OPEN ACCESS

Association between bronchopulmonary dysplasia severity and its risk factors and long-term outcomes in three definitions: a historical cohort study

Trixie A Katz ,[1,2] Anton H van Kaam,[2,3] Nicolaas P A Zuithoff,[4] S M Mugie,[5,6] Sabine Beuger,[7] Geert Jan Blok,[7] Anne A M W van Kempen,[8] Henriëtte van Laerhoven,[9] Claire A M Lutterman,[10] Maarten Rijpert,[11] Irene A Schiering,[12] Nicolien C Ran,[13] Fenna Visser,[14] Els van Straaten,[15] Cornelieke S H Aarnoudse-Moens,[16] Aleid G van Wassenaer-Leemhuis,[1,17] Wes Onland [18]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/archdischild-2024-326931).

For numbered affiliations see end of article.

Correspondence to
Dr Wes Onland;
w.onland@amsterdamumc.nl

Received 30 January 2024
Accepted 26 May 2024
Published Online First
19 June 2024

To cite: Katz TA, van Kaam AH, Zuithoff NPA, *et al*. *Arch Dis Child Fetal Neonatal Ed* 2025;**110**:F51–F56.

ABSTRACT

Objective To compare the association of the severity categories of the 2001-National Institutes of Health (NIH), the 2018-NIH and the 2019-Jensen bronchopulmonary dysplasia (BPD) definitions with neurodevelopmental and respiratory outcomes at 2 and 5 years' corrected age (CA), and several BPD risk factors.

Design Single-centre historical cohort study with retrospective data collection.

Setting Infants born between 2009 and 2015 at the Amsterdam University Medical Centers, location Amsterdam Medical Center.

Patients Preterm infants born at gestational age (GA) <30 weeks and surviving up to 36 weeks' postmenstrual age.

Interventions Perinatal characteristics, (social) demographics and comorbidities were collected from the electronic patient records.

Main outcome measures The primary outcomes were neurodevelopmental impairment (NDI) or late death, and respiratory morbidity at 2 and 5 years' CA. Using logistic regression and Brier scores, we investigated if the ordinal grade severity is associated with incremental increase of adverse long-term outcomes.

Results 584 preterm infants (median GA: 28.1 weeks) were included and classified according to the three BPD definitions. None of the definitions showed a clear ordinal incremental increase of risk for any of the outcomes with increasing severity classification. No significant differences were found between the three BPD definitions (Brier scores 0.169–0.230). Respiratory interventions, but not GA, birth weight or small for GA, showed an ordinal relationship with BPD severity in all three BPD definitions.

Conclusion The severity classification of three BPD definitions showed low accuracy of the probability forecast on NDI or late death and respiratory morbidity at 2 and 5 years' CA, with no differences between the definitions.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Bronchopulmonary dysplasia (BPD) remains the most common complication of prematurity. Studies have shown that different BPD definitions have considerable difference in reported incidences, but no evident differences in discriminating performances for long-term neurodevelopmental and respiratory outcomes.

WHAT THIS STUDY ADDS

⇒ This historical cohort study with retrospective data collection cohort study shows that no current BPD definition is superior in classifying BPD severity, and that all definitions lack a good calibration, meaning that with every incremental increase in BPD severity, the risk of long-term respiratory and neurological outcomes does not equally increase.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ If deemed important by clinicians and researchers that BPD is diagnosed as an ordinal severity-based definition, future initiatives need to focus on the calibration as well as the discriminative predictive performance of these definitions.

INTRODUCTION

The validity of the current bronchopulmonary dysplasia (BPD) definitions in preterm infants is under debate.[1 2] Improved neonatal care has led to increased survival of extremely preterm infants at earlier stages of lung development, triggering the development of various new BPD definitions.[3–5] Studies have shown that these new and different definitions have considerable differences in the reported incidences of BPD, which may hamper valid comparison of BPD rates in benchmarking projects and interpretation of neonatal trials.[1 6 7] To avoid this unwanted variation, it is imperative that the neonatal community adopts one uniform definition of BPD.

The optimal BPD definition should identify those infants at risk of long-term adverse neurodevelopmental and respiratory outcomes. As most current BPD definitions classify infants based on BPD severity using descriptive terminology (mild, moderate, severe) or grades (I–II–III), the level of

Table 1 Patient characteristics

	Included at 36 weeks' PMA (n=584)
Gestational age*, weeks	28.1 (26.7–29.0)
Birth weight*, g	1040 (850–1240)
Small for GA, n (%)	86 (14.7)
Male gender, n (%)	323 (55.3)
Singleton, n (%)	416 (71.2)
Antenatal steroids, n (%)	832 (88.5)
Caesarean section, n (%)	269 (46.1)
Apgar score at 5 min*†	8 (7–9)
PDA, n (%)	200 (34.2)
Sepsis, n (%)	205 (35.1)
NEC ≥grade II, n (%)	50 (8.6)
IVH ≥grade III, n (%)†	30 (5.2)
Surfactant, n (%)	245 (42.0)
Doxapram, n (%)	88 (15.1)
Dexamethasone, n (%)	33 (5.7)
Days of supplemental oxygen*	3 (0–35)
Duration of IMV,* days	0 .17 (0–2.23)
Total duration of NICU stay,*‡ days	31 (17–49)
BPD incidence, n (%)	
2001-NIH definition	
Mild	51 (8.7)
Moderate	43 (7.4)
Severe	57 (9.8)
2018-NIH definition	
Grade 1	44 (7.5)
Grade 2	60 (10.3)
Grade 3	19 (3.3)
2019-Jensen definition	
Grade 1	189 (32.4)
Grade 2	35 (6.0)
Grade 3	3 (1.0)
Neurodevelopmental impairment/death after discharge	
2 years' CA, n (%) (n=466)	84 (18.0)
5 years' CA, n (%) (n=351)	131 (37.3)
Respiratory morbidity	
2 years' CA, n (%) (n=456)	102 (22.4)
5 years' CA, n (%) (n=332)	61 (18.4)

BPD incidence according to the 2001-NIH definition is defined as mild, moderate and severe BPD. BPD incidence according to the 2018-NIH and the 2019-Jensen definition is defined as grade 1, grade 2 and grade 3 BPD. Neurodevelopmental impairment at 2 years' CA is defined as having a composite cognitive score <85, composite motor score <85, cerebral palsy, hearing loss despite amplification and/or visual impairment leading to blindness or light perception only in at least one eye. Neurodevelopmental impairment at 5 years' CA is defined as having a full-scale IQ <85, scoring abnormal on the Movement Assessment Battery for Children-2 <5, cerebral palsy, hearing loss despite amplification and/or visual impairment leading to blindness or light perception only in at least one eye. Severe neurodevelopmental impairment at 5 years' CA is defined as having a full-scale IQ <70, cerebral palsy, hearing loss despite amplification and/or visual impairment leading to blindness or light perception only in at least one eye. Respiratory morbidity is defined as having hospitalisation count ≥3 visits, a visit to a pulmonary specialist, oxygen usage at home, use of bronchodilators, or antibiotics or diuretics at 1 and 2 years' CA.
*Reported as median (IQR).
†n=580.
‡Preterm infants are often transferred back to regional hospitals around 32 weeks' PMA.
BPD, bronchopulmonary dysplasia; CA, corrected age; GA, gestational age; IMV, invasive mechanical ventilation; IVH, intraventricular haemorrhage; NEC, necrotising enterocolitis; NICU, neonatal intensive care unit; NIH, National Institutes of Health; PDA, persistent ductus arteriosus needing therapy; PMA, postmenstrual age.

BPD severity should ideally also correlate with the degree of risk of adverse outcomes. Furthermore, the level of BPD severity should also correlate with the level of exposure to known risk factors for BPD such as lower gestational age (GA), lower birth weight, duration of invasive mechanical ventilation (IMV) and supplemental oxygen. Some studies have shown this ordinal association between disease severity and its short-term risk factors as well as long-term adverse outcomes.[8 9] However, to date, it is unclear if this association differs between the most recently published grade-based BPD definitions.[3–5 9] This information could help select the currently best-performing BPD definition.

Therefore, the first aim of this study was to compare the association between BPD severity and long-term neurodevelopmental and respiratory outcomes at 2 and 5 years' corrected age (CA) of three severity-based BPD definitions. The second aim was to investigate the association between BPD severity and perinatal characteristics and respiratory interventions.

METHODS

Study population

In this single-centre historical cohort study with retrospective data collection, preterm infants with a GA less than 30 weeks, admitted within 24 hours after birth to the neonatal intensive care unit of the Emma Children's Hospital Amsterdam University Medical Centers, between January 2009 and December 2015, and surviving up to 36 weeks' postmenstrual age (PMA), were included.[10] Severe congenital malformations and parental refusal to reuse clinical data were exclusion criteria.

Study outcomes

Infants were categorised by BPD severity at 36 weeks' PMA following the diagnostic criteria of the 2001-National Institutes of Health (NIH), 2018-NIH and the 2019-Jensen definition.[3–5 11] High-flow nasal cannula was implemented in daily practice in our unit in 2013. Our primary outcomes were the composite outcome of neurodevelopmental impairment (NDI) or late death (after 36 weeks' PMA up until follow-up), and respiratory morbidity at 2 and 5 years' CA. The specific definitions of the primary and secondary outcomes and how these were assessed can be found in the online supplemental material.

Statistical analysis

We investigated the accuracy of the probability forecast (calibration) of the primary composite outcome NDI or late death and respiratory morbidity at 2 and 5 years' CA.[12] Patient characteristics for the cohort were described, and compared with those lost to follow-up using appropriate testing depending on their distribution. Multiple imputation was performed 10 times based on clinical characteristics and social economic status. Analyses were performed in each imputation set separately and combined using Rubin's rules.[13]

For each separate BPD definition, a logistic regression with Firth's correction was performed because of expected imbalances between the incidence of BPD per severity category.[14] We calculated ORs with 95% CIs for each severity category of the three BPD definitions, using the no BPD group of each definition as the reference category. Next, we assessed the accuracy of the probability forecast (calibration) of the three BPD definition models with Brier score which is the mean squared deviation between the predicted probabilities and their respective outcomes. The Brier score can range from 0 for a perfect model to 0.25 for a non-informative model,[15] and was calculated for each model. We compared the two recently published

Table 2 OR for neurodevelopmental impairment and respiratory morbidity at 2 years and 5 years' corrected age

	Neurodevelopmental impairment OR (95% CI)	P value	Respiratory morbidity OR (95% CI)	P value
2 years' corrected age				
2001-NIH				
Mild BPD	1.46 (0.87 to 2.42)	0.19	1.45 (0.81 to 2.61)	0.63
Moderate BPD	1.44 (0.73 to 2.70)	0.32	1.41 (0.68 to 2.92)	0.37
Severe BPD	2.78 (1.56 to 4.90)	<0.01	2.78 (1.48 to 5.23)	<0.01
2018-NIH				
Grade 1	1.51 (0.73 to 2.95)	0.28	1.24 (0.61 to 2.40)	0.48
Grade 2	1.61 (0.86 to 2.88)	0.17	1.89 (1.07 to 3.28)	0.04
Grade 3	3.94 (1.57 to 9.92)	<0.01	4.21 (1.69 to 11.03)	<0.01
2019-Jensen				
Grade 1	1.52 (0.99 to 2.30)	0.08	1.37 (0.92 to 2.01)	0.18
Grade 2	2.76 (1.31 to 5.63)	0.01	2.15 (1.05 to 4.34)	0.04
Grade 3	30.4 (2.90 to 4107)	<0.01	10.2 (1.14 to 393)	0.04
5 years' corrected age				
2001-NIH				
Mild BPD	1.73 (1.11 to 2.69)	0.03	1.13 (0.67 to 1.88)	0.26
Moderate BPD	1.43 (0.80 to 2.52)	0.31	1.44 (0.75 to 2.64)	0.27
Severe BPD	2.76 (1.61 to 4.77)	<0.01	2.18 (1.22 to 3.82)	<0.01
2018-NIH				
Grade 1	1.71 (0.91 to 3.16)	0.12	1.21 (0.57 to 2.39)	0.28
Grade 2	1.87 (1.09 to 3.20)	0.04	1.77 (0.98 to 3.13)	0.12
Grade 3	2.78 (1.12 to 7.26)	0.04	2.85 (1.10 to 7.16)	0.04
2019-Jensen				
Grade 1	1.84 (1.28 to 2.65)	<0.01	1.34 (0.89 to 2.02)	0.28
Grade 2	1.83 (0.90 to 3.66)	0.12	2.09 (0.98 to 4.28)	0.11
Grade 3	7.31 (0.82 to 282)	0.12	2.42 (0.12 to 24.4)	0.32

Neurodevelopmental impairment at 2 years' corrected age is defined as having a composite cognitive score <85 (–1 SD), composite motor score <85 (–1 SD), cerebral palsy defined as Gross Motor Function Classification System ≥2, hearing loss despite amplification and/or visual impairment.
Neurodevelopmental impairment at 5 years' corrected age is defined as having a full-scale IQ <85 (–1 SD), scoring abnormal on the Movement Assessment Battery for Children-2 ≤5, complex minor neurological dysfunction or any grade of cerebral palsy, hearing loss despite amplification and/or severe visual impairment hampering daily activities.
Respiratory morbidity is defined as having hospitalisation count >3 visits, a visit to a pulmonary specialist, oxygen usage at home, use of bronchodilators, or antibiotics or diuretics at two consecutive follow-up visits.
BPD, bronchopulmonary dysplasia; NIH, National Institutes of Health.

definitions with the 2001-NIH definition as reference category using a two-sided t-test with a null hypothesis. For our secondary objective, we calculated the median (IQR) or mean (SD) of the perinatal characteristics and respiratory interventions per BPD severity, and created three different logistic regression models showing the association between these parameters as predictors and the BPD severity of the three definitions as outcome. A p value of <0.05 was considered statistically significant. Statistical analyses were performed with R statistical software (V.3.6.3 for Windows) and RStudio (integrated development for R, Boston, 2020; R studio desktop 1.3.1093, package mice & logistf).

RESULTS

Patient characteristics

During the study period, 777 infants were eligible. Of these infants, 162 infants (20.8%) died before 36 weeks' PMA and 31 infants (4%) were not included because of either congenital abnormalities, no parental approval for use of data or admission to the research centre after 24 hours of life (online supplemental figure 1). The remaining 584 infants were classified according to the three BPD definitions. Long-term outcomes were assessed in 513 infants (87.8%) and in 380 children (65.1%) at 2 and 5 years' CA, respectively.

The median GA of the study cohort was 28.1 weeks (IQR 26.7–29.0), the median birth weight was 1040 g (IQR 850–1240) and 55.3% were males (table 1). The cumulative incidence of any BPD classification was 38.9% for the 2001-NIH definition, 21.1% for the 2018-NIH definition and 38.9% for the 2019-Jensen definition.[12] The incidence of NDI or late death was 18.0% and 37.3% at 2 and 5 years' CA, respectively. Respiratory morbidity was present in 22.4% and 18.4% at 2 and 5 years' CA, respectively. More details of the composite outcome of NDI and respiratory morbidity have been described in a previous publication.[12] Differences in antenatal corticosteroids, caesarean section, persistent ductus arteriosus, surfactant and doxapram between infants who were lost to follow-up and those who were present at follow-up are described in online supplemental table 1.

Neurodevelopmental and respiratory outcomes at 2 years' CA

Logistic regression analysis of the 2-year data for both the 2001-NIH and the 2018-NIH definition showed that the infants with severe or grade 3 BPD, but not less severe forms of BPD, had a significantly increased risk of NDI or late death, compared with the no BPD group (table 2 and online supplemental figure 2). The 2019-Jensen definition showed that the infants classified as

Table 3 Brier score for neurodevelopmental impairment and respiratory morbidity at 2 and 5 years' corrected age

	Neurodevelopmental impairment (NDI)			Respiratory morbidity		
2 years' corrected age						
	Squared error Mean (SD)	Difference in mean square Mean (SD)	P value	Squared error Mean (SD)	Difference in mean square Mean (SD)	P value
2001-NIH	0.171 (0.207)	*	*	0.199 (0.172)	*	*
2018-NIH	0.171 (0.208)	−0.0005 (0.04)	0.60	0.198 (0.173)	0.0011 (0.04)	0.53
2019-Jensen	0.169 (0.209)	0.0018 (0.04)	0.27	0.199 (0.172)	−0.0005 (0.04)	0.56
5 years' corrected age						
	NDI			Respiratory morbidity		
2001-NIH	0.227 (0.120)	*	*	0.179 (0.167)	*	*
2018-NIH	0.230 (0.114)	−0.0024 (0.0591)	0.39	0.179 (0.167)	0.0002 (0.0323)	0.53
2019-Jensen	0.228 (0.116)	−0.0010 (0.0421)	0.62	0.179 (0.168)	−0.0005 (0.0304)	0.63

NDI at 2 years' corrected age is defined as having a composite cognitive score <85 (−1 SD), composite motor score <85 (−1 SD), cerebral palsy defined as Gross Motor Function Classification System ≥2, hearing loss despite amplification and/or visual impairment.
NDI at 5 years' corrected age is defined as having a full-scale IQ <85 (−1 SD), scoring abnormal on the Movement Assessment Battery for Children-2 ≤5, complex minor neurological dysfunction or any grade of cerebral palsy, hearing loss despite amplification and/or severe visual impairment hampering daily activities.
Respiratory morbidity is defined as having hospitalisation count >3 visits, a visit to a pulmonary specialist, oxygen usage at home, use of bronchodilators, or antibiotics or diuretics at two consecutive follow-up visits.
*Reference category.
NIH, National Institutes of Health.

grade 2 and grade 3 BPD had a significant increased risk of NDI or late death at 2 years' CA, compared with infants without BPD (table 2).

Regarding the respiratory morbidity at 2 years' CA, the analysis showed that only infants classified with severe BPD according to the 2001-NIH definition had a significant increased risk compared with infants without BPD (table 2 and online supplemental figure 2). The 2018-NIH and 2019-Jensen definition showed a significant risk of respiratory morbidity at 2 years' CA for grade 2 and grade 3 BPD, compared with those without a BPD diagnosis (table 2).

All three logistic regression models resulted in comparable Brier scores for NDI or late death (0.169–0.171) and respiratory morbidity (0.198–0.199) at 2 years' CA (table 3). Analyses showed limited Brier score differences (−0.0005 to 0.0018), resulting in no difference between the accuracy of the Firth's logistic regression models for NDI/late death and respiratory morbidity at 2 years' CA for 2018-NIH and the 2019-Jensen compared with the 2001-NIH definition (reference) (table 3).

Neurodevelopmental and respiratory outcomes at 5 years' CA

Analysis at 5 years' CA showed an increased risk of NDI/late death for infants with mild BPD and severe BPD following the 2001-NIH definition, compared with those without BPD (table 2 and online supplemental figure 3). The 2018-NIH definition showed an increased risk of NDI or late death at 5 years' CA for infants with grade 2 and grade 3 BPD. In contrast to the 2-year data, the 2019-Jensen definition showed an increased risk for infants with grade 1 BPD, but not for grades 2 and 3.

Regarding respiratory morbidity at 5 years' CA, infants with severe or grade 3 BPD following 2001-NIH and 2018-NIH definition showed an increased risk for this outcome, compared with infants without BPD diagnosis. The 2019-Jensen definition showed no increased risk of respiratory morbidity at this time point compared with their peers without BPD diagnosis (table 2 and online supplemental figure 3).

In line with the Brier score results at 2 years' CA, the three BPD definitions showed similar mediocre Brier scores at 5 years' CA for NDI or late death (0.227 to 0.230) and respiratory morbidity (0.179), and limited differences (−0.0024 to − 0.0002), all of them testing non-significant (table 3).

Perinatal characteristics and respiratory interventions

The median GA had no stepwise decrease with increasing BPD severity in any of the BPD definitions, whereas the median birth weight did show a slight decrease with increasing BPD severity (online supplemental table 2). Only the 2001-NIH and 2019-Jensen definition showed an increase in small for GA percentage with increasing BPD severity. The median days of supplemental oxygen and days of IMV increased with increasing BPD severity in all BPD definitions. Ordinal regression analysis showed that GA and birth weight were significant risk factors for all BPD severity categories for all BPD definitions, with some exceptions presumably due to lack of power. The analyses showed that the ORs of GA, birth weight and small for GA between severity categories were very similar and without any signs of an ordinal decrease with increasing BPD severity. This is in contrast to the median days on supplemental oxygen and IMV showing an ordinal stepwise increase as BPD severity increases (online supplemental table 2).

DISCUSSION

This is the first study investigating the ordinal relationship between three different severity-based BPD definitions and long-term outcomes and risk factors. In this large single-centre cohort, we showed that the 2001-NIH, the 2018-NIH and the 2019-Jensen definition have similar mediocre accuracy of the probability forecast (calibration) for neurological and respiratory outcomes at 2 and 5 years' CA. Likewise, perinatal characteristics and respiratory interventions have similar associations with the three BPD definitions. An ideal grade-based or severity-based definition predicts an adverse outcome with higher odds with every increasing grade or disease severity, and therefore several findings need to be discussed. First, the logistic regression analyses showed no clear ordinal stepwise risk increment with each more severe BPD category in the three separate BPD definitions for neurodevelopmental and respiratory outcomes.

Only the infants with the most severe grade of BPD in all three definitions had a significant increased risk of NDI or respiratory morbidity, compared with the infants with no BPD. However, there are three exceptions to this statement. First, infants with grade 2 BPD following the 2019-Jensen definition also showed a significant increased risk of the NDI and respiratory morbidity, and the 2018-NIH grade 2 BPD definition for 2-year respiratory morbidity. It is unknown what the clinical relevance of this finding is since the comparisons between the definitions using the Brier scores showed similar accuracy in the probability forecast (calibration) for these long-term outcomes. Second, at 5 years' CA, we found no significant association between the infants with grade 3 BPD following the 2019-Jensen definition compared with the infants without the BPD diagnosis. We speculate that this might be due to the lack of infants in our cohort (4%) with this severe grade of BPD, defined as needing IMV at 36 weeks' PMA. Finally, our analyses at 5 years show some conflicting results compared with the 2-year data (table 2). A possible reason might be that the current neurodevelopmental tests at 2 years' CA only detect the most severe cases of NDI, resulting in a strong association between BPD severity and NDI.[16] However, as previously shown, the motor and cognitive functions are increasingly being challenged ('growing into deficit') over time, so an NDI can become more apparent, resulting in a higher prevalence of NDI and less power to detect an association at 5 years' CA.[17] In our study, we employed Bayley-III scores for the classification of NDI. It is widely recognised that the Bayley-III may lead to underestimation of developmental delay. Consequently, we have opted to use a cut-off threshold of −1 SD instead of −2 SD to better capture and address these delays.

All three BPD definitions had comparable Brier scores for both long-term outcomes at both time points. After formal statistical testing, none of the definitions appeared to be superior in probability forecasting (calibration) of the long-term outcomes. But even more striking were the high Brier scores ranging from 0.169 to 0.230, which classify the definitions as non-informative and ill-calibrating.[18] The Brier score, like accuracy and precision, is an essential part in prediction research and shows the calibration performance of a prediction model. A possible explanation for this result might be that none of the definitions under investigation were derived using calibration, but only accuracy.[5 6] Our study showed that the additional value of the currently used classification of preterm infants into three severity categories of BPD to improve risk stratification for the outcomes NDI or respiratory morbidity might be limited. Risk stratification is important for long-term prediction to inform parents on the possible outcomes their children might encounter.

In the current study, we show that only the most severe forms of BPD have a significant increased risk of poor outcome, which is in line with another retrospective cohort study.[19] However, in contrast to our study, an ordinal increased risk association for respiratory morbidity was shown, which might be explained by the profound BPD incidence (96.8%) and the lower incidence of the long-term respiratory morbidity in that study.[20]

Our study also showed a lack of incremental association between severity classification of three BPD definitions and well-established perinatal risk factors for BPD, such as GA, birth weight and small for GA.[21] However, in line with other studies investigating the Jensen definition, an ordinal increased risk of BPD severity with increasing total days of supplemental oxygen and days of IMV was found.[9 22] Unfortunately, infants diagnosed with grade 1 and grade 2 BPD were combined into one category in the study by Jensen *et al*, hampering further risk stratification and comparison with our results.

Some limitations of the current study need to be discussed. First, retrospective data from a single-centre cohort were used. Further nationwide or global studies are needed to confirm our results. Second, respiratory morbidity was determined at follow-up during outpatient visits and therefore recall bias may have occurred. Ideally, respiratory morbidity needs to also be assessed using lung function testing.[23] Third, due to the COVID-19-related lockdown, some children were assessed by telephone interviews with their parents and therefore did not undergo neurodevelopmental testing at 5 years' CA, which led to a higher loss to follow-up percentage. To correct for these missing data, we performed multiple imputation analyses to improve statistical power and minimise selection bias. Finally, some baseline differences were found between the infants lost to follow-up and those included, which may indicate selection bias. However, the impact on the comparative results is probably limited as the same groups were used to compare the three BPD definitions.

Despite these limitations, our study shows that none of these three BPD definitions is superior in classifying infants into severity categories based on the associations with long-term outcomes until 5 years' CA. If deemed important to clinicians and researchers for BPD to be an ordinal severity-based definition, future initiatives need to focus on the calibration as well as the discriminative predictive performance of BPD definitions.

In conclusion, this historical cohort study with retrospective data collection shows that the 2001-NIH, the 2018-NIH and the 2019-Jensen definition show similar, but low accuracy of probability forecast for neurodevelopmental and respiratory outcomes at 2 and 5 years' CA. In addition, the three BPD defintions show similar associations with perinatal characteristics and respiratory interventions. These results need to be confirmed in a large multicentre setting and if supported, the current BPD classifications need to be updated.

Author affiliations
[1]Neonatology, Emma Children's Hospital UMC, Amsterdam, The Netherlands
[2]Amsterdam Reproduction and Development, Amsterdam, The Netherlands
[3]Amsterdam University Medical Centers, Amsterdam, The Netherlands
[4]University Medical Center Utrecht, Utrecht, The Netherlands
[5]Department of Pediatrics, Tergooi Medical Center, Hilversum, The Netherlands
[6]Emma Children's Hospital UMC, Amsterdam, The Netherlands
[7]Department of Pediatrics and Neonatology, North West Hospital Group, Alkmaar, The Netherlands
[8]Department of Pediatrics and Neonatology, OLVG, Amsterdam, The Netherlands
[9]Department of Pediatrics, OLVG, Amsterdam, The Netherlands
[10]Department of Pediatrics, Flevoziekenhuis, Almere, The Netherlands
[11]Department of Pediatrics, Zaans Medical Center, Zaandam, The Netherlands
[12]Department of Pediatrics, Spaarne Hospital Haarlem, Haarlem, The Netherlands
[13]Department of Pediatrics, Red Cross Hospital, Beverwijk, The Netherlands
[14]Department of Pediatrics, Amstelland Hospital, Amstelveen, The Netherlands
[15]Department of Pediatrics, Bovenij Hospital, Amsterdam, The Netherlands
[16]Psychosocial Department, Emma Children's Hospital UMC, Amsterdam, The Netherlands
[17]Department of Neonatology, Emma Children's Hospital UMC, Amsterdam, The Netherlands
[18]Amsterdam UMC Locatie AMC, Amsterdam, The Netherlands

Acknowledgements This article is part of the PhD thesis of the first author.

Contributors TAK contributed to the conception and design of the study, acquisition of data, analysis and interpretation of data, and drafting of the manuscript. SB, GJB, AAMWvK, HvL, CAML, MR, IAS, NCR, FV and EvS contributed to the acquisition of data and critical revision of the manuscript. CSHA-M and AGvW-L contributed to the acquisition of data, analysis and interpretation of data and critical revision of the manuscript. WO contributed to the conception and design of the study, acquisition of data, analysis and interpretation of data, and critical revision of the manuscript. WO serves as guarantee of this paper. All authors approved the final version of the manuscript to be published and agree to be accountable for all aspects of the work.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Consent obtained from parent(s)/guardian(s).

Ethics approval This study was approved by the Institutional Review Board at the Amsterdam UMC (IRB Amsterdam UMC W22_403 # 22.478).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iDs
Trixie A Katz http://orcid.org/0009-0004-7277-5785
Wes Onland http://orcid.org/0000-0002-9623-0606

REFERENCES

1 Gomez Pomar E, Concina VA, Samide A, *et al*. Bronchopulmonary dysplasia: comparison between the two most used diagnostic criteria. *Front Pediatr* 2018;6:397.
2 Northway WH, Rosan RC, Porter DY. Pulmonary disease following respirator therapy of hyaline-membrane disease. *N Engl J Med* 1967;276:357–68.
3 Jobe AH, Bancalari E. Bronchopulmonary dysplasia. *Am J Respir Crit Care Med* 2001;163:1723–9.
4 Higgins RD, Jobe AH, Koso-Thomas M, *et al*. Bronchopulmonary dysplasia: executive summary of a workshop. *J Pediatr* 2018;197:300–8.
5 Jensen EA, Dysart K, Gantz MG, *et al*. The diagnosis of bronchopulmonary dysplasia in very preterm infants. *Am J Respir Crit Care Med* 2019;200:751–9.
6 Ehrenkranz RA, Walsh MC, Vohr BR, *et al*. Validation of the National institutes of health consensus definition of bronchopulmonary dysplasia. *Pediatrics* 2005;116:1353–60.
7 Lai NM, Yap AQY, Ong HC, *et al*. Use of composite outcomes in neonatal trials: an analysis of the cochrane reviews. *Neonatology* 2021;118:259–63.
8 Katz TA, Vliegenthart RJS, Aarnoudse-Moens CSH, *et al*. Severity of bronchopulmonary dysplasia and neurodevelopmental outcome at 2 and 5 years corrected age. *J Pediatr* 2022;243:40–6.
9 Jensen EA, Edwards EM, Greenberg LT, *et al*. Severity of bronchopulmonary dysplasia among very preterm infants in the United States. *Pediatrics* 2021;148:e2020030007.
10 Katz TA, van Kaam AH, Schuit E, *et al*. Comparison of new bronchopulmonary dysplasia definitions on long-term outcomes in preterm infants. *J Pediatr* 2023;253:86–93.
11 van Rossem MC, van de Loo M, Laan BJ, *et al*. Accuracy of the diagnosis of bronchopulmonary dysplasia in a referral-based health care system. *J Pediatr* 2015;167:540–4.
12 Katz TA, van Kaam AH, Schuit E, *et al*. Comparison of new bronchopulmonary dysplasia definitions on long term outcomes in preterm infants. *J Pediatr* 2023;253:86–93.
13 White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011;30:377–99.
14 Firth D. Bias reduction of maximum likelihood estimates. *Biometrika* 1993;80:27–38.
15 Liao L, Mark DB. Clinical prediction models: are we building better mousetraps? *J Am Coll Cardiol* 2003;42:851–3.
16 Schuit E, Groenwold RHH, Harrell FE, *et al*. Unexpected predictor-outcome associations in clinical prediction research: causes and solutions. *CMAJ* 2013;185:E499–505.
17 Taylor GL, Joseph RM, Kuban KCK, *et al*. Changes in neurodevelopmental outcomes from age 2 to 10 years for children born extremely preterm. *Pediatrics* 2021;147:e2020001040.
18 Steyerberg EW, Vickers AJ, Cook NR, *et al*. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–38.
19 Jeon GW, Oh M, Chang YS. Definitions of bronchopulmonary dysplasia and long-term outcomes of extremely preterm infants in Korean neonatal network. *Sci Rep* 2021;11:24349.
20 Jeon GW, Oh M, Lee J, *et al*. Comparison of definitions of Bronchopulmonary dysplasia to reflect the long-term outcomes of extremely preterm infants. *Sci Rep* 2022;12:18095.
21 Stoll BJ, Hansen NI, Bell EF, *et al*. Neonatal outcomes of extremely preterm infants from the NICHD neonatal research network. *Pediatrics* 2010;126:443–56.
22 Kakoo Brioso E, Moscoso J, Malveiro D, *et al*. Bronchopulmonary dysplasia: a five-year retrospective cohort study on differences in clinical characteristics and morbidities according to severity grading. *Cureus* 2023;15:e42720.
23 Greenough A, Pahuja A. Updates on functional characterization of bronchopulmonary dysplasia - the contribution of lung function testing. *Front Med (Lausanne)* 2015;2:35.

Supplementary material

Our primary outcomes were the composite outcome of neurodevelopmental impairment (NDI) or late death (after 36 weeks PMA up till follow-up), and respiratory morbidity at two- and five years corrected age (CA). Neurodevelopmental outcome was assessed at two- and five years CA by a dedicated team of child psychologists, pediatricians and child physiotherapists according to the nationwide follow-up protocol. At two years CA, NDI was defined as having either a Composite Cognitive Score (CCS) and Composite Motor Score (CMS) <-1 standard deviation (SD) (score <85) of the Dutch version of the Bayley Scales of Infant and Toddler Development-III (BSID-III)(1), any grade of cerebral palsy(2, 3), hearing loss despite amplification or severe visual impairment hampering daily activities. At five-years CA, NDI was defined as having either Full Scale Intelligence Quotient (FSIQ) <-1 SD (score<85) of the Wechsler Preschool and Primary Scale of Intelligence-III (WPPSI-III-NL)(4), movement Assessment Battery for Children-2 (mABC) score <-2 SD (≤ 5)(5), complex minor neurological dysfunction (MND)(2) or any grade of CP(6), hearing loss despite amplification or severe visual impairment hampering daily activities. Respiratory morbidity assessed by asking the parents at the outpatient clinic visits by the attending pediatrician, and was defined as occurrence of at least one the following: (1) a hospitalization count ≥ 3 for respiratory and non-respiratory reasons between discharge and follow-up visits(7) , (2) visit to a respiratory specialist, (3) oxygen usage at home between discharge and follow-up visits (4) chronic use of bronchodilators, antibiotics or diuretics at two consecutive follow-up visits(7-9).

Since the Netherlands has a referral based health system, the diagnosis of BPD was determined by data of cumulative duration of oxygen days, and mode and level of

Katz TA, *et al. Arch Dis Child Fetal Neonatal Ed* 2024;0:1–6. doi: 10.1136/archdischild-2024-326931

respiratory support at 36 weeks PMA collected by the Amsterdam UMC as well as the regional level II hospital the baby was transferred to(10).

Data collected to investigate the second objective, namely the association between the severity-based BPD definitions, and known BPD risk factors were: GA, birth weight (BW), small for gestational age (SGA) (defined as a BW below the 10th percentile of Dutch reference curves(11)), total days of supplemental oxygen, and duration of invasive mechanical ventilation (IMV). An oxygen day was counted if supplemental oxygen use was more than 12 hours that day.

We collected the following predefined additional perinatal characteristics for each patient during their primary hospitalization: gender, multiple pregnancy, antenatal corticosteroids, mode of delivery, Apgar score at 5 minutes, patent ductus arteriosus, necrotizing enterocolitis (NEC) ≥ grade 2, culture proven sepsis and intraventricular hemorrhage (IVH) ≥ grade 3. Data on respiratory characteristics during initial hospitalization, such as surfactant, dexamethasone or doxapram usage, and length of NICU stay were also collected.

References

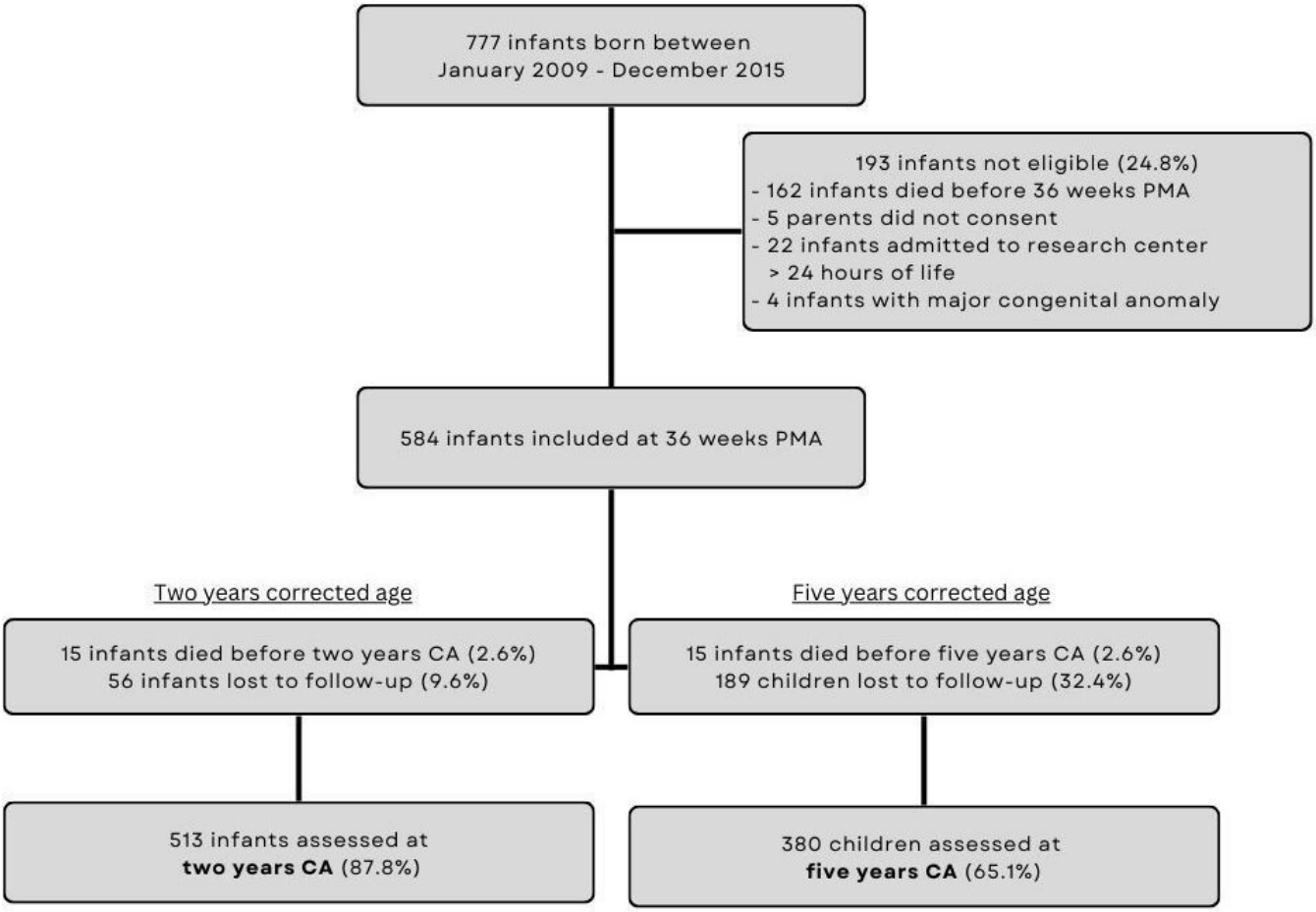
1. Bayley N. Bayley Scales of Infant Development (2nd ed.). 1993; Online.
2. Hadders-Algra M, Heineman KR, Bos AF, Middelburg KJ. The assessment of minor neurological dysfunction in infancy using the Touwen Infant Neurological Examination: strengths and limitations. *Dev Med Child Neurol*. 2010;**52**:87-92 doi: 10.1111/j.1469-8749.2009.03305.x [published Online First: 20090622].
3. Park EY. Stability of the gross motor function classification system in children with cerebral palsy for two years. *BMC Neurol*. 2020;**20**:172 doi: 10.1186/s12883-020-01721-4 [published Online First: 20200506].
4. Hendriksen JH P. WPPSI-III-NL Wechsler Preschool and Primary Scale of Intelligence, 3rd edition. Nederlandse bewerking. . Amsterdam, the Netherlands: Pearson Assessment and Information BV. ; 2009.
5. Brown T. Movement Assessment Battery for Children: 2nd Edition (MABC-2). In: Volkmar FR, ed. *Encyclopedia of Autism Spectrum Disorders*. New York, NY: Springer New York 2013:1925-39.
6. Wood E, Rosenbaum P. The gross motor function classification system for cerebral palsy: a study of reliability and stability over time. *Dev Med Child Neurol*. 2000;**42**:292-6 doi: 10.1017/s0012162200000529 [published Online.
7. Gage S, Kan P, Oehlert J, Gould JB, Stevenson DK, Shaw GM, O'Brodovich HM. Determinants of chronic lung disease severity in the first year of life; A population based study. *Pediatr Pulmonol*. 2015;**50**:878-88 doi: 10.1002/ppul.23148 [published Online First: 2015/02/06].
8. Keller RL, Feng R, DeMauro SB*, et al.* Bronchopulmonary Dysplasia and Perinatal Characteristics Predict 1-Year Respiratory Outcomes in Newborns Born at Extremely Low Gestational Age: A Prospective Cohort Study. *J Pediatr*. 2017;**187**:89-97 e3 doi: 10.1016/j.jpeds.2017.04.026 [published Online First: 2017/05/22].
9. Katz TA, van Kaam AH, Schuit E*, et al.* Comparison of new bronchopulmonary dysplasia definitions on long term outcomes in preterm infants. *J Pediatr*. 2022; doi: 10.1016/j.jpeds.2022.09.022 [published Online First: 2022/09/24].
10. van Rossem MC, van de Loo M, Laan BJ, de Sonnaville ES, Tamminga P, van Kaam AH, Onland W. Accuracy of the Diagnosis of Bronchopulmonary Dysplasia in a Referral-Based Health Care System. *J Pediatr*. 2015;**167**:540-4.e1 doi: 10.1016/j.jpeds.2015.05.008 [published Online First: 20150603].
11. Hoftiezer L, Hof MHP, Dijs-Elsinga J, Hogeveen M, Hukkelhoven C, van Lingen RA. From population reference to national standard: new and improved birthweight charts. *Am J Obstet Gynecol*. 2019;**220**:383 e1- e17 doi: 10.1016/j.ajog.2018.12.023 [published Online First: 2018/12/24].

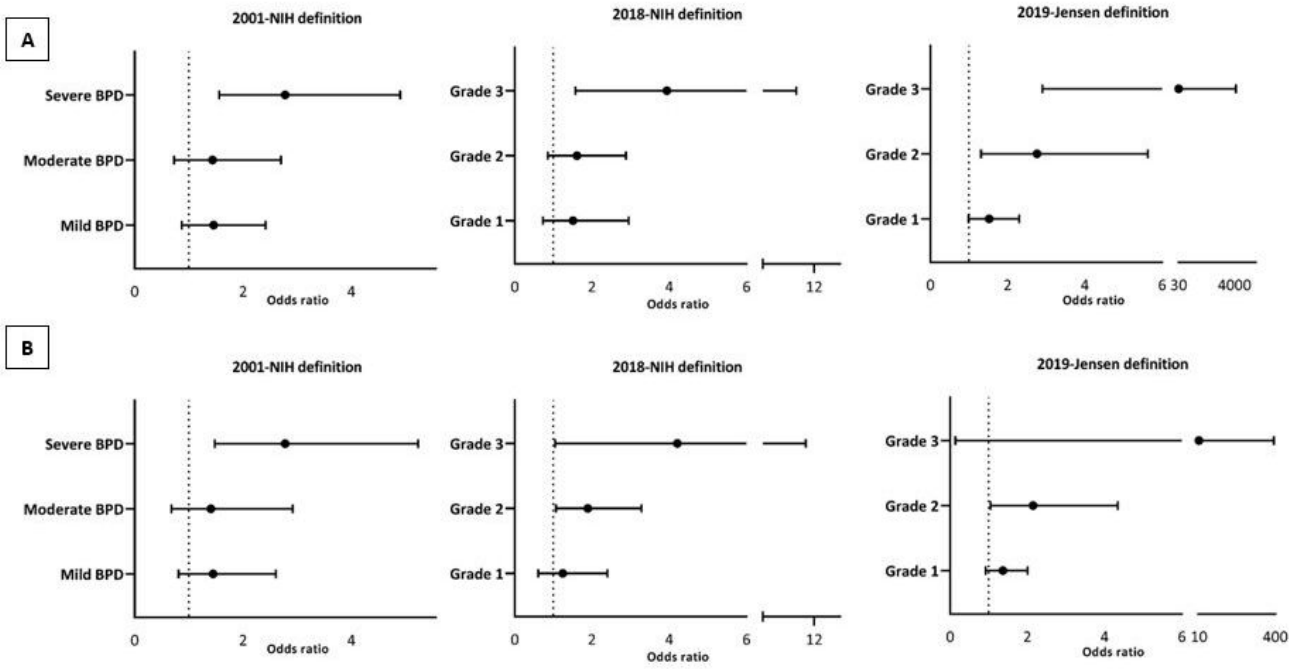
Supplementary Figures: Flowchart of included patients

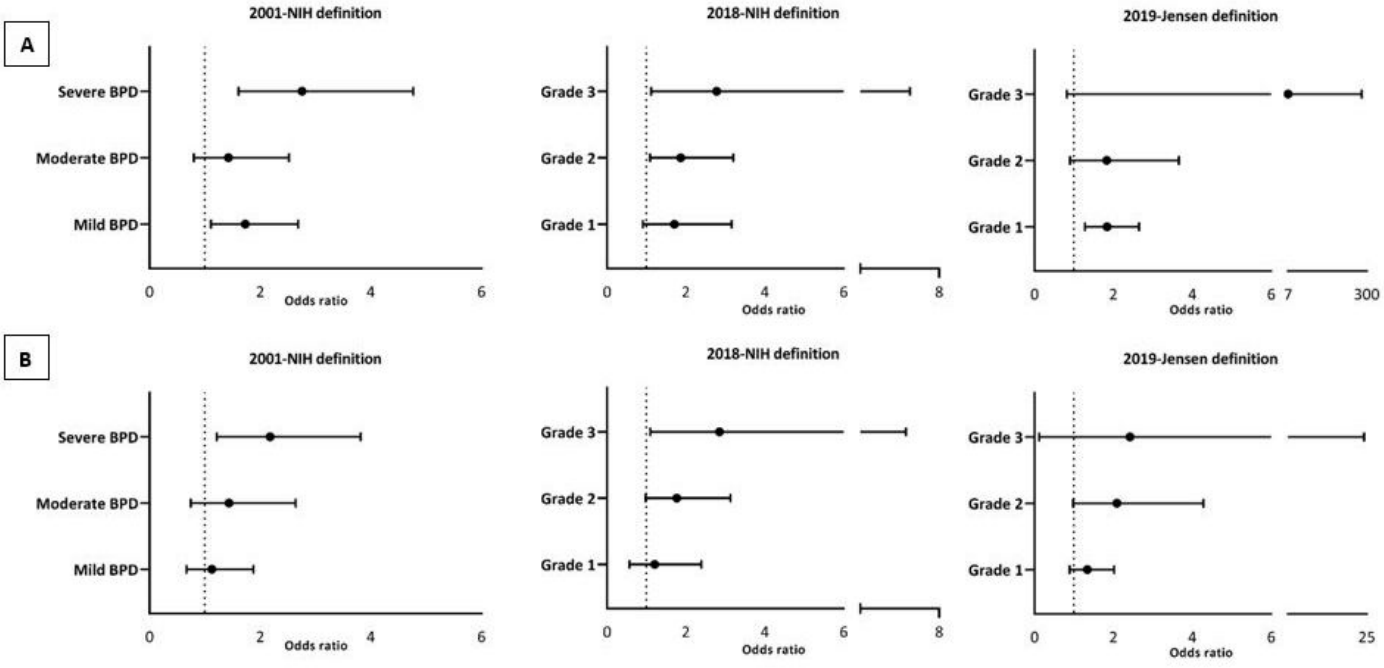
PMA: post menstrual age. *NIH:* National Institute of Health. *BPD:* bronchopulmonary dysplasia. *CA:* corrected age.

Supplementary Figures: Odds ratio plots for neurodevelopmental impairment (A) and respiratory morbidity (B) at two years corrected age per BPD definition

Supplementary Figures: Odds ratio plots for neurodevelopmental impairment (A) and respiratory morbidity (B) at five years CA per BPD definition


777 infants born between January 2009 - December 2015
193 infants not eligible (24.8%)
- 162 infants died before 36 weeks PMA
- 5 parents did not consent
- 22 infants admitted to research center > 24 hours of life
- 4 infants with major congenital anomaly
584 infants included at 36 weeks PMA
Two years corrected age
Five years corrected age
15 infants died before two years CA (2.6%)
56 infants lost to follow-up (9.6%)
15 infants died before five years CA (2.6%)
189 children lost to follow-up (32.4%)
513 infants assessed at **two years CA** (87.8%)
380 children assessed at **five years CA** (65.1%)


A
2001-NIH definition
Severe BPD
Moderate BPD
Mild BPD
0
2
4
Odds ratio
2018-NIH definition
Grade 3
Grade 2
Grade 1
0
2
4
6
12
Odds ratio
2019-Jensen definition
Grade 3
Grade 2
Grade 1
0
2
4
6 30
4000
Odds ratio
B
2001-NIH definition
Severe BPD
Moderate BPD
Mild BPD
0
2
4
Odds ratio
2018-NIH definition
Grade 3
Grade 2
Grade 1
0
2
4
6
12
Odds ratio
2019-Jensen definition
Grade 3
Grade 2
Grade 1
0
2
4
6 10
400
Odds ratio


A
2001-NIH definition
Severe BPD
Moderate BPD
Mild BPD
0
2
4
6
Odds ratio
2018-NIH definition
Grade 3
Grade 2
Grade 1
0
2
4
6
8
Odds ratio
2019-Jensen definition
Grade 3
Grade 2
Grade 1
0
2
4
6
7
300
Odds ratio
B
2001-NIH definition
Severe BPD
Moderate BPD
Mild BPD
0
2
4
6
Odds ratio
2018-NIH definition
Grade 3
Grade 2
Grade 1
0
2
4
6
8
Odds ratio
2019-Jensen definition
Grade 3
Grade 2
Grade 1
0
2
4
6
25
Odds ratio

Supplementary Table 1: Patient characteristics assessed at 36 weeks PMA compared to lost to follow up at two and five years corrected age

	Included at two years CA (n=584)	Lost to follow-up at two years CA (n=71)	p-value	Included at five years CA (n=380)	Lost to follow-up at five years CA (n=204)	p-value
Gestational age[a], weeks	28.1 (26.7, 29)	28 (26.8, 29)	0.90	28.1 (26.7, 29.1)	28.0 (26.9 – 29)	0.52
Birth weight[a], grams	1040 (850,1240)	940 (805, 1180)	0.01	1030 (870, 1250)	1040 (825, 1230)	0.47
Small for GA, n(%)	86 (14.7)	21 (24.1)	0.01	45 (11.8)	41 (20.1)	0.14
Male gender, n(%)	323 (55.3)	43 (49.4)	0.23	162 (42.6)	99 (48.5)	0.40
Singleton, n(%)	416 (71.2)	57 (67.9)	0.40	251 (66.1)	142 (69.6)	0.96
Antenatal steroids, n(%)	521 (89.2)	70 (80.5)	<0.01	321 (94.1)	200 (85.8)	0.04
Cesarean section, n(%)	269 (46.1)	41 (48.8%)	0.82	148 (38.9)	123 (52.4)	0.02
Apgar score at 5 min[a]	8 (7, 9)	8 (7,9)	0.67	8 (7,9)	8 (7,9)	0.70
PDA, n(%)	200 (34.2)	23 (27.4)	0.14	129 (33.9)	71 (34.8)	0.18
Sepsis, n(%)	205 (35.1)	28 (33.3)	0.68	132 (34.7)	73 (35.7)	0.19
NEC ≥ grade II, n(%)	50 (8.6)	15 (17.9)	<0.01	29 (7.6)	21 (10.3)	0.56
IVH ≥ grade II, n(%)	62 (10.6)	7 (8.3)	0.90	44 (11.6)	18 (8.8)	0.47
Surfactant, n(%)	245 (42.0)	31 (37.3)	0.32	155 (40.8)	90 (44.1)	0.30
Doxapram, n(%)	88 (15.1)	9 (10.3)	<0.01	55 (14.5)	33 (16.2)	0.04
Dexamethasone, n(%)	33 (5.7)	9 (10.8)	0.03	16 (4.2)	17 (8.3)	0.30
Days of supplemental oxygen[a]	3 (0, 35)	2 (0,29)	0.86	3 (0, 30)	4 (0, 39)	0.32
Duration of IMV[a], days	0.17 (0, 2.23)	0.67 (0, 3.71)	<0.01	0.40 (0.0, 2.07)	0 (0, 2.3)	0.23
Total duration of NICU stay[a], days	31 (17, 49)	32 (18, 51)	0.13	30 (18, 49)	31 (16, 53)	0.77

[a]Reported as median (interquartile range)

BPD: bronchopulmonary dysplasia. CA: corrected age. GA: gestational age. PDA: persistent ductus arteriosus needing therapy. NEC: necrotizing enterocolitis. IVH: intraventricular hemorrhage. IMV: invasive mechanical ventilation.

Table 4: Perinatal characteristics and respiratory interventions

		2001-NIH			*2018-NIH*			*2019-Jensen*		
	BPD Severity	*Median (IQR)*	*OR (95%CI)*	*p-value**	*Median (IQR)*	*OR (95%CI)*	*p-value**	*Median (IQR)*	*OR (95%CI)*	*p-value**
Gestational age	No BPD	28.6 (27.6, 29.3)	*	*	28.4 (27.0, 29.1)	*	*	28.6 (27.6, 29.3)	*	*
	Mild/Grade 1	27.6 (26.4, 28.7)	0.68 (0.58, 0.79)	<0.01	27.0 (25.7, 27.6)	0.65 (0.53-0.79)	<0.01	27.1 (26.1, 28.3)	0.61 (0.54-0.67)	<0.01
	Moderate/Grade 2	27.0 (25.9, 28.1)	0.55 (0.45-0.66)	<0.01	26.6 (25.7, 27.6)	0.56 (0.47-0.68)	<0.01	27.0 (25.3, 28.1)	0.50 (0.40-0.64)	<0.01
	Severe/Grade 3	27.0 (25.7, 28.0)	0.52 (0.43-0.63)	<0.01	27.0 (25.9, 28.2)	0.65 (0.43-0.87)	<0.01	28.3 (28.1, 29.4)	1.27 (0.46-3.50)	0.65
Birth weight[a]	No BPD	1135 (933, 1300)	*	*	1100 (890, 1290)	*	*	1135 (933, 1300)	*	*
	Mild/Grade 1	973 (809, 1176)	0.98 (0.97, 0.99)	<0.01	960 (754, 1080)	0.97 (0.96, 0.98)	<0.01	925 (780, 1105)	0.97 (0.96, 0.98)	<0.01
	Moderate/Grade 2	893 (771, 1042)	0.96 (0.95, 0.98)	<0.01	840 (751, 980)	0.96 (0.95, 0.98)	<0.01	840 (730, 1020)	0.96 (0.94, 0.98)	<0.01
	Severe/Grade 3	830 (730, 990)	0.96 (0.94, 0.97)	<0.01	800 (700, 940)	0.96 (0.93, 0.98)	<0.01	790 (410, 1140)	0.94 (0.88, 0.99)	0.03
Small for GA	No BPD	11.5%	*	*	12.6%	*	*	11.5%	*	*
	Mild/Grade 1	16.0%	1.55 (0.85, 2.86)	0.15	20.5%	1.76 (0.80, 3.85)	0.16	19.0%	1.86 (1.14, 3.03)	0.01
	Moderate/Grade 2	20.7%	2.01 (0.98, 4.11)	0.06	25.0%	2.28 (1.19, 4.35)	0.01	20.0%	1.93 (0.79, 4.70)	0.15
	Severe/Grade 3	25.4%	2.62 (1.36, 5.05)	<0.01	21.1%	1.82 (0.58, 5.70)	0.30	66.7%	15.4 (1.36, 174.7)	0.03
O2 days	No BPD	1 (0,5)	*	*	1 (0, 10)	*	*	1 (0,5)	*	*
	Mild/Grade 1	15 (3, 43)	1.06 (1.04, 1.07)	<0.01	44 (18, 50)	1.06 (1.04, 1.07)	<0.01	37 (5, 50)	1.07 (1.06, 1.08)	<0.01
	Moderate/Grade 2	45 (18, 50)	1.08 (1.07, 1.10)	<0.01	50 (36, 62)	1.07 (1.06, 1.09)	<0.01	60 (28, 99)	1.10 (1.08, 1.12)	<0.01
	Severe/Grade 3	60 (40, 76)	1.11 (1.09, 1.14)	<0.01	40 (40, 126)	1.10 (1.08, 1.12)	<0.01	140 (70, 150)	1.15 (1.10, 1.21)	<0.01
IMV days	No BPD	0 (0, 0.8)	*	*	0 (0, 1.2)	*	*	0 (0, 0.8)	*	*
	Mild/Grade 1	0.4 (0, 3,2)	1.15 (1.07, 1.23)	<0.01	2.0 (0, 1.3)	1.19 (1.12, 1.26)	<0.01	1.6 (0, 5.3)	1.21 (1.14, 1.29)	<0.01
	Moderate/Grade 2	2.1 (0, 7.3)	1.26 (1.18, 1.35)	<0.01	4.0 (0.6, 10.8)	1.24 (1.17, 1.31)	<0.01	10.8 (2.1, 18.9)	1.39 (1.29, 1.49)	<0.01
	Severe/Grade 3	8.3 (1.8, 13.8)	1.35 (1.26, 1.45)	<0.01	11.3 (3.1, 25.4)	1.30 (1.22, 1.38)	<0.01	30.6 (19.7, 72.6)	1.50 (1.36, 1.66)	<0.01

NIH: National Institute of Health. *OR:* odds ratio. *IQR:* interquartile range. *CI:* confidence interval. *GA:* gestational age. *O2:* supplemental oxygen. *IMV:* invasive mechanical ventilation. *No BPD is the reference category. [a]increments of 100-grams.