Interobserver Reliability of the Respiratory Physical Examination in Premature Infants: A Multicenter Study

Erik A. Jensen, MD, MSCE[1], Howard Panitch, MD[2], Rui Feng, PhD[3], Paul E. Moore, MD[4], and Barbara Schmidt, MD, MSc[1]

Objective To measure the inter-rater reliability of 7 visual and 3 auscultatory respiratory physical examination findings at 36-40 weeks' postmenstrual age in infants born less than 29 weeks' gestation. Physicians also estimated the probability that each infant would remain hospitalized for 3 months after the examination or be readmitted for a respiratory illness during that time.

Study design Prospective, multicenter, inter-rater reliability study using standardized audio-video recordings of respiratory physical examinations.

Results We recorded the respiratory physical examination of 30 infants at 2 centers and invited 32 physicians from 9 centers to review the examinations. The intraclass correlation values for physician agreement ranged from 0.73 (95% CI 0.57-0.85) for subcostal retractions to 0.22 (95% CI 0.11-0.41) for expiratory abdominal muscle use. Eight (27%) infants remained hospitalized or were readmitted within 3 months after the examination. The area under the receiver operating characteristic curve for prediction of this outcome was 0.82 (95% CI 0.78-0.86). Physician predictive accuracy was greater for infants receiving supplemental oxygen (0.90, 95% CI 0.86-0.95) compared with those breathing in room air (0.71, 95% CI 0.66-0.75).

Conclusions Physicians often do not agree on respiratory physical examination findings in premature infants. Physician prediction of short-term respiratory morbidity was more accurate for infants receiving supplemental oxygen compared with those breathing in room air. *(J Pediatr 2016;178:87-92).*

Premature infants are at risk for significant, persistent respiratory morbidity. Bronchopulmonary dysplasia (BPD) is the most common chronic respiratory condition associated with preterm birth and is a strong predictor of multiple adverse health outcomes, including impairments in lung function and neurodevelopmental delay[1-3]; however, many former preterm infants who do not fulfill the diagnostic criteria for BPD also experience deficits in respiratory health through school age and into adulthood.[4-6] Better methods are needed to quantify the severity of lung disease in the neonatal period and to predict which infants are likely to experience long-term respiratory complications.

The Prematurity and Respiratory Outcomes Program (PROP) is a large, multicenter observational study of infants who are born at less than 29 weeks of gestation.[7,8] One of the many aims of PROP is to evaluate whether components of the respiratory physical examination assessed between 36 and 40 weeks' postmenstrual age (PMA) in extremely preterm infants are predictive of respiratory outcomes at 1-year corrected age. A necessary step in the validation of a physical examination finding as a diagnostic test or surrogate study endpoint is assessment of interobserver agreement. This has not been measured for the respiratory examination in preterm infants.

Studies in older infants and children with acute respiratory illnesses, including asthma, bronchiolitis, croup, and pneumonia, reported levels of interobserver agreement that varied from slight to almost perfect.[9-18] Moreover, no visual or auscultatory finding consistently demonstrated high or low agreement in those studies.[9-18] The objective of the present study was to measure the interobserver reliability of several components of the respiratory physical examination in premature infants when examined between 36 and 40 weeks' PMA. In addition, we assessed whether physicians accurately could predict an infant's risk for short-term future respiratory morbidity based on findings from a single respiratory physical examination.

BPD	Bronchopulmonary dysplasia
ICC	Intraclass correlation coefficient
NICU	Neonatal intensive care unit
PMA	Postmenstrual age
PROP	Prematurity and Respiratory Outcomes Program
ROC	Receiver operating characteristic

From the [1]Department of Pediatrics, Division of Neonatology, The Children's Hospital of Philadelphia and The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; [2]Department of Pediatrics, Division of Pulmonology, The Children's Hospital of Philadelphia and The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; [3]Department of Biostatistics and Epidemiology, The Children's Hospital of Philadelphia and The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; and [4]Department of Pediatrics, Division of Pediatric Allergy, Immunology, and Pulmonary Medicine, Vanderbilt University School of Medicine, Nashville, TN

E.J. was supported by a Pediatric Hospital Epidemiology Outcomes Training grant funded by the *Eunice Kennedy Shriver* National Institutes of Child Health and Human Development (NICHD; 5T32HD060550) and a Scholar's Grant from the Prematurity and Respiratory Outcomes Program funded by the National Heart, Lung, and Blood Institute and NICHD (U01-HL-101794). The authors declare no conflicts of interest.

http://dx.doi.org10.1016/j.jpeds.2016.07.039

Methods

Infants with gestational ages between 23 and $28^{6/7}$ weeks without hemodynamically significant congenital heart disease; structural abnormalities of the upper airway, lungs, or chest; or congenital malformations or syndromes were eligible for enrollment. The proportions of infants breathing room air and receiving various levels of respiratory support at the time of the examination were a priori selected to be similar to the respiratory support characteristics of the full PROP cohort at 36-40 weeks' PMA. This was done to encourage generalizability of our findings to PROP and the broader US population of extremely preterm infants for which PROP is intended to be a representative sample. All infants were enrolled from the neonatal intensive care units (NICUs) at the Hospital of the University of Pennsylvania or the Children's Hospital of Philadelphia. The institutional review boards at both institutions approved the protocol, and written informed consent was obtained from a parent or guardian of each infant.

Attending or senior fellow physicians in the divisions of neonatology and pediatric pulmonology at the Children's Hospital of Philadelphia and physician members of the PROP steering committee were invited to participate. Written informed consent was obtained from each physician.

Physical Examination Recordings and Playback

A 3-minute audio-video recording of a structured respiratory physical examination was produced for each enrolled infant (**Video**; available at www.jpeds.com). Wide angle, cross-table lateral, and close-up views of the chest, abdomen, and face and audio recordings of 8 lung and upper airway fields were captured in each recording (**Table I**). All recordings were filmed from the infant's side at an approximately 45° angle except for a cross-table lateral view filmed level with the infant. The infants were filmed wearing a diaper only and while in a quiet awake or sleep state. All recordings were conducted in the NICU at the infant's bedside. No artificial lighting was used, and ambient sounds such as those generated by monitoring equipment and medication pumps were not silenced. To protect the privacy of those present in the NICU at the time of the examination, however, audio was muted during the visual inspection portion of the recording when a stethoscope was not in use.

Table I. Respiratory examination recording duration by section

Examination sections	Duration
Visual observation	
Full view of infant	12 s
Close-up view of the chest and abdomen	12 s
Cross table lateral view of the chest and abdomen	12 s
Close-up view of the neck and upper chest	12 s
Close-up view of the head and face	12 s
Auscultation	
Stethoscope held over the mouth and nose	12 s
Stethoscope held over the trachea	12 s
Anterior lung fields bilaterally	12 s each
Lateral lung fields bilaterally	12 s each
Posterior lung fields bilaterally	12 s each
Total examination duration	2 min, 36 s
Nonexamination duration (subject identification, video transitions)	24 s
Total recording duration	3 min

A JVC EX GZ-515 high-definition video camera (JVC, Wayne, New Jersey) mounted on a tripod and a Littmann 3200 electronic stethoscope (3M, Maplewood, Minnesota) were used to record all examinations. The Littmann 3200 captures cardiac and respiratory sounds with frequencies between 20 and 2000 Hz.[19] The majority of breath sounds have frequencies of 200-250 Hz, except for wheezes and crackles, which have frequency spectra of 100-1000 Hz and 200-2000 Hz, respectively.[20,21] All audio-video editing was done with iMovie version 10.0.4 (Apple Inc, Cupertino, California).

All physicians reviewed the examination recordings using personal computers while under the supervision of a study team member or by accessing the password protected PROP study Web site. Video resolution for all recordings was set at 720×1280 px. Each physician was provided a set of Sony MDR-EX38iP in-ear stereo headphones (Sony, New York, New York) and was requested to set the audio playback volume in a range similar to actual physical auscultation. Physicians were permitted to review the examination recordings as many times as necessary. All findings were reported to the study team via the online data capture software REDCap.[22]

Outcomes

The primary study outcomes were the physicians' ratings of 7 visual (suprasternal, intercostal, and subcostal retractions; thoracoabdominal synchrony; head bobbing; nasal flaring; and expiratory abdominal muscle use) and 3 auscultatory (wheezing or noisy breathing, crackles, and stridor) respiratory physical examination findings. Each examination finding was assessed as present or absent except for synchronous or asynchronous thoracoabdominal movement. All infants were followed for 3 months after the recording to determine whether they remained hospitalized or were readmitted for a respiratory illness after discharge from the NICU. Physicians were asked to estimate the probability of this respiratory morbidity for each infant solely on the basis of the physical examination findings observed in the recording. No additional demographic or clinical information was provided.

Statistical Analyses

Physician and infant characteristics were summarized with standard descriptive statistics. The interobserver reliability for each examination finding was measured via the intraclass correlation coefficient (ICC). The ICC was calculated with a mixed-effects logistic regression model that included infants as a random effect and physicians as a fixed effect.[23] According to standard criteria, an ICC < 0.4 indicated "poor" reliability, 0.4-0.74 indicated "fair" reliability, and ≥0.75 indicated "excellent" reliability.[24] To assess the accuracy of the physicians' predictions of future respiratory morbidity, we computed a single best linear unbiased prediction for each infant using a generalized linear model. Infants were included in the model as a first order random effect, and physicians were included