European Journal of Radiology 169 (2023) 111168

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Chest computed tomography in severe bronchopulmonary dysplasia: Comparing quantitative scoring methods

S. Fontijn [a,1], S.J.A. Balink [b,1], M. Bonte [b], E.R. Andrinopoulou [c,d], L. Duijts [b,e], A.A. Kroon [e], P. Ciet [b,f,g], M.W. Pijnenburg [b,*]

[a] Post-graduate School of Paediatrics, University of Modena and Reggio Emilia, Modena, Italy
[b] Erasmus MC - Sophia Children's Hospital, University Medical Centre Rotterdam, Department of Paediatrics, Division of Respiratory Medicine and Allergology, Rotterdam, the Netherlands
[c] Erasmus MC, University Medical Centre Rotterdam, Department of Biostatistics, Rotterdam, the Netherlands
[d] Erasmus MC, University Medical Centre Rotterdam, Department of Epidemiology, Rotterdam, the Netherlands
[e] Erasmus MC - Sophia Children's Hospital, University Medical Centre Rotterdam, Department of Paediatrics, Division of Neonatology, Rotterdam, the Netherlands
[f] Erasmus MC, University Medical Centre Rotterdam, Department of Radiology and Nuclear Medicine, Rotterdam, the Netherlands
[g] Policlinico Universitario, University of Cagliari, Cagliari, Italy

ARTICLE INFO

Keywords:
Chest-computed tomography scan
Child
Bronchopulmonary dysplasia
Quantitative scores
(semi-) automated
Bronchodilator use
Birth weight

ABSTRACT

Purpose: Bronchopulmonary dysplasia (BPD) is the most common complication of extreme preterm birth and structural lung abnormalities are frequently found in children with BPD. To quantify lung damage in BPD, three new Hounsfield units (HU) based chest-CT scoring methods were evaluated in terms of 1) intra- and inter-observer variability, 2) correlation with the validated Perth-Rotterdam-Annotated-Grid-Morphometric-Analysis (PRAGMA)-BPD score, and 3) correlation with clinical data.

Methods: Chest CT scans of children with severe BPD were performed at a median of 7 months corrected age. Hyper- and hypo-attenuated regions were quantified using PRAGMA-BPD and three new HU based scoring methods (automated, semi-automated, and manual). Intra- and inter-observer variability was measured using intraclass correlation coefficients (ICC) and Bland-Altman plots. The correlation between the 4 scoring methods and clinical data was assessed using Spearman rank correlation.

Results: Thirty-five patients (median gestational age 26.1 weeks) were included. Intra- and inter-observer variability was excellent for hyper- and hypo-attenuation regions for the manual HU method and PRAGMA-BPD (ICCs range 0.80–0.97). ICC values for the semi-automated HU method were poorer, in particular for the inter-observer variability of hypo- (0.22–0.71) and hyper-attenuation (-0.06–0.89). The manual HU method was highly correlated with PRAGMA-BPD score for both hyper- (ρs0.92, $p < 0.001$) and hypo-attenuation (ρs0.79, $p < 0.001$), while automated and semi-automated HU methods showed poor correlation for hypo- ($\rho s < 0.22$) and good correlation for hyper-attenuation (ρs0.72–0.74, $p < 0.001$). Several scores of hyperattenuation correlated with the use of inhaled bronchodilators in the first year of life; two hypoattenuation scores correlated with birth weight.

Conclusions: PRAGMA-BPD and the manual HU method have the best reproducibility for quantification of CT abnormalities in BPD.

Abbreviations: BPD, Bronchopulmonary Dysplasia; CF, Cystic Fibrosis; CI, Confidence Interval; CT, Computed Tomography; HU, Hounsefield Unit; ICC, Intraclass Correlation Coefficients; PRAGMA, Perth Rotterdam Annotated Grid Morphometric Analysis.

* Corresponding author at: Room SP 3454, Erasmus MC-Sophia Children's Hospital, University Medical Centre Rotterdam, Department of Paediatrics/Paediatric Respiratory Medicine and Allergology, PO Box 2060, 3000 CB Rotterdam, the Netherlands.
E-mail address: m.pijnenburg@erasmusmc.nl (M.W. Pijnenburg).

[1] Both first authors contributed equally.

https://doi.org/10.1016/j.ejrad.2023.111168
Received 1 August 2023; Received in revised form 19 October 2023; Accepted 21 October 2023
Available online 24 October 2023

1. Introduction

Bronchopulmonary dysplasia (BPD) is a complex condition characterized by an arrest in the development of the lung and pulmonary vascular systems [1]. It is a common complication of premature birth, affecting 60 % of infants born before 26 weeks of gestation [2]. BPD can result in lifelong pulmonary sequelae including respiratory symptoms, reduced lung function and an increased risk of pulmonary hypertension [3,4].

Chest Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) can detect structural abnormalities in the lung parenchyma and airways in up to 85 % of BPD patients, including hypo-attenuated and hyper-attenuated areas, subpleural opacities, and architectural distortion [3,5–7]. Early evaluation of these abnormalities can help to assess the severity of disease and predict which infants are at high risk of developing severe respiratory problems [8]. However, imaging does not necessarily change clinical management and therefore the European Respiratory Society Task Force on long-term management of BPD suggests that lung imaging should be only performed in subgroups of children with clinically severe BPD [9].

Chest CT is the most sensitive imaging technique for detecting structural lung abnormalities in patients with BPD, but there is currently no gold standard for quantifying these abnormalities [3,10]. Nine different semi-quantitative CT scoring methods in patients with BPD have been identified in a systematic review, but none have been validated and universally accepted [3]. Therefore, we adapted the Perth-Rotterdam Annotated Grid Morphometric Analysis (PRAGMA) scoring system, based upon a method developed and validated for cystic fibrosis, for BPD [11]. This method effectively quantifies BPD features, with a good correlation to clinical parameters, but is time consuming (up to 30 min per CT scan) and requires an experienced radiologist [12]. For clinical implementation, automated and validated methods should be used to assess CT scans in BPD patients. Therefore, we aimed to compare the manual PRAGMA-BPD scoring method with three new Hounsfield Units (HU) based scoring methods to quantify BPD structural abnormalities on chest CT in children with severe BPD. Objectives were to assess the intra- and inter-observer agreement of these HU based scoring methods and their reproducibility against the PRAGMA-BPD scoring method and to correlate CT-scores as assessed with the three new HU based scoring methods and PRAGMA-BPD with clinical data.

2. Materials and methods

2.1. Patients

Our hospital operates a long-term follow up program for children with severe BPD. As part of this program, children undergo a chest CT scan at approximately 6 months corrected age.

For this study, BPD was defined as need for supplemental oxygen for $\geq$ 28 days between birth and 36 weeks postmenstrual age. Severe BPD was defined as need for $\geq$ 30 % oxygen and/or positive pressure ventilation and/or nasal continuous positive airway pressure at 36 weeks postmenstrual age or discharge [13]. Patients were eligible for the study if they had severe BPD and a chest CT scan was performed between August 2016 and April 2019. We considered 30 CT scans sufficient for comparison of the scoring methods.

We obtained data from patient records on gestational age, birth weight, days on mechanical ventilation and on respiratory support, use of inhaled bronchodilators, inhaled corticosteroids or antibiotics, and hospital admissions between 0 and 6 months and between 6 and 12 months corrected age.

Parents of all infants provided written informed consent for using these data for research purposes. The medical ethical committee of Erasmus MC approved the study (MEC-2015–694).

2.2. CT image acquisition and scoring

Free breathing chest CT scans were acquired in supine position without sedation using a standardized protocol on SOMATOM Drive and SOMATOM Force scanners (Siemens, Erlangen, Germany). Acquisition parameters included: tube voltage: 90 kV / single collimation width: 0.6 / total collimation width: 57.6 / spiral pitch factor 3 (Force) and tube voltage: 100 kV/ single collimation width: 0.6 / total collimation width: 38.4 / spiral pitch factor 3 (Drive). Scans were reconstructed using a sharp reconstruction kernel, with slice thickness of 1.0 mm and slice increments of 0.8 mm (for SOMATOM Drive: I70, for SOMATOM Force: Bl57).

Scans were scored using three new HU based segmentation methods using Myrian software (version 2.6.0, Intrasense 2018, France, Montpellier): a fully automated, semi-automated and manual HU method, and with the PRAGMA-BPD scoring method [12].

The three new HU based scoring methods are explained in Table 1. Lung segmentation began with the selection of the whole lung and separation from bordering structures, namely chest wall and mediastinum. For the automated HU method, the segmentation was fully automated, whereas for semi-automated and manual HU methods the first step of automated segmentation was followed by manual editing with 3D tools. Manual editing enabled inclusion of those areas automatically excluded by the automated segmentation, such as the hilum (only for semi-automated method) and the peripheral hyper-attenuated areas (semi-automated and manual method). The central airways, namely trachea and main stem bronchi, were not included in the segmentation.

In the automated and semi-automated HU methods, hyper- and hypo-attenuation regions were quantified using adapted HU thresholds based on the method proposed by *Chassagnon et al* [14,15]. For hypo-attenuation areas, we used five adapted thresholds obtained by the difference between average attenuation value and a series of fixed values: $\leq$ Average −0 HU, −50 HU, −100 HU, −150 HU, and −200 HU. The average was the mean attenuation between six density measurements (expressed in HU) collected using an elliptical region of interest in different areas of the lung (three measurements on normal lung and three measurements on hypo-attenuated lung) [15]. For hyper-attenuation regions, we used nine adapted thresholds obtained by the sum between Mode or Mean Intensity (MI) and a fixed value or a standard deviation (SD) multiple: $\geq$ Mode + 500 HU, +400 HU, +300 HU, +1SD, +2SD, +3SD; $\geq$ MI + 1SD, +2SD, +3SD. The Mode, expressed as the average HU value between the most highly represented attenuation value, MI and SD were extrapolated from the histogram HU distribution of the segmented lung for each CT scan. In the manual HU method, diseased areas were manually selected using 3D tools, including the vessels (hilum and main intrapulmonary vessels).

A CT-density score (CT-DS) was calculated for each method, representing the ratio of hyper- or hypo-attenuated lung volume to total lung volume. For instance, a hyper-attenuation CT-DS of 2 meant that 2 % of the total lung volume was hyper-attenuated.

The PRAGMA-BPD scoring method consists of a visual analysis of CT scans, where abnormalities are scored as hypo-attenuation (mosaic perfusion, emphysema, and bullae), hyper-attenuation (linear and/or subpleural triangular opacities, septal thickening, consolidations, atelectasis), and bronchial wall thickening [3]. The PRAGMA-BPD score for each category is expressed as a percentage of the total lung volume or in millilitre (ml). In our study, we used a modified version of PRAGMA-BPD (Fig. 1) which included bronchial wall thickening in the hyper-attenuation areas, which was necessary to allow for comparison with the other scoring methods.

Scoring was performed by a certified paediatric resident (S.F.) who received standardized training and was blinded to clinical information. Intra-observer agreement was evaluated by scoring 20 scans 4–6 weeks apart, while inter-observer agreement was assessed by a second experienced observer (M.B.) with 5 years of experience in PRAGMA-BPD

Table 1
Scoring methods details (segmentation, pathology detection and scoring time).

	Slices analyzed	Sections analyzed	LS	Areas included in LS	Areas excluded from LS	Dection of patological areas	Vessles quantification	Average time (CI 95 %)
PRAGMA-BPD	10 each scan	Axial	Manual	Lung parenchyma Intrapulmonary bronchi Pulmonary vessels	Main airways Hilum	Manual	No	40.9′ (35.2–46.7)
Automated HU	All	Axial Coronal	Automated	Lung parenchyma Intrapulmonary bronchi Pulmonary vessels	Main airways Hilum Periferal hyper-attenuated areas	Hyperattenuation: automated (threshold method) Hypoattenuation: manual* and automated (threshold method)	No	11.1′ (10.3–11.9)
Semi Automated HU	All	Axial Coronal	Automated and manual	Lung parenchyma Intrapulmonary bronchi Pulmonary vessels Hilum	Main airways	Hyperattenuation: automated (threshold method) Hypoattenuation: manual* and automated (threshold method)	No	18.4′ (15.2–21.6)
Manual HU	All	Axial Coronal	Automated and manual	Lung parenchyma Intrapulmonary bronchi Pulmonary vessels	Main airwaysHilum (quantified as vessles)	Manual	Yes	33.3′ (29–37.6)

LS Lung segmentation. *Thresholds are obtained by the manual selection of areas (six in total) with normal and low density and the subsequent calculation of media between their HU values.

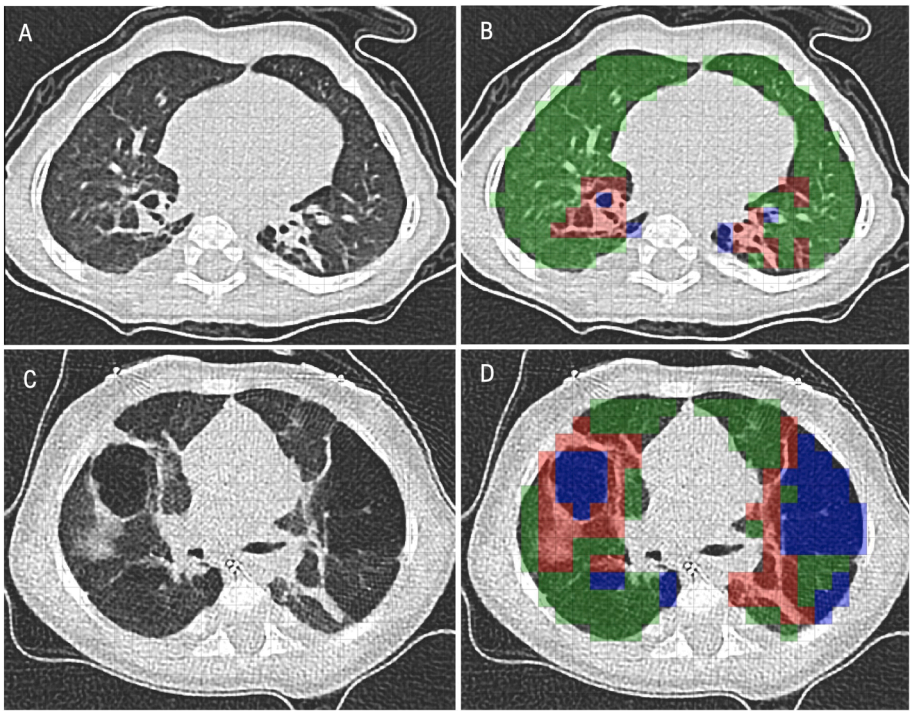

Fig. 1. Example of PRAGMA-BPD scoring. (A, B) Slices from two CT scans with overlaid grid. (C, D) Same slices annotated with PRAGMA-BPD: green = normal lung tissue, blue = hypo-attenuated areas, red = hyper-attenuated areas.

scoring. Time needed for each scoring method was also measured.

2.3. Statistical analysis

Continuous data are presented as median values (interquartile range), and categorical data are presented as counts (percentages). Intra-observer and inter-observer reproducibility were assessed using the Intraclass correlation (ICC) and Bland-Altman plots. ICC between 0.4 and 0.6 was considered moderate, between 0.6 and 0.8 good and above 0.8 excellent [16]. The Spearman rank correlation (ρ_s) was used to compare the new scoring methods to PRAGMA-BPD scores and clinical data. Bonferroni's correction was used to correct for multiple testing. All analyses were performed using Stata version 14.1 and SPSS version 15.0.

3. Results

We included 35 patients with severe BPD. Characteristics of the patients are reported in Table 2. The median gestational age was 26.1 weeks (IQR 25.4 – 27.3) and median birth weight was 820 g (IQR 700–940 g). Chest CT scans were performed at a mean age corrected for

Table 2
Patient characteristics.

Gestational age (weeks)	26.1 (IQR 25.2 – 27.3)
Birth weight (g)	800 (IQR 667 – 930)
Sex, n (%)	20
Male	(54 %)17
Female	(46 %)
Mechanical ventilation (days)	18.0 (IQR 6.0 – 24.5)
Respiratory support (days)	135 (IQR 87–250)
Corrected age at CT scan (months)	7.3 (SD 2.9)
Admissions first year of life	8/34
Inhaled bronchodilators 0–6 months corrected age	6/34
Inhaled corticosteroids 0–6 months corrected age	2/34
Maintenance antibiotics 0–6 months corrected age	1/34
Inhaled bronchodilators 6–12 months corrected age	12/34
Inhaled corticosteroids 6–12 months corrected age	3/34
Maintenance antibiotics 6–12 months corrected age	5/34

the number of weeks children were born preterm of 7.4 months (SD 3.0).

3.1. Agreement of scoring methods (Table 3)

Automated, semi-automated and manual HU scores: Intra- and inter-observer repeatability was not assessed for the automated HU method as we consider this should be 1.0. For the semi-automated HU method, the intra-observer repeatability was excellent for volume of hyper-attenuation (ICC ranging from 0.93 to 0.99) and poor for hypo-attenuation regions which varied between 0.37 and 0.85 depending on the threshold used. The highest agreements were found for the threshold Average −150 and −200 HU.

The inter-observer repeatability of the semi-automated HU method varied between poor and good for both volume of hypo-attenuation (from 0.22 to 0.71), with the highest agreement for the threshold Average −200 and the lowest agreement for the threshold Average, and hyper-attenuation volume (from −0.06 to 0.89) with the highest agreement for the threshold Mean Intensity + 1SD.

For the manual HU method, good to excellent intra- and inter-observer repeatability was shown in both the volumes for hypo-attenuation (ICC respectively 0.80 (range 0.51–0.92) and 0.91 (0.78–0.96)) and hyper-attenuation (ICC respectively 0.97 (0.93–0.98) and 0.80 (0.50–0.92)).

Bland-Altman plots showing agreement are shown in Fig. 2.

PRAGMA-BPD: The intra- and inter-observer repeatability for the PRAGMA-BPD scores were excellent. The intra-observer repeatability was 0.95 for both volumes of hyper- and hypo-attenuation (range respectively 0.88–0.98 and 0.89–0.98); the inter-observer variability was 0.83 (0.57–0.93) for hypo- and 0.93 (0.84–0.97) for hyper-attenuation.

3.2. Scoring time

The average time needed for scoring one scan was 33.3 min (95 % CI 29–37.6) for the manual HU method and 40.9 min (95 % CI 35.2–46.7) for the PRAGMA-BPD method. The scoring was faster with semi-automated and automated HU methods (18.4 (95 % CI 15.2–21.6) and 11.1 min (95 % CI 10.3–11.9), respectively).

3.3. Correlation of new HU based scoring methods with PRAGMA-BPD

The HU based method that correlated best with PRAGMA-BPD was the manual HU method both for the hypo-attenuation (ρ_s 0.79, $p < 0.001$) and for the hyper-attenuation (ρ_s 0.92, $p < 0.001$) scores. The automated and semi-automated HU methods had a poor correlation with PRAGMA-BPD for hypo-attenuation scores (ρ_s 0.12 and 0.22, p 0.508 and 0.208). For the hyper-attenuation scores, the correlation with PRAGMA-BPD was strongest with threshold Mode + 300 for the automated method (ρ_s 0.72, $p < 0.001$) and with threshold Mode + 2SD for the semi-automated method (ρ_s 0.74, $p < 0.001$). All methods had an excellent agreement for total lung volume quantification (see Table 4). Fig. 3 is an example of PRAGMA scoring versus manual HU scoring; Fig. 4 gives an example of automated HU versus manual HU scoring.

3.4. Correlation between new HU based scores and PRAGMA-BPD scores with clinical data

There were significant correlations between several hyper-attenuation scores and use of bronchodilators between 0 and 6, and between 6 and 12 months corrected age (ρ_s ranging from 0.48 to 0.59, $p = 0.004$ to < 0.001) (Table 5. 5). We also found a negative correlation between birthweight and hypo-attenuation CT scores with the semi-automated and automated methods at −200 HU (ρ_s −0.54, $p < 0.001$ for both methods).

4. Discussion

We compared three new HU based CT scoring methods with PRAGMA-BPD to assess lung structural abnormalities in infants with severe BPD. We found strong correlations between the manual HU scoring method and PRAGMA-BPD, while the automated and semi-automated HU scoring systems showed poor correlation. Significant correlations were found between several hyperattenuation scores and

Table 3
Scoring methods repeatability.

	PRAGMA-BPD ICC (95 % CI)		Semi Automated HU ICC (95 % CI)		Manual HU ICC (95 % CI)	
	Intraobserver	*Interobserver*	*Intraobserver*	*Interobserver*	*Intraobserver*	*Interobserver*
Hypoattenuation	**0.95 (0.88–0.98)**	**0.83 (0.57–0.93)**			**0.80 (0.51–0.92)**	**0.91 (0.78–0.96)**
Average			**0.37 (-0.58–0.75)**	**0.22 (-0.98–0.68)**		
Average −50			**0.80 (0.50–0.92)**	**0.51 (-0.23–0.80)**		
Average −100			**0.83 (0.58–0.93)**	**0.58 (-0.06–0.83)**		
Average −150			**0.85 (0.63–0.94)**	**0.65 (0.12–0.86)**		
Average −200			**0.85 (0.61–0.94)**	**0.71 (0.27–0.89)**		
Hyperattenuation	**0.95 (0.89–0.98)**	**0.93 (0.84–0.97)**			**0.97 (0.93–0.98)**	**0.80 (0.50–0.92)**
Mode + 500			**0.99 (0.97–0.99)**	**0.63 (0.06–0.85)**		
Mode + 400			**0.98 (0.96–0.99)**	**0.52 (-0.20–0.81)**		
Mode + 300			**0.98 (0.95–0.99)**	**0.33 (-0.67–0.74)**		
Mode + 1SD			**0.96 (0.90–0.98)**	**−0.06 (-1.69–0.58)**		
Mode + 2SD			**0.95 (0.88–0.98)**	**0.36 (-0.60–0.75)**		
Mode + 3SD			**0.98 (0.96–0.99)**	**0.73 (0.33–0.89)**		
Mean Intensity + 1SD			**0.96 (0.89–0.98)**	**0.89 (0.72–0.96)**		
Mean Intensity + 2SD			**0.97 (0.93–0.99)**	**0.37 (-0.58–0.75)**		
Mean Intensity + 3SD			**0.93 (0.82–0.97)**	**0.77 (0.42–0.91)**		

ICC: intraclass correlation coefficient, CI: confidence interval.

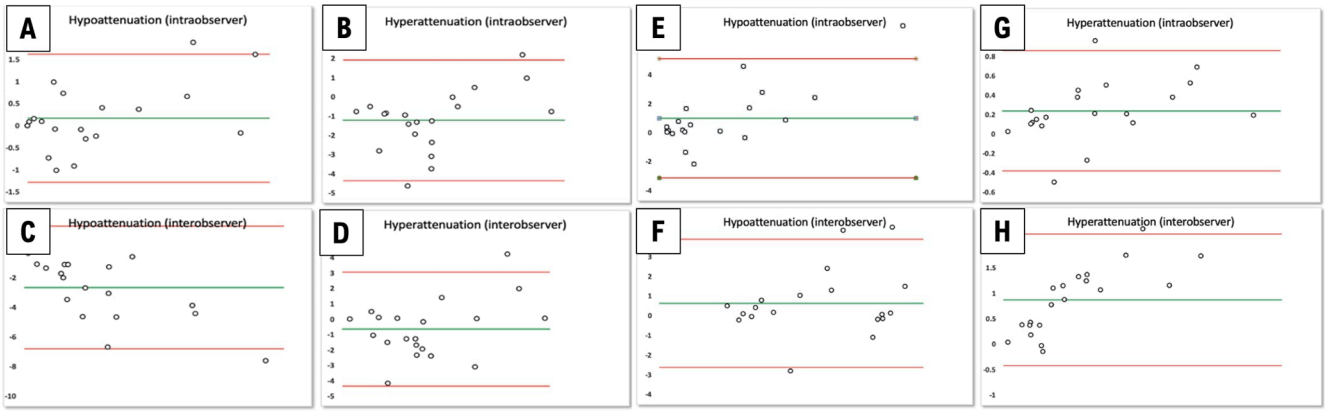

Fig. 2. Bland-Altman Plots. The figure shows PRAGMA-BPD intra-observer agreement for hypo-attenuation (A) and for hyper-attenuation (B), PRAGMA-BPD inter-observer agreement for hypo-attenuation (C) and for hyper-attenuation (D), manual HU method intra-observer agreement for hypo-attenuation (E) and for hyper-attenuation (F) and manual HU method inter-observer agreement for hypo-attenuation (G) and for hyper-attenuation (H). X-axis = average of the two scores of the same scan, Y-axis = bias difference between the two scores of the same scan.

Table 4
Relationships between PRAGMA-BPD score and other scoring methods.

	Automated HU ρ_s (P)	Semi Automated HU ρ_s (P)	Manual HU ρ_s (P)
Hypoattenuation			0.80 (<0.001)
Average	0.09 (0.584)	0.11 (0.521)	
Average −50	0.17 (0.311)	0.16 (0.354)	
Average −100	0.16 (0.326)	0.17 (0.327)	
Average −150	0.17 (0.312)	0.16 (0.337)	
Average −200	0.07 (0.670)	0.07 (0.672)	
Hyperattenuation			0.92 (<0.001)
Mode + 500	0.42 (0.009)	0.63 (<0.001)	
Mode + 400	0.57 (<0.001)	0.70 (<0.001)	
Mode + 300	0.72 (<0.001)	0.72 (<0.001)	
Mode + 1SD	0.37 (0.025)	0.43 (0.007)	
Mode + 2SD	0.42 (0.009)	0.72 (<0.001)	
Mode + 3SD	0.06 (0.719)	0.20 (0.217)	
Mean Intensity + 1SD	0.31 (0.064)	0.51 (0.001)	
Mean Intensity + 2SD	0.26 (0.113)	0.52 (<0.001)	
Mean Intensity + 3SD	−0.10 (0.550)	−0.28 (0.082)	
Total Lung Volume	0,98	0,98	0,98

Automated HU, Semi Automated HU and Manual HU scores are presented as Spearman ρs (P value).

the use of inhaled bronchodilators in the first year of life and between hypoattenuation scores at −200 HU and birth weight. PRAGMA-BPD and the manual HU scoring method required nearly the same scoring time, which was considerably longer than the semi-automated and automated HU methods.

4.1. Comparison with previous studies

Currently, there are no universally accepted scoring systems for the quantitative evaluation of lung abnormalities in children with BPD [3,6,7,10,17–20]. PRAGMA-BPD has been validated against clinical parameters in 49 children and showed significant association with gestational age and weight at 6 months corrected age, but its long scoring time hinders clinical implementation [12]. Automation with Artificial Intelligence (AI) and machine learning could overcome this limitation in the future.

Semi quantitative scoring systems for BPD are also time consuming and not routinely used in clinical practice [7,10,19,20]. To date, only one study in BPD patients assessed high and low attenuation regions on CT with a quantitative analysis and showed that children with BPD had a greater amount of these pathological findings than controls [7]. This difference was found with both a manual HU and automated HU method, although in the latter case the difference between the two groups was significant only for hyper-attenuation regions and not for hypo-attenuation. This is in line with our data where scores for hyper-attenuation as assessed with automated and semi-automated HU methods correlated better with manual HU methods than scores for hypo-attenuation. This suggests that hyper-attenuated regions are more easily recognized by (semi-)automated methods.

The advantage of automatic quantification is the faster scoring time and the operator independency. Nevertheless, this automated HU method also has several limitations. First, when analysing hyper-attenuation regions, the detection of these areas is hampered by the presence of non-pathological hyper-dense elements, such as vessels. However, the differences in volume of the pulmonary vessels between patients probably has little influence on the differences in volume of hyper-attenuated regions of the lung parenchyma. Second, there is a risk of underestimation of the hyper-attenuated areas, which happens when a hyper-attenuation area reaches the peripheral lung and is not recognized as lung parenchyma by the automated segmentation tool. Third, it is difficult to score the hypo-attenuation regions as these may vary based on different inflation levels during CT scanning in young children during spontaneous breathing and related to age [21,22]. Age and lung volume have a strong influence on the CT lung density of young children, which is higher and more variable than in older ages [23,24]. Therefore, it is difficult to choose a single fixed threshold for hypo-attenuated areas for children of different ages. To solve this problem, we decided to use adaptive thresholds, instead of fixed thresholds, which were manually obtained by measuring the density of three low-density areas of the whole lung. Unfortunately this resulted in poor reproducibility with high variability within and between observers. A way to overcome this problem could be to increase the number of measured areas of hypo-attenuation.

Despite these limitations, the automated method still has great potential, being fast and not requiring training of observers. Among the several thresholds tested in our study, Mode + 300 (for automated method) and Mode + 2SD (for semi-automated method) showed the best results for hyper-attenuation regions. At present, especially for hypo-attenuation, the HU threshold analysis is disturbed by the noise level of our CT images and by breathing artefacts, which can affect image quality and makes the automated HU method still too imprecise compared to the manual one. Improving the HU-based scoring method through machine learning and improving image quality with deep learning denoising filters could overcome these limitations.

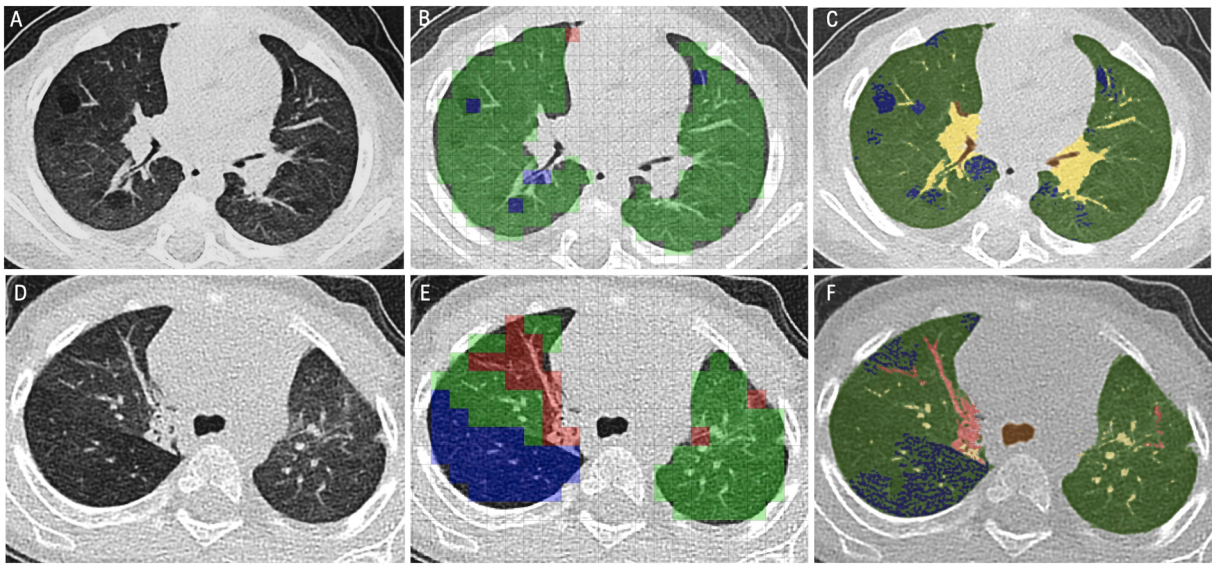

Fig. 3. Comparison between PRAGMA-BPD and manual HU method. (A, D) Slices from two computed tomography scans. (B, E) Slices annotated with PRAGMA-BPD: green = normal lung tissue, blue = hypo-attenuated areas, red = hyper-attenuated areas. (C, F) Slices annotated with manual HU method: green = normal lung tissue, blue = hypo-attenuated areas, red = hyper-attenuated areas, yellow = hilum and main intrapulmonary vessels.

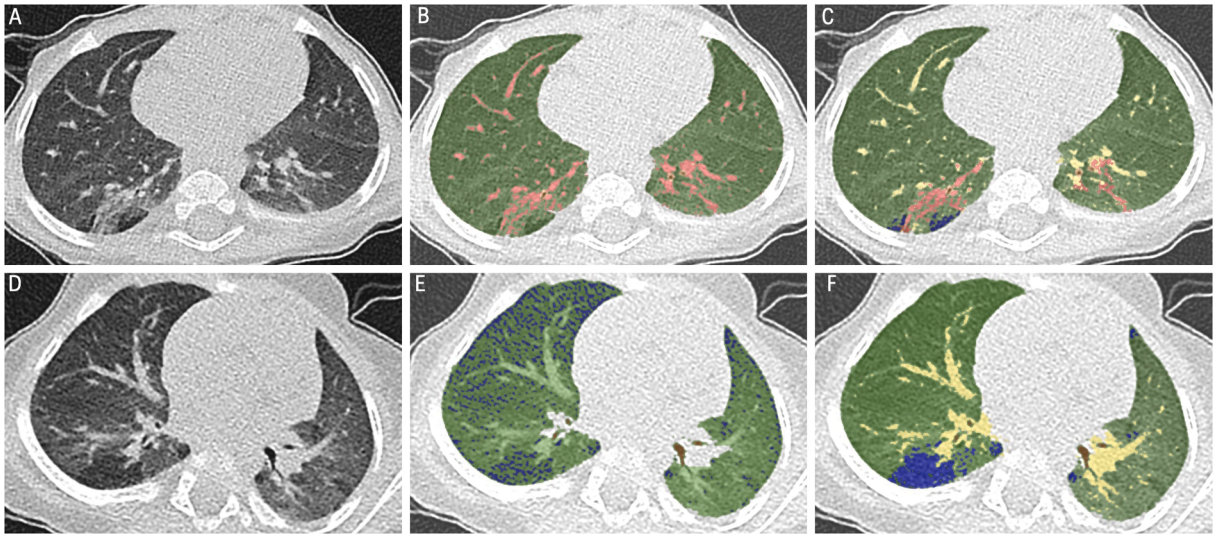

Fig. 4. Comparison between automated HU and manual HU methods. (A, D) Slices from two computed tomography scans. (B) Slice A annotated with automated HU method using threshold Mode + 2SD. (C) Slice A annotated with manual HU method. (E) Slice D annotated with automated HU method using threshold Average −50. There is a peripheral noise that compromises an accurate detection of low-attenuated regions. (F) Slice D annotated with manual HU method. green = normal lung tissue, blue = hypo-attenuated areas, red = hyper-attenuated areas, yellow (manual HU method) = hilum and main intrapulmonary vessels.

In our study, several hyperattenuation scores correlated with the use of inhaled bronchodilators during the first year of life. Although we do not have lung function data in our infants with BPD, the use of inhaled bronchodilators suggests that the infants had signs of broncho-obstruction such as wheezing episodes. In contrast, in a study in 163 preterm children (99 with BPD), obstruction was associated with hypoattenuated areas on inspiratory chest CT scans [5]. Also in older studies the forced expiratory volume in 1 s (FEV_1) correlated with hypoattenuated instead of hyperattenuated regions [3]. This discrepancy may be explained by the fact that we obtained free breathing CT scans instead of in- and expiratory scans, children were much younger and had no lung function data available and the indication for bronchodilators was not clear in all children. Only two scores for hypo-attenuation showed a negative correlation with birth weight, which is in line with earlier data from our cohort where we found a negative correlation with gestational age and with data from a study in adults where lower birth weight correlated with more emphysema [12,25].

Finally, it has become increasingly clear that the resolution of today's CT scanners is relatively poor to detect structural relevant changes in the first years of life. Introducing recently developed innovations like photon counting CT is likely to substantially improve resolution and the ability to pick up clinically relevant structural changes while reducing radiation dose [26]. In addition, over the years radiation dose has come down considerably with more advanced CT-scanners and with the use of artificial intelligence this radiation dose can be reduced further without losing diagnostic accuracy. At the same time, advancements in (neonatal) MRI scanning also improved the potential to quantify lung disease in patients with BPD without radiation, to phenotype disease, and to predict future outcomes [10,27]. Recently, a quantitative scoring system for BPD abnormalities on MRI successfully identified moderate

Table 5
Significant correlations between radiological parameters and clinical parameters (spearman rho and p-values).

	Birth weight	Bronchodilator use 0–6 months corrected age	Bronchodilator use 6–12 months corrected age	Antibiotic use 6–12 months corrected age
Hypoattenuation				
Automated, average −200	−0.54 ($p < 0.001$)			
Semi-automated, average −200	−0.54 ($p < 0.001$)			
Hyperattenuation				
Automated, mode + 2SD		0.57 ($p < 0.001$)		
Semi-automated, mode + 500		0.49 ($p = 0.003$)	0.50 ($p = 0.003$)	
Semi-automated, mode + 400		0.54 ($p = 0.001$)	0.50 ($p = 0.002$)	
Semi-automated, mode + 300		0.52 ($p = 0.002$)	0.49 ($p = 0.003$)	
Semi-automated, mi* +2SD		0.58 ($p < 0.001$)	0.48 ($p = 0.004$)	0.50 ($p = 0.003$)
Manual		0.59 ($p < 0.001$)		
PRAGMA		0.58 ($p < 0.001$)		

*mi = mean intensity.

and severe disease with good interobserver reproducibility [10].

4.2. Limitations

Our study has a few limitations to consider. First, we did not compare our scoring methods to others reported in literature [3,7,19]. Second, we lacked longitudinal CT data to better understand the scoring methods' sensitivity to detect disease progression. Third, during the inclusion period, we used two different CT scanners, which may have influenced the agreement between the scoring methods. Lastly, we did not have long-term prospective clinical data, such as lung function and quality of life. In the future, our long-term follow up program will increase our sample size, allow us to monitor children for a longer period and examine the association of lung structural abnormalities and extensive clinical data. This will make it possible to improve our scoring methods and to assess the predictive value of CT scans in children with BPD.

5. Conclusions

We aimed to develop a new, accurate, time efficient fully automated HU scoring method for CT scans in children with BPD. Although the HU-based automated and semi-automated methods improved scoring speed, they showed lower consistency in scoring, especially for areas with low density, and weak correlation with the previously validated PRAGMA-BPD scoring method. Therefore, these methods are not currently considered reliable enough for quantifying structural lung abnormalities in children with BPD. In the future, we aim to enhance these scoring methods using AI and machine learning, and validate their effectiveness by comparing CT results with clinical outcomes and patient prognosis.

CRediT authorship contribution statement

S. Fontijn: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **S.J.A. Balink:** . **M. Bonte:** Writing – review & editing, Resources, Methodology, Data curation. **E.R. Andrinopoulou:** . **L. Duijts:** Writing – review & editing, Formal analysis. **A.A. Kroon:** . **P. Ciet:** Writing – review & editing, Supervision, Software, Methodology, Formal analysis, Conceptualization. **M.W. Pijnenburg:** .

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] B. Thébaud, K.N. Goss, M. Laughon, J.A. Whitsett, S.H. Abman, R.H. Steinhorn, J. L. Aschner, P.G. Davis, S.A. McGrath-Morrow, R.F. Soll, A.H. Jobe, Bronchopulmonary Dysplasia, Nat. Rev. Dis Primers. 5 (1) (2019) 78.

[2] B.J. Stoll, N.I. Hansen, E.F. Bell, M.C. Walsh, W.A. Carlo, S. Shankaran, A. R. Laptook, P.J. Sánchez, K.P. Van Meurs, M. Wyckoff, A. Das, E.C. Hale, M.B. Ball, N.S. Newman, K. Schibler, B.B. Poindexter, K.A. Kennedy, C.M. Cotten, K. L. Watterberg, C.T. D'Angio, S.B. DeMauro, W.E. Truog, U. Devaskar, R.D. Higgins, Trends in care practices, morbidity, and mortality of extremely preterm neonates, J. Am. Med. Assoc. 314 (10) (2015) 1039–1051.

[3] E. van Mastrigt, K. Logie, P. Ciet, I.K. Reiss, L. Duijts, M.W. Pijnenburg, H. A. Tiddens, Lung CT imaging in patients with bronchopulmonary dysplasia: a systematic review, Pediatr Pulmonol. 51 (9) (2016 Sep) 975–986.

[4] S.J. Kotecha, J.T.D. Gibbons, C.W. Course, E.E. Evans, S.J. Simpson, W.J. Watkins, S. Kotecha, Geographical differences and temporal improvements in forced expiratory volume in 1 second of preterm-born children: a systematic review and meta-analysis, J. Am. Med. Assoc. Pediatr. 176 (9) (2022) 867–877.

[5] S.J. Simpson, K.M. Logie, C.A. O'Dea, G.L. Banton, C. Murray, A.C. Wilson, J. J. Pillow, G.L. Hall, Altered lung structure and function in mid-childhood survivors of very preterm birth, Thorax 72 (8) (2017) 702–711.

[6] T.J. Sung, S.M. Hwang, M.Y. Kim, S.G. Park, K.Y. Choi, Relationship between clinical severity of "new" bronchopulmonary dysplasia and HRCT abnormalities in VLBW infants, Pediatr. Pulmonol. 53 (10) (2018) 1391–1398.

[7] D.R. Spielberg, L.L. Walkup, J.M. Stein, E.J. Crotty, M.S. Rattan, M.M. Hossain, A. S. Brody, J.C. Woods, Quantitative CT scans of lung parenchymal pathology in premature infants ages 0–6 years, Pediatr. Pulmonol. 53 (3) (2018) 316–323.

[8] N.S. Higano, A.J. Bates, C.C. Gunatilaka, E.B. Hysinger, P.J. Critser, R. Hirsch, J. C. Woods, R.J. Fleck, Bronchopulmonary dysplasia from chest radiographs to magnetic resonance imaging and computed tomography: adding value, Pediatr. Radiol. 52 (4) (2022) 643–660.

[9] L. Duijts, E.R. van Meel, L. Moschino, E. Baraldi, M. Barnhoorn, W.M. Bramer, C. E. Bolton, J. Boyd, F. Buchvald, M.J. Del Cerro, A.A. Colin, R. Ersu, A. Greenough, C. Gremmen, T. Halvorsen, J. Kamphuis, S. Kotecha, K. Rooney-Otero, S. Schulzke, A. Wilson, D. Rigau, R.L. Morgan, T. Tonia, C.C. Roehr, M.W. Pijnenburg, European Respiratory Society guideline on long-term management of children with bronchopulmonary dysplasia, Eur. Respir. J. 55 (1) (2020) 1900788.

[10] K. Förster, H. Marchi, S. Stöcklein, O. Dietrich, H. Ehrhardt, M.O. Wielpütz, A. W. Flemmer, B. Schubert, M.A. Mall, B. Ertl-Wagner, A. Hilgendorff, Magnetic resonance imaging-based scoring of the diseased lung in the preterm infant with bronchopulmonary dysplasia: UNiforme Scoring of the disEAsed Lung in BPD (UNSEAL BPD), Am. J. Physiol. Lung Cell Mol. Physiol. 324 (2) (2023 Feb 1) L114–L122.

[11] T. Rosenow, M.C. Oudraad, C.P. Murray, L. Turkovic, W. Kuo, M. de Bruijne, S.C. Ranganathan, H.A. Tiddens, S.M. Stick, Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST CF). PRAGMA-CF. A Quantitative Structural Lung Disease Computed Tomography Outcome in Young Children with Cystic Fibrosis, Am. J. Respir. Crit. Care Med. 191(10) (2015) 1158-65.

[12] E. van Mastrigt, E. Kakar, P. Ciet, H.T. den Dekker, K.F. Joosten, P. Kalkman, R. Swarte, A.A. Kroon, H.A.W.M. Tiddens, J.C. de Jongste, I. Reiss, L. Duijts, M. W. Pijnenburg, Structural and functional ventilatory impairment in infants with severe bronchopulmonary dysplasia, Pediatr. Pulmonol. 52 (8) (2017) 1029–1037.

[13] A.H. Jobe, E. Bancalari, Bronchopulmonary dysplasia, Am. J. Respir. Crit. Care Med. 163 (2001) 1723–1729.

[14] G. Chassagnon, C. Martin, P.R. Burgel, D. Hubert, I. Fajac, N. Paragios, E. I. Zacharaki, P. Legmann, J. Coste, M.P. Revel, An automated computed tomography score for the cystic fibrosis lung, Eur. Radiol. 28 (12) (2018) 5111–5120.

[15] T.N. Hoang-Thi, M.P. Revel, P.R. Burgel, L. Bassinet, I. Honoré, T. Hua-Huy, C. Martin, B. Maitre, G. Chassagnon, Automated computed tomographic scoring of lung disease in adults with primary ciliary dyskinesia, BMC Pulm. Med. 18 (1) (2018) 194.

[16] J.M. Bland, D.G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, Lancet 1 (1986) 307–310.

[17] K. Vanhaverbeke, A. Van Eyck, K. Van Hoorenbeeck, B. De Winter, A. Snoeckx, T. Mulder, S. Verhulst, Lung imaging in bronchopulmonary dysplasia: a systematic review, Respir. Med. 171 (2020), 106101.

[18] E. Ronkainen, M. Perhomaa, L. Mattila, M. Hallman, T. Dunder, Structural pulmonary abnormalities still evident in schoolchildren with new bronchopulmonary dysplasia, Neonatology 113 (2) (2018) 122–130.

[19] M. Ochiai, S. Hikino, H. Yabuuchi, H. Nakayama, K. Sato, S. Ohga, T.A. Hara, New scoring system for computed tomography of the chest for assessing the clinical status of bronchopulmonary dysplasia, J. Pediatr. 152 (2008) 90–95.

[20] R.R. De Mello, M.V. Dutra, J.R. Ramos, P. Daltro, M. Boechat, J.M. de Andrade Lopes, Lung mechanics and high-resolution computed tomography of the chest in very low birth weight premature infants, Sao Paulo Med. J. 121 (2003) 167–172.
[21] P.J. Robinson, L. Kreel, Pulmonary tissue attenuation with computed tomography: comparison of inspiration and expiration scans, J. Comput. Assist. Tomogr. 3 (6) (1979) 740–748.
[22] R. Yuan, J.R. Mayo, J.C. Hogg, P.D. Paré, A.M. McWilliams, S. Lam, H.O. Coxson, The effects of radiation dose and CT manufacturer on measurements of lung densitometry, Chest 132 (2) (2007) 617–623.
[23] E.E. Sarria, R. Mattiello, L. Rao, C.J. Tiller, B. Poindexter, K.E. Applegate, J. Granroth-Cook, C. Denski, J. Nguyen, Z. Yu, E. Hoffman, R.S. Tepper, Quantitative assessment of chronic lung disease of infancy using computed tomography, Eur. Respir. J. 39 (4) (2012) 992–999.
[24] Y.S. Tsai, Y.S. Liu, Y.H. Shih, M.T. Chuang, Y.J. Lin, C.H. Lin, Y.C. Lin, Lung density standard deviations obtained using high-pitch dual-source computed tomography are valid predictors of bronchopulmonary dysplasia in preterm infants, Clin Imaging. 40 (4) (2016) 594–600.
[25] P.M. Wong, A.N. Lees, J. Louw, F.Y. Lee, N. French, K. Gain, C.P. Murray, A. Wilson, D.C. Chambers, Emphysema in young adult survivors of moderate-to-severe bronchopulmonary dysplasia, Eur Respir J. 32 (2) (2008) 321–328.
[26] M.J. Willemink, M. Persson, A. Pourmorteza, N.J. Pelc, D. Fleischmann, Photon-counting CT: technical principles and clinical prospects, Radiology 289 (2) (2018) 293–312.
[27] L.L. Walkup, J.A. Tkach, N.S. Higano, R.P. Thomen, S.B. Fain, S.L. Merhar, R. J. Fleck, R.S. Amin, J.C. Woods, Quantitative magnetic resonance imaging of bronchopulmonary dysplasia in the neonatal intensive care unit environment, Am. J. Respir. Crit. Care Med. 192 (10) (2015) 1215–1222.