


Interobserver Agreement on Clinical Judgment of Work of Breathing in Spontaneously Breathing Children in the Pediatric Intensive Care Unit

Marcel G. de Groot¹ Marjorie de Neef¹ Marieke H. Otten¹  Job B. M. van Woensel¹
Reinout A. Bem¹

¹ Pediatric Intensive Care Unit, Emma Children's Hospital, Amsterdam University Medical Centers, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

Address for correspondence Reinout A. Bem, MD, PhD, Pediatric Intensive Care Unit, Emma Children's Hospital, Amsterdam University Medical Centers, Location AMC, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands (e-mail: r.a.bem@amsterdamumc.nl).

J Pediatr Intensive Care 2020;9:34–39.

Abstract

Clinical assessment of the work of breathing (WOB) remains a cornerstone in respiratory support decision-making in the pediatric intensive care unit (PICU). In this study, we determined the interobserver agreement of 30 observers (PICU physicians and nurses) on WOB and multiple signs of effort of breathing in 10 spontaneously breathing children admitted to the PICU. By reliability analysis, the agreement on overall WOB was poor to moderate, and only three separate signs of effort of breathing (breathing rate, stridor, and grunting) showed moderate-to-good interobserver reliability. We conclude that the interobserver agreement on the clinical WOB judgment among PICU physicians and nurses is low.

Keywords

- ▶ respiratory distress
- ▶ children
- ▶ critical care

Introduction

Worldwide, severe respiratory illness is among the most common reasons for children to be admitted to a pediatric intensive care unit (PICU).^{1–3} Respiratory support for children in the PICU currently includes an increasing variety of non-invasive and invasive modalities. In the day-to-day choice among these respiratory support modalities during escalating and deescalating critical care, PICU clinicians use measures of gas exchange, such as blood gas analysis and pulse oximetry, cardiovascular monitoring, and assessment of the work of breathing (WOB).

WOB is defined by the energy expenditure during the entire breathing cycle, and is expressed as work per unit volume or as a work rate (power). Objective assessment of the WOB can be performed by pleural pressure measurements (e.g., by an esophageal catheter) with calculation of the WOB from the Campbell volume-pressure diagram or the pressure-rate/-time product.^{4,5} However, this invasive, complex, and laborious technique is not readily available at the

bedside. As such, subjective clinical judgment of the WOB by critical care professionals remains a cornerstone in the treatment decision of respiratory support in the PICU.

Clinical WOB scores have been constructed and validated for several pediatric respiratory diseases, such as asthma, upper airway disease, and bronchiolitis.^{5–9} However, many of these scores do not solely incorporate pure clinical signs of the effort of breathing. In addition, they have been developed for specific diseases and thus may not apply to many PICU patients, spanning a wide age and in disease spectrum. Clearly, a generalizable clinical WOB score for children in the PICU may prove a very helpful instrument in respiratory critical care decision making.

An important challenge in developing clinical WOB scores is to minimize interobserver variability. A score with relatively high interobserver variability may, even when validated, still prove to be of limited use in daily practice. Further insight into how similar or different critical care physicians and nurses judge the effort of breathing in pediatric patients

received

May 16, 2019

accepted after revision

August 21, 2019

published online

October 7, 2019

Copyright © 2020 by Georg Thieme
Verlag KG, Stuttgart · New York

DOI <https://doi.org/10.1055/s-0039-1697679>.
ISSN 2146-4618.

may contribute to the process of developing a PICU-specific clinical WOB score. The primary goal of this study was to determine the interobserver agreement on the clinical judgment of the WOB and its separate, specific signs of effort of breathing in spontaneously breathing children admitted to the PICU.

Materials and Methods

We designed a two-center (Amsterdam UMC, location AMC, and VUmc) tertiary PICU study in which multiple observers were asked to rate the overall WOB and multiple signs of effort of breathing by watching patient movies of spontaneously breathing critically ill children. Both PICU nurses and physicians were asked to participate as an observer. Upon acceptance to participate, they received a short instruction and scoring form together with the patient movies. The observers did not receive any special training or learning module for clinical WOB judgment prior to the study.

Patient Movies

After written consent of the parents, videotaping of 10 randomly selected critically ill, spontaneously breathing children admitted to our PICU was performed. The movies were shot with focus on the visibility of the specific signs of effort of breathing, which in some cases necessitated removal of clothing from the thorax. The movies were processed so that facial characteristics (e.g., eyes) were made invisible. A waiver of the local medical ethical committee (Amsterdam UMC, location AMC) was obtained.

Overall WOB and Signs of Effort of Breathing

Observers were requested to rate the overall WOB on a 4-point scale. In addition to this assessment of the overall WOB, the scoring form contained multiple ordinal/binary items representing signs of effort of breathing (► **Table 1**). These signs of effort of breathing were selected based on a systematic search of the pediatric WOB literature (see ► **Supplementary Material**, available in the online version). From this systematic search of signs of effort of breathing in the pediatric literature, we selected 12 items categorized in four WOB domains (breathing rate, inspiratory effort, expiratory effort, and general signs of effort of breathing) after a consensus meeting by a local panel of experts, consisting of one PICU physician, one research nurse/clinical epidemiologist, and two PICU nurses with specific respiratory expertise.

Primary Outcome

Interobserver agreement of the clinical judgment of the overall WOB and separate signs of effort of breathing.

Statistical Analysis

The primary outcome was determined by reliability analysis calculating the intraclass correlation coefficient (ICC) for each item,¹⁰ which incorporates both observer and subject variability. A two-way random ICC model was used. Because, ultimately, we are interested in the use of clinical WOB scoring in the daily practice, thus in the context of a single observer for a single patient over time, we used the most stringent approach of calculating the single measures ICC for absolute agreement. As a secondary outcome, average measure ICCs are also reported. Values for ICC less than 0.4

Table 1 Scoring system of clinical judgment of the overall WOB and signs of effort of breathing

Points		1	2	3	4
Overall WOB		Normal	Mild	Moderate	Severe
Signs of effort of breathing Domains					
Rate	Breathing rate (compared with normal for age ^a)	Normal	>20%	20–50%	>50%
Inspiratory effort	Inspiration time	Normal	–	Abnormal	–
	Retractions ^b	Absent	Mild	Moderate	Severe
	Stridor	Absent	Mild	Moderate	Severe
	Nasal flaring	Absent	–	Moderate	Severe
	Head bobbing	Absent	–	–	Present
Expiratory effort	Expiration time	Normal	–	Abnormal	–
	Active use of abdominal muscles	Absent	–	Moderate	Severe
	Grunting	Absent	–	Moderate	Severe
	Wheeze/rales (audible without stethoscope)	Absent	–	–	Present
General effort	Limited awareness/feeding/communication/activity	No	Mild	Moderate	Severe
	Abnormal/fixed posture	No	Mild	Moderate	Severe

Abbreviation: WOB, work of breathing.

^aNormal breathing rate predefined and available for the observers: 30–60/min for age < 1 year; 24–60/min for age 1–3 years; 22–34/min for age 3–5 years; 18–30/min for age 5–12 years (adapted from Qureshi et al¹⁶ and Fleming et al¹⁷).

^bRetractions at four locations: suprasternal, supraclavicular, intercostal, and subcostal/substernal. Mild: one location; moderate: two locations; severe: at more than locations.

Table 2 Patient and observer characteristics

Patients, n = 10 (%)		
Primary diagnosis	Upper airway disease	1 (10)
	Bronchiolitis	4 (40)
	Pneumothorax	1 (10)
	Failure to thrive with metabolic derangement	1 (10)
	Diaphragm paralysis	1 (10)
	Pneumonia	2 (20)
Respiratory/airway support	None	3 (30)
	High flow nasal cannula	5 (50)
	Tracheostomy	1 (10)
	Nasopharyngeal tube	1 (10)
Observers, n = 30 (%)		
PICU professional	Physician	6 (20)
	Nurse	24 (80)
Years of PICU experience	≤ 5 y	12 (40)
	> 5 y	18 (60)

Abbreviation: PICU, pediatric intensive care unit.

indicate poor agreement, 0.4 to 0.75 indicate moderate agreement, and values greater than 0.75 indicate good agreement between observers. For items with missing values, we excluded those observers who did not complete

the full assessment of the 10 patients for that particular item. With a prespecified value of α with 0.05 and power of at least 0.8, we determined a minimal sample size of 10 observations per patient ($n = 10$) to detect the smallest possible value of 0.2 for ICC, when initially assumed there is no agreement.¹¹ All analyses were performed using SPSS (version 24, IBM SPSS Statistics, Chicago, Illinois, United States).

Results

Patient and Observer Characteristics

The patient cohort consisted of young children (age below 5 years), with seven (70%) being infants. Primary underlying conditions and type of respiratory support of the patients are shown in ►Table 2. Of the total 110 invited PICU professionals from the two centers (20 physicians and 90 nurses), 30 observers responded (response rate 27.3%). Observer characteristics are shown in ►Table 2.

Interobserver Agreement

There was considerable variability in the clinical judgment of all items for the 10 patients, except for the item head bobbing (►Fig. 1). In addition, examples of the overall WOB rating for two patients are shown in ►Fig. 2. Together, this reflects patient and/or observer variability, which is a prerequisite for performing the reliability analysis to calculate the ICC. While the pure interobserver agreement for the item head bobbing was high, variability was too low to calculate the ICC for this item.

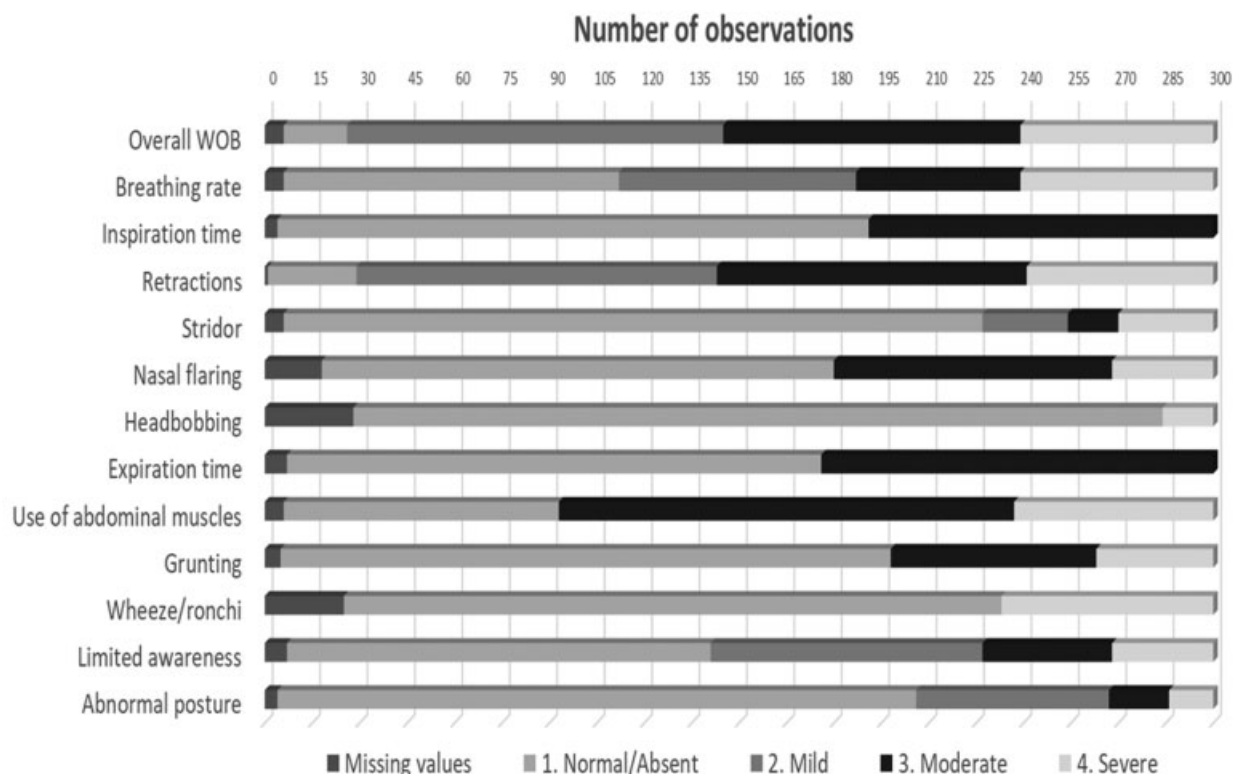


Fig. 1 Variability in the clinical judgment of work of breathing (WOB) per item scored. The total number and judgment of observations of the 13 items assessed by the observers ($n = 30$) in the pediatric intensive care unit patients ($n = 10$).

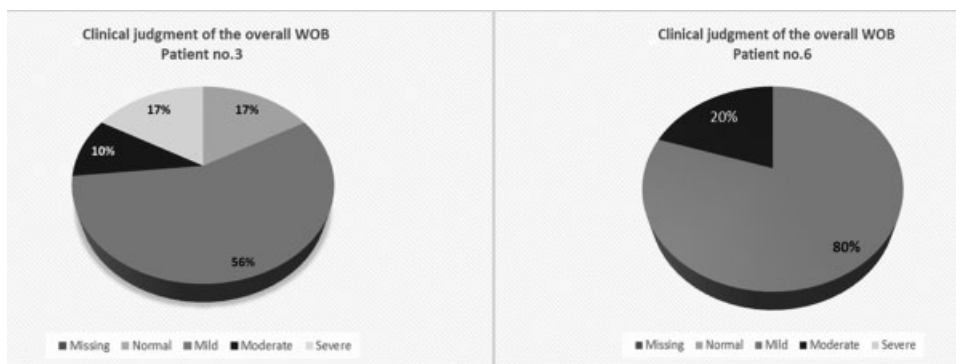


Fig. 2 Variability in the clinical judgment of work of breathing (WOB) per patient (examples of two patients). Percentages of observations scoring the overall WOB in two pediatric intensive care unit patients. Note the relatively low interobserver agreement in patient no. 3 as compared with patient no. 6.

The calculated single-measure ICCs, our primary outcome, for rating of the overall WOB and separate effort of breathing items are shown in ►Table 3. The ICC (95% confidence interval) of rating the overall WOB was 0.482 (0.291–0.762), reflecting poor to moderate interobserver agreement. There was no substantial change in this interobserver agreement when calculating the ICC for the overall WOB scored by PICU physicians or nurses, or by observers with limited or extensive experience in the PICU: the ICC (95% confidence interval) was 0.347 (0.122–0.684) for PICU physicians and 0.519 (0.319–0.789) for PICU nurses, and 0.423 (0.230–0.723) for observers with limited (≤ 5 years) experience and 0.550 (0.332–0.813) for observers with extensive (> 5 years) experience.

There was moderate to good agreement (lower bound 95% confidence interval above 0.4) for only three items (breathing rate, stridor, and grunting). In contrast, the average measure ICCs were much higher for all items tested (see ►Table 3).

Discussion

In this study, we aimed to determine the interobserver agreement on the clinical judgment of WOB in spontaneously breathing children admitted to the PICU. The main finding of this study is that the interobserver agreement among PICU clinicians on rating the overall WOB is poor to moderate. Only three signs of effort of breathing (breathing rate, stridor, and grunting) show moderate to good agreement.

In the PICU, a clinical WOB score used by both physicians and nurses may prove a very helpful instrument in respiratory support decision making. The ideal clinical WOB score is a simple and relatively short list of signs of effort of breathing, performing with high absolute interobserver agreement and good discrimination between patients with varying respiratory distress. It should correlate with objective measurements of WOB and, evidently, should be validated in a cohort of critically ill children for relevant patient outcomes, such as need for

Table 3 Intraclass correlation coefficients (ICC) for absolute agreement

	Number of observers	ICC single measures (95% confidence interval)	ICC average measures (95% confidence interval)
Overall WOB	27	0.482 (0.291–0.762)	0.962 (0.917–0.989)
Signs of effort of breathing			
Breathing rate	27	0.810 (0.662–0.935)	0.991 (0.981–0.997)
Inspiration time	26	0.137 (0.052–0.379)	0.804 (0.590–0.941)
Retractions	29	0.276 (0.139–0.574)	0.917 (0.824–0.975)
Stridor	24	0.733 (0.554–0.903)	0.985 (0.968–0.996)
Nasal flaring	19	0.374 (0.198–0.680)	0.919 (0.824–0.976)
Expiration time	25	0.243 (0.114–0.537)	0.889 (0.763–0.967)
Active use of abdominal muscles	25	0.252 (0.121–0.547)	0.894 (0.774–0.968)
Grunting	28	0.672 (0.482–0.875)	0.983 (0.963–0.995)
Wheeze/rales	20	0.453 (0.262–0.744)	0.943 (0.877–0.983)
Limited awareness/feeding/communication/activity	25	0.267 (0.130–0.565)	0.901 (0.790–0.970)
Abnormal/fixed posture	28	0.157 (0.067–0.407)	0.839 (0.669–0.951)

Abbreviations: ICC, intraclass correlation coefficients; WOB, work of breathing.

escalation of respiratory support as well as weaning success. As a first step, our study contributes to this process of developing a clinical WOB score by determining the reliability of judging the WOB in children admitted to the PICU. The strength of our study is the very high number of observers, including both PICU physicians and nurses, who varied in their clinical experience, and use of an unbiased, large set of included signs of effort of breathing based on a systematic search of the current literature.

Given the high number of clinical judgments of the WOB in children that PICU clinicians make on a day-to-day basis, it is quite disturbing that the interobserver agreement on rating the overall WOB in our study was low. Similar findings have previously been reported for subjectively assessing the severity of acute dyspnea in children with wheezing conditions such as asthma^{12,13} and postextubation upper airway obstruction.¹⁴ Apparently, even in a setting with clinicians highly specialized in pediatric acute pulmonary medicine, such as the PICU in our study, there is large variability in judgment of the degree of respiratory distress.

One could hypothesize that breaking up the judgment of the overall WOB into a score of several separate signs of effort of breathing will increase the interobserver agreement, as the observers are forced to rate the separate parameters of the WOB more specifically. Yet, in our study only three signs were found to be judged with acceptable interobserver agreement in a reliability analysis. Of these, only two (stridor and grunting) are pure subjective signs of effort of breathing. Interestingly, Shein et al recently derived a clinical three-item (stridor, pulsus paradoxus, and retractions) score in the PICU from objective WOB measurements by esophageal manometry in a secondary analysis from a previous prospective cohort focused on pediatric postextubation upper airway obstruction.⁵ This score acceptably predicted the need for escalating respiratory support, showing that a clinical WOB score may still be of value even when consisting of only a few signs of effort of breathing. However, the external validity of such a simple clinical WOB score in a PICU population including a variety of underlying illnesses, remains to be determined.

An important observation from the Shein study is that the prediction model worked best when the summated WOB score from (at least) three observers was used,⁵ thus in the situation that a patient is observed by a team instead of one observer. In line with this, high interobserver agreement has been reported previously in a reliability analysis of the pediatric asthma score using the average measures of multiple observers.¹⁵ However, we believe that to function well in daily practice, interobserver agreement of a clinical WOB score should be evaluated in the context of observations by *single* raters (e.g., at various time points before and after physician/nurse rotations). Indeed, in our study calculated average measure ICCs were high, contrasting with the relatively low single-measure ICCs (primary outcome), suggesting poor reliability of individual clinical judgment of WOB in our cohort.

Our study has several limitations. First, we used movie clips of patients instead of “live” patients, which may result in limited or altered assessment of clinical WOB by the observers. However, the use of movie clips enabled us to include a uniquely high number of observers scoring the same patient

at exactly the same time point/phase of disease, which was most relevant for the scope of the study. In addition, the use of movie clips precluded bias based on availability of any prior information on the primary diagnosis or patient outcome, enabling us to assess interobserver agreement purely on subjective findings. Second, the inclusion of patients in our study was random, resulting in a selection of children with relatively young age. Although this cohort bias reflects the age distribution in a general PICU population, it is possible that interobserver reliability analysis differs among older children. Third, for the item “head bobbing” the variability in rating was too low (based on little variation in the children) to be able to discriminate between patients, and thus we were not able to reliably calculate the ICC. Head bobbing may be an important sign of effort of breathing in infants, and additional reliability analysis should be performed on this parameter. Finally, the primary goal of our study was to determine the interobserver agreement on judgment of a large set of subjective clinical items of the WOB in children. Although our findings may aid future development of a simple clinical WOB score in the PICU, we must stress that assessment of the validity of such a clinical WOB scoring instrument against objective measures of WOB or patient outcomes is a prerequisite in this future process.

In conclusion, the interobserver agreement on the clinical judgment of the WOB in spontaneously breathing children admitted to the PICU among physicians and nurses is disappointingly low. These results should be taken into account in daily respiratory support decision making in critically ill children and future development of clinical WOB scores designed specifically for the PICU.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

None declared.

References

- 1 Crow SS, Undavalli C, Warner DO, et al. Epidemiology of pediatric critical illness in a population-based birth cohort in Olmsted County, MN. *Pediatr Crit Care Med* 2017;18(03):e137–e145
- 2 Ibiebele I, Algert CS, Bowen JR, Roberts CL. Pediatric admissions that include intensive care: a population-based study. *BMC Health Serv Res* 2018;18(01):264
- 3 Punchak M, Hall K, Seni A, et al. Epidemiology of disease and mortality from a PICU in Mozambique. *Pediatr Crit Care Med* 2018;19(11):e603–e610
- 4 Cabello B, Mancebo J. Work of breathing. *Intensive Care Med* 2006;32(09):1311–1314
- 5 Shein SL, Hotz J, Khemani RG. Derivation and validation of an objective effort of breathing score in critically ill children. *Pediatr Crit Care Med* 2019;20(01):e15–e22
- 6 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(01):98–112
- 7 Maue DK, Krupp N, Rowan CM. Pediatric asthma severity score is associated with critical care interventions. *World J Clin Pediatr* 2017;6(01):34–39

- 8 Justicia-Grande AJ, Pardo Seco J, Rivero Calle I, Martín-Torres F. Clinical respiratory scales: which one should we use? *Expert Rev Respir Med* 2017;11(12):925–943
- 9 Wang EE, Milner RA, Navas L, Maj H. Observer agreement for respiratory signs and oximetry in infants hospitalized with lower respiratory infections. *Am Rev Respir Dis* 1992;145(01):106–109
- 10 de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;59(10):1033–1039
- 11 Bujang MD, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intra-class correlation coefficient: a review. *Arch Orophac Sci* 2017;12:1–11
- 12 Bekhof J, Reimink R, Bartels IM, Eggink H, Brand PL. Large observer variation of clinical assessment of dyspnoeic wheezing children. *Arch Dis Child* 2015;100(07):649–653
- 13 Eggink H, Brand P, Reimink R, Bekhof J. Clinical scores for dyspnoea severity in children: a prospective validation study. *PLoS One* 2016;11(07):e0157724
- 14 Khemani RG, Schneider JB, Morzov R, Markovitz B, Newth CJ. Pediatric upper airway obstruction: interobserver variability is the road to perdition. *J Crit Care* 2013;28(04):490–497
- 15 Biondi EA, Gottfried JA, Dutko Fioravanti I, Schriefer JA, Aligne CA, Leonard MS. Interobserver reliability of attending physicians and bedside nurses when using an inpatient paediatric respiratory score. *J Clin Nurs* 2015;24(9-10):1320–1326
- 16 Qureshi F, Pestian J, Davis P, Zaritsky A. Effect of nebulized ipratropium on the hospitalization rates of children with asthma. *N Engl J Med* 1998;339(15):1030–1035
- 17 Fleming S, Thompson M, Stevens R, et al. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. *Lancet* 2011;377(9770):1011–1018