



# Agreement of an echocardiogram-based diagnosis of pulmonary hypertension in infants at risk for bronchopulmonary dysplasia among masked reviewers

Andrew W. McCrary<sup>1</sup> · Piers C. A. Barker<sup>1</sup> · Rachel D. Torok<sup>1,2</sup> · Tracy G. Spears<sup>2,3</sup> · Jennifer S. Li<sup>1,2</sup> · Christopher P. Hornik<sup>1,2</sup> · Matthew M. Laughon<sup>4</sup>

Received: 31 July 2018 / Revised: 26 October 2018 / Accepted: 30 October 2018 / Published online: 21 November 2018  
© Springer Nature America, Inc. 2018

## Abstract

**Objective** To evaluate the agreement of an echocardiogram-based pulmonary hypertension diagnosis in premature infants at risk for bronchopulmonary dysplasia (BPD).

**Study design** Echocardiograms from infants born  $\leq 28$  weeks post menstrual age were retrospectively reviewed with a standardized reading protocol by three pediatric cardiologists masked to patient's clinical history to determine the presence of pulmonary hypertension.

**Results** A total of 483 echocardiograms from 49 unique patients were each reviewed by three pediatric cardiologists. Overall there was an 82.9% agreement on the presence of pulmonary hypertension among the three readers (95% CI: 78.4%, 85.4%) with a modified Fleiss' kappa of 0.759 (95% CI: 0.711, 0.801). Percent agreement between rereads was 92.4%, and modified Fleiss' kappa was 0.847 (95% CI: 0.750, 0.931).

**Conclusions** Using a standardized reading protocol and echocardiogram-based definition of pulmonary hypertension, there is high inter- and intra-rater agreement for the diagnosis of pulmonary hypertension in at-risk premature infants, suggesting echocardiography can be successfully used for clinical and research monitoring of pulmonary hypertension in infants.

## Introduction

More than 50,000 infants are born  $\leq 28$  weeks post menstrual age (PMA) each year in the US and ~40–55% develop bronchopulmonary dysplasia (BPD) [1]. Premature infants with BPD are challenging to treat, have a high mortality rate, and survivors suffer from multiple morbidities, including pulmonary hypertension, prolonged hospitalization, and neurodevelopmental impairment [2–4]. Approximately 14–28%

of premature infants with BPD develop pulmonary hypertension, and there is increased risk of mortality among those with BPD who develop pulmonary hypertension during the neonatal admission, usually due to cardiorespiratory compromise and right-sided heart failure. These infants are often treated with multiple drugs in an attempt to reduce pulmonary vascular resistance [5]. However, the US Food and Drug Administration has not approved any drugs for infants with BPD and pulmonary hypertension, and to date, no drug has demonstrated efficacy.

The lack of a robust, reproducible non-invasive diagnostic marker for pulmonary hypertension is a major barrier to therapeutic trials in this population [6–8]. Cardiac catheterization is the gold standard for quantifying pulmonary hypertension in adults and children, but is associated with a high incidence of mortality and morbidities in infants, particularly those weighing less than 2 kg [9–11]. Echocardiography is the ideal non-invasive alternative but cannot be validated against the gold standard due to risks and limited feasibility of performing a high volume of cardiac catheterizations in the neonatal population. Typically, echocardiographic diagnosis of pulmonary

✉ Matthew M. Laughon  
matt\_laughon@med.unc.edu

<sup>1</sup> Division of Pediatric Cardiology, Department of Pediatrics, Duke University Medical Center, Durham, NC, USA

<sup>2</sup> Duke Clinical Research Institute, Durham, NC, USA

<sup>3</sup> Department of Pediatrics Center for Predictive Medicine, Duke University Medical Center, Durham, NC, USA

<sup>4</sup> Division of Neonatal-Perinatal Medicine, Department of Pediatrics, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

hypertension is reliant on tricuspid regurgitation jet velocity (TRJV) to estimate pulmonary arterial pressure, but the measurement frequently fails due to the lack of a sufficient regurgitation jet [9, 12]. Without a sufficient tricuspid regurgitation jet, there is a reliance on a combination of qualitative and quantitative measures of pulmonary hypertension by reading cardiologists. Historically, there has been poor agreement in diagnostic interpretation among cardiologists relying on such combinations of qualitative and quantitative measures for diagnosis in different disease states [13, 14]. In the neonatal population, recent work has shown that assessment of septal geometry by end-systole eccentric index improves diagnostic consistency of pulmonary hypertension [15]. Recognizing the current limitations, echocardiograms are used clinically in the intensive care nursery for monitoring pulmonary hypertension progression and treatment effects [12, 16, 17]. Echocardiography has not been well-validated in premature infants and needs to be studied prior to using it in a therapeutic clinical trial for the identification and monitoring of pulmonary hypertension in premature infants with BPD.

Recognizing this critical need, our objective was to determine whether echocardiograms are sufficiently reliable to diagnose and monitor pulmonary hypertension in at risk premature infants. We report our evaluation of the agreement of pulmonary hypertension diagnosis using clinically obtained echocardiograms read by pediatric cardiologists in premature infants at risk for BPD.

## Methods

### Study population

Infants were selected from two sites (University of North Carolina Medical Center and Duke University Medical Center) as a part of this retrospective cohort study. Inclusion criteria included: (1) Post-menstrual age at birth  $\leq 28$  weeks; (2) At least two echocardiograms performed per standard of care after day of life 14 and at least 1 week apart; and (3) Born between 2005 and 2015. Exclusion criteria included: (1) Known congenital heart disease (except patent ductus arteriosus, patent foramen ovale, or secundum atrial septal defect); (2) Known congenital diaphragmatic hernia; or (3) Known chromosome abnormality.

The study protocol was approved by The University of North Carolina at Chapel Hill Institutional Review Board and the Duke University Medical Center Institutional Review Board with a waiver of informed consent. Study coordinators at participating sites abstracted data from the medical record. Echocardiograms were transferred to Duke Clinical Research Institute (DCRI), stripped of all protected health information, and loaded onto a central server for reading.

Echocardiographic measures were obtained from two-dimensional and Doppler echocardiograms using a calibrated off-line analysis system (Digisonics Digiview, Houston, TX, USA). Data quality was assessed by the DCRI study team through frequent structured data review meetings throughout the echocardiogram-reading period. Additionally, DCRI core laboratory sonographers checked all pediatric cardiologist echocardiogram report forms for appropriateness and completeness of measurement and interpretation reports.

### Echocardiographic measurements

Three pediatric cardiologists (1 senior and 2 juniors) masked to any patient-specific clinical information each interpreted all study echocardiograms using a standardized imaging protocol with an additional random 10% re-read. The pediatric cardiologists were aware of the risk of pulmonary hypertension but were not aware of any clinical information regarding the patients, including diagnoses, clinical status, and treatments, and timing of the echocardiograms. End-systolic eccentricity index was used to quantify septal geometry changes related to elevated pulmonary arterial pressure [18]. Right ventricular structure and function were assessed by the following qualitative and quantitative measurements: subjective composite scoring assessment (Table 1), direction of cardiac shunts, right ventricular fractional area change percent, and tricuspid annular plane systolic excursion measured according the pediatric quantification guidelines of the American Society of Echocardiography [19]. Measurements were obtained in triplicate of three different beats when possible and averaged together [20]. Blood pressure measurement was obtained from the echocardiogram reports; blood pressure measurement method was not recorded.

### Definition of pulmonary hypertension

We defined pulmonary hypertension as follows:

1. Echocardiograms with a tricuspid regurgitation jet complete Doppler envelope:
  - a. Estimated right ventricular systolic pressure/systolic blood pressure ratio  $\geq 0.67$  without pulmonary outflow tract obstruction was considered pulmonary hypertension [9]. Right ventricular systolic pressures were estimated from a continuous wave Doppler tracing of the peak tricuspid valve regurgitation velocity and converted to pressure using modified Bernoulli's equation ( $4 \times \text{TRJV}^2$ ) without consideration of central venous pressure.
  - b. A peak tricuspid valve regurgitation jet velocity  $> 3.0$  m/s was also considered consistent with pulmonary hypertension.

**Table 1** Pulmonary hypertension composite score of echocardiographic measures

Subjective quantification of RV pressures	Absolute criteria Septal geometry in systole	Supportive criteria		
		Shunting <sup>a</sup>	RV function <sup>b</sup>	RV/RA size <sup>c</sup>
Normal (<1/3 systemic)	Septum round Eccentricity index <sup>d</sup> ≤1.0	L to R	Normal	Normal
Mildly elevated (1/3–2/3 systemic)	Septal geometry mildly distorted, Eccentricity index 1.01–1.20	L to R	Normal	Mildly dilated
Moderately elevated (≥2/3 systemic)	Septum moderately distorted but not flat, Eccentricity index 1.21–1.4	L to R	Mildly depressed	Mod. dilated
Severely elevated (≥systemic)	Septum flat or bowing into left ventricle, Eccentricity index > 1.4	R to L	≥ Moderately depressed	Severely dilated

RV right ventricle, RA right atrium, L left, R right

<sup>a</sup>Across a PFO/ASD or PDA

<sup>b</sup>Measures of RV function to include: subjective estimates, RV fractional area change, tricuspid annular plane systolic excursion

<sup>c</sup>RV/RA size based on subjective appearance

<sup>d</sup>LV eccentricity index (LVEI) = LV antero-posterior diameter/LV septo-lateral diameter

## 2. Echocardiograms without a tricuspid regurgitation jet:

A composite scoring system previously developed by the study team from prior data was used (Table 1) [21]. Validation of the scoring system was not possible in this study as only one infant had a cardiac catheterization.

- a. Using the composite scoring system, a subjective right ventricular pressure estimate was assigned to all echocardiograms, regardless of presence of tricuspid regurgitation jet. For each echocardiogram, the severity level was assigned based on the echocardiogram possessing at least two of the four criteria for that severity level (e.g. eccentricity index of 1.3 and moderately depressed right ventricular function would qualify for a subjective right ventricular pressure estimate of moderately elevated). The most severe level in which the 2 of 4 criteria were met was assigned. Moderately or severely elevated pressure indices were considered consistent with pulmonary hypertension.

## Analysis

The primary outcome measure was inter-rater agreement of the diagnosis of pulmonary hypertension among the panel of pediatric cardiologists (dichotomous variable: present or absent). We report the proportion of infants diagnosed with pulmonary hypertension using echocardiography for each pediatric cardiologist. Inter-rater agreement was measured using modified Fleiss' kappa coefficients with bootstrap 95% confidence intervals [22].

Three secondary outcomes were analyzed. First, the intra-rater agreement of the diagnosis of pulmonary

hypertension was assessed by comparing a random sample of 10% of the echocardiograms read twice, in random order. This intra-rater agreement of the diagnosis of pulmonary hypertension was measured by percent agreement and modified Fleiss' kappa with 95% confidence intervals. Second, the performance of the consensus panel over time was measured by comparing percent agreement and modified Fleiss' kappa calculated at three sequential study time points (study reading period divided into thirds). Third, we assessed the internal consistency of the pulmonary hypertension composite score using Cronbach's alpha. Cronbach's alpha if one item deleted was also calculated to assess for redundancy in composite score items. Additionally, percent agreement and modified Fleiss' kappa was calculated for each component of the composite score to assess inter-rater agreement within components. These measures were calculated with the sample population stratified by clinical factors (weight > or ≤ 1500 g, day of life [DOL] > or ≤ 60, PMA > or ≤ 40 weeks, mechanically ventilated or not).

## Sample size

A sample size of least 450 echocardiograms was deemed necessary to estimate Fleiss' kappa with a precision ≤0.05, assuming a kappa of 0.7, prevalence of pulmonary hypertension of 0.3, a type I error rate of 0.05, and three raters [23].

## Results

A total of 483 echocardiograms from 49 unique patients meeting inclusion criteria were selected from two institutions and read by all three pediatric cardiologists. Forty-nine echocardiograms were randomly selected to be read a

second time by each reader. Infants with the most echocardiograms at each institution were selected to minimize the number of unique patients and patients most likely to carry a clinical diagnosis of pulmonary hypertension. Included patients had a median of 9 echocardiograms (range 6–29), of which a median of four were per infant prior to 36 weeks PMA. Overall, 188 of 483 (38.9%) echocardiograms were from infants <36 weeks PMA.

Patient characteristics are shown in Table 2. Forty-six of the 49 infants were diagnosed with BPD prior to hospital discharge defined as oxygen use at 36 weeks' post menstrual age [1]. Of these, 12 of 49 (24.5%) were treated with nitric oxide, 11 of 49 (22.5%) were treated with sildenafil, and 19 of 49 (38.8%) were treated with one of the two.

The clinical echocardiograms generally did not have quantifiable measures of right ventricular structure or function. Pulmonary arterial pressure estimates based on TRJV were measured in only 6% of echocardiograms (88 of 1449 total studies interpreted). The median TRJV was 40 mmHg (range 13–133). For echocardiograms with measurable TRJV, the Kappa correlation coefficient between TRJV and the composite score diagnosis of pulmonary

hypertension was 0.6 (95% CI: 0.42–0.79). EI was measured by all three readers in 344 echocardiograms. The interclass correlation for EI among the readers was 0.59. Right ventricular fractional area change was reported in 73% of echocardiograms (1059 of 1449 studies). Agreement of fractional area change among readers was very low with an interclass correlation of 0.17. No patients had echocardiographic evidence of pulmonary vein stenosis.

Only one infant underwent a cardiac catheterization, which demonstrated the diagnosis of pulmonary hypertension. For this infant, there was an echocardiogram 7 days prior to the catheterization, and all reviewers reported no evidence of pulmonary hypertension. Additionally, there was another echocardiogram 14 days after the catheterization, and all reviewers reported evidence of pulmonary hypertension.

### Agreement among three pediatric cardiologists on the echocardiographic definition of pulmonary hypertension

Of the 483 echocardiograms reviewed, 470 (97.3%) contained results sufficient to determine the presence or absence of pulmonary hypertension using the primary study definition (elevated TRJV or composite score) from all three pediatric cardiologists and were included in the analysis. Thus, the composite score defined most cases of pulmonary hypertension. The percentage of pulmonary hypertension diagnosis by each pediatric cardiologist was 12.3, 5.5, and 18.5%, respectively. Percent agreement among the panel of cardiologists for echocardiogram-based diagnosis of pulmonary hypertension was 81.9% (95% confidence interval: 78.4–85.4%), demonstrating overall very good agreement (Table 3). The modified Fleiss' kappa was 0.759 (95% confidence interval: 0.771–0.801) reflecting substantial agreement among the cardiologists [24]. Furthermore, the percent agreement over time was calculated to assess changing individual reading patterns over time. There was consistent percent agreement among the readers throughout the study (range 79.5–83.88%) and modified Fleiss' kappa range was 0.726–0.784.

Assessing intra-rater reliability, duplicates of ~10% (49/483) of the echocardiograms were randomly incorporated into the readers' reading list. Percent agreement between re-read echocardiograms was 92.4% overall (range 89.8–95.7%), and modified Fleiss' kappa was 0.847 (95% confidence interval: 0.750–0.931), demonstrating high agreement of initial interpretation with re-read echocardiograms.

Additionally, agreement among the three readers was stratified by several clinical factors (Table 4). There tended to be higher agreement among the readers with echocardiograms obtained on younger (less than 60 days old), PMA <40 weeks, and sicker (mechanically ventilated) infants.

**Table 2** Patient characteristics

Characteristic	Patients (N = 49)
Post-menstrual age at birth (weeks)	25.0 (22.0, 28.0)
Birth weight (grams)	660.0 (420.0, 915.0)
<i>Sex</i>	
Female	28/49 (57%)
<i>Race</i>	
Asian	1/49 (2%)
Black/African American	27/49 (55%)
Other	5/49 (10%)
White/caucasian	16/49 (33%)
<i>Ethnicity</i>	
Hispanic or Latino	3/49 (6%)
Not Hispanic or Latino	44/49 (90%)
Not reported	2/49 (4%)
<i>Inborn/outborn</i>	
Inborn	40/49 (82%)
<i>Mode of delivery</i>	
Vaginal	18/49 (37%)
Caesarian section	31/49 (63%)
5 min APGAR	6 (1, 9)
Antenatal steroids	39/49 (80%)
Surfactant therapy	47/49 (96%)
<i>Site</i>	
UNC	27/49 (55%)
Duke	22/49 (45%)

Values presented are counts (%) for categorical variables or median (min, max) for continuous variables

**Table 3** Agreement among three pediatric cardiologists on echocardiographic diagnosis of pulmonary hypertension

Pulmonary hypertension diagnosis	Overall	Over time <sup>a</sup>		
		First read period	Second read period	Last read period
Overall	171/1410 (12%)	65/472 (14%)	70/467 (15%)	38/477(8%)
Percent agreement among three readers (95% CI)	81.9% (78.4%, 85.4%)	120/151 (79.5%)	126/154 (81.8%)	129/154 (83.88%)
Modified Fleiss' $\kappa$ (95%CI)	0.759 (0.711, 0.801)	0.726 (0.638, 0.806)	0.758 (0.671, 0.835)	0.784 (0.706, 0.861)

<sup>a</sup>Read periods defined by dividing the total echocardiograms read into three groups. The order of the echocardiograms was the same for all readers. Absolute time to completion in each period was different for each reader

**Table 4** Stratified agreement by clinical factors at time of echocardiogram

Clinical factor	Count (% agreement)	Modified Fleiss' $\kappa$ (95% CI)
Weight $\leq$ 1500 g	125/140 (89.3)	0.857 (0.781, 0.914)
Weight $>$ 1500 g	222/282 (78.7)	0.716 (0.655, 0.778)
DOL $\leq$ 60	140/161 (87.0)	0.826 (0.743, 0.892)
DOL $>$ 60	245/309 (79.3)	0.724 (0.663, 0.784)
PMA $\leq$ 40 weeks	211/245 (86.1)	0.815 (0.755, 0.875)
PMA $>$ 40 weeks	174/225 (77.3)	0.698 (0.621, 0.763)
Mechanically ventilated	211/245 (86.1)	0.780 (0.711, 0.843)
No ventilator support	208/258 (80.6)	0.742 (0.674, 0.809)

DOL day of life, PMA post-menstrual age

### Agreement on each composite score component

Table 5 demonstrates the agreement among all three pediatric cardiologists for each component of the composite score definition for pulmonary hypertension. There was a high level of agreement for the direction of shunting at the level of inter-atrial connection (percent agreement 95.8%, modified Fleiss' kappa 0.944). The remaining three components revealed wider variability but overall good agreement (percent agreement 76.7–82.9%, modified Fleiss' kappa 0.743–0.785). For the composite score, there was a Cronbach's alpha for all items of 0.640. Cronbach's alpha increased to 0.712 with the shunting item deleted (Table 5).

### Discussion

Diagnosis of pulmonary hypertension in infants with BPD is challenging. Diagnostic cardiac catheterizations are not routinely performed in the intensive care nursery due to significant risk of complications. Thus, clinicians and researchers are left to diagnose and monitor disease course with non-invasively derived estimates of pulmonary arterial pressure. Echocardiography is currently the clinical modality for the diagnosis of pulmonary hypertension and for disease monitoring. In this current study, we present the first publication to evaluate the agreement of pulmonary

hypertension diagnosis from clinically acquired echocardiograms in at-risk infants using an agreed definition for pulmonary hypertension. The main findings of our study are [1]; there was a high level of inter- and intra-rater agreement in the diagnosis of pulmonary hypertension among the panel of pediatric cardiologists masked to patient-specific clinical information; and [2] there was a high level of agreement among readers in the composite score definition of pulmonary hypertension despite intermediate levels of internal consistency.

These findings are an important step forward in refining a research- and clinically-deliverable diagnosis of pulmonary hypertension. An advantage of this study approach included a pre-defined definition of pulmonary hypertension. As seen in this study and others, TRJV without consideration of central venous pressure is frequently inadequate for estimation of pulmonary pressures in this population, possibly related to low-intravascular volume, as diuretics use is common in these infants [9]. Recently published work shows that there are additional concerns about TRJV's poor correlation with invasively derived mean pulmonary artery pressure and suggest that other echocardiographically derived estimates of mean pulmonary artery pressure may perform better maximum TRJV [25]. Our composite score definition of pulmonary hypertension had advantages of clear categories, recognizing that evaluation of septal geometry may be the next most important feature of the echocardiogram in defining pulmonary hypertension in infants following TRJV [15, 26–28]. Additionally, the inter- and intra-reliability of the readers' interpretations performed well. In other pediatric disease states relying on subjective assessments, inter-observer agreements in the interpretation of echocardiograms have been poor [13, 14]. Finally, the consistency of pulmonary hypertension diagnosis across the study timeline suggests that there was no significant maturation bias.

Cronbach's alpha is a measure of a scale's internal consistency. Values below 0.7 indicate poor internal consistency, whereas values  $>$ 0.9 indicate redundancy in scale components [29]. Cronbach's alpha is also calculated for the scale with each component deleted. As Cronbach's alpha increased from 0.640 on the full scale to 0.712 with

**Table 5** Agreement and internal consistency of the composite score definition of pulmonary hypertension

Composite score component	Reader agreement		Cronbach's alpha <sup>a</sup> (overall 0.640)
	Count (% agreement)	Modified Fleiss' $\kappa$ (95% CI)	
Septal geometry in systole	283/369 (76.7)	0.689 (0.631, 0.743)	0.555
Shunting	275/287 (95.8)	0.944 (0.912, 0.972)	0.712
RV function	374/466 (80.3)	0.737 (0.688, 0.785)	0.468
RV/RA size	388/468 (82.9)	0.716 (0.655, 0.778)	0.531

RV right ventricle, RA right atrium

<sup>a</sup>Cronbach's alpha if composite score item is removed

the shunting item deleted, this measure suggests that the shunting item should be removed, possibly due to the overly common result of shunting direction (i.e. left to right). Further work should be undertaken to improve and validate the echocardiographic definition, as it is apparent that a consensus definition is needed clinically and for therapeutic research trials to proceed.

Agreement among the three cardiologists tended to be best with the younger, smaller infants, even when the same infants have multiple echocardiograms. There are a few plausible explanations for this finding. First, as the child in the intensive care nursery ages, there is a lower likelihood of the presence of shunt direction indicators such as a patent foramen ovale or patent ductus arteriosus. As these physiologic shunts close with age or intervention, there are fewer sources for reliable indicators of relative pressure differences between the left and right ventricles in the echocardiogram. Related to presence of echocardiographic findings, acoustic windows may deteriorate as a children with BPD ages due to worsening lung condition. Second, agreement may worsen as the child ages due to cardiac changes from long-standing elevated pulmonary pressure. For instance, an older infant could have a large, thick right ventricle due to past elevation in pulmonary arterial pressure, but due to successful treatment the pressure is not elevated at the time of the exam. The discordance between the appearance of the right sided structures and the current pulmonary pressure could be another source of disagreement. Finally, younger infants may have had a greater proportion of severe disease. Thus, the relationship of younger infants and greater agreement between readers may reflect greater agreement with more severe cases of pulmonary hypertension.

There are several limitations to the current study. First, the study is retrospective in nature and relied on clinically-acquired echocardiograms across two institutions with slightly different institutional echocardiogram protocols. Although this limitation is important, it also suggests that standardized research echocardiograms performed according to a strict protocol may enjoy even higher agreement

and better differentiation of pulmonary hypertension. Likewise, the high degree of agreement among the readers should be interpreted within the context of adherence to study definitions of measurement and interpretations detailed in this manuscript. Second, while the composite score performed well by agreement, it did not provide distinction between two structural and physiologic states in children with pulmonary hypertension: [1] current elevation of pulmonary arterial pressure vs [2]. cardiac changes (i.e. dilated right ventricle, main pulmonary artery, etc.) associated with prolonged elevation of pulmonary arterial pressure but not current elevation of pulmonary arterial pressure. Due to the lack of cardiac catheterization data on these infants, there is no feasible gold standard test to validate these diagnoses or validate the scoring system proposed in this study. Third, the selection of patients into this study increased the pretest probability of the diagnosis of pulmonary hypertension among these infants. Thus, these results should be interpreted in the context of high risk infants with BPD. With that stated, less than 20% of individual echocardiograms were determined to meet the study definition of pulmonary hypertension. Finally, we had intended to compare cardiac catheterization measurements to the echocardiograms. However, only one infant meeting inclusion criteria at either institution had a cardiac catheterizations performed. This underscores the relative rarity of cardiac catheterizations in the intensive care nursery, presumably due to the high risk of a procedural adverse event in small infants.

Additionally, the total number of echocardiograms are from only 49 patients thus may not represent independent observations. Whereas correlation between the echocardiograms of the same infant may exist, the study was designed to evaluate the correlation of the assessment between different reviewers of each individual echocardiogram. The readers were not aware of studies from the same infant. Further, we acknowledge that correlation between observations may violate the assumption of independence used in the sample size calculation. We did not measure the quantitative measures of the correlation of agreement and its

impact on the confidence interval due to limited statistical research in this area as to the appropriate methods. In this study, the potential correlation of interest is between the agreement of one echocardiogram with agreement on another echocardiogram from the same infant. As the echocardiograms reviewed for each infant in this study represent routine medical care across a variety of clinical conditions (ventilatory status, PMA, pulmonary vasodilator medication exposure), no two echocardiograms within an infant were expected to be highly similar. Within infants, a median of 9 echocardiograms over a median 18 weeks of life were included. No infant had two echocardiograms from the same day.

## Conclusion

Pediatric cardiologists, masked to patient-specific clinical information, can reliably agree on the diagnosis of pulmonary hypertension by echocardiography. Further work should focus on defining and staging pulmonary hypertension in infants using prospectively acquired research echocardiograms validated with cardiac catheterization.

**Funding** This work was supported by the National Institutes of Health, NHLBI R34 HL124038. Dr. Laughon receives support from the US government for work in neonatal clinical pharmacology and clinical trials (FDA R01FD005101; PI Laughon, NHLBI R34 HL124038; PI Laughon, and Pediatric Trials Network Government Contract HHSN267200700051C; PI Benjamin Duke), for trial coordination (NIH OD 1UC2-OD023375; PI Smith); and as the satellite site PI for NICHD Neonatal Research Network (5U10HD040492; PI Cotton, Duke). Dr. Hornik receives salary support for research from National Institute for Child Health and Human Development (NICHD) (1K23HD090239) and the U.S. government for his work in pediatric and neonatal clinical pharmacology (Government Contract HHSN267200700051C, PI: Benjamin under the Best Pharmaceuticals for Children Act).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Stoll BJ, Hansen NI, Bell EF, Walsh MC, Carlo WA, Shankaran S, et al. Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993-2012. *JAMA*. 2015;314:1039–51.
2. Fily A, Pierrat V, Delporte V, Breart G, Truffert P. Group EN-P-d-CS. Factors associated with neurodevelopmental outcome at 2 years after very preterm birth: the population-based Nord-Pas-de-Calais EPIPAGE cohort. *Pediatrics*. 2006;117:357–66.
3. Vohr BR, Wright LL, Dusick AM, Mele L, Verter J, Steichen JJ, et al. Neurodevelopmental and functional outcomes of extremely low birth weight infants in the National Institute of Child Health

- and Human Development Neonatal Research Network, 1993-1994. *Pediatrics*. 2000;105:1216–26.
4. Wood NS, Costeloe K, Gibson AT, Hennessy EM, Marlow N, Wilkinson AR, et al. The EPICure study: associations and antecedents of neurological and developmental disability at 30 months of age following extremely preterm birth. *Arch Dis Child Fetal Neonatal Ed*. 2005;90:F134–40.
5. Altit G, Lee HC, Hintz S, Tacy TA, Feinstein JA, Bhombal S. Practices surrounding pulmonary hypertension and bronchopulmonary dysplasia amongst neonatologists caring for premature infants. *J Perinatol*. 2018;38:361–367.
6. Rich JD, Shah SJ, Swamy RS, Kamp A, Rich S. Inaccuracy of Doppler echocardiographic estimates of pulmonary artery pressures in patients with pulmonary hypertension: implications for clinical practice. *Chest*. 2011;139:988–93.
7. Janda S, Shahidi N, Gin K, Swiston J. Diagnostic accuracy of echocardiography for pulmonary hypertension: a systematic review and meta-analysis. *Heart*. 2011;97:612–22.
8. Bossone E, D'Andrea A, D'Alto M, Citro R, Argiento P, Ferrara F, et al. Echocardiography in pulmonary arterial hypertension: from diagnosis to prognosis. *J Am Soc Echocardiogr*. 2013; 26:1–14.
9. Mourani PM, Sontag MK, Younoszai A, Ivy DD, Abman SH. Clinical utility of echocardiography for the diagnosis and management of pulmonary vascular disease in young children with chronic lung disease. *Pediatrics*. 2008;121:317–25.
10. Caputo S, Furcolo G, Rabuano R, Basilicata AM, Pilla LM, De Simone A, et al. Severe pulmonary arterial hypertension in a very premature baby with bronchopulmonary dysplasia: normalization with long-term sildenafil. *J Cardiovasc Med*. 2010;11:704–6.
11. Hon KL, Cheung KL, Siu KL, Leung TF, Yam MC, Fok TF, et al. Oral sildenafil for treatment of severe pulmonary hypertension in an infant. *Biol Neonate*. 2005;88:109–12.
12. Skinner JR, Boys RJ, Hunter S, Hey EN. Non-invasive assessment of pulmonary arterial pressure in healthy neonates. *Arch Dis Child*. 1991;66(4 Spec No):386–90.
13. Dragulescu A, Mertens L, Friedberg MK. Interpretation of left ventricular diastolic dysfunction in children with cardiomyopathy by echocardiography: problems and limitations. *Circ Cardiovasc Imaging*. 2013;6:254–61.
14. Colan SD, Shirali G, Margossian R, Gallagher D, Altmann K, Canter C, et al. The ventricular volume variability study of the Pediatric Heart Network: study design and impact of beat averaging and variable type on the reproducibility of echocardiographic measurements in children with chronic dilated cardiomyopathy. *J Am Soc Echocardiogr*. 2012;25:842–54 e6.
15. Abraham S, Weismann CG. Left ventricular end-systolic eccentricity index for assessment of pulmonary hypertension in infants. *Echocardiography*. 2016;33:910–5.
16. Currie PJ, Seward JB, Chan KL, Fyfe DA, Hagler DJ, Mair DD, et al. Continuous wave Doppler determination of right ventricular pressure: a simultaneous Doppler-catheterization study in 127 patients. *J Am Coll Cardiol*. 1985;6:750–6.
17. Skinner JR, Stuart AG, O'Sullivan J, Heads A, Boys RJ, Hunter S. Right heart pressure determination by Doppler in infants with tricuspid regurgitation. *Arch Dis Child*. 1993;69:216–20.
18. Ryan T, Petrovic O, Dillon JC, Feigenbaum H, Conley MJ, Armstrong WF. An echocardiographic index for separation of right ventricular volume and pressure overload. *J Am Coll Cardiol*. 1985;5:918–27.
19. Lopez L, Colan SD, Frommelt PC, Ensing GJ, Kendall K, Younoszai AK, et al. Recommendations for quantification methods during the performance of a pediatric echocardiogram: a report from the Pediatric Measurements Writing Group of the American Society of Echocardiography Pediatric and Congenital Heart

- Disease Council. *J Am Soc Echocardiogr.* 2010;23:465–95. quiz576-7
20. Margossian R, Chen S, Sleeper LA, Tani LY, Shirali G, Golding F, et al. The reproducibility and absolute values of echocardiographic measurements of left ventricular size and function in children are algorithm dependent. *J Am Soc Echocardiogr.* 2015;28:549–58 e1.
  21. Mourani PM, Sontag MK, Ivy DD, Abman SH. Effects of long-term sildenafil treatment for pulmonary hypertension in infants with chronic lung disease. *J Pediatr.* 2009;154:379–84. 84 e1-2
  22. Falotico R, Quatto P. Fleiss' kappa statistic without paradoxes. *Qual Quant.* 2015;49:463–70.
  23. Donner A, Rotondi MA. Sample size requirements for interval estimation of the Kappa statistic for interobserver agreement studies with a binary outcome and multiple raters. *Int J Biostat.* 2010;6:1–13.
  24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–74.
  25. Hellenkamp K, Unsold B, Mushemi-Blake S, Shah AM, Friede T, Hasenfuss G, et al. Echocardiographic estimation of mean pulmonary artery pressure: a comparison of different approaches to assign the likelihood of pulmonary hypertension. *J Am Soc Echocardiogr.* 2018;31:89–98.
  26. Altit G, Dancea A, Renaud C, Perreault T, Lands LC, Sant'Anna G. Pathophysiology, screening and diagnosis of pulmonary hypertension in infants with bronchopulmonary dysplasia - A review of the literature. *Paediatr Respir Rev.* 2017;23:16–26.
  27. King ME, Braun H, Goldblatt A, Liberthson R, Weyman AE. Interventricular septal configuration as a predictor of right ventricular systolic hypertension in children: a cross-sectional echocardiographic study. *Circulation.* 1983;68:68–75.
  28. McCrary AW, Malowitz JR, Hornick CP, Hill KD, Cotten CM, Tatum GH, et al. Differences in eccentricity index and systolic-diastolic ratio in extremely low-birth-weight infants with bronchopulmonary dysplasia at risk of pulmonary hypertension. *Am J Perinatol.* 2016;33:57–62.
  29. De Vet H, Terwee C, Mokkink L, Knol D. Field testing: item reduction and data structure. *Measurement in Medicine: A Practical Guide (Practical Guides to Biostatistics and Epidemiology)*. Cambridge: Cambridge University Press; 2011. p. 65-95.