, we grouped respondents with different job positions as follows: (1) *doctors* - includes all medicine and critical care physicians across all levels of training; (2) *nurses* - includes critical care nurses, licensed vocational nurses, nurse managers and charge nurses; and (3) *ancillary personnel* - includes pharmacists, nursing aids and assistants, ward clerks, and respiratory therapists. Respondents designated as “other” ($n=34$) were excluded from the analyses.

statistical analysis

We conducted a secondary data analysis of prospectively collected data. The NICU was the unit of analysis. Negatively worded items were reverse scored. Response scores were transformed to a 100-point scale using the following equation:

$$\text{Scale score for a respondent} = (\text{Mean of the scale items} - 1) * 25^{12}$$

Justification for aggregation—Given that culture is a unit-level phenomena (i.e., shared perceptions of different aspects of the work environment), it is necessary for researchers to demonstrate that aggregation of individual survey responses within each unit is warranted. Two statistical conditions for aggregation need to be satisfied: 1) respondents from each NICU reported similar scores for the NICU on a given item and 2) there is significant variance for a given item between NICUs. The four main metrics to determine whether aggregation is appropriate are - ANOVA (one way), ICC(1), ICC(2), and $r_{wg(j)}$.² These metrics were computed for each scale of the SAQ domains. We then calculated a “composite score” from the arithmetic mean of the six SAQ scale scores.

Basic descriptive analyses of unadjusted scale scores are published elsewhere.¹³ Here we present the variation in scale scores across NICUs adjusted for site and respondent mix, with which they are significantly associated. For each scale we ranked NICUs according to their composite score.

To what extent can the SAQ detect consistency of performance across NICUs?—We examined the degree to which superior NICU caregiver assessments in one key scale (safety climate) was associated with superior perceptions in the other scales. We used two approaches to test for consistency. First, we transformed ratings on individual scales into ranks and tested for correlation across ranks using the Spearman Rank Correlation coefficient. Second, we compared the distribution of ranking in the top four NICUs across scales to a binomial distribution using a chi-square test.⁹ A negative test indicates independence of performance among scales. For all analyses, we considered two-sided P values of $< .05$ as statistically significant.

RESULTS

characteristics of respondents and NICUs

We received completed surveys from 547 of 639 respondents for a response rate of 86% (range, 69 – 100%). Table 1 describes the characteristics of the survey sample. Of note is the strong preponderance of the aggregated category of nursing respondents (82%) in this sample, especially when compared to doctors (5%). NICUs A and I had only ten and twelve respondents, respectively.

justification of aggregation of scale scores to a composite score

Table 2 shows the details of the psychometric results justifying aggregation to a composite score. For all of the SAQ dimensions except stress recognition, these metrics were acceptable. For stress recognition, the one-way ANOVA was not significant ($p < .06$), ICC(1) = .02, was outside typical values of .05 to .30, and ICC (2) = .69 and $r_{wg(j)}$ = .64 were just below the traditional .70 cutoff level needed. While we decided to aggregate all scales, we are cautious in our interpretation of stress recognition.

performance on composite safety culture score and individual domains

Table 3 describes the range of NICU performance on the studied quality domains. We display adjusted results and NICU ranks. Rankings across domains were quite stable

indicating that performance tracks across domains. NICU performance within domains was quite variable except for “stress recognition” (range, 54 – 64).

to what extent can the SAQ detect consistency of performance across NICUs?

Table 3 displays the NICU level rank correlation matrix among quality domains. Except for the “stress recognition” domain, correlations were moderate to strong. Of fifteen NICU level rank correlations, six were significant at $p < .05$. Correlations between pairs of safety culture domains were strong ($\rho = 0.7$) for two pairs, moderate ($\rho = 0.4 - 0.69$) for seven pairs, weak for ($\rho = 0.2 - 0.39$) for three pairs, and absent ($\rho = 0.2$) for three pairs.

Consistency of high performance across domains of safety culture—High performance of NICUs was consistent across SAQ domains. The number of times NICUs were among the top four NICUs (a “high performer”) for the six safety attitudes domains ranged from none (never in the top four) to five. Figure 1 shows the observed and expected distribution under an assumption that “high performance” on different domains occurs at random (according to a binomial distribution in which the probability of success on each trial is 0.25 and the six trials are independent). There was a trend towards significance between the actual and the binomial distributions ($p = 0.05$), indicating that one can infer high overall performance based on performance on individual domains.

DISCUSSION

In this study, we examined the SAQ as a tool for comparative performance assessment of safety culture among twelve NICUs. The most notable conclusion is that while there is wide variation of performance within domains of the SAQ, NICUs were quite consistent in their performance across domains.

The consistency of NICU performance across domains of the SAQ implies that performance on one subscale predicts performance on another. This suggests that the different scales of the SAQ may measure a cohesive underlying construct. NICUs with high performance on safety, value teamwork, have better working conditions, relations with management, and job satisfaction. This result makes the SAQ an attractive tool for comparative measurement of safety culture among NICUs.

Comparative measurement of safety culture in the NICU setting may be particularly salient as preterm infants are fragile, often very ill, and exposed to complex and prolonged intensive health care interventions. These circumstances make preterm infants vulnerable to lapses in patient safety.¹⁴ In a study of voluntarily reported errors in the NICU setting, poor teamwork and poor communication contributed to errors in 9 and 22%, respectively.¹⁵ In the labor and delivery setting poor teamwork and communication breakdowns were a root cause of perinatal deaths and injuries in 55 and 72%, respectively.¹⁶ Team performance is especially important in emergent situations where a rescue team must assemble quickly, communicate clearly and collaborate effectively to avoid needless morbidity or mortality.¹⁷

In the NICU setting, safety culture has not been widely studied. Despite a clear rationale to improve safety culture and encouraging literature on positive associations with improved

clinical outcomes in other areas of health care⁴⁻⁸, it is not yet known whether and how improvements in NICU safety culture will translate into improved quality of care and outcomes for infants. In this study two of the SAQ domains, stress recognition and perceptions of management, did not link well to the others. This finding may be explained in a number of ways. Realistically acknowledging threats to safety and quality (stress recognition) and having the requisite trust in leadership to engage meaningfully in QI efforts (perceptions of management) may act as gatekeepers that subsequently facilitate better teamwork and safety-related norms to flourish. As such, we could expect associations between these two domains and the remaining four domains to be lower. In specific NICUs, where intense and successful QI has taken place over many years, we would expect the relationships to be higher for perceptions of management in particular. Second, improvements in stress recognition and perceptions of management may only represent a first step in a series of actions a NICU needs to take to improve clinical outcomes. For example, one study found associations between a non-punitive approach to error, hospital management support for patient safety, and overall perceptions of safety with incident reporting behavior in the NICU.¹⁸ Possibly, organizations which facilitate openness in error detection and encourage learning may eventually achieve better clinical results. Third, the questions asked in this version of the SAQ related to hospital management, not unit management (current versions of the SAQ distinguish between various levels of leadership).

In an accompanying paper we demonstrated wide variations in safety culture among this sample of NICUs.¹³ On the other hand, in previous work, we found little performance consistency among NICUs across various common measures of clinical quality.¹⁹ Clearly, more work in the NICU setting, including prospective hypothesis testing, is required to better understand the correlation between safety culture, clinical processes, operational processes and health outcomes.

Despite these unresolved areas of inquiry, the ability of the SAQ to capture NICU's safety culture makes it attractive for comparative measurement, especially given that individual scales and items of the SAQ can be linked to specific safety interventions. For example, collaborative rounds²⁰, aviation based crew resource management training²¹, or improved communication in hierarchy^{22,23} improve teamwork whereas Leadership WalkRounds^{24,25} or a Comprehensive Unit-based Safety Program²⁶ improve safety. In addition, intensive care unit caregiver safety culture assessments have shown to predict their ability to implement complex safety practices.²⁷

Since the SAQ measures frontline worker assessments of safety culture, we believe its use for comparative performance measurement is most valid for purposes of internal benchmarking and quality improvement. By internal benchmarking we mean here the use of the SAQ within individual NICUs or within neonatal quality collaboratives that already collect and compare clinical data. In this environment, the SAQ provides useful and complementary data to clinical quality of care measures.

Traditionally, NICUs have focused on disease-specific aspects of clinical care and devised remedies for improvement.^{28,29} Although this approach is intuitive and necessary, it may not address underlying preconditions which may enable many adverse outcomes. In contrast,

systematic monitoring and efforts to improve safety culture may improve the system of care delivery by promoting safe and teamwork-based care of infants throughout their hospital stay.

We emphasize the importance of interpreting our results in light of the intended context of the study. For this proof of concept study, we used the mean score across the SAQ's domains as a composite index for benchmarking. Although aggregation works technically, such a score implies that all domains are equally important and that poor performance in one domain (safety climate) can be offset with good performance in another (stress recognition). A better solution would be a composite that encourages high performance in all domains. Methods are available to accomplish this³⁰, and we are testing these in our work on a clinical composite index for NICU care.³¹ In order to ensure that an SAQ composite score would be actionable, reliable, and valid in the eyes of frontline workers, future research will need to test the links between safety culture domains scores and NICU outcomes that include clinical and operational metrics.

Finally, our study sample was quite small and from a single health system. Although it is possible that our results do not generalize to the wider universe of NICUs, our study is strengthened by finding consistency among even a small sample.

CONCLUSION

We found moderate to strong correlations in NICU caregiver safety culture assessments across the dimensions of the SAQ. Results of the SAQ may provide a useful starting point for assessing and trending of safety culture among NICUs and can serve as a yardstick by which to assess the need for, pace of, and opportunities underlying quality improvement initiatives.

Acknowledgments

Grant support: Jochen Profit's contribution is supported in part by the Eunice Kennedy Shriver National Institute of Child Health and Human Development #1 K23 HD056298-01 (PI: Profit). Dr. Petersen was a recipient of the American Heart Association Established Investigator Award (#0540043N) at the time that this work was carried out. Drs. Petersen, Hysong, and Mr. Mei also receive support from a Veterans Administration Center Grant (VA HSR&D CoE HFP90-20). Dr. Hysong's contribution is supported in part by the Department of Veterans Affairs Health Services Research and Development Program (#CD 2-07-0818). Dr. Thomas' effort is supported in part by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development #1 K24 HD053771-01 (PI: Thomas) and #1 PO1 HS1154401 (PI: Thomas). Dr. Sexton received support from an Agency for Healthcare Research and Quality (AHRQ) grant # 1UC1HS014246. Dr. Etchegaray's effort is supported by a K02 award from AHRQ #1 K02 HS017145-02 and the University of Texas at Houston-Memorial Hermann Center for Quality and Safety.

In addition to thanking the NICU personnel who participated by sharing their assessments, we would like to acknowledge the contribution of the study staff, Christen Fullwood, Chris Holzmüller, Angelina Barbosa and Linda Marcellino.

Abbreviations

NICU	neonatal intensive care unit
SAQ	safety attitudes questionnaire

Reference List

1. National Initiative for Children's Health Care Quality Project Advisory Committee. Principles of Patient Safety in Pediatrics. *Pediatrics*. 2001; 107:1473–5. [PubMed: 11389280]
2. Vogus TJ, Sutcliffe KM. The Safety Organizing Scale: development and validation of a behavioral measure of safety culture in hospital nursing units. *Med Care*. 2007; 45:46–54. [PubMed: 17279020]
3. Sexton JB, Helmreich RL, Neilands TB, et al. The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res*. 2006; 6:44. [PubMed: 16584553]
4. Thomas EJ, Sexton JB, Helmreich RL. Discrepant attitudes about teamwork among critical care nurses and physicians. *Crit Care Med*. 2003; 31:956–9. [PubMed: 12627011]
5. Colla JB, Bracken AC, Kinney LM, et al. Measuring patient safety climate: a review of surveys. *Qual Saf Health Care*. 2005; 14:364–6. [PubMed: 16195571]
6. Kho ME, Carbone JM, Lucas J, et al. Safety Climate Survey: reliability of results from a multicenter ICU survey. *Qual Saf Health Care*. 2005; 14:273–8. [PubMed: 16076792]
7. Sexton JB, Holzmüller CG, Pronovost PJ, et al. Variation in caregiver perceptions of teamwork climate in labor and delivery units. *J Perinatol*. 2006; 26:463–70. [PubMed: 16775621]
8. Modak I, Sexton JB, Lux TR, et al. Measuring safety culture in the ambulatory setting: the safety attitudes questionnaire--ambulatory version. *J Gen Intern Med*. 2007; 22:1–5. [PubMed: 17351834]
9. Wilson IB, Landon BE, Marsden PV, et al. Correlations among measures of quality in HIV care in the United States: cross sectional study. *BMJ*. 2007; 335:1085–91. [PubMed: 17967826]
10. Rosenthal GE. Weak associations between hospital mortality rates for individual diagnoses: implications for profiling hospital quality. *Am J Public Health*. 1997; 87:429–33. [PubMed: 9096547]
11. Pronovost PJ, Sexton BJ. Assessing safety culture: guidelines and recommendations. *Qual Saf Health Care*. 2005; 14:231–3. [PubMed: 16076784]
12. Sexton, BJ.; Helmreich, RL.; Thomas, EJ. [accessed 2 Mar 2010] Safety Attitude Questionnaire Scale Computation Instructions. 2011. http://www.uth.tmc.edu/schools/med/imed/patient_safety/questionnaires/SAQBibliography.html
13. Profit J, Etchegaray J, Petersen LA, Sexton JB, Hysong SJ, Mei M, Thomas EJ. Neonatal intensive care unit safety culture varies widely. *Archives of Disease in Childhood Fetal and Neonatal Edition*.
14. Gray JE, Davis DA, Pursley DM, et al. Network analysis of team structure in the neonatal intensive care unit. *Pediatrics*. 2010; 125:e1460–e1467. [PubMed: 20457681]
15. Suresh G, Horbar JD, Plsek P, et al. Voluntary anonymous reporting of medical errors for neonatal intensive care. *Pediatrics*. 2004; 113:1609–18. [PubMed: 15173481]
16. JCAHO. [accessed 25 Jan 2008] Sentinel Event Alert #30: Preventing infant death and injury during delivery. 2005. http://www.jointcommission.org/SentinelEvents/SentinelEventAlert/sea_30.htm
17. Clarke SP, Aiken LH. Failure to rescue. *Am J Nurs*. 2003; 103:42–7. [PubMed: 12544057]
18. Snijders C, Kollen BJ, van Lingen RA, et al. Which aspects of safety culture predict incident reporting behavior in neonatal intensive care units? A multilevel analysis. *Crit Care Med*. 2009; 37:61–7. [PubMed: 19050606]
19. Profit J, Pietz K, Gould JB, Zupancic JA, Petersen LA. Top performance in one measure of quality does not predict top performance in others. *E-PAS*. 2008:634457.13.
20. Uhlig, PN.; Haan, CK.; Nason, AK.; Niemann, PL.; Camelio, A.; Brown, J. Improving Patient Care by the Application of Theory and Practice from the Aviation Safety Community. Columbus, OH: 2001.
21. McCulloch P, Mishra A, Handa A, et al. The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre. *Qual Saf Health Care*. 2009; 18:109–15. [PubMed: 19342524]

22. Leonard M, Graham S, Bonacum D. The human factor: the critical importance of effective teamwork and communication in providing safe care. *Qual Saf Health Care*. 2004; 13:i85–i90. [PubMed: 15465961]
23. Karima Velji GRB. Effectiveness of an adapted SBAR communication tool for a rehabilitation setting. *Healthcare Quarterly*. 2008; 11:72–9. [PubMed: 18382165]
24. Thomas EJ, Sexton JB, Neilands TB, et al. The effect of executive walk rounds on nurse safety climate attitudes: a randomized trial of clinical units[ISRCTN85147255] [corrected]. *BMC Health Serv Res*. 2005; 5:28. [PubMed: 15823204]
25. Frankel A, Grillo SP, Pittman M, et al. Revealing and resolving patient safety defects: the impact of Leadership WalkRounds on frontline caregiver assessments of patient safety. *Health Serv Res*. 2008; 43:2050–66. [PubMed: 18671751]
26. Pronovost P, Weast B. Implementing and validating a comprehensive unit-based safety program. *J Patient Saf*. 2005:33–40.
27. Sexton JB, Berenholtz SM, Goeschel CA, et al. Assessing and improving safety climate in a large cohort of intensive care units. *Crit Care Med*. 2011
28. Horbar JD, Rogowski J, Plsek PE, et al. Collaborative quality improvement for neonatal intensive care. NIC/Q Project Investigators of the Vermont Oxford Network. *Pediatrics*. 2001; 107:14–22. [PubMed: 11134428]
29. Horbar JD, Carpenter JH, Buzas J, et al. Collaborative quality improvement to promote evidence based surfactant for preterm infants: a cluster randomised trial. *BMJ*. 2004; 329:1004. [PubMed: 15514344]
30. Profit J, Typo KV, Hysong SJ, et al. Improving benchmarking by using an explicit framework for the development of composite indicators: an example using pediatric quality of care. *Implement Sci*. 2010; 5:13. [PubMed: 20181129]
31. Profit J, Gould JB, Zupancic JA, et al. Formal selection of measures for a composite index of NICU quality of care: Baby-MONITOR. *J Perinatol*. 2011 Feb 24. Epub ahead of print.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Safety culture varies across NICUs and may become a target for comparative performance measurement.

WHAT THIS STUDY ADDS

NICU caregiver safety culture assessments showed moderate to strong correlations across the dimensions of the Safety Attitudes Questionnaire.

The Safety Attitudes Questionnaire can be used to compare safety culture across NICUs.

Figure 1.

Distribution of rankings in the top 4 NICUs across scales compared to a binomial distribution. The binomial distribution indicates expected random performance of NICUs across safety culture scales. If actual NICUs are consistently high or poor performers the distribution should be U-shaped. We indeed found an approximate U-shaped distribution among our study sample (χ^2 value = 9.43, $P = .051$).

Table 1

Respondent and NICU characteristics

Characteristic Respondents	n (%)
Job positions	
ancillary staff	71 (14.2)
nurse	404 (82.0)
doctor	24 (4.8)
Experience	
5 years	151 (30.0)
6 – 19 years	142 (28.2)
20 years	210 (41.8)
Race	
black	27 (5.4)
white	455 (91.2)
asian	17 (3.4)
Gender	
male	20 (3.9)
female	492 (96.1)
Work Shifts	
days	217 (48.0)
evenings	22 (4.9)
nights	159 (35.2)
variable shifts	54 (12.0)
NICUs (n=12)	
A	10 (1.8)
B	57 (10.4)
C	25 (4.6)
D	42 (7.7)
E	71 (13.0)
F	22 (4.0)
G	39 (7.1)
H	39 (7.1)
I	12 (2.2)
J	71 (13.0)
K	50 (9.1)
L	109 (20.0)

NICU - neonatal intensive care unit

Table 2

Statistical justification for aggregation of scale scores to a composite score

SAQ scales	ICC (1)	ICC (2)	ANOVA P-Value	Median of Rwg (j)
Teamwork Climate	0.08	0.90	< 0.01	0.84
Safety Climate	0.07	0.89	< 0.01	0.88
Job Satisfaction	0.06	0.88	< 0.01	0.87
Stress Recognition	0.02	0.69	0.06	0.64
Perception of Management	0.11	0.93	< 0.01	0.74
Working Conditions	0.06	0.86	< 0.01	0.77
<i>Expected range to justify aggregation</i>	<i>0.05 – 0.3</i>	<i>> 0.7</i>	<i>< 0.05</i>	<i>> 0.7</i>

ICC (1): proportion of total variance explained by NICU membership

ICC (2): overall estimate of the reliability of NICU means

One-way ANOVA: indicates that responses differ between NICUs

Rwg (j) is the within-NICU inter-rater agreement index

ICC – intraclass correlation coefficient

Table 3

Adjusted composite and mean scale scores and (ranks)

NICUs	Composite Score	Teamwork Climate	Safety Climate	Job Satisfaction	Stress Recognition	Perception of Management	Working Conditions
A	76.8 (1)	78.8 (6)	80.1 (5)	87.9 (1)	60.1 (5)	80.1 (1)	72.3 (4)
B	75.0 (4)	84.1 (2)	82.6 (2)	81.4 (7)	56.5 (9)	69.6 (5)	75.9 (1)
C	75.2 (3)	80.7 (4)	81.2 (4)	84.2 (3)	61.4 (4)	71.8 (4)	70.0 (6)
D	76.4 (2)	85.6 (1)	85.7 (1)	85.5 (2)	57.6 (7)	68.1 (6)	75.2 (2)
E	74.1 (5)	73.2 (9)	76.8 (7)	81.6 (6)	64.9 (1)	76.3 (2)	70.7 (5)
F	71.7 (6)	80.6 (5)	77.9 (6)	81.2 (8)	54.7 (11)	65.4 (7)	72.9 (3)
G	69.6 (8)	72.8 (10)	74.1 (9)	77.8 (9)	56.6 (8)	72.4 (3)	65.4 (8)
H	69.8 (7)	83.9 (3)	73.8 (10)	83.1 (4)	57.9 (6)	60.8 (9)	59.9 (10)
I	68.1 (10)	74.2 (8)	75.4 (8)	83.1 (5)	54.1 (12)	62.6 (8)	64.4 (9)
J	68.9 (9)	76.4 (7)	81.7 (3)	74.9 (10)	54.7 (10)	58.9(10)	67.5 (7)
K	63.6 (11)	61.7 (12)	68.7 (11)	72.9 (11)	63.6 (2)	54.8 (11)	59.2 (11)
L	60.6 (12)	68.2 (11)	66.7 (12)	64.3 (12)	62.7 (3)	43.7 (12)	57.8 (12)

Scale scores range from 0–100. Multivariate model with NICU as random effect. Adjustment made for job position. Ranks are shown in brackets. The composite score is the average of the mean scale scores. Abbreviations: NICU, neonatal intensive care unit. NICUs (k = 12), responses (n = 499)

Table 4

Correlation of NICU ranks across scales

	TC	SC	JS	SR	PM	WC
Teamwork Climate (TC)	1	0.74**	0.66*	-0.25	0.22	0.69*
Safety Climate (SC)		1	0.57	-0.10	0.44	0.92***
Job Satisfaction (JS)			1	0.00	0.57	0.55
Stress Recognition (SR)				1	0.24	-0.03
Perception of Management (PM)					1	0.61*
Working Conditions (WC)						1

Spearman rank correlation coefficient.

* $p < .05$;
** $p < .01$;
*** $p < .001$.