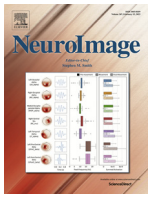
NeuroImage 267 (2023) 119815

Contents lists available at ScienceDirect

NeuroImage

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



Early brain morphometrics from neonatal MRI predict motor and cognitive outcomes at 2-years corrected age in very preterm infants

Alex M. Pagnozzi[a,*], Liza van Eijk[a,b], Kerstin Pannek[a], Roslyn N. Boyd[c], Susmita Saha[a], Joanne George[c,d], Samudragupta Bora[e], DanaKai Bradford[a], Michael Fahey[f], Michael Ditchfield[g,h], Atul Malhotra[f,i], Helen Liley[e], Paul B. Colditz[j], Stephen Rose[a], Jurgen Fripp[a]

[a] CSIRO Health and Biosecurity, The Australian e-Health Research Centre, Royal Brisbane and Women's Hospital, Herston, Brisbane, QLD 4029, Australia
[b] Department of Psychology, James Cook University, Townsville, Queensland, Australia
[c] Child Health Research Centre, Queensland Cerebral Palsy and Rehabilitation Research Centre, Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia
[d] Physiotherapy Department, Queensland Children's Hospital, Children's Health Queensland Hospital and Health Service, Brisbane, Australia
[e] Mothers, Babies and Women's Health Program, Mater Research Institute, Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia
[f] Monash Health Paediatric Neurology Unit and Department of Paediatrics, School of Clinical Sciences, Monash University, Clayton, Victoria, Australia
[g] Monash Imaging, Monash Health, Melbourne, Victoria, Australia
[h] Department of Medicine, Monash University, Melbourne, Victoria, Australia
[i] Monash Newborn, Monash Children's Hospital, Melbourne, Victoria, Australia
[j] Perinatal Research Centre, Faculty of Medicine, The University of Queensland Centre for Clinical Research, The University of Queensland, Brisbane, Queensland, Australia

ARTICLE INFO

Keywords:
Preterm birth
Structural MRI
Biomarkers
Neurodevelopment
Motor
Cognition
Language
Neonates

ABSTRACT

Infants born very preterm face a range of neurodevelopmental challenges in cognitive, language, behavioural and/or motor domains. Early accurate identification of those at risk of adverse neurodevelopmental outcomes, through clinical assessment and Magnetic Resonance Imaging (MRI), enables prognostication of outcomes and the initiation of targeted early interventions. This study utilises a prospective cohort of 181 infants born <31 weeks gestation, who had 3T MRIs acquired at 29-35 weeks postmenstrual age and a comprehensive neurodevelopmental evaluation at 2 years corrected age (CA). Cognitive, language and motor outcomes were assessed using the Bayley Scales of Infant and Toddler Development – Third Edition and functional motor outcomes using the Neuro-sensory Motor Developmental Assessment. By leveraging advanced structural MRI pre-processing steps to standardise the data, and the state-of-the-art developing Human Connectome Pipeline, early MRI biomarkers of neurodevelopmental outcomes were identified. Using Least Absolute Shrinkage and Selection Operator (LASSO) regression, significant associations between brain structure on early MRIs with 2-year outcomes were obtained (r = 0.51 and 0.48 for motor and cognitive outcomes respectively) on an independent 25% of the data. Additionally, important brain biomarkers from early MRIs were identified, including cortical grey matter volumes, as well as cortical thickness and sulcal depth across the entire cortex. Adverse outcome on the Bayley-III motor and cognitive composite scores were accurately predicted, with an Area Under the Curve of 0.86 for both scores. These associations between 2-year outcomes and patient prognosis and early neonatal MRI measures demonstrate the utility of imaging prior to term equivalent age for providing earlier commencement of targeted interventions for infants born preterm.

Abbreviations: AUC, Area Under the Curve; CA, Corrected Age; CP, Cerebral Palsy; CPAP, Continuous Positive Airway Pressure; CSF, Cerebrospinal Fluid; CT, Cortical Thickness; dHCP, Developing Human Connectome Project; ETT, Endotracheal Tube; GA, Gestational Age; GI, Gyrification Index; GM, Grey Matter; GMA, General Movements Assessment; LASSO, Least Absolute Shrinkage and Selection Operator; MAE, Mean Absolute Error; MRI, Magnetic Resonance Imaging; NSMDA, Neuro-Sensory Motor Developmental Assessment; PMA, Postmenstrual Age; PPREMO, Prediction of PREterm Motor Outcomes; PREBO, Prediction of Preterm Brain Outcomes; SA, Surface Area; SD, Sulcal Depth; SES, Socioeconomic Status; TEA, Term Equivalent Age; TPN, Total Parenteral Nutrition; WM, White Matter.

* Corresponding author.
E-mail address: Alex.Pagnozzi@csiro.au (A.M. Pagnozzi).

https://doi.org/10.1016/j.neuroimage.2022.119815.
Received 6 July 2022; Received in revised form 5 December 2022; Accepted 13 December 2022
Available online 15 December 2022.

1. Introduction

The current global prevalence of preterm birth is estimated at around 10% (Chawanpaiboon et al., 2019). Very preterm infants (gestational age [GA] at birth < 32 weeks) are at greater risk of cognitive, language, behavioural and motor problems. From the end of the second and beginning of the third trimesters, cortical neurogenesis, migration and folding are rapidly occurring up until Term Equivalent Age (TEA) (Dubois et al., 2008). This development may be disrupted by the types of injury that can ensue from premature birth, potentially resulting in altered white and grey matter volumes, destructive lesions in the white matter, neuronal and axonal injury involving the cerebral cortex and deep grey matter, as well as axonal degeneration, hypomyelination, abnormal apoptosis and diminution of late migrating neurons (Volpe, 2009).

Brain magnetic resonance imaging (MRI) is improving the ability to detect evidence of brain injury, with MRI at TEA playing an important role in the identification of infants at risk of adverse neurodevelopmental outcomes (George et al., 2018; van't Hooft et al., 2015). Increasingly, brain MRI is technically feasible at earlier time points, such as 30-32 weeks postmenstrual age (PMA) (Counsell et al., 2003; George et al., 2015; Ibrahim et al., 2018) and from 26 weeks PMA (Makropoulos et al., 2018), potentially allowing an even earlier prognostication, opening a new window for therapeutic interventions at a time of rapid brain development while the infant is still receiving care in a Neonatal Intensive Care Unit (NICU). Many very preterm infants are discharged prior to TEA, so MRI before TEA may be a more practical time-point at which imaging biomarkers can be measured to assist clinical assessment (Malhotra et al., 2017). Earlier imaging can help to ensure infants who are at greatest risk of adverse outcomes are identified and receive appropriate surveillance and early interventions (Morgan et al., 2021).

Volumetric information (from structural MRI) and microstructural information (from diffusion MRI) provide insights into brain development and maturation, and in turn are associated with neurodevelopmental outcomes at 2 years and beyond (Anderson et al., 2015; De Bruïne et al., 2013; Thompson et al., 2014), however this has largely been based on MRI acquired at TEA. Using the state-of-the-art developing Human Connectome Pipeline (dHCP) pipeline (Dubois et al., 2021), Kline et al. (2020) found cortical curvature and surface area to be associated with cognitive and language scores at 2 years, and the volume of the thalamus in addition to surface area and cortical gyrification to be associated with motor outcomes at 2 years (Kline et al., 2020) in a very preterm cohort scanned at TEA. Qualitative findings from earlier MRI before TEA have been shown to be associated with abnormal neurodevelopmental outcome (Miller et al., 2005). Quantitative findings including tissue volumes and cortical morphology have been associated with motor and cognitive outcomes at 2-3 years (Moeskops et al., 2017). As yet, quantitative biomarkers from Early MRI (29-35 weeks PMA) have been relatively under-investigated compared to imaging at TEA, despite potentially providing earlier detection for infants at risk of impairments and reducing burden on families as imaging can be performed while the infant is still in the NICU.

Our team has previously used our large unique prospective Prediction of PREterm Motor Outcomes (PPREMO, n = 121) (George et al., 2015) and Prediction of Preterm Brain Outcomes (PREBO, n = 150) cohorts to demonstrate identification of infants at risk of adverse motor outcomes and CP at 2 years corrected age (CA) using *semi-quantitative* Early and TEA MRI measures (George et al., 2021), based on the existing manual scoring system (Kidokoro et al., 2013) modified for infants born very preterm. These studies indicate that the brain changes responsible for these 2-year outcomes may already be present as early as 29-35 weeks PMA (Early time point). In this study we aim to investigate whether heterogeneity in *quantitative* neonatal brain measures such as anatomical volumes and cortical morphometrics from 3T MRI acquired at this early time point are associated with cognitive, language, and motor outcomes at 2 years CA.

2. Methods

2.1. Participants

This study analyzed data from two longitudinal prospective cohort studies of very preterm infants: the Prediction of PREterm Motor Outcomes (PPREMO) study (George et al., 2015) and the Preterm Brain Outcomes (PREBO) study. Both the PPREMO and PREBO cohorts recruited infants born at <31 weeks' gestation with no congenital or chromosomal abnormality whose parents/caregivers were English-speaking and lived within a 200km radius of the recruiting hospital. Infants were recruited at the Royal Brisbane and Women's Hospital (RBWH) and Monash Children's Hospital (MCH), Mater Mothers Hospital (MMH) and Queensland Children's Hospital (QCH); PPREMO between February 2013 and February 2016 (RBWH only), and PREBO between February 2016 and December 2019 (RBWH, MCH, MMH, QCH).

For PPREMO and PREBO cohorts, ethics approval was obtained from the Royal Brisbane and Women's Hospital Human Research Ethics Committee (HREC/12/QRBW/245), the Queensland Children's Hospital (HREC/15/QRCH/7), The University of Queensland (2012001060, 2015000290) and Monash University/Monash Children's Hospital (SSA/15/MonH/41). Both cohort studies were registered with the Australian New Zealand Clinical Trials Registry (PPREMO: ACTRN12613000280707; PREBO: ACTRN12615000591550). For both cohorts, informed written parental/caregiver consent was obtained for each infant.

2.2. Image acquisition

Infants were scanned at 29-35 weeks PMA during natural sleep on a 3T MRI scanner (Siemens Tim Trio [PPREMO], Siemens Skyra [PREBO RBWH and QLD Children's Hospital for MMH infants], Erlangen, Germany; Philips Ingenia [PREBO MCH]) (total n = 271) using an MR compatible incubator with a dedicated 8-channel neonatal head coil (Lammers LMT, Lübeck, Germany) [PPREMO, PREBO RBWH and QCH], or paediatric 32-channel head coil with no incubator (Phillips, Amsterdam, Netherlands) [PREBO Monash]. Infants were placed on an immobilisation pillow to minimise movement. Noise from the MRI scanner was attenuated using mini muffs (Natus Medical Inc., San Carlos, CA). All infants were monitored with pulse oximetry and electrocardiographic monitoring. No sedation or anaesthesia was used. The PPREMO study used multi-echo T2-weighted turbo spin-echo (TSE) volumes acquired in the axial plane with the following parameters at RBWH: TR/TE1/TE2/TE3 10,580/27/122/189ms; flip angle 150°, field of view 144 × 180mm, matrix 204 × 256, in-plane resolution 0.7 × 0.7mm, slice thickness 2mm, scan time 5:40 min. In the later PREBO study at RBWH and QCH, we acquired three orthogonal T2-weighted images (in axial, coronal, and sagittal plane) to improve image quality in relation to motion, with the following parameters: at RBWH T2 HASTE TR/TE 2,280/117ms; flip angle 120°, field of view 160 × 206mm, matrix 252 × 198mm, voxel size 0.8 × 0.8mm, slice thickness 1.8mm, scan time 3:11 min each, and at Monash: T2 TSE TR/TE 2,280/117ms; flip angle 120°, field of view 160 × 200mm, matrix 256 × 256, in-plane resolution 0.80 × 0.80mm, slice thickness 1.8mm, scan time 2:05 min each.

2.3. Image processing

2.3.1. Super-resolution reconstruction

To enable accurate cortical reconstruction and volumetry, the three T2 HASTE or TSE images (PREBO cohort only) were reconstructed into one super-resolution 3D image with resolution matching the in-plane resolution (0.8mm^3 isotropic [PREBO RBWH, QCH and Monash]), as illustrated in Fig. 1. For each individual HASTE image, the brain was extracted using FSL's 'bet' tool (Smith, 2002). Approximately 50% of the automatically generated brain masks were manually corrected to

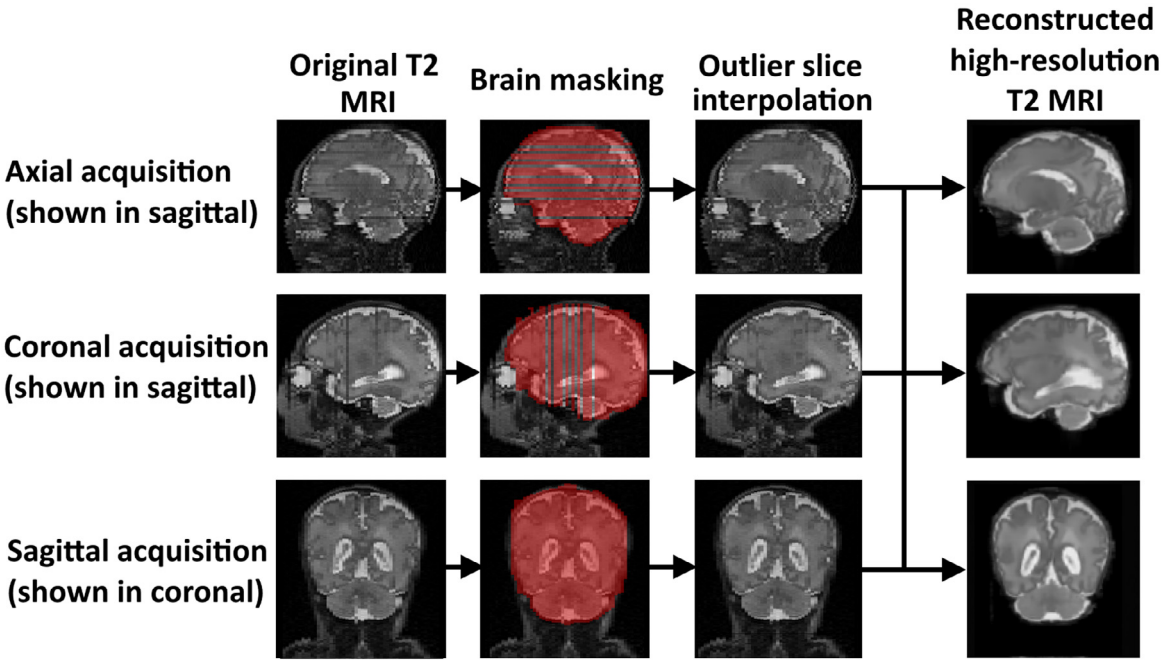


Fig. 1. Original T2 MRI (first column) for one participant (age 33 weeks PMA) from the PREBO study, illustrating the thick-slice acquisitions in the axial, coronal and sagittal planes in the respective rows. Each acquisition underwent brain masking (shown in red in the second column), with slices containing artefact removed. These slices are replaced in the third column, with interpolated data used to reconstruct the final high-resolution T2 MRI.

prevent the mask extending into the orbital cavity and below the cerebellum, requiring about 15 min per mask to correct. Slices impacted by subject motion were detected and replaced using an in-house automated structural MRI Python workflow. In this workflow, a slice was detected as an outlier if its mean signal intensity was less than 92% of the mean intensity of the adjacent two slices (previous and next). This threshold was determined empirically from a subset (20%) of the data. In cases where one of the adjacent slices had a substantially lower intensity (80% of the other adjacent slice), the mean slice intensity was only compared to the slice with the higher intensity. In cases where detected outlier slices were adjacent to two normal slices (unimpacted by motion artefact), they were then replaced by interpolated anatomy derived as the mid-point between two adjacent slices using 'ANTs unbiased pairwise registration' (Yushkevich et al., 2012). Adjacent outlier slices were not corrected by this approach due to a lack of a normative slice comparison. As a result, dark slices will then be visible in the reconstructed output, and a poorer image quality rating. After slice interpolation, images were denoised in the Baby Brain Toolkit (BTK) version 2 (Rousseau et al., 2013) using the non-local means algorithm to improve the quality of subsequent super-resolution reconstruction. These denoised images were used as inputs for the super-resolution reconstruction also performed using BTK. Super-resolution reconstructed images were visually inspected and rated as 'good' (no or negligible artefacts), 'fair' (minor artefacts) or 'unusable' (reconstruction failed or contained major artefacts) in accordance with previous recommendations (Backhausen et al., 2016). Of the total 271 participants with MRI, 192 were rates as 'good' (71%), 67 were rated as 'fair' (25%) and specific attention was drawn to whether the subtle artefacts impacted the subsequent segmentation, and 12 had significant artefacts precluding further analysis (4%).

2.3.2. Image segmentation

Reconstructed images then underwent several pre-processing operations, including slice interpolation up-sampling to $0.7 \times 0.7 \times 1\text{mm}^3$ (for PPREMO only) and N4 bias field correction (Tustison et al., 2010). Neck slices were manually removed for the PREBO images covering a large portion of the neck, using FSL's 'robustfov' tool. Images were reoriented to RAS (Right, Anterior, Superior) space to reduce the failure rate for the subsequent segmentation step. Resulting images were processed with the state-of-the-art dHCP parcellation pipeline (Makropoulos et al., 2018). The dHCP pipeline automatically performs segmentation on several cortical and subcortical areas using the DrawEM algorithm (Makropoulos et al., 2014), segmenting the brain into 87 regions using the 20 ALBERT atlas (Gousias et al., 2013, 2012). This segmentation was facilitated by alignment of the MRIs to an age-appropriate preterm template (Schuh et al., 2018). The dHCP pipeline additionally measures cortical thickness (CT), sulcal depth (SD), surface area (SA) and gyrification index (GI) of each region of the cerebral cortex, as defined by the Gousias neonatal atlas (Gousias et al., 2013). As the dHCP pipeline generates many variables, particularly for such a small brain, to reduce the risk of inflation of false positive (type 1) errors cortical measures were grouped into bilateral frontal, parietal, temporal and occipital lobes. In addition, 8 tissue volumes were measured (CSF, cortical GM, WM, ventricles, cerebellum, DGM, brainstem, hippocampi/amygdala), resulting in a total of 24 morphometric variables of interest. All segmentations were visually checked for transformation or segmentation accuracy errors, with instances of substantial errors in the segmentation being removed. Four example segmentation provided by the dHCP pipeline is illustrated in Fig. 2. In the illustrated case 4, the combination of mild but visible super-resolution artefacts as well as ventriculomegaly led to the segmentation failure.

2.4. Neurodevelopmental outcomes

Experienced paediatric physiotherapists, blinded to all earlier clinical and MRI findings, assessed participants at 2 years' CA. Neurodevelopmental outcomes were assessed with the Bayley Scales of Infant and Toddler Development – Third Edition (Bayley-III) (Bayley, 2006), with the motor, cognitive and language composite scores being used in this study. The Bayley-III is a norm-referenced, discriminative measure used to describe the developmental functioning of the child with good to strong validity and reliability (Albers and Grieve, 2007). Subtest

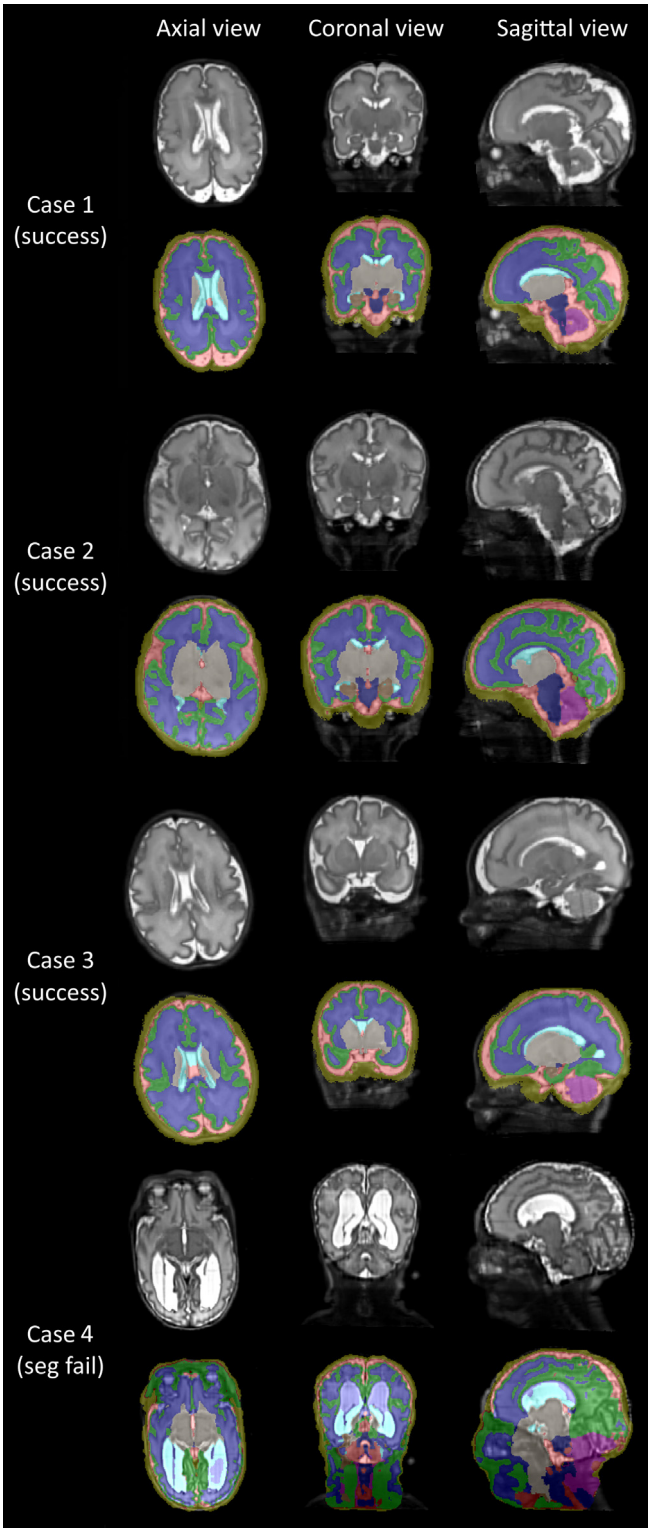


Fig. 2. Illustration of tissue parcellation performed by the dHCP processing pipeline, illustrating three successful segmentations (ages at scan; case 1 32 weeks + 6 days PMA, case 2 34 weeks + 2 days PMA, case 3 31 weeks + 5 days PMA) and one failure (case 4, age at scan 35 weeks + 2 days PMA).

composite scores range from 40 to 160, with a mean of 100 (and standard deviation ± 15) based on normative US data (Piñon, 2010), with higher scores indicating better outcomes.

Motor outcomes were also assessed using the standardized Neuro-Sensory Motor Developmental Assessment (NSMDA) (Burns et al., 1989). The NSMDA is criterion-referenced and has no normative data, rather infants are given a categorical score of motor performance with higher scores indicating increased disability (total score 6-8 normal, 9-11 minimal disability, 12-14 mild disability, 15-19 moderate disability, 20-25 severe disability, >25 profound disability). The NSMDA has been found to be associated with quality of life (Boswell et al., 2017), and more severe motor impairment in children without CP (Danks et al., 2012).

Infants were assessed at the Early time point (29-35 weeks PMA, concurrent with Early MRI) with the General Movements Assessment (GMA) (Einspieler and Prechtl, 2004) that has previously been shown to be predictive of a later CP diagnosis (sensitivity 75-100%, specificity 40-48%) (Bosanquet et al., 2013; Darsaklis et al., 2011) and neuromotor deficits (Einspieler et al., 2016).Spontaneous movements were categorised as either Normal, Poor Repertoire or Cramped Synchronised during the Writhing period by two advanced GMs raters, with cases of non-agreement reviewed until consensus reached and advice sought from a third blinded rater. In addition, a measure of socioeconomic status (SES) was determined at TEA using a combination of factors including family living situation, parental relationship, languages spoken at home parental education and work status, which were collected through a questionnaire completed by infants' primary care giver (Hack et al., 1991; Roberts et al., 2008). Summing each question (0 no risk, 1 potential risk) provides a total raw score from 0 to 12, with scores of 2 and above being considered high social risk in line with other research in this population (Caesar et al., 2016; Spittle et al., 2009).

2.5. Statistical analyses

To examine the relationship between structural measures from MRIs taken at the Early time point with 2-year outcomes, Least Absolute Shrinkage and Selection Operator (LASSO) regression was used (Tibshirani, 1996). Covariates were included in all models, which include sex, gestational age (GA) at birth, PMA at MRI scan, SES raw score, and age at NSMDA assessment, study cohort, site, and GMA, as well as measures of clinical care including days of total parenteral nutrition (TPN), hours of phototherapy, days of endotracheal tube ventilation, days of continuous positive airway pressure (CPAP) and hours of oxygen therapy. Age at Bayley-III assessment was not included in any model as the measure inherently accounts for this. Given the large number of structural measures (24 regional measures) relative to the number of participants (n = 181 included in final analysis), model simplification (i.e. variable reduction) is important to prevent overfitting and to identify the most predictive measures. Variable reduction is performed implicitly in LASSO with a penalty term (alpha), with higher values enforcing sparser (fewer non-zero) model coefficients. In addition, models with only the covariates included (i.e. no MRI measures) were also constructed to determine the potential benefit of adding MRI information. Trained models were compared using the pseudo R2, which is equivalent to the multiple R2 for LASSO regression and provide the percent of variance explained by the model.

Data was split into training (75%) and test (25%) sets that were comparable in GA at birth, PMA at MRI, and sex. Models were generated on the training set, with the optimal sparsity term (alpha) determined using mean-square error and 5-fold cross-validation. The alpha of the best performing model was then applied to the entire training set, yielding an optimal training model (which includes only the retained features from LASSO as well as all covariates). This optimal training model was applied to the test set, with the correspondence between the actual scores and the predicted 2-year outcome measured with both Pearson's r correlation as well as a mean absolute error (MAE). This was repeated for each outcome score, (Bayley-III motor, cognitive, language, and NSMDA), with each using the same train and test splits. Additionally, outcome measures on the test set were dichotomised as 'normal' or 'poor outcome', which for Bayley-III was defined as 1 standard deviation below the mean (scores <85) and for NSMDA was dichotomised the NSMDA normal/minimal (6-11) vs mild-profound (12 or above). From this the

area under the curve (AUC) was quantified using the 'glmnet' library in R, which measures the overall classification agreement between the predicted model and the dichotomised outcome of the predictive models.

3. Results

3.1. Participants

A total number of 271 recruited participants in the PPREMO/PREBO studies had Early MRIs available. Of these, 181 participants were available for the analyses in this study, with the breakdown of data loss illustrated in Fig. 3.

The overall demographics and clinical assessments of the cohort are detailed in Table 1. The cohort includeds (n = 181) and excluded (n = 90) from the analysis did not differ significantly on baseline characteristics or outcome assessment except on the socioeconomic status risk raw score, and MR image quality, as expected as MRI quality was one of the factors for being excluded from the analysis. We note that in this table, not all 2-year assessment scores from the n = 90 excluded group were available for comparison. Investigating the covariates using a logistic regression model on the whole n = 271 cohort, SES risk score was the only significant predictor of participants returning for 2-year assessments (Supplementary Table 1) with those at higher social risk being less likely to return for follow up.

3.2. Prediction of 2-year functional outcomes

For each of the three Bayley-III models and the NSMDA functional grade model, the predictive features retained from LASSO regression are provided in Table 2. As higher NSMDA scores indicate more impaired function, unlike Bayley-III composite scores where higher scores indicate improved function, NSMDA scores were inverted for regression coefficients of all four models to indicate associations with better outcome. The regression coefficients for all 4 models are provided in Table 2, however confidence intervals are not provided by penalised regression such as LASSO.

Of the retained MRI features, cortical grey matter volume was retained in three models with cortical grey matter volume being positively associated with outcome scores. Conversely CSF volumes and ventricle volumes (present in four and two models, respectively) were negatively associated with outcome scores. Of the cortical features retained by LASSO, cortical thickness of the frontal, occipital and parietal lobes was a commonly retained feature in all four models, with a thicker cortex being associated with increased assessment scores for all retained CT features. Sulcal depth of the frontal, temporal and occipital lobes were also frequently retained, with larger sulcal depths being predominantly positively associated with Bayley's score (in 4 of 5 retained SD features).

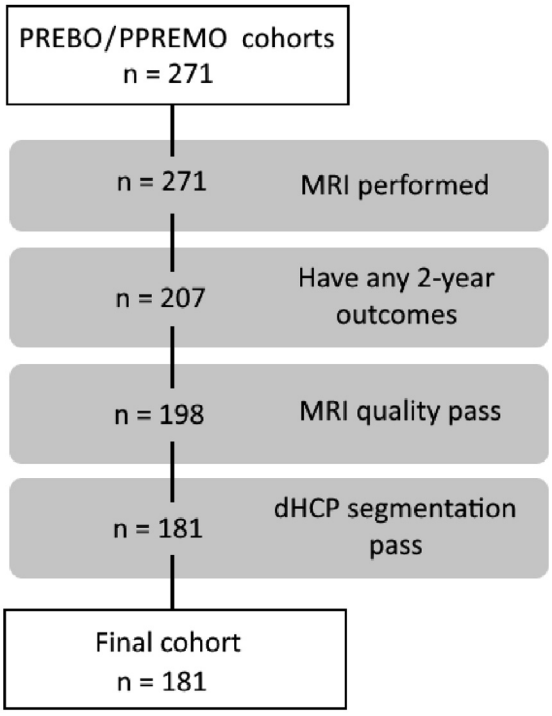


Fig. 3. Diagram illustrating data available for statistical analysis, including reasons for exclusion.

Table 1
Baseline characteristics of the very preterm cohort with 2-year follow-up assessments, divided into those included in the analysis (n = 181), and those excluded (n = 90) due to MRI being of poor quality (n = 12), segmentation was inaccurate (n = 17), or 2-year assessments were not available (n = 64). There was n = 3 excluded participants who had no MRI or 2-year assessments available.

	Included in analysis N = 181	Excluded from analysis N = 90	P-value
Postmenstrual age at MRI, weeks^{+days}, median (range)	32^{+4} (29^{+1}- 35^{+2})	32^{+2} (30^{+2}- 35^{+0})	0.972
Gestational age at birth, weeks^{+days}, median (range)	28^{+3} (23^{+1}- 30^{+6})	27^{+3} (23^{+4}- 30^{+6})	0.599
Male, n (%)	98 (54%)	27 (30%)	0.680
SES risk score, mean (SD)	1.79 (1.67)	2.58 (1.58)	<0.001
High social risk (risk score > 2), n(%)	52 (29%)	25 (21%)	0.253
GMAs classification at Early time point			
Normal (%)	70 (39%)	37 (41%)	0.885
Poor Repertoire (%)	98 (54%)	44 (49%)	0.571
Cramped Synchronised (%)	13 (7%)	9 (10%)	0.612
Bayley-III motor composite, mean (SD)	97.37 (16.33)	95.15 (17.49)	0.488
Bayley-III cognitive composite, mean (SD)	95.51 (16.02)	94.04 (17.49)	0.499
Bayley-III language composite, mean (SD)	92.87 (17.01)	93.07 (16.02)	0.907
NSMDA functional grade, mean (SD)	8.61 (2.99)	9.05 (3.82)	0.429
MR image quality			
No MRI (%)	0 (0%)	0 (0%)	1.0
Unusable (%)	0 (0%)	12 (13%)	0.001
Fair (%)	34 (19%)	33 (37%)	0.007
Good (%)	147 (81%)	45 (50%)	<0.001

Bayley-III, Bayley Scales of Infant and Toddler Development 3rd Edition; GMA, General Movements Assessment; NSMDA, Neuro-Sensory Motor Developmental Assessment; SD, Standard deviation; SES, Socioeconomic status; TEA, Term Equivalent Age.

Table 2
The retained morphometric biomarkers and covariates for each of the 2-year Bayley-III outcomes (motor, cognitive and language composite scores) and NSMDA functional grade, and their corresponding regression coefficient.

Model 1: Bayley-III motor composite	
Feature	Coefficient
CSF	-0.00036
Cortical grey matter	0.00001
Ventricles	-0.00010
Occipital lobe CT	1.64
Frontal lobe SD	0.17818
Occipital lobe SD	-0.05580
Temporal lobe SD	0.04080
GA at birth	0.16967
Sex (male reference)	0.24845
PMA at MRI	-1.0701
SES risk score	-2.39768
GMAs classification	-0.79964
Cohort (PREBO reference)	-4.31271
Site	<0.00001
Days of ETT ventilation	-0.29369
Days of CPAP	-0.37486
Hours of oxygen therapy	-0.00358
Days of TPN	0.28418
Hours of phototherapy	-0.02337
Model 2: Bayley-III cognitive composite	
Feature	Coefficient
CSF	-0.00030
Cortical grey matter	0.00004
Frontal lobe CT	1.64
Parietal lobe CT	1.06
Frontal lobe GI	0.95901
GA at birth	0.05028
Sex (male reference)	2.83693
PMA at MRI	-4.00087
SES risk score	-2.94637
GMAs classification	-0.41402
Cohort (PREBO reference)	-11.10821
Site	<0.00001
Days of ETT ventilation	0.21381
Days of CPAP	-0.09152
Hours of oxygen therapy	-0.00616
Days of TPN	0.11960
Hours of phototherapy	0.06188
Model 3: Bayley-III language composite	
Feature	Coefficient
CSF	-0.00036
Frontal lobe CT	1.96
Frontal lobe SA	-0.01542
Frontal lobe SD	0.33345
Occipital lobe SD	0.24407
GA at birth	0.25547
Sex (male reference)	0.70911
PMA at MRI	-1.4561
SES risk score	-4.8492
GMAs classification	-2.34220
Cohort (PREBO reference)	0.38724
Site	<0.00001
Days of ETT ventilation	-0.09814
Days of CPAP	0.10227
Hours of oxygen therapy	-0.00047
Days of TPN	0.38442
Hours of phototherapy	0.01127
Model 4: NSMDA functional grade	
Feature	Coefficient
CSF	-0.00002
Cortical grey matter	0.00001
Ventricles	-0.00017
Frontal lobe CT	0.36541
GA at birth	0.01977
Sex (male reference)	0.47920
PMA at MRI	-0.60005
SES risk score	0.45359
GMAs classification	-0.48023
Cohort (PREBO reference)	-0.77474
Site	<0.00001
Age at NSMDA assessment	-0.56597
Days of ETT ventilation	-0.00128
Days of CPAP	-0.02750
Hours of oxygen therapy	0.00029
Days of TPN	0.04703
Hours of phototherapy	-0.00530

(*continued on next column*)

Table 2 (*continued*)

CPAP, Continuous Positive Airway Pressure; CT, Cortical Thickness; ETT, Endotracheal Tube; GI, Gyrification Index; GMA, General Movements Assessment; PMA, Postmenstrual Age; SA, Surface Area; SD, Sulcal Depth; GA, Gestation Age; SES, Socio-economic status; TPN, Total Parenteral Nutrition.

Table 3
Pseudo r^2 of the four optimal LASSO models indicating variance in clinical outcome explained by MRI measures and covariates, as well as the covariates only.

Model	Pseudo r^2 with MRI	Pseudo r^2 no MRI
M1: Bayley-III motor composite	0.757	0.465
M2: Bayley-III cognitive composite	0.546	0.431
M3: Bayley-III language composite	0.773	0.482
M4: NSMDA functional grade	0.660	0.492

Bayley-III, Bayley Scales of Infant and Toddler Development 3rd Edition; LASSO, Least Absolute Shrinkage and Selection Operator; MRI, Magnetic Resonance Images; NSMDA, Neuro-sensory Motor Developmental Assessment.

Covariates were manually retained in all four models in line with previous studies (Kline et al., 2020). Higher socio-economic status risk scores were negatively associated with all developmental outcome scores (average reduction -2.66 per point in SES risk raw score), suggesting that this is one of the biggest drivers of 2-year outcomes. Similarly, abnormal writhing movements classified by the GMAs classification were negatively associated with all outcome scores (average reduction -1.01 per GMAs risk category). Higher GA at birth was associated with better outcomes (score increase 0.12 per week), while PMA at MRI was negatively associated with outcomes (score decrease 1.78 per week). Girls performed better than boys in all models (average increase 1.07). Babies in the PPREMO cohort performed worse on average that the PREBO cohort for three of the four 2-year outcomes (average Bayley-III composite score reduction -5.01), however site played a negligible role in all models (effect size <1e-10). Measures of care were predominantly negatively associated with 2-year outcomes, including days of ventilation (average reduction -0.04 per day), days of CPAP (average reduction -0.09 per day), hours of oxygen therapy (average reduction -0.002 per hour), and hours of phototherapy (average reduction -0.02 per hour), with the exception of days of total parenteral nutrition which was positively associated with 2-year outcomes (average increase 0.21 per day).

3.3. MRI scores explain much of the variance in 2-year assessments

The model fit for each of the LASSO models is provided by the Pseudo r^2 in Table 3. When the four models were run without MRI correlates, the factors accounted for less of the variance in Bayley-III and NSMDA scores, as indicated by the Pseudo r^2.

3.4. Model validation on the test set

The trained LASSO models were used to predict assessments scores for the independent test data. Pearson's r correlations between actual and predicted assessment scores are provided in Table 4. Moderate correlations were observed for all assessments, however only three of the

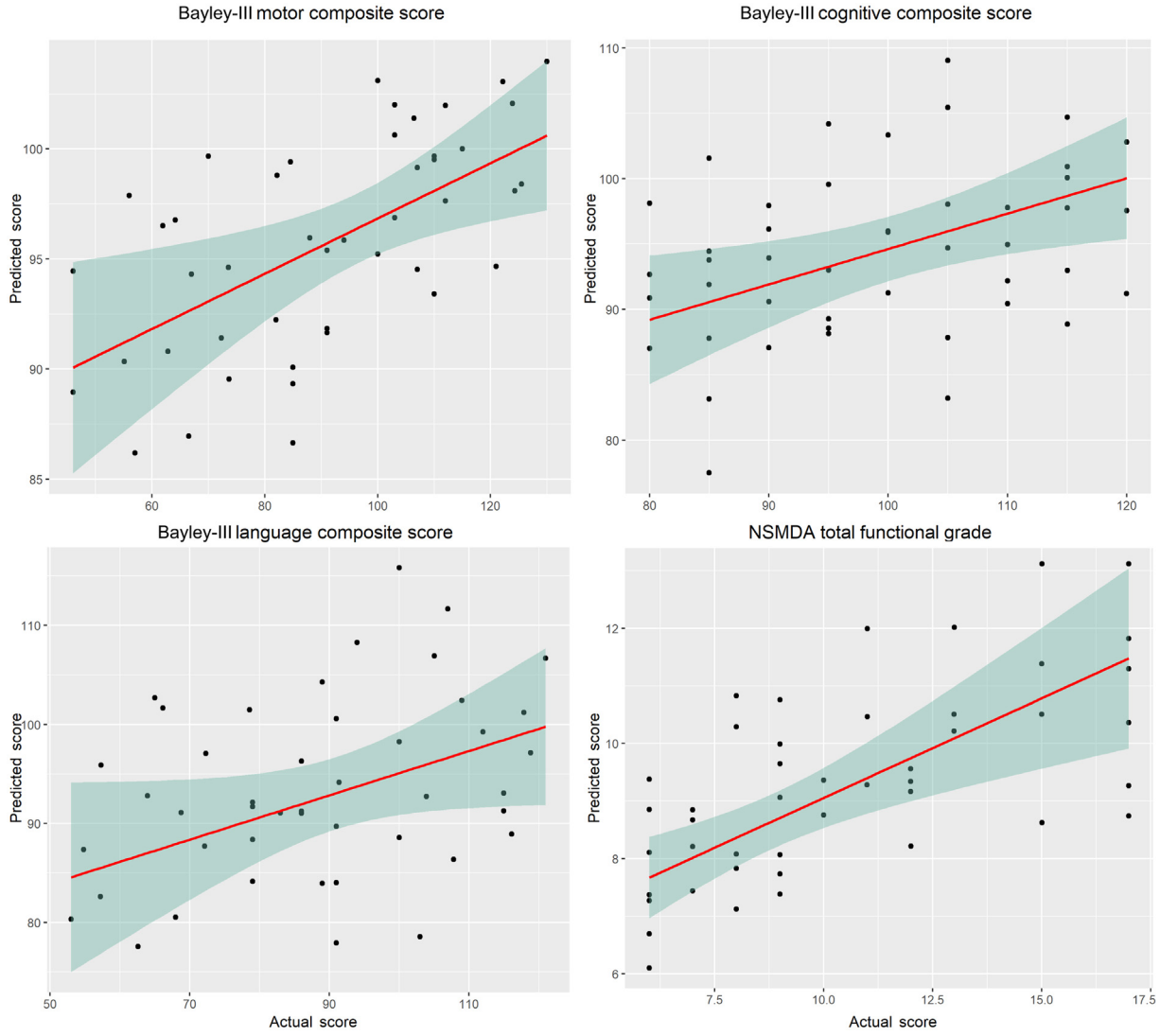


Fig. 4. Test set correlations between the best performing LASSO model and the actual Bayley-III assessment scores (motor and cognitive composite score, top row; language compositive score, bottom left) and NSMDA functional grade (bottom right) using measures from the Early MRI acquired between 29 and 35 weeks PMA.

Table 4
Correlations between the best performing models with and without MRI measures, and the test set assessment scores independent by the model. Significance level was corrected using Bonferroni correction ($\alpha = 0.05 / 4$ models $= 0.0125$).

Model	Pearson's r	MAE	AUC (Range)
M1: Bayley-III motor composite	0.508* (p=0.008)	9.25	0.864 (0.856-0.874)
M2: Bayley-III cognitive composite	0.477* (p=0.011)	11.33	0.863 (0.846-0.880)
M3: Bayley-III language composite	0.367 (p=0.059)	12.77	0.677 (0.664-0.690)
M4: NSMDA functional grade	0.624* (p=0.001)	1.34	0.852 (0.846-0.858)

* $p < 0.00125$. AUC, Area Under the Curve; Bayley-III, Bayley Scales of Infant and Toddler Development 3rd Edition; MAE, Mean Absolute Error; MRI, Magnetic Resonance Images; NSMDA, Neuro-sensory Motor Developmental Assessment.

four models were statistically significant in the test set ($p < 0.0125$ Bonferroni corrected), including the Bayley-III motor and cognitive composite scores and NSMDA functional grade with Early MRI features, but not the Bayley-III language composite score. The scatter plot illustrating these test set correlations are provided in Fig. 4. In addition, the AUC metrics using the dichotomised Bayley-III composite scores (<1 SD below mean) and the NSMDA functional grade (scores above 8 indicating disability) are also shown in Table 4.

4. Discussion

In this study, we demonstrate that measures derived from Early structural MRI taken between 29-35 weeks PMA are associated with motor, cognitive and language ability at two-years CA in a very preterm born cohort. By combining Early brain morphometrics with key covariates in a LASSO prediction model, estimates of 2-year Bayley-III motor and cognitive composite scores, and NSMDA functional grade were obtained (in an independent test set), demonstrating the predictive ability of these early MRI measures. General trends across the models revealed higher GM volume and lower CSF and ventricle volumes to be associated with better 2-year outcomes across multiple functional domains, supporting previous studies that reported these early brain measures in particular may be an important early biomarkers of abnormal brain development (Murphy et al., 2020). In all four models with MRI measures, increased sulcal depth and thicker cortex across all lobes (frontal, parietal, occipital and temporal) were frequently retained among models as predictors of better outcome, albeit with different regions impacting different functional outcomes Interestingly cerebellar volumes were not retained in any model, despite evidence in the literature that they are correlated with motor and cognitive development (Matthews et al., 2018; Shah et al., 2006). We suspect that this may be because brain volume reduction occurs globally, leading to volumes of the cortical grey matter

and the cerebellum to covary, and as a result LASSO retains only one of these measures for sparsity. Additionally the covariates were important drivers of 2-year outcomes, including socioeconomic status which has been demonstrated previously (Kline et al., 2020), GMA classification which has been shown to be predictive of neurodevelopment at both 2 and 4 years CA (Spittle et al., 2013), and sex, with the male sex being an important risk factor both for poor neonatal outcome as well as cognition and Motor outcomes at 2-years (Peacock et al., 2012). Measures of care were predominantly associated with adverse 2-year outcomes, which may be a result that babies with more complicated births due to prematurity require more care, and also are more likely to have adverse neurodevelopmental outcomes (Kiechl-Kohlendorfer et al., 2009). In contrast, the number of days of TPN was positively associated with all four 2-year outcomes, supporting previous studies that found greater nutritional intake to be positively associated with growth and neurodevelopment (Christmann et al., 2017). In fact, patient demographics and measures of clinical care combined explained more variance in 2-year outcomes than was explained by to MRI measures (Table 3). This may be due to covariates being retained by LASSO with only MRI measures potentially being removed, and in fact the number of covariates exceeded the number of retained MRI biomarkers in all four models.

Prediction scores derived from our model (which includes the MRI biomarkers and covariates) were correlated with actual outcome measures at 2-year follow-up (r = 0.36-0.62, Table 4). This is consistent with previous findings following up preterm cohorts imaged at TEA, with a previous study showing that cortical features (such as surface area, gyrification index, sulcal depth and curvature) were correlated with both Bayley's cognitive and language scores (r = 0.51-0.53, n = 110), respectively (Kline et al., 2020). We note that in the study by Kline et al. (2020), models were validated across the whole dataset using leave-one-out cross validation. In our present study model performance was estimated on the training subset only (with 5-fold cross-validation), and then optimal models were validated on the test subset of data completely unseen by the model, which should provide a more unbiased estimate of the model error. In another study of 86 infants scanned at the Early time point (30-32 weeks PMA), volumetrics and global cortical morphology were found to be predictive of both low Bayley-III motor and cognitive outcomes at 2 years (AUC of 0.80 and 0.78 respectively, n = 86) (Moeskops et al., 2017). Our findings are in line with this, showing that brain morphometrics including regional cortical morphometrics are predictive of motor outcomes (AUC = 0.86) and cognitive outcomes (AUC = 0.86). Expanding on previously reported findings, we also found associations with language ability (AUC = 0.67) and NSMDA functional grade (AUC = 0.85) (Table 4). While mean Bayley-III scores for healthy Australian infants are statistically significantly higher than the US standardised means for all subtests (Chinta et al., 2014), the chosen cut-off of <85 was found to provide an accurate definition of moderate-severe neurodevelopmental delay (Johnson et al., 2014). These models demonstrate that both the NSMDA and Bayley-III clinical outcomes can be predicted at Early (30-32 weeks PMA) time points, with these clinical assessments in turn being strongly predictive of later impairment at 4 years and beyond (Griffiths et al., 2018).

A limitation of this study is that, for the PPREMO study (n = 121), we only used morphology measured from the multi-echo T2-TSE MRIs to predict 2-year outcomes, which has a thick slice (2mm) to produce a faster sequence. This sequence is also more susceptible to "ring" artefacts when there is motion when compared to HASTE sequences, resulting in greater data loss due to image quality. Ideally a 3D T1-weighted sequence would be added, which could provide an additional MR contrast to assist the segmentation provided by the dHCP pipeline (Makropoulos et al., 2018). In our cohort however, given the long duration of this T1-weighted scan (~5 min), many were affected by motion artefact, and were subsequently removed from the analysis. Similarly, diffusion MRI was not included in the present analysis, which would probe the microstructure of the white matter not detectable by structural imaging. Several studies have found associations between diffusion tensor imaging (DTI) or other diffusion models, and 2-year Bayley-III outcomes (Vassar et al., 2020), including studies using this cohort, which found stronger associations with 1-year Bayley-III outcomes over 2-year outcomes, and greater associations using fixel-based rather than DTI metrics (Pannek et al., 2020). In future work we plan to include MRI derived measures from images acquired at Early and TEA time-points, as well as to combine metrics of brain structure and microstructure to see if there is an enhanced prediction of later outcomes. In future, we also plan to replace the FSL 'bet' brain masking with artificial neural-network approaches (Isensee et al., 2019) to minimize the amount of manual mask correction. A further limitation of this study is that the dHCP tissue atlas groups all anatomies bilaterally, groups the functionally distinct hippocampus and amygdala as one label, and additionally groups cortical measures by lobe. As in previous studies (Moeskops et al., 2017), this was done to minimize model complexity and avoid overfitting given the current sample size (n = 181), however as a result it ignored functional differences between the hippocampus and amygdala, and structures bilaterally in general. Another potential limitation of this work was that the sample may have been skewed towards infants at lower risk of adverse neurodevelopmental outcomes, because infants with the highest severity of illness during the 29-35 week PMA interval were less likely to be recruited for reasons of feasibility and safety of performing the MRI. Furthermore, for those participants whose adverse outcomes were associated with severe brain abnormality, these MRIs were also more likely to fail at the dHCP segmentation. Additionally, the PMA at MRI range of this cohort at this 'Early' time point is quite wide (median age = 32 weeks, range = 29-35 weeks), and although accounted for in the model, there is a known bias in our cohort where infants scanned later at this time point had poorer outcomes as they were not clinically stable enough to have the MRI earlier. This may explain why PMA at MRI was found to be consistently negatively associated with Bayley's composite outcomes, and positively associated with NSMDA functional grade (Table 2). One final limitation with the prediction LASSO models is that they tend to underestimate high assessment scores and overestimate low assessment scores, making predictions much closer to the mean (Fig. 4). What this means is that prediction of the numeric Bayley-III composite score is not accurate in terms of MAE (Table 3), however clinically these models could accurately predict those with adverse outcomes (<1 SD from the mean) based on the AUC (Table 3), potentially allowing earlier follow-up and interventions for those infants.

There is a need to look beyond 2-year outcomes and see if Early MRI imaging can reveal developmental outcomes at childhood age. This is important as intellectual, learning and behavioural outcomes cannot be fully determined at 1-3 years CA. Later follow up beyond 3 years is required to evaluate the true predictive accuracy of Early MRI. While the relationship between MRI at TEA and childhood outcomes has been demonstrated at preschool age (Vanes et al., 2021) and at 7 years (Anderson et al., 2017; Thompson et al., 2014) in preterm infants, these studies however have been limited by sample size, MR field strength, limiting image resolution and quality (Anderson et al., 2017; Loh et al., 2017), and only investigated neonatal MRI taken at TEA (Anderson et al., 2015; van't Hooft et al., 2015). Integration of diffusion MRI-based white matter measures, including fixel-based measures, will be important in addition to the macro measures used in this study to explain the impact of preterm birth on brain development and later outcomes (Pannek et al., 2018). We are currently following up this cohort at 6 years as part of the PREBO-6 study (George et al., 2020), when a diagnosis of CP, ASD and school readiness outcomes can be reliably obtained. This longitudinal information data is key, increasing the value of the PPREMO/PREBO cohort and potentially the predictive potential of the early MRI.

5. Conclusions

In this study we perform state-of-the-art structural image analysis on one of the largest cohorts of infants born preterm who underwent

Early 3T MRI (29-35 weeks PMA). Using LASSO regression, we identified several biomarkers predictive of motor, cognitive and language function at 2-years of age, including GM volume, CSF and ventricular volume, cortical thickness and sulcal depth across the cortex, and key covariates including socio-economic status, Early GMAs classification, sex and GA at birth, which has been identified previously as important factors driving outcomes. Furthermore, using these models revealed accurate prediction of outcomes on an independent test set, indicating that Bayley-III motor and cognitive composite scores and NSMDA functional grades could be estimated as early as 29 weeks PMA. Accurate prediction of outcomes derived from early MRI would facilitate interventions to be provided to at-risk infants much earlier and may assist with more effective intervention strategies if delivered during this optimal early period of neuroplasticity. Ultimately, we hope to use this information to improve clinical outcomes for infants born preterm.

Data availabililty statement

The study team are available to collaborate with other research teams on receipt of a reasonable request to access study data. Expressions of interest to access study data, made out to the corresponding author, will be considered and then group level or individual level deidentified data could be shared as appropriate.

Data Availability

Data will be made available on request.

Credit authorship contribution statement

Alex M. Pagnozzi: Conceptualization, Methodology, Software, Validation, Formal analysis, Visualization, Writing – original draft. **Liza van Eijk:** Data curation, Methodology, Software, Investigation, Writing – review & editing. **Kerstin Pannek:** Data curation, Methodology, Supervision, Writing – review & editing. **Roslyn N. Boyd:** Project administration, Writing – review & editing. **Susmita Saha:** Data curation, Methodology, Software, Writing – review & editing. **Joanne George:** Conceptualization, Project administration, Funding acquisition, Data curation, Writing – review & editing. **Samudragupta Bora:** Project administration, Funding acquisition, Writing – review & editing. **DanaKai Bradford:** Supervision, Writing – review & editing. **Michael Fahey:** Data curation, Writing – review & editing. **Michael Ditchfield:** Data curation, Writing – review & editing. **Atul Malhotra:** Data curation, Writing – review & editing. **Helen Liley:** Writing – review & editing. **Paul B. Colditz:** Project administration, Funding acquisition, Data curation, Writing – review & editing. **Stephen Rose:** Writing – review & editing. **Jurgen Fripp:** Supervision, Project administration, Writing – review & editing.

Funding

The following support has been provided by the Australian National Health and Medical Research Council (NHMRC): Project grant (PREBO NHMRC 1084032) and Investigator Fellowship (RB: NHMRC 1195602) and an NHMRC Clinical CRE (NHMRC 1116442 CRE Australasian CP Clinical Trials Network). PPREMO was funded through grants from the Cerebral Palsy Alliance Research Foundation (IRG1413), the Financial Markets Foundation for Children (2014–074), the Queensland Government (Smart State; Health Practitioner Stimulus Grant) and the Royal Brisbane and Women's Hospital. We would like to acknowledge the families of preterm infants who participated in this study, the medical and nursing staff of the neonatal unit, radiology and administrative staff in the department of medical imaging and the Physiotherapists and Psychologists who assessed 24-month outcomes, and Julie Trinder for her management of the imaging, demographic, and clinical data for this cohort.

The study team are available to collaborate with other research teams on receipt of a reasonable request to access study data. Expressions of interest to access study data, made out to the corresponding author, will be considered and then group level or individual level deidentified data could be shared as appropriate.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119815.

References

Albers, C.A., Grieve, A.J., 2007. Test review: Bayley, N. (2006). Bayley scales of infant and toddler development– Third Edition. San Antonio, TX: Harcourt Assessment. J. Psychoeduc. Assess. 25, 180–190. doi:10.1177/0734282906297199.

Anderson, P.J., Cheong, J.L.Y., Thompson, D.K., 2015. The predictive validity of neonatal MRI for neurodevelopmental outcome in very preterm children. Semin. Perinatol. 39, 147–158. doi:10.1053/J.SEMPERI.2015.01.008.

Anderson, P.J., Treyvaud, K., Neil, J.J., Cheong, J.L.Y., Hunt, R.W., Thompson, D.K., Lee, K.J., Doyle, L.W., Inder, T.E., 2017. Associations of newborn brain magnetic resonance imaging with long-term neurodevelopmental impairments in very preterm children. J. Pediatr. 187, 58–65. doi:10.1016/j.jpeds.2017.04.059, e1.

Backhausen, L.L., Herting, M.M., Buse, J., Roessner, V., Smolka, M.N., Vetter, N.C., 2016. Quality control of structural MRI images applied using FreeSurfer-a hands-on workflow to rate motion artifacts. Front. Neurosci. 10, 558. doi:10.3389/FNINS.2016.00558/BIBTEX.

Bayley, N., 2006. Bayley Scales of Infant and Toddler Development. PsychCorp, Pearson.

Bosanquet, M., Copeland, L., Ware, R., Royd, R., 2013. A systematic review of tests to predict cerebral palsy in young children. Dev. Med. Child Neurol. 55, 418–426. doi:10.1111/dmcn.12140.

Boswell, L., Weck, M., Santella, M., Patrick, C., Russow, A., Deregnier, R., 2017. Neurosensory motor developmental assessment at 18-24 months predicts quality of life at 3-1/2 to 5 years. Dev. Med. Child Neurol. 59, 61–62. doi:10.1111/dmcn.93_13511.

Burns, Y.R., Ensbey, R.M., Norrie, M.A., 1989. The Neuro-sensory motor developmental assessment part 1: development and administration of the test. Aust. J. Physiother. 35, 141–149. doi:10.1016/S0004-9514(14)60503-1.

Caesar, R., Boyd, R.N., Colditz, P., Cioni, G., Ware, R.S., Salthouse, K., Doherty, J., Jackson, M., Matthews, L., Hurley, T., Morosini, A., Thomas, C., Camadoo, L., Baer, E., 2016. Early prediction of typical outcome and mild developmental delay for prioritisation of service delivery for very preterm and very low birthweight infants: a study protocol. BMJ Open 6, e010726. doi:10.1136/BMJOPEN-2015-010726.

Chawanpaiboon, S., Vogel, J.P., Moller, A.B., Lumbiganon, P., Petzold, M., Hogan, D., Landoulsi, S., Jampathong, N., Kongwattanakul, K., Laopaiboon, M., Lewis, C., Rattanakanokchai, S., Teng, D.N., Thinkhamrop, J., Watananirun, K., Zhang, J., Zhou, W., Gülmezoglu, A.M., 2019. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. Lancet Glob. Health 7, e37–e46. doi:10.1016/S2214-109X(18)30451-0.

Chinta, S., Walker, K., Halliday, R., Loughran-Fowlds, A., Badawi, N., 2014. A comparison of the performance of healthy Australian 3-year-olds with the standardised norms of the Bayley Scales of Infant and Toddler Development (version-III). Arch. Dis. Child. 99, 621–624. doi:10.1136/ARCHDISCHILD-2013-304834.

Christmann, V., Roeleveld, N., Visser, R., Janssen, A.J.W.M., Reuser, J.J.C.M., van Goudoever, J.B., van Heijst, A.F.J., 2017. The early postnatal nutritional intake of preterm infants affected neurodevelopmental outcomes differently in boys and girls at 24 months. Acta Paediatr. 106, 242. doi:10.1111/APA.13669.

Counsell, S.J., Rutherford, M.A., Cowan, F.M., Edwards, A.D., 2003. Magnetic resonance imaging of preterm brain injury. Arch. Dis. Child. Fetal Neonatal Ed. 88, F269–F274. doi:10.1136/fn.88.4.F269.

Danks, M., Maideen, M.F., Burns, Y.R., O'Callaghan, M.J., Gray, P.H., Poulsen, L., Watter, P., Gibbons, K., 2012. The long-term predictive validity of early motor development in "apparently normal" ELBW survivors. Early Hum. Dev. 88, 637–641. doi:10.1016/J.EARLHUMDEV.2012.01.010.

Darsaklis, V., Snider, L.M., Majnemer, A., Mazer, B., 2011. Predictive validity of Prechtl's method on the qualitative assessment of general movements: a systematic review of the evidence. Dev. Med. Child Neurol. 53, 896–906. doi:10.1111/J.1469-8749.2011.04017.X.

De Bruïne, F.T., Van Wezel-Meijler, G., Leijser, L.M., Steggerda, S.J., Van Den Berg-Huysmans, A.A., Rijken, M., Van Buchem, M.A., Van Der Grond, J., 2013. Tractography of white-matter tracts in very preterm infants: a 2-year follow-up study. Dev. Med. Child Neurol. 55, 427–433. doi:10.1111/dmcn.12099.

Dubois, J., Alison, M., Counsell, S.J., Hertz-Pannier, L., Hüppi, P.S., Benders, M.J.N.L., 2021. MRI of the neonatal brain: a review of methodological challenges and neuroscientific advances. J. Magn. Reson. Imaging 53, 1318–1343. doi:10.1002/jmri.27192.

Dubois, J., Benders, M., Borradori-Tolsa, C., Cachia, A., Lazeyras, F., Leuchter, R.H.-V., Sizonenko, S.V., Warfield, S.K., Mangin, J.F., Hüppi, P.S., 2008. Primary cortical folding in the human newborn: an early marker of later functional development. Brain 131, 2028–2041. doi:10.1093/brain/awn137.

Einspieler, C., Bos, A.F., Libertus, M.E., Marschik, P.B., 2016. The general movement assessment helps us to identify preterm infants at risk for cognitive dysfunction. Front. Psychol. 7. doi:10.3389/FPSYG.2016.00406.

Einspieler, C., Prechtl, H.F.R., 2004. Prechtl's method on the qualitative assessment of general movements in preterm, term, and young infants 91.

George, J.M., Boyd, R.N., Colditz, P.B., Rose, S.E., Pannek, K., Fripp, J., Lingwood, B.E., Lai, M.M., Kong, A.H., Ware, R.S., Coulthard, A., Finn, C.M., Bandaranayake, S.E., 2015. PPREMO: a prospective cohort study of preterm infant brain structure and function to predict neurodevelopmental outcome. BMC Pediatr. 15, 123. doi:10.1186/s12887-015-0439-z.

George, J.M., Colditz, P.B., Chatfield, M.D., Fiori, S., Pannek, K., Fripp, J., Guzzetta, A., Rose, S.E., Ware, R.S., Boyd, R.N., 2021. Early clinical and MRI biomarkers of cognitive and motor outcomes in very preterm born infants. Pediatr. Res. 90, 1243–1250. doi:10.1038/S41390-021-01399-5.

George, J.M., Pagnozzi, A.M., Bora, S., Boyd, R.N., Colditz, P.B., Rose, S.E., Ware, R.S., Pannek, K., Bursle, J.E., Fripp, J., Barlow, K., Iyer, K., Leishman, S.J., Jendra, R.L., 2020. Prediction of childhood brain outcomes in infants born preterm using neonatal MRI and concurrent clinical biomarkers (PREBO-6): study protocol for a prospective cohort study. BMJ Open 10, e036480. doi:10.1136/bmjopen-2019-036480.

George, J.M., Pannek, K., Rose, S.E., Ware, R.S., Colditz, P.B., Boyd, R.N., 2018. Diagnostic accuracy of early magnetic resonance imaging to determine motor outcomes in infants born preterm: a systematic review and meta-analysis. Dev. Med. Child Neurol. 60, 134–146. doi:10.1111/dmcn.13611.

Gousias, I.S., Edwards, A.D., Rutherford, M.A., Counsell, S.J., Hajnal, J.V, Rueckert, D., Hammers, A., 2012. Magnetic resonance imaging of the newborn brain: manual segmentation of labelled atlases in term-born and preterm infants. Neuroimage 62, 1499–1509. doi:10.1016/j.neuroimage.2012.05.083.

Gousias, I.S., Hammers, A., Counsell, S.J., Srinivasan, L., Rutherford, M.A., Heckemann, R.A., Hajnal, J.V, Rueckert, D., Edwards, A.D., 2013. Magnetic resonance imaging of the newborn brain: automatic segmentation of brain images into 50 anatomical regions. PLoS One 8, e59990. doi:10.1371/journal.pone.0059990.

Griffiths, A., Toovey, R., Morgan, P.E., Spittle, A.J., 2018. Psychometric properties of gross motor assessment tools for children: a systematic review. BMJ Open 8, e021734. doi:10.1136/bmjopen-2018-021734.

Hack, M., Breslau, N., Weissman, B., Aram, D., Klein, N., Borawski, E., 1991. Effect of very low birth weight and subnormal head size on cognitive abilities at school age. N. Engl. J. Med. 325, 231–237. doi:10.1056/nejm199107253250403.

Ibrahim, J., Mir, I., Chalak, L., 2018. Brain imaging in preterm infants <32 weeks gestation: a clinical review and algorithm for the use of cranial ultrasound and qualitative brain MRI. Pediatr. Res. 846 (84), 799–806. doi:10.1038/s41390-018-0194-6, 2018.

Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.P., Heiland, S., Wick, W., Bendszus, M., Maier-Hein, K.H., Kickingereder, P., 2019. Automated brain extraction of multisequence MRI using artificial neural networks. Hum. Brain Mapp. 40, 4952–4964. doi:10.1002/HBM.24750.

Johnson, S., Moore, T., Marlow, N., 2014. Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used? Pediatr. Res. 755 (75), 670–674. doi:10.1038/pr.2014.10, 2014.

Kidokoro, H., Neil, J.J., Inder, T.E., 2013. New MR imaging assessment tool to define brain abnormalities in very preterm infants at term. AJNR. Am. J. Neuroradiol. 34, 2208–2214. doi:10.3174/ajnr.A3521.

Kiechl-Kohlendorfer, U., Ralser, E., Peglow, U.P., Reiter, G., Trawöger, R., 2009. Adverse neurodevelopmental outcome in preterm infants: risk factor profiles for different gestational ages. Acta Pædiatr. 98, 792–796. doi:10.1111/J.1651-2227.2009.01219.X.

Kline, Julia E, Illapani, V.S.P., He, L., Altaye, M., Logan, J.W., Parikh, N.A., 2020. Early cortical maturation predicts neurodevelopment in very preterm infants. Arch. Dis. Child. Fetal Neonatal Ed. 105, 460–465. doi:10.1136/archdischild-2019-317466.

Loh, W.Y., Anderson, P.J., Cheong, J.L.Y., Spittle, A.J., Chen, J., Lee, K.J., Molesworth, C., Inder, T.E., Connelly, A., Doyle, L.W., Thompson, D.K., 2017. Neonatal basal ganglia and thalamic volumes: very preterm birth and 7-year neurodevelopmental outcomes. Pediatr. Res. 82, 970–978. doi:10.1038/pr.2017.161.

Makropoulos, A., Gousias, I.S., Ledig, C., Aljabar, P., Serag, A., Hajnal, J.V., Edwards, A.D., Counsell, S.J., Rueckert, D., 2014. Automatic whole brain MRI segmentation of the developing neonatal brain. IEEE Trans. Med. Imaging 33, 1818–1831. doi:10.1109/TMI.2014.2322280.

Makropoulos, A., Robinson, E.C., Schuh, A., Wright, R., Fitzgibbon, S., Bozek, J., Counsell, S.J., Steinweg, J., Vecchiato, K., Passerat-Palmbach, J., Lenz, G., Mortari, F., Tenev, T., Duff, E.P., Bastiani, M., Cordero-Grande, L., Hughes, E., Tusor, N., Tournier, J.-D., Hutter, J., Price, A.N., Teixeira, R.P.A.G., Murgasova, M., Victor, S., Kelly, C., Rutherford, M.A., Smith, S.M., Edwards, A.D., Hajnal, J.V, Jenkinson, M., Rueckert, D., 2018. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. Neuroimage 173, 88–112. doi:10.1016/j.neuroimage.2018.01.054.

Malhotra, A., Fahey, M.C., Davies-Tuck, M., Wong, F., Carse, E., Whiteley, G., Ditchfield, M., 2017. Comparison of preterm and term equivalent age MRI for the evaluation of preterm brain injury. J. Perinatol. 37, 864–868. doi:10.1038/JP.2017.39.

Matthews, L.G., Inder, T.E., Pascoe, L., Kapur, K., Lee, K.J., Monson, B.B., Doyle, L.W., Thompson, D.K., Anderson, P.J., 2018. Longitudinal preterm cerebellar volume: perinatal and neurodevelopmental outcome associations. Cerebellum 17, 610. doi:10.1007/S12311-018-0946-1.

Miller, S.P., Ferriero, D.M., Leonard, C., Piecuch, R., Glidden, D.V., Partridge, J.C., Perez, M., Mukherjee, P., Vigneron, D.B., Barkovich, A.J., 2005. Early brain injury in premature newborns detected with magnetic resonance imaging is associated with adverse early neurodevelopmental outcome. J. Pediatr. 147, 609–616. doi:10.1016/j.jpeds.2005.06.033.

Moeskops, P., Išgum, I., Keunen, K., Claessens, N.H.P., van Haastert, I.C., Groenendaal, F., de Vries, L.S., Viergever, M.A., Benders, M.J.N.L., 2017. Prediction of cognitive and motor outcome of preterm infants based on automatic quantitative descriptors from neonatal MR brain images. Sci. Rep. 7, 2163. doi:10.1038/s41598-017-02307-w.

Morgan, C., Fetters, L., Adde, L., Badawi, N., Bancale, A., Boyd, R.N., Chorna, O., Cioni, G., Damiano, D.L., Darrah, J., de Vries, L.S., Dusing, S., Einspieler, C., Eliasson, A.-C., Ferriero, D., Fehlings, D., Forssberg, H., Gordon, A.M., Greaves, S., Guzzetta, A., Hadders-Algra, M., Harbourne, R., Karlsson, P., Krumlinde-Sundholm, L., Latal, B., Loughran-Fowlds, A., Mak, C., Maitre, N., McIntyre, S., Mei, C., Morgan, A., Kakooza-Mwesige, A., Romeo, D.M., Sanchez, K., Spittle, A., Shepherd, R., Thornton, M., Valentine, J., Ward, R., Whittingham, K., Zamany, A., Novak, I., 2021. Early Intervention for Children Aged 0 to 2 Years With or at High Risk of Cerebral Palsy: International Clinical Practice Guideline Based on Systematic Reviews. JAMA Pediatr. 175, 846–858. doi:10.1001/jamapediatrics.2021.0878.

Murphy, V.A., Shen, M.D., Kim, S.H., Cornea, E., Styner, M., Gilmore, J.H., 2020. Extra-axial CSF relationships to infant brain structure, cognitive development, and risk for schizophrenia. Biol. Psychiatry Cogn. Neurosci. Neuroimaging 5, 651. doi:10.1016/J.BPSC.2020.03.008.

Pannek, K., Fripp, J., George, J.M., Fiori, S., Colditz, P.B., Boyd, R.N., Rose, S.E., 2018. Fixel-based analysis reveals alterations is brain microstructure and macrostructure of preterm-born infants at term equivalent age. NeuroImage Clin. 18, 51–59. doi:10.1016/J.NICL.2018.01.003.

Pannek, K., George, J.M., Boyd, R.N., Colditz, P.B., Rose, S.E., Fripp, J., 2020. Brain microstructure and morphology of very preterm-born infants at term equivalent age: Associations with motor and cognitive outcomes at 1 and 2 years. Neuroimage 221, 117163. doi:10.1016/J.NEUROIMAGE.2020.117163.

Peacock, J.L., Marston, L., Marlow, N., Calvert, S.A., Greenough, A., 2012. Neonatal and infant outcome in boys and girls born very prematurely. Pediatr. Res. 71, 305–310. doi:10.1038/pr.2011.50.

Piñon, M., 2010. Theoretical background and structure of the bayley scales of infant and toddler development, 3rd ed.. Bayley-III Clin. Use Interpret. 1–28. 10.1016/B978-0-12-374177-6.10001-7

Roberts, G., Howard, K., Spittle, A.J., Brown, N.C., Anderson, P.J., Doyle, L.W., 2008. Rates of early intervention services in very preterm children with developmental disabilities at age 2 years. J. Paediatr. Child Health 44, 276–280. doi:10.1111/J.1440-1754.2007.01251.X.

Rousseau, F., Oubel, E., Pontabry, J., Schweitzer, M., Studholme, C., Koob, M., Dietemann, J.-L., 2013. BTK: An open-source toolkit for fetal brain MR image processing. Comput. Methods Progr. Biomed. 109, 65. doi:10.1016/J.CMPB.2012.08.007.

Schuh, A., Makropoulos, A., Robinson, E.C., Cordero-Grande, L., Hughes, E., Hutter, J., Price, A.N., Murgasova, M., Teixeira, R.P.A.G., Tusor, N., Steinweg, J.K., Victor, S., Rutherford, M.A., Hajnal, J. V., Edwards, A.D., Rueckert, D., 2018. Unbiased construction of a temporally consistent morphological atlas of neonatal brain development. bioRxiv 251512. 10.1101/251512

Shah, D.K., Anderson, P.J., Carlin, J.B., Pavlovic, M., Howard, K., Thompson, D.K., Warfield, S.K., Inder, T.E., 2006. Reduction in cerebellar volumes in preterm infants: relationship to white matter injury and neurodevelopment at two years of age. Pediatr. Res. 60, 97–102. doi:10.1203/01.pdr.0000220324.27597.f0.

Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155. doi:10.1002/hbm.10062.

Spittle, A.J., Ferretti, C., Anderson, P.J., Orton, J., Eeles, A., Bates, L., Boyd, R.N., Inder, T.E., Doyle, L.W., 2009. Improving the outcome of infants born at <30 weeks' gestation - a randomized controlled trial of preventative care at home. BMC Pediatr. 9, 1–14. doi:10.1186/1471-2431-9-73/TABLES/2.

Spittle, A.J., Spencer-Smith, M.M., Cheong, J.L.Y., Eeles, A.L., Lee, K.J., Anderson, P.J., Doyle, L.W., 2013. General movements in very preterm children and neurodevelopment at 2 and 4 years. Pediatrics 132, e452–e458. doi:10.1542/peds.2013-0177.

Thompson, D.K., Lee, K.J., Egan, G.F., Warfield, S.K., Doyle, L.W., Anderson, P.J., Inder, T.E., 2014. Regional white matter microstructure in very preterm infants: predictors and 7 year outcomes. Cortex 52, 60–74. doi:10.1016/j.cortex.2013.11.010.

Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B 58, 267–288.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29, 1310–1320. doi:10.1109/TMI.2010.2046908.

van't Hooft, J., van der Lee, J.H., Opmeer, B.C., Aarnoudse-Moens, C.S.H., Leenders, A.G.E., Mol, B.W.J., de Haan, T.R., 2015. Predicting developmental outcomes in premature infants by term equivalent MRI: Systematic review and meta-analysis. Syst. Rev. 4, 1–10.

Vanes, L.D., Hadaya, L., Kanel, D., Falconer, S., Ball, G., Batalle, D., Counsell, S.J., Edwards, A.D., Nosarti, C., 2021. Associations between neonatal brain structure, the home environment, and childhood outcomes following very preterm birth. Biol. Psychiatry Glob. Open Sci. 1, 146–155. doi:10.1016/j.bpsgos.2021.05.002.

Vassar, R., Schadl, K., Cahill-Rowley, K., Yeom, K., Stevenson, D., Rose, J., 2020. Neonatal brain microstructure and machine-learning-based prediction of early language development in children born very preterm. Pediatr. Neurol. 108, 86–92. doi:10.1016/j.pediatrneurol.2020.02.007.

Volpe, J.J., 2009. Brain injury in premature infants: a complex amalgam of destructive and developmental disturbances. Lancet Neurol 8, 110–124. doi:10.1016/S1474-4422(08)70294-1.

Yushkevich, P.A., Wang, H., Pluta, J., Avants, B.B., 2012. From label fusion to correspondence fusion: a new approach to unbiased groupwise registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 956–963. doi:10.1109/CVPR.2012.6247771.