Published in final edited form as:

Pediatr Crit Care Med. 2019 January ; 20(1): e15–e22. doi:10.1097/PCC.0000000000001760.

Derivation and validation of an objective effort of breathing score in critically ill children

Steven L. Shein, MD¹, Justin Hotz, BSRT, RRT-NPS², and Robinder G. Khemani, MD, MsCI^{2,3}

¹Rainbow Babies and Children's Hospital, Division of Pediatric Critical Care Medicine, Cleveland, OH

²Children's Hospital of Los Angeles, Department of Anesthesiology and Critical Care Medicine, Los Angeles, CA

³University of Southern California, Keck School of Medicine

Abstract

Objective: To derive and validate a score that correlates with an objective measurement of a child's effort of breathing (EOB).

Design: Secondary analysis of a previously conducted observational study.

Setting: The pediatric and cardiothoracic Intensive Care Units of a quaternary-care children's hospital.

Patients: Patients more than 37 weeks gestational age to age 18 years who were undergoing extubation.

Interventions: EOB was measured in patients following extubation using esophageal manometry to calculate pressure rate product (PRP). Simultaneously, members of a multidisciplinary team (nurse, physician and respiratory therapist) assessed respiratory function using a previously validated tool. Elements of the tool that were significantly associated with PRP in univariate analysis were identified and included in a multivariate model. An EOB score was derived from the results of the model using data from half of the subjects (Derivation cohort), and then validated using data from the remaining subjects (Validation cohort) by calculating the area under the receiver-operator characteristic curve (ROC AUC) for PRP >90th percentile and for the need for reintubation.

Measurements and Main Results: Among 409 subjects, the median age was 5 months and nearly half were cardiac surgery patients (49.1%). Retractions, stridor and pulsus paradoxus were included in the Simple Score. ROC AUC for PRP >90th percentile was 0.8359 (95% CI: 0.7996–0.8722) in the Derivation cohort and 0.7930 (0.7524–0.8337) in the Validation cohort. ROC AUC for reintubation was 0.7280 (0.6807–0.7752) when all scores were analyzed individually, and was

Corresponding Author: Steven L. Shein, MD, Address: 11100 Euclid Avenue, Rainbow Babies and Children's Hospital, 3rd Floor, Cleveland, OH 44106, Telephone: 218-844-3310, steven.shein@uhhospitals.org.

Copyright form disclosure: Mr. Hotz has disclosed that he does not have any potential conflicts of interest

0.7548 (0.6644–0.8452) if scores from three clinicians from different disciplines were summated. Results were similar regardless of provider discipline or training.

Conclusions: A scoring system was derived and validated, performed acceptably to predict increased EOB or need for advanced respiratory support, and may function best when used by a team.

Keywords

pediatric; respiratory; assessment; pressure-rate product; multidisciplinary

Introduction

Respiratory illnesses such as asthma, bronchiolitis and pneumonia are the most common indication for children to require hospitalization in the United States(1). The clinician's assessment of a child's effort of breathing (EOB) is important to guide provision of pharmacologic agents and intensification of respiratory support(2–4). EOB is also a component of several pediatric respiratory scores used for both clinical care and research(5). Objective measurements of EOB have been reported in studies of critically ill children, but may be limited for routine clinical care by invasiveness, cost, requisite technical expertise, and necessary equipment(6–9). Instead, clinicians use assessments of EOB that may be limited by inherent subjectivity (e.g. degree of retractions) and poor specificity (e.g. tachypnea secondary to metabolic acidosis and not primary respiratory pathology)(10–12). It is unknown if clinical assessments correlate with objective measures of EOB in critically ill children.

Pressure-rate product (PRP) is an objective measure of EOB that has been reported in critically ill children(7, 9). It is the mathematical product of a patient's respiratory rate and the change in esophageal pressure during the respiratory cycle(9). Our group and others have shown that PRP is increased in children who subsequently require mechanical ventilation (MV)(13) and decreases with initiation and titration of high-flow nasal cannula flow rate in children with bronchiolitis(7). In a prior study, 409 children prospectively underwent simultaneous measurement of PRP and clinical assessment of EOB by a multidisciplinary team of pediatric intensive care unit (PICU) physicians, nurses and respiratory therapists(9, 14). The clinical assessment of EOB was done with a scoring tool with established inter-rater reliability(14), but limited associations with the need for subsequent therapeutic interventions(13). We hypothesized that these existing data could be used to evaluate for associations between clinical assessments and objective measures of EOB, and to modify the original clinical assessment tool to be more strongly associated with EOB and subsequent therapeutic interventions. The purposes of this secondary analysis of those data were 1) to evaluate if clinical assessments correlate with an objective measure of EOB and 2) to develop a novel weighted score using the clinical assessments to predict EOB and subsequent therapeutic interventions.

Materials and Methods

This is a secondary analysis of prospectively collected data, and detailed methods have been previously reported(9). Briefly, with approval of the Institutional Review Board, patients in the pediatric and cardiothoracic intensive care units at Children's Hospital of Los Angeles undergoing planned extubation between July 2012 and April 2015 were screened.

Extubation readiness was determined by the clinical team, typically based on clinical parameters during a 2-hour spontaneous breathing trial performed on CPAP of 5 cm-H₂O without pressure support. After parental consent, we included children more than 37 weeks gestation and younger than 18 years who did not have a contraindication to placement of an esophageal catheter. Data collected included demographics, the primary reason for MV, respiratory parameters before and after extubation, and re-intubation or non-invasive ventilation (NIV) in the 48 hours following extubation.

Study Protocol:

An age-appropriate esophageal balloon catheter was placed prior to extubation, and position verified by pressure deflections during endotracheal tube occlusion or by chest radiography. At two time points after extubation (5min and 60min), subjects underwent simultaneous measurement of PRP and clinical assessment using a previously described tool(14). Clinical assessments were performed by at least 3 bedside clinicians from different disciplines (physician, nurse and respiratory therapist), at least one of whom had previously undergone a training course on the assessment tool ("core trained"). Other providers were not trained. Elements of the tool were air entry (scored as normal, decreased, severely decreased), level of consciousness (depressed or normal), pulsus paradoxus (assessed qualitatively by inspecting the pulse oximeter waveform [and arterial line waveform, when present] and palpating the brachial or radial pulse; scored as absent or present), retractions (none, mild, moderate or severe), and stridor (none, with agitation, at rest). Decreased air entry and severely decreased air entry were combined into one group. Stridor with agitation was termed "mild" and stridor at rest was termed "severe." In general, the assessors were the bedside clinicians caring for the child, though some assessments were done by study personnel not involved in clinical care. All clinicians and non-clinical assessors were blinded to each other's assessments and the PRP measurement.

Analysis and Outcomes:

Data obtained at both time points were combined and univariate associations between PRP and each tier of each element of the scoring tool (e.g. no stridor, mild stridor or severe stridor) were tested using either Wilcoxon rank-sum (2 tiers) or Kruskal-Wallis with Dunn's test (3 tiers). To derive and validate the EOB score, data obtained at the 5min assessment were used. The subjects were randomly divided 1:1 into a Derivation cohort and a Validation cohort. PRP measurements were transformed (natural log) to improve skewness and kurtosis(15), and a multivariate linear regression model was created with PRP as the dependent variable. Independent variables were included in the model if they were significantly associated with PRP ($p < 0.05$) in the initial univariate analysis. Variables significantly associated with PRP in the multivariate model were then included in the EOB

score, and the coefficients (β) were used to determine the number of “points” each finding contributed to the score.

The predictive ability of the score was assessed by calculating the area under the receiver-operator characteristic curve (ROC AUC). Physiologic outcomes were PRP >50th percentile, PRP >75th percentile and PRP >90th percentile and were assessed for the Derivation and Validation cohorts separately. Data from both groups were combined to measure ROC AUC for the clinical outcomes of reintubation and post-extubation respiratory failure (defined as use of either MV or NIV for 48 hours following extubation). Patients who were on NIV for <48hrs after extubation were treated as not having received NIV. NIV could include high flow nasal cannula, nasal CPAP or intermittent mandatory ventilation, or oro-nasal CPAP or BiPAP.

“Summated scores” were created by adding together the scores simultaneously obtained by one physician, one nurse and one respiratory therapist. Because there were situations when more than 3 clinicians performed the assessment, we created two separate summated scores. The “Summated Score (most experienced)” used scores from the most experienced provider for each discipline that had more than one assessor. A separate “Summated Score (least experienced)” was created using scores from the least experienced provider from each discipline. ROC AUC was calculated for clinical outcomes for each of the Summated Scores. The median PRP values for each Summated Score (most experienced) were compared with Kruskal-Wallis with Dunn’s test. Associations between Summated Score (most experienced) and clinical outcomes were analyzed with chi squared.

Intraclass correlation coefficients (ICC) were calculated using a one-way random effects model to assess agreement of scores measured by different providers. In general ICC less than 0.5 is considered poor, between 0.5 and 0.75 is considered moderate, between 0.75 and 0.9 is considered good, and greater than 0.9 is considered excellent(16). All analyses were performed with SigmaPlot v12.5 (Systat Software, San Jose CA) or Stata v15 (StataCorp, College Station TX; ICC calculation only). Data shown as n (%) or median (IQR) unless otherwise specified and alpha was set at 0.05.

Results

There were 409 children who were included in this analysis. As shown in Table 1, most were male (58.7%), the median age was 5 (1–16) months and nearly half were cardiac surgery patients (49.1%). In 92 cases, the clinical team decided *a priori* to initiate NIV immediately after extubation. Reintubation was performed in 34 (8.3%) children and an additional 38 (10.3%) children required NIV for 48 hours after extubation, including 19 subjects who had been extubated directly to NIV. A total of 1614 measurements using the original assessment tool were performed at 5min after extubation and 1517 assessments were performed at 60min after extubation.

As shown in Figures 1a-1c, PRP was significantly associated with decreased air entry, pulsus paradoxus, retractions and stridor (also see Supplemental Table 1). Results were similar regardless of Core training or professional discipline (see Supplemental Tables 2 and 3).

Those four elements of the clinical assessment were included in the multivariate linear regression model for subjects randomly selected to be in the Derivation cohort. In that model, pulsus paradoxus ($\beta = 0.191$, $p = 0.001$), mild stridor ($\beta = 0.215$, $p = 0.002$), severe stridor ($\beta = 0.446$, $p < 0.001$), mild retractions ($\beta = 0.332$, $p < 0.001$), moderate retractions ($\beta = 0.773$, $p < 0.001$) and severe retractions ($\beta = 1.218$, $p < 0.001$) were significantly associated with PRP, but decreased air entry ($\beta = 0.096$, $p = 0.053$) was not.

Three scoring systems were created using the variables that were significantly associated with PRP in the linear regression model (Table 2). The “Raw Score” included all variables that were significant in the model and the points allotted for each variable were the actual β coefficient values calculated in the model. A “Simple Score” was created using the same variables, but the points were allotted in whole numbers chosen to preserve the approximate relative weight of each variable. A “Retractions Only Score” used the same points system as the Simple Score, but severity of retractions was the only variable included. For both the Derivation cohort and the Validation cohort, the ROC AUC for all three scores for elevated PRP ranged from ≈ 0.7 for PRP $> 50^{\text{th}}$ percentile to ≈ 0.8 for PRP $> 90^{\text{th}}$ percentile (Table 2).

When the 1614 assessments performed by an individual clinician at the 5min time-point were analyzed separately, the ROC AUC for reintubation was ≈ 0.72 for all three scores (Table 3). The ROC AUC for post-extubation respiratory failure (re-intubation or NIV for 48 hours) was modestly lower for each score. For the 403 children who were assessed by at least one provider from each of the three disciplines at the 5min time-point, the ROC AUC for both outcomes were ≈ 0.76 for both of the Summated Scores (Table 3). Higher values of the Summated Score (most experienced) were associated with increased PRP (Figures 2a and 2b) and more frequent usage of additional respiratory support (Figures 2c and 2d).

The agreement of each of the three scoring systems between providers is shown in Supplemental Table 4. For all three scoring systems, ICC-Individual was ≈ 0.55 and ICC-Average was ≈ 0.80 ($p < 0.001$ between the types of ICC for all three scoring systems). ROC AUC values were also calculated for each type of provider for both the Simple Score and the Retractions Only score and no consistent differences were observed between provider groups (see Table 4 and Supplemental Table 5).

Discussion

In this secondary analysis of prospectively collected data in critically ill children, we found that critical care clinicians’ assessments of decreased air entry, pulsus paradoxus, retractions and stridor were significantly associated with invasively measured EOB. Using separate subgroups, scoring systems to predict EOB were both derived and validated, and found to have fair to good accuracy(17) for predicting increased EOB and the need for subsequent intensification of respiratory support. While we found variability in any individual provider’s assessments of the score against other individuals (ICC ≈ 0.55), the scoring systems appear to have good validity when used by a team of three clinicians. In total, these findings support that clinicians’ assessments are good surrogates for objective measures of EOB, especially when assessments are team-based.

We aimed to evaluate if clinical assessments correlate with an objective measure of EOB. Clinician assessments of EOB have been previously shown to correlate with reduced forced expiratory volume in 1 second (FEV₁) and oxygen saturation in children with asthma(18, 19). Clinician-assessed EOB is also a component of pediatric asthma severity scores that have been shown to correlate with FEV₁, hypoxia and hypercarbia(20, 21). In adult patients, clinician assessment of EOB strongly correlates with patients' self-report of dyspnea, and also with hypoxia and the need for subsequent MV(11, 12, 22). However, we are unaware of prior reports that show that clinical assessments of critically ill children correlate with increased objective measures of EOB. Our group has previously shown that an increased objective measure of EOB, specifically a PRP greater than 500, is associated with an increased need for reintubation(13). Coincidentally, the 75th percentile of PRP in this analysis was 500, and we found that all three derived scoring systems performed acceptably (ROC AUC: 0.71–0.78) for predicting this PRP(23).

Assessments in this study were done by a multidisciplinary team with providers from three different professions. Our data do not support that a clinician's discipline or prior training with the assessment tool systematically influences ability to assess EOB or predict the need for subsequent respiratory support. We found that, in aggregate, clinical assessments are associated with PRP regardless of discipline or prior training, and no provider type had consistently better predictive ability based on the calculated ROC AUC values. For example, nurses had the highest predictive ability for PRP >50th percentile, but the lowest for PRP >90th percentile. It is not certain if findings would be similar if non-ICU clinicians had been included, and our findings may be best applied by clinicians with experience assessing critically ill children. The lack of an observed effect of our training suggests that our training methods were ineffective or that clinicians have sufficient experience in respiratory assessments to render additional training superfluous.

Our data support a team-based approach to assessing EOB in a critically ill child as team-based scores had superior predictive ability and reliability. The ROC AUC values for subsequent respiratory support were consistently higher with the Summated Scores compared to the individual scores, regardless of whether the most experienced or least experienced providers were considered. Based on the calculated ICC values, the reliability of the scoring systems was moderate amongst individual providers. This implies that if the score is evaluated by two providers, agreement between those two providers on the actual score was only moderate. However, when looking at all three providers and the average score amongst those providers, the validity of the scoring system was generally in the good range(16). These findings support that simultaneous evaluation by multiple providers may provide maximal ability to assess EOB and predict need for further respiratory support.

Summated Simple Scores totaling between 1 and 7 had generally similar findings in regards to median PRP and the need for subsequent MV or NIV. Scores greater than 7, and especially scores greater than 11, were indicative of substantially higher PRP and frequent use of advanced respiratory support. Scores of those magnitudes require at least one clinician to give a maximum score of 8 points or, more likely, at least two clinicians to assess the child to have moderate-severe symptoms. Similarly, a Retractions Only Score greater than 5 was associated with increased PRP and frequent use of NIV or MV, and requires

either one maximum score or two moderate scores. Simply, if any two clinicians out of a team of three assess a child to have at least moderate respiratory pathology – or if one clinician believes the child has maximally severe symptoms – that child would be likely to have elevated PRP and may be at increased risk of needing higher levels of respiratory support.

Strengths of this study include that the data were collected prospectively; assessors were blinded to each other and to PRP measurement; a large cohort with a wide range of measured PRP, including a substantial fraction (~25%) with a PRP suggesting severe respiratory distress and increased risk of intubation; a similar re-intubation rate compared to multicenter studies(24, 25); and the similar findings regardless of cohort (Derivation vs. Validation) or the discipline, experience or core training of the clinician. We also developed three scores, each with relative strengths. Though the Raw Score had the best accuracy for predicting both increased PRP and subsequent respiratory failure, the Simple Score may be the most appropriate for clinical use given the relative ease of using whole numbers and the similar ROC AUC values. Performance of the Retractions Only Score was less strong, though this score may be considered in situations where patients are not at risk of having upper airway obstruction or when pulsus paradoxus is difficult to assess due to lack of expertise or invasive blood pressure monitoring.

There are also several limitations. First, all data were collected from a single center and exclusively in an ICU, and clinicians in other settings may differ in their clinical assessment skills, especially in detection of pulsus paradoxus. Second, this is a secondary analysis, and other variables that may have predictive value in an EOB score were not collected in the original study, including heart rate, blood pressure, agitation and nasal flaring. Third, agreement between providers at this institution for assessing the variables in the tool is generally fair to moderate(14, 26, 27). Prior studies have shown pediatric clinicians have moderate(28, 29) to substantial(30) agreement in the assessment of retractions(27). External validation of this scoring system and further assessment of inter-rater agreement is needed, especially in non-ICU patients, older children, non-cardiac surgery patients, and those with new-onset respiratory illness. Fourth, the association between an assessment-based score and subsequent MV or NIV may have been self-fulfilling, and the fact that many of the assessors were also clinicians involved in the decision to initiate advanced respiratory support may have biased our results and may hinder interpretation of that observed association. However, the assessors were blinded to PRP readings, so the associations between our scores and elevated PRP were not at similar risk of bias and support the validity of our scoring systems. Nevertheless, a follow-up study that includes appropriate blinding is needed to confirm our findings. Fifth, like all observational studies, it is possible that study methods like measuring PRP and using the scoring tool influenced behavior (i.e. the Hawthorne effect), though it is not possible to know if such an influence would have been toward more aggressive therapies or towards more conservative ones. Sixth, we used a conservative definition for NIV that required >48hrs of usage. This was chosen in part because some providers decide *a priori* to use NIV immediately following extubation based on perceived risks, often for short courses, therefore limiting its utility as a marker of severe respiratory insufficiency unless a conservative definition is used. Finally, we used a systematic but qualitative definition for pulsus paradoxus, which may have limited the reliability of this finding.

Conclusion

In conclusion, in this secondary analysis of prospective data, we showed that clinical assessments are associated with objectively measured EOB, supporting the use of clinical assessments of EOB for both clinical and research purposes. The scoring system that was derived and validated performed acceptably to predict increased EOB or need for advanced respiratory support, and may function best with a team of clinicians regardless of providers' experience level. Further work is needed to assess non-ICU clinicians, confirm our findings at other centers, and evaluate additional factors that may strengthen the performance of the EOB score.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding source: Supported by National Institutes of Health/Eunice Kennedy Shriver National Institute of Child Health and Human Development grant 1K23HL103785 (RGK) and the Los Angeles Basin Clinical Translational Science Institute.

Dr. Shein received funding from Accelerate Diagnostic. Drs. Shein and Khemani received support for article research from the National Institutes of Health (NIH). Dr. Khemani's institution received funding from the NIH.

Abbreviations

BiPAP	bi-level positive airway pressure
CPAP	continuous positive airway pressure
EOB	effort of breathing
ICC	intraclass correlation coefficients
MV	mechanical ventilation
NIV	non-invasive ventilation
PRP	pressure-rate product
ROC AUC	area under the receiver-operator characteristic curve

References

1. Witt WP, Weiss AJ, Elixhauser A. Overview of Hospital Stays for Children in the United States, 2012: Statistical Brief #187 In: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville (MD); 2006.
2. Wood PR, Hill VL. Practical management of asthma. *Pediatr Rev* 2009;30(10):375–385; quiz 385. [PubMed: 19797480]
3. Malhotra A, Krilov LR. Viral croup. *Pediatr Rev* 2001;22(1):5–12. [PubMed: 11139641]
4. Gupta P, Rettiganti M, Gossett JM, et al. Risk factors for mechanical ventilation and reintubation after pediatric heart surgery. *J Thorac Cardiovasc Surg* 2016;151(2):451–458 e453. [PubMed: 26507405]

5. Bekhof J, Reimink R, Brand PL. Systematic review: insufficient validation of clinical scores for the assessment of acute dyspnoea in wheezing children. *Paediatr Respir Rev* 2014;15(1):98–112. [PubMed: 24120749]
6. Essouri S, Durand P, Chevret L, et al. Optimal level of nasal continuous positive airway pressure in severe viral bronchiolitis. *Intensive Care Med* 2011;37(12):2002–2007. [PubMed: 21993811]
7. Pham TM, O'Malley L, Mayfield S, et al. The effect of high flow nasal cannula therapy on the work of breathing in infants with bronchiolitis. *Pediatr Pulmonol* 2015;50(7):713–720. [PubMed: 24846750]
8. Milesi C, Baleine J, Matecki S, et al. Is treatment with a high flow nasal cannula effective in acute viral bronchiolitis? A physiologic study. *Intensive Care Med* 2013;39(6):1088–1094. [PubMed: 23494016]
9. Khemani RG, Hotz J, Morzov R, et al. Evaluating Risk Factors for Pediatric Post-extubation Upper Airway Obstruction Using a Physiology-based Tool. *Am J Respir Crit Care Med* 2016;193(2):198–209. [PubMed: 26389847]
10. Benbassat J, Bauml R. Narrative review: should teaching of the respiratory physical examination be restricted only to signs with proven reliability and validity? *J Gen Intern Med* 2010;25(8):865–872. [PubMed: 20349154]
11. Tulaimat A, Gueret RM, Wisniewski MF, et al. Association between rating of respiratory distress and vital signs, severity of illness, intubation, and mortality in acutely ill subjects. *Respir Care* 2014;59(9):1338–1344. [PubMed: 24847098]
12. Tulaimat A, Patel A, Wisniewski M, et al. The validity and reliability of the clinical assessment of increased work of breathing in acutely ill patients. *J Crit Care* 2016;34:111–115. [PubMed: 27288621]
13. Khemani RG, Sekayan T, Hotz J, et al. Risk Factors for Pediatric Extubation Failure: The Importance of Respiratory Muscle Strength. *Crit Care Med* 2017;45(8):e798–e805. [PubMed: 28437378]
14. Khemani RG, Schneider JB, Morzov R, et al. Pediatric upper airway obstruction: interobserver variability is the road to perdition. *J Crit Care* 2013;28(4):490–497. [PubMed: 23337481]
15. Paul SR, Zhang X. Testing for normality in linear regression models. *Journal of Statistical Computation and Simulation* 2010;80(10):1101–1113.
16. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016;15(2):155–163. [PubMed: 27330520]
17. Safari S, Baratloo A, Elfil M, et al. Evidence Based Emergency Medicine; Part 5 Receiver Operating Curve and Area under the Curve. *Emerg (Tehran)* 2016;4(2):111–113. [PubMed: 27274525]
18. Kerem E, Canny G, Tibshirani R, et al. Clinical-physiologic correlations in acute asthma of childhood. *Pediatrics* 1991;87(4):481–486. [PubMed: 2011424]
19. Yung M, South M, Byrt T. Evaluation of an asthma severity score. *J Paediatr Child Health* 1996;32(3):261–264. [PubMed: 8827548]
20. Wood DW, Downes JJ, Lecks HI. A clinical scoring system for the diagnosis of respiratory failure. Preliminary report on childhood status asthmaticus. *Am J Dis Child* 1972;123(3):227–228. [PubMed: 5026202]
21. Becker AB, Nelson NA, Simons FE. The pulmonary index. Assessment of a clinical score for asthma. *Am J Dis Child* 1984;138(6):574–576. [PubMed: 6720644]
22. Campbell ML. Psychometric testing of a respiratory distress observation scale. *J Palliat Med* 2008;11(1):44–50. [PubMed: 18370892]
23. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol* 2010;5(9):1315–1316. [PubMed: 20736804]
24. Kurachek SC, Newth CJ, Quasney MW, et al. Extubation failure in pediatric intensive care: a multiple-center study of risk factors and outcomes. *Crit Care Med* 2003;31(11):2657–2664. [PubMed: 14605539]
25. Curley MA, Wypij D, Watson RS, et al. Protocolized sedation vs usual care in pediatric patients mechanically ventilated for acute respiratory failure: a randomized clinical trial. *JAMA* 2015;313(4):379–389. [PubMed: 25602358]

26. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37(5):360–363. [PubMed: 15883903]
27. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22(3):276–282. [PubMed: 23092060]
28. Angelilli ML, Thomas R. Inter-rater evaluation of a clinical scoring system in children with asthma. *Ann Allergy Asthma Immunol* 2002;88(2):209–214. [PubMed: 11868927]
29. Biondi EA, Gottfried JA, Dutko Fioravanti I, et al. Interobserver reliability of attending physicians and bedside nurses when using an inpatient paediatric respiratory score. *J Clin Nurs* 2015;24(9–10):1320–1326. [PubMed: 25420627]
30. Gajdos V, Beydon N, Bommenel L, et al. Inter-observer agreement between physicians, nurses, and respiratory therapists for respiratory clinical evaluation in bronchiolitis. *Pediatr Pulmonol* 2009;44(8):754–762. [PubMed: 19598273]

Figure 1:

PRP was significantly associated with retractions (A), air entry (B), pulsus paradoxus (B), and stridor (C). For retractions and stridor, each tier was significant vs. all other tiers in pairwise comparisons ($p < 0.05$ with Dunn's test). Lines represent median, bars represent IQR, whiskers represent 10th/90th percentiles, and circles represent 5th/95th percentiles.

*, $p < 0.001$ vs. none/normal tier

Figure 2:

PRP was significantly associated with higher Summated Score for both the Simple Score (A) and Retractions Only Score (B). Lines represent median, bars represent IQR, whiskers represent 10th/90th percentiles, and circles represent 5th/95th percentiles.

Reintubation and post-extubation respiratory failure (reintubation or NIV) were associated with Summated Scores for both the Simple Score (C) and Retractions Only Score (D), both $p < 0.001$ by Chi squared.

*, $p < 0.05$ vs. 0 points

^, $p < 0.05$ vs. 0-1 points

v, $p < 0.05$ vs. 0-2 points

#, $p < 0.05$ vs. 0-5 points

Table 1 -**Demographics and Outcomes**

Variable	n (%) or median (IQR)
Male	240 (58.7%)
Female	169 (41.3%)
Age (months)	5 (1–16)
Primary reason for mechanical ventilation	
- Cardiac surgery	201 (49.1%)
- Respiratory failure	96 (23.5%)
- Shock/cardiovascular failure	39 (9.5%)
- Neurologic failure	38 (9.3%)
- Other	35 (8.6%)
Reintubated	34 (8.3%)
Post-extubation respiratory failure (MV or NIV)	76 (18.6%)

MV, mechanical ventilation

NIV, non-invasive ventilation

Table 2 -

Scoring systems: points allotted and predictive ability for physiologic outcomes

	Raw Score	Simple Score	Retractions Only Score
Variable	Points Alloted		
Pulsus Paradoxus	0.191	1	0
Retractions - Mild	0.332	1	1
Retractions - Moderate	0.773	3	3
Retractions - Severe	1.218	5	5
Stridor - Mild	0.215	1	0
Stridor - Severe	0.446	2	0
Outcome	ROC AUC		
<u>Derivation Cohort</u>			
PRP >50th percentile	0.6834 (0.6569–0.7098)	0.6790 (0.6525–0.7056)	0.6665 (0.6396–0.6933)
>75th	0.7821 (0.7536–0.8105)	0.7783 (0.7497–0.8069)	0.7562 (0.7272–0.7853)
>90th	0.8384 (0.8023–0.8745)	0.8359 (0.7996–0.8722)	0.7896 (0.7516–0.8275)
<u>Validation Cohort</u>			
PRP >50th percentile	0.7156 (0.6904–0.7409)	0.7098 (0.6844–0.7351)	0.6822 (0.6559–0.7084)
>75th	0.7403 (0.7121–0.7685)	0.7366 (0.7082–0.7650)	0.7052 (0.6764–0.7340)
>90th	0.7954 (0.7548–0.8361)	0.7930 (0.7524–0.8337)	0.7329 (0.6893–0.7766)

ROC AUC shown with 95% confidence interval in parentheses

PRP, pressure rate product

ROC AUC, area under the receiver-operator characteristic curve

Table 3 -

Scoring systems: predictive ability for clinical outcomes

Score/outcome	Raw Score	Simple Score ROC AUC	Retractions Only Score
Individual scores			
Reintubation	0.7293 (0.6820–0.7767)	0.7280 (0.6807–0.7752)	0.7152 (0.6674–0.7630)
Post-extubation respiratory failure	0.7001 (0.6668–0.7333)	0.6989 (0.6657–0.7321)	0.6883 (0.6547–0.7219)
Summated scores			
Reintubation			
Most experienced	0.7564 (0.6638–0.8490)	0.7548 (0.6644–0.8452)	0.7458 (0.6511–0.8406)
Least experienced	0.7704 (0.6748–0.8661)	0.7689 (0.6760–0.8619)	0.7549 (0.6566–0.8532)
Post-extubation respiratory failure			
Most experienced	0.7547 (0.6933–0.8162)	0.7537 (0.6928–0.8146)	0.7488 (0.6860–0.8116)
Least experienced	0.7633 (0.7020–0.8246)	0.7638 (0.7033–0.8243)	0.7555 (0.6939–0.8171)

ROC AUC shown with 95% confidence interval in parentheses

NIV, non-invasive ventilation

ROC AUC, area under the receiver-operator characteristic curve

Table 4

Outcome	Attendings	Fellows	Nurses	Respiratory Therapists	Core	Not Core
PRP >50th	0.7014 (0.6458–0.7569)	0.6769 (0.6384–0.7155)	0.7041 (0.6746–0.7337)	0.6997 (0.6702–0.7292)	0.7029 (0.6728–0.7331)	0.6910 (0.6697–0.7124)
>75th	0.7816 (0.7236–0.8395)	0.7702 (0.7293–0.8110)	0.7596 (0.7265–0.7926)	0.7367 (0.7017–0.7718)	0.7671 (0.7331–0.8010)	0.7518 (0.7281–0.7755)
>90th	0.8586 (0.7890–0.9282)	0.8081 (0.7459–0.8702)	0.8055 (0.7620–0.8491)	0.8059 (0.7569–0.8548)	0.8208 (0.7723–0.8694)	0.8104 (0.7784–0.8424)
Reintubation	0.7262 (0.5713–0.8811)	0.7529 (0.6527–0.8531)	0.7425 (0.6677–0.8173)	0.6965 (0.6137–0.7794)	0.7471 (0.6652–0.8291)	0.7187 (0.6628–0.7746)
PERF	0.7618 (0.6655–0.8580)	0.6892 (0.6171–0.7613)	0.6857 (0.6279–0.7435)	0.6904 (0.6371–0.7436)	0.7319 (0.6758–0.7880)	0.6834 (0.6441–0.7228)

Data for residents not shown

ROC AUC shown with 95% confidence interval in parentheses

PERF, post-extubation respiratory failure

PRP, pressure rate product