



Published in final edited form as:

*J Perinatol.* 2012 April ; 32(4): 247–252. doi:10.1038/jp.2011.199.

## Do Practicing Clinicians Agree with Expert Ratings of Neonatal Intensive Care Unit Quality Measures?

Marc Kowalkowski, M.S.<sup>1,2</sup>, Jeffrey B Gould, M.D., M.P.H.<sup>3,4</sup>, Carl Bose, M.D.<sup>5</sup>, Laura A Petersen, M.D., M.P.H.<sup>1,2</sup>, and Jochen Profit, M.D., M.P.H.<sup>1,2,6</sup>

<sup>1</sup>Section of Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, TX, United States

<sup>2</sup>Health Services Research and Development Center of Excellence, Health Policy and Quality Program, Michael E. DeBakey VA Medical Center, Houston, TX

<sup>3</sup>California Perinatal Quality Care Collaborative, Palo Alto, CA

<sup>4</sup>Division of Neonatology, Perinatal Epidemiology and Health Outcomes Research Unit, Stanford University, Palo Alto, CA

<sup>5</sup>Division of Neonatal-Perinatal Medicine, Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC

<sup>6</sup>Department of Pediatrics, Baylor College of Medicine, Texas Children's Hospital, Houston, TX

### Abstract

**Objective**—To assess the level of agreement when selecting quality measures for inclusion in a composite index of neonatal intensive care quality (Baby-MONITOR) between two panels: one comprised of academic researchers (Delphi) and another comprised of academic and clinical neonatologists (Clinician).

**Design/Methods**—In a modified Delphi process, a panel rated twenty eight quality measures. We assessed clinician agreement with the Delphi panel by surveying a sample of forty eight neonatal intensive care practitioners. We asked the clinician group to indicate their level of agreement with the Delphi panel for each measure using a five- point scale (much too high, slightly too high, reasonable, slightly too low, and much too low). In addition, we asked clinicians to select measures for inclusion in the Baby-MONITOR based on a yes or no vote and a pre-specified two-thirds majority for inclusion.

**Results**—Twenty three (47.9%) of the clinicians responded to the survey. We found high levels of agreement between the Delphi and clinician panels, particularly across measures selected for the Baby-MONITOR. Clinicians selected the same nine measures for inclusion in the composite as the Delphi panel. For these nine measures, 74% of clinicians indicated that the Delphi panel rating was ‘reasonable’.

---

**Corresponding Author:** Jochen Profit, M.D., M.P.H., Houston VA HSR&D Center of Excellence, 2002 Holcombe Boulevard (152), Houston, TX 77030, profit@bcm.edu.

#### Conflict of interest

Drs. Profit and Gould are serving as Expert Consultants with the Vermont Oxford Networks NICQ 7 Quality Improvement Collaborative. Dr. Gould is the Principal Investigator for the California Perinatal Quality Care Collaborative.

Author Contributions: MAK, JP, JBG, CB, and LAP led the design, data analysis, and interpretation of results. MAK coordinated the study, prepared study materials, and assisted with data acquisition. MAK and JP undertook data analysis and interpretation. MAK and JP wrote the first draft of the manuscript. CB, JBG, and LAP reviewed and edited the manuscript. JP is guarantor of the paper.

**Conclusions**—Practicing clinicians agree with an expert panel on the measures that should be included in the Baby-MONITOR, enhancing face validity.

### Keywords

infant; newborn; quality of health care; measurement; composite indicator

## INTRODUCTION

Composite indicators are being used increasingly to track provider performance in adult health care settings (1–3). Composite indicators aggregate individual measures into a single summary score. They can provide broad insights and trends of quality for external benchmarking against other providers' institutions and facilitate the tracking of quality improvement efforts within institutions. However, the process of developing composite indicators is complex and developers have to make choices in their construction that may significantly influence performance ratings. Thus, a standardized and explicit approach to indicator development is necessary and has been described elsewhere.(4;5)

Briefly, composite indicator development begins with a theoretical framework which provides the foundation for the selection and aggregation of variables. Data selection targets variables with strong analytical soundness, measurability, and relevance to the measured event. Additionally, developers examine the completeness, structure, and comparability of each selected variable, as well as the methods for weighting and aggregation. Finally, developers evaluate uncertainty (i.e., around aggregate ranks), transparency (i.e., explaining relative significance of individual domains), and linkages (i.e., to other indicators), and provide coherent visual tools for interpretation.

We are working to develop a CI of quality delivered to very low birth weight infants (VLBW), specifically the Measure Of Neonatal InTensive care Outcomes Research, or Baby-MONITOR. In a previous study, a panel of neonatal outcome researchers utilized a modified Delphi method to assign scores from 1–9 (9 = best) to 28 measures of quality routinely collected at the NICU-level by the California Perinatal Quality Care Collaborative (CPQCC) and Vermont Oxford Network (VON) (6). Based on the panel's ratings, nine measures were selected for inclusion in the Baby-MONITOR. It is possible that a sample of largely academic researchers and quality improvement experts, which composed our original panel, would have a biased view regarding quality of care measurement. We address this concern by examining agreement between the Delphi panel's selections of quality measures with those of clinical providers of neonatal intensive care.

## METHODS

### Physician Nominations

Clinical neonatologists were nominated for study participation via executive channels within the American Academy of Pediatrics (AAP) Section on Perinatal Pediatrics. We invited the District Chairs from each of the ten AAP districts to provide three to four nominations of clinical neonatologists from within their district for study participation. Nominees were Board-certified in Neonatal Perinatal Medicine and had experience in quality improvement. They were judged by the nominator to be respected by their peers. Nominees were selected to create a sample that represented both public and private hospitals from a geographic distribution throughout the districts.

Nominees were contacted electronically. Consent was implied by those who responded with intent to participate. The study was approved by the Baylor College of Medicine Institutional Review Board.

### Measure Selection

Details of the clinical measures and the process for selecting a candidate measure set have been previously described (6). Briefly, 28 routinely reported process-and-outcome quality measures were selected by consensus of a panel of quality improvement researchers. These measures were selected from CPQCC and VON operations manuals. The selection process was designed for inclusiveness and only excluded measures of surgical quality. For the current study, one of the 28 previously selected measures (duration of ventilation) was excluded because of inconsistent recording in the CPQCC database.

### Survey Instrument

Participants were provided with a summary of each quality measure along with ratings from the Delphi panel. The Delphi panel's ratings were presented as median scores across five measurement domains: importance, reliability, validity, scientific soundness, and usability (rated on a scale of 1–9; 9 = best). In addition, we provided the Delphi panel's overall median rating.

### Data Collection

Clinicians utilized the measure summary information, as well as the ratings generated by the previous Delphi panel, to assess their level of agreement with the panel's ratings. Clinicians reported one measure of agreement, using a five-point scale (the measure is rated much too high, slightly too high, reasonably, slightly too low, or much too low), to represent their assessment of the panel's full complement of ratings, including individual domain and overall scores. In addition to structured multiple-choice response options, respondents provided additional free-text feedback regarding the measure itself and their rationale for selecting a specific response option. Following the evaluation of each measure independently, respondents designated which of the 27 measures to incorporate into the Baby-MONITOR. Prior to data collection, we specified that measures must receive support from a two-thirds majority to be selected for the composite. Lastly, clinicians provided information regarding practice setting (i.e., academic versus private; NICU-level designation [using established criteria (7)]; years in practice; level of research participation; quality improvement involvement; and location of medical training [e.g., foreign versus US medical graduate]).

### Statistical Evaluation

SAS version 9.2 (SAS Institute Inc., Cary, North Carolina, USA) was utilized for data analyses. Descriptive statistics were calculated to determine the distribution of characteristics in the Clinician panel. Additionally, frequency distributions were generated to determine the percentage of agreement with Delphi panel ratings for each measure. Because the information provided by the respondents was being used to determine the appropriateness of the measures included in the BABY-Monitor, we defined "agreement" as all ratings that did not indicate a strong divergence from the expert panel's opinions (e.g., much too high/much too low). Potential differences in the reported percentage of agreement were evaluated in comparisons of clinicians practicing in academic versus private settings, using chi-square tests. Additional comparisons were conducted around the percentage of agreement between measures selected versus not selected for the composite by the Delphi panelists.

## RESULTS

### Demographics

District chairs nominated 48 neonatologists, representing all ten AAP districts. Of the 48 nominees, 30 clinicians consented to participate. At the completion of the study, 23 of the 30 participants consented had completed and submitted responses. Nominations and responses were received over a period of nine months, from January to September 2009.

Clinician characteristics are shown in Table 1. Frequency distributions for practice setting; years in practice since completion of subspecialty training; NICU level designation; involvement in quality improvement; clinical research participation; and foreign medical education are displayed. Respondents represented geographically diverse regions of the US, including nine out of ten districts from the AAP Section on Perinatal Pediatrics. None of the nominees from District 10, which includes Alabama, Florida, Georgia, and Puerto Rico, provided responses. Overall, a majority of the respondents were academic clinicians (67%); in practice for ten years or more (86%); involved in quality improvement (95%); and US medical graduates (100%). Approximately one half of respondents reported practicing medicine at a NICU with a Level 3C (7) subspecialty designation and participating in clinical research as either a principal or co-investigator. Similar to respondents, a majority of non-respondents were from academic hospitals (60%; data not shown).

### Clinician Agreement

Surveyed clinicians reported a high degree of agreement with the previous Delphi panel's ratings (see Table 2). Overall, 62% of clinician responses across the 27 measures indicated the ratings provided by the Delphi panel were 'quite reasonable,' and only 6% indicated extreme disagreement with the Delphi panel's ratings ('much too high'/'much too low'). Furthermore, the Delphi panel's ratings for one measure, timely ROP examination, received almost unanimous support from clinicians (96%). However, there were also measures that engendered disagreement between the panels. The three measures with the most discordant responses were feeding with human milk only at discharge; surfactant administration within two hours of birth; and intracranial hemorrhage severity > grade 2. For example, clinician opinion regarding the Delphi panel's ratings of 'feeding with human milk only at discharge' was distributed across all response options, including 'much too high' and 'much too low' ('reasonable'=46%). However, when taking into account slight agreement and disagreement, even this most "controversial" measure achieved an overall agreement of 80%.

Appendix 1, available electronically as supplementary material, summarizes clinician feedback regarding aspects of each of the measures assessed by the panel. On average, there were seven comments provided for each measure. Comments were categorized by measure definition, usability, and validity. With the exception of duplicate content, all of the clinician comments were presented in Appendix 1.

Respondents reported several recurring criticisms of the measures under evaluation. Frequently, clinicians indicated that patient transfers in and out of the NICU may significantly impact inter-unit comparisons. Thus, to facilitate accurate assessments, definitions must appropriately adjust for patient transfers. Additionally, clinicians reported concerns about quality of care being measured on variables significantly influenced by factors outside of their control. Furthermore, respondents expressed apprehension around the variability in policies, practices, and techniques that may obscure comparisons between different units. Finally, respondents suggested that several measures were susceptible to "gaming" and that risk-adjustment models have not been sufficiently tested and validated.

## Differences by Clinician Practice Setting

Clinicians from different practice settings responded similarly to the ratings for most measures ( $p>0.1$ ). However, results indicate clinicians practicing in private settings viewed some of the panel's ratings differently than those practicing in academic settings. Most notably, clinicians in private settings reported that surfactant administration was rated too high by the research panel while those affiliated with academic institutions reported the panel's ratings were reasonable ( $p=0.05$ ). Additionally, no private practice setting clinicians ( $n=7$ ) voted to include surfactant in the composite. Conversely, compared to academic clinicians, those practicing in private settings tended to report that length of stay was rated too low ( $p=0.06$ ) and to include the measure in the composite indicator more frequently than academic clinicians ( $p<0.1$ ). Neither of these measures was included in the Baby-MONITOR by the Delphi panel.

## Measure Selection

The Delphi panel had selected nine measures of quality for inclusion in the Baby-MONITOR. In the current study, we pre-specified a two-thirds majority of clinician responses as evidence for agreement. Using this criterion, the clinicians independently selected the same nine measures as the Delphi panel. The percentage agreement ranged from 66% to 100% across the selected measures. Clinicians reported significantly higher agreement with the Delphi panel's ratings across the measures included in the composite, compared to measures not included ( $p<0.001$ ). Table 3 presents the median ratings for each of the measures selected by the Delphi panel; the percentage of clinician agreement with the panel's ratings; and the percentage of clinicians in favor of including the measure in the Baby-MONITOR.

## DISCUSSION

We report results from an electronic survey to assess agreement with the selection of measures of NICU quality-of-care for inclusion in the Baby-MONITOR. Our principal finding was a high level of agreement with the Delphi panel's selection of measures into the Baby-MONITOR by a sample of clinical neonatologists.

Comparative performance measurement has become a national health policy priority in an attempt to promote better quality of care and decrease health care expenditures (8;9). Composite indicators have become popular because they are able to summarize otherwise complex information (5;10). However, the selection of quality measures into composite indicators must undergo a careful vetting process in order to avoid undue bias on part of the developers. We have undertaken a careful, explicit, and structured measure selection process for the Baby-MONITOR (6). Results from the current study demonstrate a high degree of agreement between the clinical neonatologists surveyed and the Delphi panel's ratings. The Delphi and clinical panels identified identical measure sets for the Baby-MONITOR. We, therefore, believe this measure set to represent a state-of-the-art selection of robust measures of neonatal intensive care quality. This measure set is based on available measures within the CPQCC and VON consortia. These organizations include more than nine hundred member institutions; therefore, the measure set has the potential for broad application.

The substantial agreement between the Delphi and Clinical panels among the measures selected for the Baby-MONITOR provides powerful support for its acceptability as a measurement tool. The panels also agreed strongly with regard to the measures least suitable to measure NICU quality of care (e.g., VLBW volume, any intracranial hemorrhage, oxygen at discharge), which calls into question the utility of allocating resources to modifying the outcomes described in these measures. Lastly, the disagreement observed across measures

with moderate ratings (e.g., only human milk at time of discharge, surfactant within two hours of birth, severe intracranial hemorrhage) suggests that these measures may be functional but require modifications or additional research to enhance their value as quality measures.

Comments in Appendix 1 emphasize potential subjective weaknesses of each of the quality measures. However, empirical evidence will be needed to assess whether changes in measure definitions would improve ratings of measures not currently included in the Baby-MONITOR (11). Additionally, these comments highlight the struggle involved in the development of standardized quality measures. Inherent practice-based differences create challenges in balancing measurement uniformity with specificity and applicability. These factors support the utilization of longitudinal comparisons within organizations, rather than comparisons across organizations.

Clinicians from different practice settings agreed upon most of the ratings of the Delphi panel. Most importantly, clinicians from different practice settings reported agreement for all measures selected for inclusion in the Baby-MONITOR. However, in our sample, clinicians practicing in private hospitals indicated length of stay was rated too low by the previous researcher panel, compared to clinicians from academic hospitals. There is a large degree of variation in length of stay for VLBW newborns (12). This variation may be explained by a variety of factors not directly associated with quality of care (e.g., financial incentives, transfer bias) (8;13).

Clinicians also reported disagreement over the rating for surfactant administration, which may reflect the current lack of consensus regarding surfactant utilization. Studies conducted during the 1990's firmly established intubation and early surfactant administration as the best therapy for the treatment of respiratory distress syndrome in premature babies (14;15). However, more recently there has been a movement to reduce the negative long-term effects associated with mechanical ventilation, such as chronic lung disease and ventilator-associated pneumonia. Specifically, researchers have evaluated the benefits of avoiding intubation through the early use of nasal continuous positive airway pressure (16–19). A best-practice consensus has yet to be established, potentially explaining the variation in responses across academic and private settings.

The findings from the current study should be viewed within the context of the study design. We collected responses from a small number of clinicians with membership in the AAP. While the majority of the country's neonatologists are members of AAP, our results may not generalize to other samples. In addition, our sample size was relatively small. The survey required time-consuming consideration and commentary that was not suitable for mass distribution. Our final sample included fewer clinicians from the private practice setting than intended. However, the current sample does represent a much more practice-focused group than the initial Delphi panel. Therefore, we think the strong agreement between the clinician and Delphi panels enhances the generalizability of our results. Overall, the Baby-MONITOR has now undergone an extensive formal measure selection process.

## CONCLUSION

Our results indicate a high level of agreement between a panel of researchers and clinical neonatologists concerning the selection of appropriate quality measures for the Baby-MONITOR, enhancing its face validity.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We gratefully acknowledge the contributions and effort by all panelists listed here, alphabetically: Francis J. Bednarek, MD (UMass Memorial Health Care, University of Massachusetts Medical School); Richard E. Bell, MD (Northbay Medical Center); Carl J. Bodenstein, MD (Neonatology Associates Spokane, Sacred Heart Children's Hospital); Robert Boyle, MD (University of Virginia); James A. Cook, MD (Geisinger Health System); Leandro Cordero, MD (The Ohio State University Medical Center); David Fisher, MD (Levine Children's Hospital at Carolinas Medical Center); Ronnie Guillet, MD (University of Rochester Medical Center); Victor Herson, MD (Connecticut Children's Medical Center); Michael J Horgan, MD (Children's Hospital at Albany Medical Center); Priscilla Joe, MD (Children's Hospital & Research Center Oakland); Jonathan NedreLOW, MD (Cook Children's Medical Center); Mary Revenis, MD (Children's National Medical Center); William D. Rhine, MD (Stanford University); Renate Savich, MD (University of New Mexico); Lawrence Skolnick, MD (MidAtlantic Neonatology Associates, Morristown Memorial Hospital); Susan F. Townsend, MD (Memorial Health System); Robert Ursprung, MD (Cook Children's Medical Center). *Note: additional panelists did not release their information to receive acknowledgement.*

**Funding 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). At the time this work was conducted, Dr. Petersen was a recipient of the American Heart Association Established Investigator Award (#0540043N). Drs. Petersen, Profit, and Mr. Kowalkowski also receive support from a Veterans Administration Center Grant (VA HSR&D CoE HFP90-20). The results of this study were presented as a poster at the Pediatric Academic Societies' Annual Meeting in Vancouver, British Columbia on May 4, 2010.

## Abbreviations

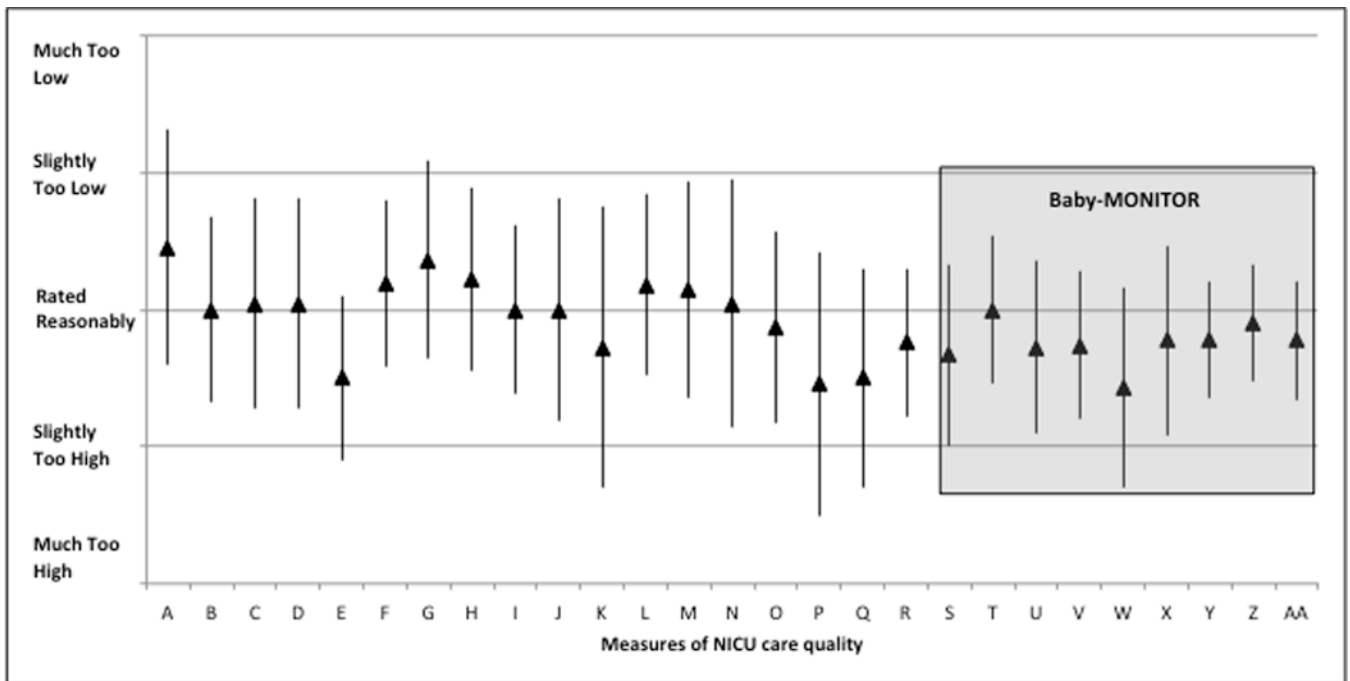
<b>CI</b>	Composite Indicator
<b>CLD</b>	Chronic lung disease
<b>CPQCC</b>	California Perinatal Quality Care Collaborative
<b>GA</b>	gestational age
<b>IH</b>	intracranial hemorrhage
<b>IQR</b>	interquartile range
<b>NEC</b>	necrotizing enterocolitis
<b>NICU</b>	neonatal intensive care unit
<b>ROP</b>	retinopathy of prematurity
<b>VLBW</b>	very low birth weight
<b>VON</b>	Vermont Oxford Network

## Reference List

1. Campbell SM, Roland MO, Middleton E, Reeves D. Improvements in quality of clinical care in English general practice 1998–2003: longitudinal observational study. *BMJ*. 2005; 331:1121. [PubMed: 16257992]
2. Premier Inc. Summary of the composite indicator scoring methodology. Available at: [www.premierinc.com/all/quality/hqi/methodology.jsp](http://www.premierinc.com/all/quality/hqi/methodology.jsp). Accessed on 11.22.2007
3. Roland M. Linking physicians' pay to the quality of care—a major experiment in the United kingdom. *N Engl J Med*. 2004; 351:1448–54. [PubMed: 15459308]
4. Nardo, M.; Saisana, M.; Saltelli, A.; Tarantolo, S.; Hoffman, A.; Giovanini, E. Handbook on Constructing Composite Indicators: Methodology and User Guide. Paris, France: OECD Publishing; 2005.
5. Profit J, Typpo KV, Hysong SJ, Woodard LD, Kallen MA, Petersen LA. 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]

6. Profit J, Gould JB, Zupancic JA, Stark AR, Wall KM, Kowalkowski MA, 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].
7. Committee on Fetus and Newborn. Levels of neonatal care. *Pediatrics*. 2004; 114:1341–7. [PubMed: 15520119]
8. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med*. 2006; 145:265–72. [PubMed: 16908917]
9. Rosenthal MB, Fernandopulle R, Song HR, Landon B. Paying for quality: providers' incentives for quality improvement. *Health Aff*. 2004; 23:127–41.
10. Freudenberg, M. Composite indicators of country performance: a critical assessment. Available at: <http://www.sourceoecd.org/10.1787/405566708255>. Accessed on 2.22.2006
11. Profit J, Gould JB, Draper David, Zupancic JAF, Woodard LD, Pietz K, Petersen LA. Variations in definitions of mortality have little influence on NICU outcome performance ratings. *JAMA*. Under review at.
12. Lemons JA, Bauer CR, Oh W, Korones SB, Papile LA, Stoll BJ, et al. Very low birth weight outcomes of the National Institute of Child Health and Human Development Neonatal Research Network, January 1995 through December 1996. NICHD Neonatal Research Network. *Pediatrics*. 2001; 107:e1. [PubMed: 11134465]
13. Profit J, Zupancic JA, Gould JB, Petersen LA. Implementing pay-for-performance in the neonatal intensive care unit. *Pediatrics*. 2007; 119:975–82. [PubMed: 17473099]
14. Yost CC, Soll RF. Early versus delayed selective surfactant treatment for neonatal respiratory distress syndrome. *Cochrane Database Syst Rev*. 2000:CD001456. [PubMed: 10796266]
15. Soll RF, Morley CJ. Prophylactic versus selective use of surfactant in preventing morbidity and mortality in preterm infants. *Cochrane Database Syst Rev*. 2001:CD000510. [PubMed: 11405966]
16. Morley CJ, Davis PG, Doyle LW, Brion LP, Hascoet JM, Carlin JB, et al. Nasal CPAP or intubation at birth for very preterm infants. *N Engl J Med*. 2008; 358:700–8. [PubMed: 18272893]
17. Rojas MA, Lozano JM, Rojas MX, Laughon M, Bose CL, Rondon MA, et al. Very early surfactant without mandatory ventilation in premature infants treated with early continuous positive airway pressure: a randomized, controlled trial. *Pediatrics*. 2009; 123:137–42. [PubMed: 19117872]
18. Sandri F, Plavka R, Ancora G, Simeoni U, Stranak Z, Martinelli S, et al. Prophylactic or early selective surfactant combined with nCPAP in very preterm infants. *Pediatrics*. 2010; 125:e1402–e1409. [PubMed: 20439601]
19. SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network. Finer NN, Carlo WA, Walsh MC, Rich W, Gantz MG, et al. Early CPAP versus surfactant in extremely preterm infants. *N Engl J Med*. 2010; 362:1970–9. [PubMed: 20472939]





**Figure 1. Deviation among clinicians' responses evaluating ratings of 27 measures of NICU care quality**

Each midpoint (triangle) represents the mean rating from the clinician panel. The tails represent  $\pm 1$  standard deviation. Among the measures selected for the Baby-MONITOR were those with the smallest standard deviation.

**Measures not selected for the Baby-MONITOR:** A. VLBW volume; B. IVH; C. Cystic PVL; D. Assisted ventilation (AV); E.  $O_2$  at 28 days; F.  $O_2$  at discharge; G. AV at discharge; H. Retinopathy of Prematurity Surgery; I. Any NEC; J. NEC surgery; K. Human milk only at discharge; L. Length of stay; M. Retinopathy of Prematurity > stage 2; N. IVH > grade 2; O. Temp 1 hour; P. Surfactant within 2 hours; Q. Steroids for chronic lung disease; R. Mortality < 28 days

**Measures selected for the Baby-MONITOR:** S.  $O_2$  at 36 weeks; T. Any human milk at discharge; U. Hospital mortality; V. Hypothermia at admission; W. Pneumothorax; X. Growth velocity; K. Antenatal steroids; Z. Retinopathy of Prematurity Exam; AA. Late onset sepsis

**Table 1**

## Characteristics of respondents

	N (%)
<b>Practice Setting</b>	
Academic	14 (66.7)
Private	7 (33.3)
<b>Years in Practice</b>	
0–5 yrs	2 (9.5)
6–10 yrs	1 (4.8)
>10 yrs	18 (85.7)
<b>NICU level</b>	
3A	1 (4.8)
3B	8 (38.1)
3C	12 (57.1)
<b>Involved in Quality Improvement</b>	
Yes	20 (95.2)
No	1 (4.8)
<b>Research Participation</b>	
PI/Co-PI	9 (45.0)
Active participation	6 (30.0)
Limited participation	5 (25.0)
<b>Foreign Medical Graduate</b>	
Yes	0 (0.0)
No	21 (100.0)

Table 2

Frequency Distribution of Clinician Responses for Each Measure Evaluated for Inclusion in Composite Indicator of NICU Quality of Care (Baby-MONITOR)

Measure	Much too high		Slightly too high		Reasonable		Slightly too low		Much too low	
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
VLBW Volume	0	2 (9.1)	11 (50.0)	6 (27.3)	3 (13.6)					
Antenatal Steroids	0	5 (21.7)	18 (78.3)	0	0					
Temperature measured within 1hr	1 (4.4)	4 (17.4)	15 (35.2)	3 (13.0)	0					
Hypothermia at admission	1 (4.4)	4 (17.4)	18 (78.3)	0	0					
Surfactant	4 (18.2)	7 (31.8)	8 (36.4)	3 (13.6)	0					
ROP Exam	1 (4.6)	0	21 (95.5)	0	0					
ROP > Grade 2	0	4 (19.0)	11 (52.4)	5 (23.8)	1 (4.8)					
ROP Surgery	0	2 (8.7)	15 (65.2)	5 (21.7)	1 (4.4)					
Any ICH	0	5 (21.7)	13 (56.5)	5 (21.7)	0					
ICH > Grade 2	0	7 (31.8)	8 (36.4)	6 (27.3)	1 (4.5)					
Periventricular leukomalacia	1 (4.4)	3 (13.0)	13 (56.5)	6 (26.1)	0					
Assisted ventilation	0	5 (21.7)	13 (56.5)	4 (17.4)	1 (4.4)					
Pneumothorax	3 (13.0)	7 (30.4)	13 (56.5)	0	0					
Postnatal steroids for CLD	3 (13.6)	6 (27.3)	12 (54.6)	1 (4.6)	0					
O <sub>2</sub> at 28 days	1 (4.6)	9 (40.9)	12 (54.6)	0	0					
O <sub>2</sub> at 36 weeks GA	2 (9.5)	3 (14.3)	16 (76.2)	0	0					
O <sub>2</sub> at discharge	0	2 (9.5)	13 (61.9)	6 (28.6)	0					
Mechanical ventilation at discharge	0	1 (4.6)	14 (63.6)	5 (22.7)	2 (9.1)					
NEC	0	4 (18.2)	14 (63.6)	4 (18.2)	0					
NEC surgery	0	7 (31.8)	8 (36.4)	7 (31.8)	0					
Human milk only	3 (13.6)	5 (22.7)	10 (45.5)	3 (13.6)	1 (4.6)					
Any human milk	0	3 (13.6)	16 (72.7)	3 (13.6)	0					
Growth velocity	1 (4.6)	5 (22.7)	14 (63.6)	2 (9.1)	0					
Healthcare associated infection	0	5 (22.7)	17 (77.3)	0	0					
Length of stay	0	2 (9.1)	15 (68.2)	4 (18.2)	1 (4.6)					
NICU mortality <28 days	1 (4.8)	3 (14.3)	17 (80.9)	0	0					

Measure	Much too high N (%)	Slightly too high N (%)	Reasonable N (%)	Slightly too low N (%)	Much too low N (%)
NICU mortality during hospitalization	1 (4.6)	5 (22.7)	15 (68.2)	1 (4.6)	0

**Table 3**

Clinician Agreement with Research Panel Ratings and Selection of Measures for Inclusion in Composite Indicator of NICU Quality of Care (Baby-MONITOR)

<b>METRIC</b>	<b>Panel Median Rating* (IQR)</b>	<b>Clinician Agreement, % (SD)</b>	<b>Clinician Vote, %</b>
Antenatal steroids	9 (0)	78.3 (0.42)	95
Timely ROP exam	9 (0)	95.5 (0.43)	95
Nosocomial infection	9 (1)	77.3 (0.43)	100
Hypothermia at admission	8 (1)	78.3 (0.54)	95
Pneumothorax	8 (2)	56.5 (0.73)	66
Growth velocity	8 (2)	63.6 (0.69)	82
Oxygen at 36 weeks	7 (2)	76.2 (0.66)	77
Any human milk at discharge	7 (2)	72.7 (0.54)	82
NICU mortality during hospital stay	7 (2)	68.2 (0.63)	77

IQR – Interquartile Range; SD – Standard Deviation

\* Range 1–9, 9 is best. Clinician Agreement – % reporting metric reasonably rated

Clinician Vote – % voting in favor of metric inclusion in composite index (2/3 majority = Include)