ARTICLE IN PRESS

Nature Communications

https://doi.org/10.1038/s41467-026-69657-w

Article in Press

Normative growth trajectories of fetal brain regions validated by satisfactory maturation of neurodevelopmental domains at 2 years of age

Received: 14 July 2025

Accepted: 3 February 2026

Cite this article as: Wyburd, M.K., Kennedy, S.H., Fernandes, M. *et al.* Normative growth trajectories of fetal brain regions validated by satisfactory maturation of neurodevelopmental domains at 2 years of age. *Nat Commun* (2026). https://doi.org/10.1038/s41467-026-69657-w

Madeleine K. Wyburd, Stephen H. Kennedy, Michelle Fernandes, Nicola K. Dinsdale, Linde S. Hesse, Robert B. Gunier, Leila Cheikh Ismail, Eric O. Ohuma, Michael G. Gravett, Manorama Purwar, Wu Qingqing, Adele Winsey, Enrico Bertino, Yasmin Jaffer, Maria Carvalho, Fernando C. Barros, Alan Stein, Alison J. Noble, Zoltán Molnár, Mark Jenkinson, Thomas E. Nichols, Stephen Smith, Zulfiqar A. Bhutta, Aris T. Papageorghiou, Jose Villar & Ana I. L. Namburete

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

ARTICLE IN PRESS

Normative growth trajectories of fetal brain regions validated by satisfactory maturation of neurodevelopmental domains at 2 years of age

Madeleine K. Wyburd PhD[1*°], Prof Stephen H. Kennedy MD[2*], Michelle Fernandes MD[2,3,4], Nicola K. Dinsdale PhD[1] ,Linde S Hesse PhD[1], Robert B Gunier PhD[5], Leila Cheikh Ismail PhD[6], Prof Eric O Ohuma PhD[7], Prof Michael G Gravett MD[8], Prof Manorama Purwar MD[9], Wu Qingqing MD[10], Adele Winsey PhD[2,4], Prof Enrico Bertino MD[11], Yasmin Jaffer MD[12], Maria Carvalho MD[13], Prof Fernando C. Barros[14], Prof Alan Stein FRCPsych[15,16,17], Prof Alison J Noble PhD[18], Prof Zoltán Molnár DPhil[19], Prof Mark Jenkinson[20,21,22], Prof Thomas E Nichols PhD[20,23], Prof Stephen Smith DPhil[20], Prof Zulfiqar A Bhutta PhD[24,25], Prof Aris T Papageorghiou MD[2,4], Prof Jose Villar MD[2,4]†, Prof Ana I. L. Namburete DPhil[1,20]† °

°corresponding author, madeleine.wyburd@cs.ox.ac.uk, ana.namburete@cs.ox.ac.uk
* These authors contributed equally
† These authors jointly supervised this work

[1]Oxford Machine Learning in Neuroimaging Laboratory, Department of Computer Science, University of Oxford, Oxford, United Kingdom
[2]Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford, United Kingdom
[3]Department of Paediatrics, Institute of Developmental and Regenerative Medicine, University of Oxford, Oxford, United Kingdom
[4]Oxford Maternal & Perinatal Health Institute, Green Templeton College, University of Oxford, Oxford, United Kingdom
[5]School of Public Health, University of California, Berkeley, CA, United States of America
[6]Department of Clinical Nutrition and Dietetics, College of Health Sciences, University of Sharjah, Sharjah, United Arab Emirates
[7]Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, United Kingdom
[8]Departments of Obstetrics and Gynecology and of Global Health, University of Washington, Seattle, WA, United States of America
[9]Nagpur INTERGROWTH-21st Research Centre, Ketkar Hospital, Nagpur, India
[10]School of Public Health, Peking University, Beijing, China
[11]Dipartimento di Scienze Pediatriche e dell'Adolescenza, SCDU Neonatologia, Universita di Torino, Turin, Italy
[12]Department of Family and Community Health, Ministry of Health, Muscat, Sultanate of Oman
[13]Department of Obstetrics and Gynaecology, Faculty of Health Sciences, Aga Khan University Hospital, Nairobi, Kenya
[14]Post-graduate Course in Health in the Life Cycle, Universidade Católica de Pelotas, Pelotas, Brazil
[15]Blavatnik School of Government, University of Oxford, Oxford, United Kingdom
[16]African Health Research Institute, KwaZulu-Natal, South Africa
[17]MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[18]Department of Engineering Science, University of Oxford, Oxford, United Kingdom
[19]Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom
[20]Oxford Centre for Integrative Neuroimaging (OxCIN), FMRIB Centre, Nuffield Department of Clinical Neurosciences (NDCN), University of Oxford, JR Hospital, Oxford, United Kingdom
[21]Australian Institute for Machine Learning (AIML), School of Mathematical and Computer Sciences, The University of Adelaide, Adelaide, Australia
[22]South Australian Health and Medical Research Institute (SAHMRI), Adelaide, Australia
[23]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom
[24]Center for Global Child Health, Hospital for Sick Children, Toronto, Ontario, Canada
[25]Institute for Global Health and Development at Aga Khan University, South-Central Asia & East Africa, Karachi, Pakistan

ARTICLE IN PRESS

ARTICLE IN PRESS

We previously constructed a qualitative, 3D ultrasound derived atlas of the normative spatiotemporal dynamics of fetal brain maturation. Here, using the same healthy multi-national cohort, we applied deep learning methods to 4,205 fetal brain scans from 18–27 weeks' gestation, to produce an extensive, quantitative description of the growth of 16 fetal brain structures associated with satisfactory domain-specific neurodevelopmental-scores at 2 years of age. The methodology, which is publicly available, takes less than 10 seconds per scan. We define 28 region-specific, functionally relevant, normative growth trajectories, a ratio between the relative volumes of the insular (rILV) and parietal (rPLV) lobes reflecting asynchronous maturation of fetal brain regions, and introduce a fetal brain maturation index that quantifies biological age and deviations from chronological age. Finally, the very low percentage of variance explained by between site differences (0.6% to 5.8% of the total variance) reinforces a fundamental biological principle: fetal growth and development across populations with diverse ancestries is similar provided that environmental constraints on growth are minimal.

ARTICLE IN PRESS

INTRODUCTION

The use of imaging tools to describe the normative early maturation of human brain structures is essential for monitoring differences over time and linking structural organisation to developmental functions in later life. To date, the most comprehensive description of human brain growth was compiled from over 100 studies that collected magnetic resonance imaging (MRI) data between 115 days post-conception and 100 years of age [1]. Unfortunately, only three studies contributed fetal data to those brain growth charts.

All are limited by their cross-sectional design, small sample size during fetal life, methodological heterogeneity, and lack of postnatal follow-up which means the normative neurodevelopment of the fetal sample cannot be confirmed. In addition, although MRI scans have been performed more than once during pregnancy to produce fetal brain volumetric trajectories [2], healthy women with low-risk pregnancies rarely have repeated MRI scans and certainly not in large numbers. Including only healthy subjects in studies to define normative growth and development (i.e. a prescriptive approach) is essential, as recommended by the World Health

ARTICLE IN PRESS

Organization (WHO) [3]. However, none of the recent brain charts across the human lifespan appear to have included pregnant women from low-risk populations who meet WHO's strict criteria [1,4].

Using the 3D US scans obtained from over 4,000 healthy pregnancies in the international, population-based INTERGROWTH-21st Fetal Growth Longitudinal Study (FGLS), and by following WHO recommendations [3], we recently constructed a qualitative digital atlas of the normative spatiotemporal dynamics of fetal brain maturation [5]. We now report, from the same dataset, a quantitative description of the growth of specific brain regions with corresponding normative ranges. To do so at scale, intracranial structures need to be labelled manually, which is extremely time consuming [6]. Although deep-learning (DL) approaches have been proposed to automate structure labelling [7,8] they do not always work on fetal US images, as large acoustic shadows obstruct key anatomical regions, leading to poor quality segmentations [6].

To avoid these limitations, we apply a DL method that guides structural labelling in regions with acoustic shadows to produce anatomical segmentations of fetal brains [9,10]

. From the standardised FGLS 3D US dataset [5], we labelled 11 brain structures and five cortical regions and extracted 28 image-derived phenotypes (IDPs) that were also used to produce a fetal brain maturation index, which quantitatively describes normative development. We focused on the gestational period between 18 and 27 weeks as we have previously shown that fetal cranial growth trajectories diverge within the 20-25-week gestational age window predicting differential growth and neurodevelopmental outcomes at 2 years of age [11].

RESULTS

Image-derived phenotypes of the fetal brain

A total of 4,321 women, enrolled in FGLS during the first trimester of pregnancy, had live singleton births in the absence of severe maternal conditions or congenital abnormalities detected by US or at birth, all of whom had very accurate estimates of gestational age. (Fig. 1a). Thereafter, 765 (17.7%) of these fetuses were excluded from the analysis due to missing postnatal follow-up data (n= 681) or evidence of severe morbidity (n= 84). Of the remaining

ARTICLE IN PRESS

3,556 fetuses, 2,906 had at least one 3D US scan between 18 and 27 weeks, the gestational period of interest.

To validate the postnatal development at 2 years of age of the normative cohort, we studied a sub-sample of 1,112 children (38.3% of the eligible fetuses) using the INTERGROWTH-21st Neurodevelopmental Assessment (INTER-NDA) tool [12] and Cardiff Visual Acuity and Contrast Sensitivity tests for binocular vision [13].

As we have measured small size differences within different regions of the fetal brain, we adopted a very conservative approach to the analysis of the neurodevelopmental and vision assessments. Thus, we excluded the 3D US scans of the 101 eligible fetuses that scored in the bottom 3% on any one of the INTER-NDA domains or had low scores for visual acuity (<24 at 50cm for 6/m) or sensitivity (<33.3) at 2 years of age.

Hence, to construct normative growth charts, we included a total sample of 2,805 fetuses (prospectively monitored from early pregnancy onwards) that had 4,205 3D US scans in the 9-week period, which was the focus of the present study. Sensitivity analysis confirmed the appropriateness of combining the IDPs of the fetuses with and without a developmental assessment at 2 years of age (Supplementary Table 1 and Supplementary Fig. 1). Finally, the sub-sample (n=1011/2906) and the proportion of the total FGLS population (n=1181/4321) evaluated at 2 years of age were similar on the full set of variables measured (Supplementary Table 2).

The median interval between scanning sessions was 5 weeks (mean 4.61; SD 0.58) as per protocol, and 49.6% (n=1,390) of the women had at least two scans (median 1.0; mean 1.50; SD 0.5; range 1-3 weeks) (Fig. 1b, Supplementary Table 3). Using two DL convolutional networks [9,10], each of the 4,205 3D US scans was labelled into 11 structures. From each scan, the total brain volume (TBV) was extracted across both hemispheres. However, as the proximal hemisphere is frequently obscured by acoustic shadowing from the skull, only the distal hemisphere (i.e. the side furthest from the US probe) was labelled into the cortical plate (CoP), white matter (WM), deep grey matter (DGM), cerebellum (CB), thalamus (Th), lateral posterior

ARTICLE IN PRESS

ventricle horns (LV), choroid plexus (ChP), frontal horns (FH), brainstem (BS) and cavum septum (CSP), as shown in Fig. 2. Each structure's volume was then computed.

As the Sylvian fissure (SF) is the epicentre of change in the second trimester of pregnancy, an additional measure of SF depth (SFD) was included. We further characterised the complex expansion and folding of the CoP, a process driven by neuron migration and differentiation [14], by measuring four CoP components: volume (CoPV), surface area (CoPSA), thickness (CoPT) and depth (CoPD). CoPV, CoPT and CoPD were extended into five functionally delimited lobes: frontal (FL), temporal (TL), parietal (PL), occipital (OL) and insular (IL) (Fig. 3), enabling average regional measures of each lobe.

Thus, in total, 28 IDPs were automatically extracted measuring volume, surface area, cortical depth and thickness from the 4,205 3D US scans of 2,805 fetuses, taking a total of 15 hours on a single Nvidia A10 GPU (Nvidia Corporation, Santa Clara, CA, USA). Table 1 presents an overview of the IDPs studied and their acronyms. Except for TBV and CSPV, the measurement of each structure was possible only on the distal hemisphere (the side furthest from the US probe), as the shape of the skull creates an acoustic cavity that often obstructs the proximal hemisphere with strong shadows and thus, too much structural information is lost to guide the segmentation algorithm.

ARTICLE IN PRESS

Agreement with magnetic resonance imaging studies

Our algorithms labelled structures based on the brain features manually annotated on an US atlas[5] by a researcher (MKW) experienced in the field of brain imaging, and then verified by AILN, the senior investigator. The structures were selected based on overlap with existing MRI atlases [8,15] and those clearly visible in our digital atlas [5].

To ensure the structures were accurately delineated, the volumes of the atlas labels were compared to those from five, publicly available, labelled fetal MRI atlases [8,15-18], as shown in Fig. 4. The volumes of the CoP, CB, Th, WM and CSP corresponded strongly to the previous MRI labels, validating the accuracy of these tissue segmentations. A large difference was seen for the BS volume, which is expected as the US field of view is much higher in the BS (Fig. 4). Differences are also seen in the ventricular system (VS) and DGM. In the US atlas, there is a

ARTICLE IN PRESS

clear separation between the FH, LV and ChP [5], whereas, in MRI studies these structures are usually grouped [8,15-18]. The VS volume varied across the MRI atlases due to different definitions, with some including the third ventricle and others only segmenting the ChP [18]. In the US atlas, the VS only includes the FH, LV and ChP, missing the inferior horn: hence, a lower volume compared to MRI measures is expected, as seen in Fig. 4c. The DGM shows a similar pattern, with inconsistent labelling or grouping between MRI atlases, leading to large uncertainty in DGM volume across the methods.

Brain structural similarities across study sites

The contribution of fetuses (n=2,805) from each site was 11.2% Turin (Italy), 11.4% Oxford (UK), 11.8% Pelotas (Brazil), 14.2% Nairobi (Kenya), 15.6% Muscat (Oman), 17.1% Beijing (China) and 18.8% Nagpur (India). A full breakdown of the subject and scan contribution from each study site is shown in Supplementary Table 3 and Fig. 1c.

To ensure the data could be pooled to enable normative growth trajectories to be constructed, we conducted sensitivity analyses to assess the effect of removing a single site's data, one at a time for each of the 28 IDPs. There were no substantive effects on the remaining pooled sample's 3rd, 50th and 97th centiles for any brain structure. Fig. 5a shows a representative example of the sensitivity analyses derived for TBV by excluding, one at a time, the samples from China, India, Kenya and the UK, whose general populations are popularly believed to differ in anthropometric sizes [19]. Excluding data from each of these four different populations had minimal impact on the centile estimates derived from the remaining pooled sample. Excluding the other countries similarly had little effect on the centiles (data not shown).

Supplementary Table 4 presents the percentage of the total variance attributed to differences amongst study sites (between sites differences) estimated using variance component analysis. Amongst the 28 IDPs, the proportion of the total variance adjusted by sex and gestational age at assessment explained by study site differences was only 1.6% for TBV; for the other intracranial structures, site differences explained less than 5.8% of the total variance for each brain region (Fig. 5b). Consistent with this minimal variance explained, site effects were not statistically significant in the mixed-effects models once sex and gestational age were included.

ARTICLE IN PRESS

ARTICLE IN PRESS

Finally, we calculated the standardised site difference (SSD), which allows for direct comparisons of volumetric brain measures across study sites, standardised by the corresponding pooled SD and adjusted for gestational age. SSD was computed for each of the seven sites and each IDP at three 3-week time-windows: 18^{+0} to 20^{+6}, 21^{+0} to 23^{+6}, and 24^{+0} to 26^{+6} weeks' gestation, to account for longitudinal growth. Out of the possible 588 SSD calculations (Fig. 5c; Supplementary Tables 5), 532 (90.5%) were within -0.5 and 0.5 units of SD, the interval prespecified in the FGLS protocol as adequate for combining data from all sites based on the cut-off point used in the WHO Multicentre Growth Reference Study (MGRS) that produced the prescriptive WHO Child Growth Standards [20]. For TBV, three out of 21 comparisons were outside the -0.5 to 0.5 units of SD; however, these can be explained by the skewed mean average gestational age in each sample within the 3-week time-window, shown in Supplementary Table 5. For instance, the mean gestational age in the Beijing dataset was 18.5 weeks in the 18^{+0} to 20^{+6} weeks' gestation window, a week less than the value for the sites combined (19.5 weeks). Therefore, we would expect the volume for this study site to be smaller than the sites combined. The results of these analyses clearly show that the seven study populations were sufficiently similar for the data to be pooled.

ARTICLE IN PRESS

ARTICLE IN PRESS

Normative fetal brain growth

Scatter plots of observed growth data for each IDP, and the fitted 3rd, 50th and 97th smoothed centiles according to gestational age (in weeks) are shown in Fig. 6 and Fig. 7. To produce these normative equations the IDPs were pooled across fetal sex, cerebral hemisphere, and whether or not neurodevelopment was assessed at 2 years of age. Out of the possible 504 SSD calculations (Supplementary Table 1,6,7), 503 (99.8%) were within -0.5 and 0.5 units of SD. We excluded any measures that were more than 4 SD of all sites' gestational age-specific mean. Any such outlier values were excluded from subsequent analyses to ensure that spurious measures did not bias the growth models. The number of excluded scans per IDP is shown in Supplementary Table 8, alongside the corresponding equations for the mean and SD from the fractional polynomial regression models for each IDP, allowing readers to calculate any centiles or z-scores according to gestational age (in weeks). Table 2 shows the equation for TBV; the remaining 27 equations are shown in Supplementary Table 8.

Gestational week-specific comparisons between empirical centiles and observed values showed excellent agreement (Supplementary Fig. 2 and Supplementary Fig. 3). Overall, there was increasing, non-linear growth over the 9-week study period and the mean differences between smoothed and observed centiles for the 3rd, 50th and 97th centiles were small for all brain structures.

The centiles reveal the trajectories of absolute volumetric growth of the brain structures, which all increased with gestational age. However, when the volume IDPs were normalised by TBV, as shown in Supplementary Fig. 4, the relative size of most structures decreased with advancing gestational age. Notably, the relative volume of the ChP decreased approximately three-fold.

Insular to parietal lobe volume ratio, each relative to the total cortical plate volume

The relative growth of each CoP lobe, normalised by total CoPV, is shown in Supplementary Fig. 5. The relative volume of the IL (rILV) increased with gestational age (Fig 8a), consistent with the opercularisation of the Sylvian fissure from mid-gestation [21-23]. In contrast, the proportional decrease in relative PL volume (rPLV) likely reflects region-specific, temporally hierarchical maturational processes, including dendritic remodeling and synaptic pruning [24-26]. Although

ARTICLE IN PRESS

most synaptic pruning occurs postnatally [27], histological and MRI studies indicate this process, particularly within the transient subplate zone, commences prenatally and displays marked regional heterogeneity [24,28,29].

The subplate and intermediate zones (often collectively segmented as 'white matter' in MRI analyses) exhibit distinct developmental trajectories, reflecting regional differences in neuronal and axonal dynamics that likely underpin the observed relative volumetric changes, particularly in association areas such as the PL [24,29].

As the rPLV progressively declined between 18 and 27 weeks' gestation, the rILV steadily increased, resulting in a concomitant rise in the rILV:rPLV ratio. We have estimated normative values for rILV:rPLV with the corresponding centiles according to gestational age (Fig 8b). Thus, this ratio provides a snapshot of the cortical maturation of the fetal brain and captures the asynchrony in growth between these two cortical lobes.

Fetal brain maturation index

Comparing biological and chronological ages is commonly used in the adult MRI literature to determine if cohorts or individuals are ageing at the expected rate [30], and differences between the two are associated with disease and cognitive decline [30-32].

Here, we compare fetal brain maturation (as a marker of biological age) with our very accurate estimate of gestational (chronological) age by training a random forest to assess each fetus' gestational age at each scan using the 28 IDPs. The method assumes that brain development is consistent across healthy populations, as shown above.

Three additional factors (fetal sex, study site and distal hemisphere) were included as confounding factors in the model. Consistent with our previous findings [5], their contribution to the total variance was minimal (0.50%, 1.06% and 1.25%, respectively), and was therefore regressed out of the data.

The full set of 28 deconfounded features was used as the input for the random forest model, which accurately predicted gestational age with a mean absolute error of only 3.9 days in relation to fetal chronological age. The result was similar to the 3-day accuracy previously

ARTICLE IN PRESS

ARTICLE IN PRESS

achieved at 20-30 weeks' gestation using the 2D US FGLS data and a machine learning approach [33]. Hence, in this dataset of healthy accurately dated fetuses, we demonstrate that biological and chronological age align very closely (R = 0.954; ICC = 0.955), enabling us to produce a fetal brain maturation index based on regional brain structure growth.

The six most informative features that contribute to the model across the gestational age range studied: TBV, CoPSA, WMV, DGMV, ILV and CBV (Supplementary Fig.6). Including only these IDPs in the model led to an agreement between the two predictions of 4.2 days (R = 0.95), indicating that although these are the most informative features, the remaining 22 IDPs still provide a useful indication of maturation.

The contribution of each of the 28 IDPs at the three individual time-points (18, 22 and 26 weeks' gestation) is shown in Supplementary Fig.6.

DISCUSSION

Using DL methods, we have presented here a detailed, quantitative description of the normative growth and maturation of 16 key fetal brain structures during the second trimester of pregnancy; a critical window during which faltering growth of the whole cranium is associated with delayed neurodevelopment at 2 years of age [34].

We also describe, for the first time, a ratio (rILV:rPLV) representative of the temporally asynchronous, relative volumetric growth of the IL and PV lobes between 18 and 27 weeks' gestation. IL US measures in fetuses with early onset growth restriction are reduced compared to appropriately grown fetuses [35], and reduced ILV (measured with MRI) in late onset small-for-gestational-age fetuses is associated with poorer neurobehavioural outcomes in early neonatal life [36]. Hence, given the interconnected functional roles of these two cortical lobes, we speculate that deviation from the normative values of the rILV:rPLV ratio, because of an adverse intrauterine environment, may lead to differences in sensory integration, self-awareness and higher-order cognition in early child development.

In addition, we have produced a fetal brain maturation index that is a learned, weighted summary of all 28 IDPs based upon the features most informative of healthy maturation. We plan to use the index in the future to quantify the degree of departure from chronological age

ARTICLE IN PRESS

ARTICLE IN PRESS

across the participants in the INTERBIO-21st Fetal Study [34] who have different risk profiles at this key stage of brain development.

The inclusion of CoPSA and ILV in the list of informative features is interesting because cortical folding is an important neurodevelopmental process that yields increased CoPSA and ILV, allowing more neurons to be arranged within a limited skull volume, which is associated with increased cognitive abilities [14,37-39]. The emergence and accelerated growth of some secondary fissures, especially SF operculisation, occurs in the second trimester [40,41], the period of study here, and might therefore explain why CoPSA and ILV are so informative of gestational age.

Importantly, the children studied at 2 years of age had satisfactory growth and neurodevelopment particularly in domains involved in language acquisition and comprehension that are associated with FL and TL function. These healthy children, whose intrauterine brain growth was measured during FGLS, were evaluated at 2 years of age as performance then is predictive of intelligence, school attainment, adult nutrition and human capital in high-, middle- and low-income settings [42-46].

Reference charts describing volumetric growth of fetal intracranial structures using MR [47,48] and US [49,50] imaging have been published previously, but from single centres with much smaller sample sizes (n=68 to 344) and only one study included neurodevelopmental assessment at 2 years of age [47]. Interestingly, those authors reported that, in fetuses with congenital heart disease, global fetal brain volumetric growth was the most consistent predictor of the variance attributable to the known risk factors across neurodevelopmental domains at 2 years of age, accounting for 10% to 21% of the total 18% to 45% [47].

In our study, the percentage of variance explained by between-site differences ranged from 0.6% to 5.8% of the total variance. Of the 588 comparisons using SSDs, 91% were between ±0.50 units of the pooled SD for the corresponding item. Therefore, our findings complement those studies of skeletal growth from early pregnancy to 2 years of age [51], newborn size [19], infant neurodevelopment and associated behaviours [52], brain size [5] and child growth [20], which consistently demonstrate, amongst different ethnic populations in diverse regions of the world, that only a very small percentage of the total variance in these measures is explained by

ARTICLE IN PRESS

ARTICLE IN PRESS

differences among populations (Fig. 5b) when constraints on human growth are minimal. In addition, our use of GAMLSS explicitly models non-linear developmental trajectories and variance across gestation, thereby mitigating potential site-by-age confounding effects without the need for additional harmonisation methods.

The manual atlas labels describing average normative growth corresponded well with previous MRI atlases for most structures (Fig. 4), and individual trajectories also overlapped. For example, the median CoPV range of 3 to 18 cm^3 reported here (Fig. 7) is comparable to the 5 to 22 cm^3 range in a previous MRI study [53], with lobar volumes also showing high compatibility. Similar agreement was also seen for CBV and DGMV. Further, this approach was recently validated on an independent dataset of 90 fetal brain ultrasound volumes with paired, same-day MRI[54], where volume measures were highly correlated, with good to excellent agreement in the subcortical structures.

We observed a median cortical plate thickness of 2-3 mm between 18 and 27 weeks, which is consistant with 1.92-2.84 mm reported in an MRI-based study[55]. However, cortical thickness estimates are variable across studies[16,55,56]. Differences in absolute values likely reflect modality- and method-related factors: in ultrasound, cortical plate boundaries are defined by echogenicity, whereas in MRI they are delineated using tissue relaxation contrasts. Such cross-modality variation has also been described for other fetal brain structures[54]. We note, however, that cortical thickness estimates show greater variability than other measures, consistent with previous MRI reports[56], reflecting both methodological differences and the intrinsic difficulty of delineating this boundary during early development. Because cortical thickness was computed voxel-wise from the CoP label, precision is inherently limited by voxel size (±0.6mm in this study), segmentation accuracy, and potential partial-volume effects[57]. In future work, adapting surface-based techniques to operate robustly on single-hemisphere ultrasound data may help address these limitation[58]. Importantly, the developmental trajectories across gestation are consistent with MRI findings[16], supporting the biological validity of our measures.

Importantly, some structures are more clearly distinguished in US compared to MRI [5], enabling us to extract previously unreported IDPs and describe their longitudinal growth for the first time. In US studies, the FH, LV and ChP are reported separately[5], whereas, in MRI studies, these

ARTICLE IN PRESS

ARTICLE IN PRESS

structures are usually grouped [8,15]. Disentangling the VS into FH, LV and ChP, shows that each structure grows at a very different rate (Fig. 6), information previously lost from MRI analysis. Notably, although ChP increases with gestational age, the growth is more gradual than for the other structures, and when normalised by TBV, it actually decreases (Supplementary Fig. 4).[59]

Our automated pipeline analysis extracted 28 IDPs from the fetal brain in less than 10 seconds per 3D US scan, as compared to approximately 20 hours when done manually. Our approach not only segments 11 structures and 5 functionally delineated cortical lobes from each scan, but also performs complex analyses, such as measures of CoPSA, SFD, CoPT, CoPD and CoPSA, which were highly informative of gestational age and could, therefore, be used clinically to estimate gestational age in the second trimester of pregnancy if a first-trimester US scan has not been performed. This potential use is only possible because of the speed achieved by the analysis pipeline, which will be made open access, enabling future research to: a) perform in depth neuroanatomical research from 3D US and b) compare individuals to the normative growth trajectories and fetal brain maturation index presented here.

The strengths of the study were: 1) The DL algorithms delineated the whole brain (Dice overlap = 0.95) and individual brain structures (Dice overlap = 0.78) accurately in less than 10 seconds per scan, overcoming the time-consuming task of manual segmentations previously required to study volumetric growth; 2) The IDPs studied were highly informative of chronological age, demonstrating their potential use as a “fetal maturation index” for the study of less healthy cohorts, provided that gestational age is accurately determined; 3) We adopted the same prescriptive approach as the WHO MGRS, which produced the WHO Child Growth Standards [20], in keeping with the recommendation of the 1995 special WHO Expert Committee [3]. Thus, we prospectively enrolled a cohort of healthy, educated and adequately nourished pregnant women whose risk of adverse maternal and perinatal outcomes (including preterm birth and fetal growth restriction) was low based on their individual clinical profiles and the socioeconomic, demographic and environmental characteristics of the underlying seven geographically diverse populations [19]; 4) We followed a sample of 36% of mother-infant eligible dyads from early pregnancy to 2 years of age using standardised measures, feeding practices and data collection methods. We also, very conservatively, excluded from the analyses 101

ARTICLE IN PRESS

ARTICLE IN PRESS

children that had scores, on any of the INTER-NDA domains, below the 3rd centile of the normative values. Detailed information was also obtained about the socioeconomic, education and environmental backgrounds of the families selected at both population and individual levels [60].

The principal limitation of the study were: 1) We pooled the size measures between hemispheres rather than report values for each hemisphere. However, this was justified as the standardised differences between the hemispheres were within the WHO recommended range for comparing human growth patterns (± 0.5 units of SD) and confirmed by consistent growth trajectories across both hemispheres (Supplementary Fig. 7). Importantly, this strategy is supported by our previous findings showing that the asymmetries in fetal brain structures during the gestational age window studied here are driven by shape, not size[5]; 2) While the segmentation methods were thoroughly tested and verified on a representative subset of scans (2% of the total dataset; 4205 scans in total), their performance was not explicitly verified across the entire dataset. Although DL-based segmentation could in principle hallucinate anatomy in poor-quality scans, we did not observe this. Manual review of 70 representative scans and independent validation in 90 volumes with same-day MRI demonstrated anatomically faithful labels, supports the reliability of our approach [54]. Given the scale of the data and the practical constraints on manual review, this approach reflects a considered and pragmatic balance between methodological rigour and feasibility. The analysis therefore rests on the reasonable assumption that the model generalises comparably across the full dataset, supported by its strong performance on the verified subset.

In summary, we have presented evidence that, when environmental, nutritional and health care constraints on growth are minimal, functionally important areas of the human brain develop similarly across diverse populations in the second trimester of pregnancy, a key period of neurogenesis. The findings are consistent with previously reported similarities in the same cohort in neurodevelopmental outcomes at 2 years of age [52], and skeletal growth during fetal life, at birth and in early childhood [51]. The pooled IDPs enabled us to construct normative growth trajectories, a marker of asynchronous volumetric growth of the IL and PV lobes and a fetal brain maturational index (biological fetal age), which when combined with our published

ARTICLE IN PRESS

ARTICLE IN PRESS

digital atlas [5], that shows the spatiotemporal dynamics of the maturation of the fetal brain (https://intergrowth21.com/research/brain-atlas-project), should be valuable as a tool to investigate factors that modify developmental processes and affect cognitive function in childhood [34].

Finally, our findings reinforce a fundamental biological principle: most of the observed differences in growth and neurodevelopment across populations or regions of the world are primarily due to socioeconomic, educational and class disparities, rather than genetic ancestry [61-63]. The notion that brain size and neurodevelopment are related to skin colour or 'race' [64] is simply false.

METHODS

The INTERGROWTH-21st Study and its ancillary studies were approved by the Oxfordshire Research Ethics Committee "C" (reference: 08/H0606/139), the research ethics committees of the individual participating institutions, as well as the corresponding regional health authorities where the project was implemented. All mothers provided written informed consent for the use of their clinical data.

An overview of the final dataset used here (4,205 scans from 2,805 subjects), including sex, age at scan and site, is provided in Supplementary Table 3.

ARTICLE IN PRESS

Study participants and fetal ultrasound scans

A detailed description of the study protocol has been published elsewhere, including the selection criteria at cluster and individual level [19,65]. In brief, the population-based Fetal Growth Longitudinal Study (FGLS) of the INTERGROWTH-21st Project was conducted between 2009 and 2016 in eight delimited, geographically diverse, urban areas: Pelotas (Brazil), Turin (Italy), Muscat (Oman), Oxford (UK), Seattle (USA), Shunyi County in Beijing (China), the central area of Nagpur (India), and the Parklands suburb of Nairobi (Kenya). However, 3D US scans from the Seattle site were not included in this analysis because of missing follow-up data.

ARTICLE IN PRESS

Participating women, who initiated antenatal care before 14 weeks' gestation, were selected based on WHO criteria for optimal health, nutrition, education and socioeconomic status needed to construct international growth standards [3]. Hence, they had low-risk pregnancies that fulfilled well-defined and strict inclusion criteria at both population and individual levels [19].

The last menstrual period (LMP) was used to calculate gestational age (±7 days) provided that: a) the date was certain; b) the woman had a regular 24-32 day menstrual cycle; c) she had not been using hormonal contraception or breastfeeding in the preceding 2 months, and d) any discrepancy between the gestational ages based on LMP and crown-rump length (CRL), measured by ultrasound at 9+0 to 13+6 weeks from the LMP was ≤7 days, using the formula described by Robinson & Fleming [66]. The CRL technique was standardised across sites and all ultrasonographers were trained uniformly [67].

3D US scans of the fetal head were acquired every 5±1 weeks from 14 weeks' gestation to delivery (i.e. 14-18, 19-23, 24-28, 29-33, 34-38 and 39-42 weeks' gestation) to explore their use for undertaking 2D measurements [68], and for additional exploratory analyses [69,70].

Dedicated ultrasonographers performed the scans using identical, commercially available, equipment (Philips HD-9, Philips Ultrasound, Bothell, WA, USA), with a curvilinear abdominal 3D transducer (V7-3). The 3D US probe was positioned such that the central axial view was collected at the level of the thalami, and the angle of insonation was adjusted to include the entire skull (approximately 70°) for a typical volume acquisition time of 4 seconds. We conducted centralised hands-on training of sonographers, and the Oxford-based Ultrasound Quality Control (QC) Unit regularly carried out site-specific standardisation procedures to ensure proper use of the US equipment and protocol adherence.

FGLS has produced international growth standards for fetal head, abdomen and femur [65]. Crucially, at 2 years of age, these infants had satisfactory growth, neurodevelopmental milestones and associated behaviours [51,52], confirming the cohort's suitability for producing growth standards.

Fetal brain analysis

Fetal Brain Volume Acquisition

ARTICLE IN PRESS

ARTICLE IN PRESS

Post-acquisition processing of the 3D volumes followed a minimal pipeline for processing fetal brain US scans. This involved a previously published sequence of manual and automated procedures.

In brief, the 3D volumes were first visually inspected to ensure they were of good quality and free from motion artifacts and acoustic shadows that occluded the structures of interest, ultimately excluding scans from 6% (259/4321) of the participants (Fig. 1a).

The images were then resampled to an isotropic voxel size of 0.6×0.6×0.6 mm^3 using trilinear interpolation. Across the dataset, initial voxel sizes ranged from 0.14-0.76 mm (SD 0.04) in the x -dimension, 0.30-1.01 mm (SD 0.15) in the y -dimension, and 0.27-1.91 mm (SD 0.23) in the z -dimension, with a median of 0.32×0.51×0.85 mm^3. As the scans were collected at the axial view of the fetal head, the x -direction in the raw volumes consistently corresponds to the axial plane. This was confirmed empirically by examining the rotation parameters from the alignment step across all 4,205 scans, which showed two narrow peaks for separated by π (or 180°), corresponding to fetuses facing left versus right. The narrowness of these peaks indicates only modest deviation from the intended axial orientation (Supplementary Fig. 8). In 3D ultrasound, the reported "slice thickness" reflects voxel spacing in the reconstructed volume, not physical slice separation as in MRI, and therefore does not imply angular misalignment.

This interpolation approach avoids pixel skipping and minimises the risk of aliasing artifacts. Given that most acquisitions had finer in-plane resolution and only modestly coarser through-plane resolution, the resampling primarily involved upsampling rather than downsampling, thereby further reducing the likelihood of aliasing. Voxel size distributions were highly uniform across imaging sites and gestational ages, indicating effective cross-site harmonisation of acquisition protocols (Supplementary Fig. 8).

All images were then intensity-normalised using per-volume min-max scaling, such that the minimum voxel intensity was set to 0 and the maximum to 1, thereby standardising all inputs while preserving relative contrast. Per-volume min-max normalisation results in minor variation in the absolute normalised intensity values of specific tissues across scans, but preserves the relative intra-volume contrast (e.g., between cortical plate and CSF). Tissue intensities were

ARTICLE IN PRESS

found to be relatively consistent across the dataset (see Supplementary Fig. 8), reflecting the standardised scanning protocol and the fact that ultrasound contrast arises from the physical acoustic properties of tissues. Therefore, more complex MRI-style intensity standardisation was not required; simple per-volume min–max normalisation was sufficient to standardise input scale while preserving relative contrast. Importantly, our downstream analysis models are robust to small inter-scan intensity differences and reliably detect tissue boundaries independent of absolute signal levels.

To ensure that all fetal brains were included whole, the images were cropped to 160×160×160 voxels around the centre of the brain. To align the brain to a pre-defined coordinate space, each scan was processed using the deep learning-based BEAN tool [59]. This convolutional neural network automatically localises the brain in the image and outputs seven affine transformation parameters (3 rotation, 3 translation, and 1 isotropic scale factor) to align the scan rigidly.

At each stage, QC was performed to ensure that the processing steps did not visibly distort the brain scans.

Atlas Structural Parcellation Protocol

As there is no established protocol for labelling US-derived structures in the fetal brain, we present the first US protocol for doing so using knowledge from previous fetal MRI parcellations maps [8,15,71,72] and histology studies [73].

We performed parcellation on nine weekly US atlases between 18 and 27 weeks' gestation, and included structures that were clearly visible in these atlases, as described previously [5]. The atlases were derived from a subset of the INTERGROWTH-21st FGLS cohort, which was found to be demographically representative of the full FGLS cohort in terms of maternal characteristics, fetal sex and study site (see Supplementary Tables 3–5 in [5]).

Segmentation was performed on ITK-snap [74] by MKW, a researcher with over 6 years' experience in fetal brain US imaging, and verified by AN. For structures previously segmented in MRI fetal atlases (i.e. CoP, CSP, CB, Th and BS), we used the Computational Radiology Laboratory atlas as a guide for our segmentation maps [15] . For structures not previously described by MRI (i.e. LV, ChP and FH), MKW followed the tissue boundaries and referred to

ARTICLE IN PRESS

histology reports [73] to delineate the structure accurately. The DGM label encompassed the hippocampus, putamen, caudate, amygdala, internal capsules and ventricular zone. The labelling was performed iteratively, with each structure inspected and updated in the axial, coronal and sagittal planes. Thus, the cortical and subcortical structures manually labelled on each atlas were: *CoP, WM, DGM, CB, Th, LV, ChP, FH, BS, and CSP.*

CoP lobe atlas parcellation (i.e., TL, OL, IL,PL,FL) was performed by aligning the CoP atlas masks to the MRI-derived Computational Radiology Laboratory cortical lobe parcellations [15] (see Supplementary Table 9 for description of sulcal regions included within each lobe demarcation), and mislabelled regions were identified by visual inspection and manually corrected (by MKW).

The structures and lobe parcellations were labelled in both hemispheres on the atlas. Fig. 2 shows an example at 22 weeks' gestation.

Population Structural Labelling

The atlas-based segmentation maps were propagated to the subset of the FGLS fetal brain scans from the atlas cohort (number of scans = 404) using the nonlinear registration fields learnt in atlas creation [5]. This inclusion was based on GA at scan and availability of registration fields. The propagated labels were then inspected and manually refined by MKW, when required. In practice, 46 of the 404 propagated scans required correction (16 substantial, 30 minor), yielding a curated labelled set that combined atlas propagation with expert manual review.

The refined labelled dataset was divided into two subsets: $\mathcal{D}_{training}$ for model training and $\mathcal{D}_{testing}$ for evaluation. The split was performed at the subject level, allocating 80% of the data to training (334 scans). The split was randomised while preserving the distribution of scans across gestational weeks and visible hemisphere. The resulting subsets were found to be evenly balanced with respect to study site and newborn sex (Supplementary Table 10).

To perform segmentation, two DL algorithms were trained in a supervised fashion using $\mathcal{D}_{training}$. First, in order to measure TBV, the whole brain was extracted using a U-Net [75]. This network achieved an average Dice overlap of 0.96 ± 0.01 on a previously unseen $\mathcal{D}_{testing}$. This

ARTICLE IN PRESS

Dice value refers specifically to the whole-brain extraction step, and not the individual anatomical labels.

A key challenge in segmenting structures in both 2D and 3D US scans is that visibility of complete tissue boundaries is hampered by the presence of strong acoustic shadows. Consequently, conventional segmentation tools result in topological mistakes, such as isolated segments, gaps, or holes in regions of poor contrast or occlusions. This necessitates manual or semi-automated post-hoc corrective steps to produce complete surfaces for extracting brain measures (e.g. lobe volumes). Here, we extracted topologically correct segmentations of all 10 brain regions from each 3D US volume using the Topology-Enforcing Diffeomorphic Segmentation Network (TEDS-Net) [9,10].

In brief, TEDS-Net is a convolutional neural network that warps a pre-defined set of binary masks (i.e., the priors) to match each target structure of the imaged brain. Each initial binary mask is constructed with the desired topological properties of the structure of interest. In this work, we used the atlas labels at 22 weeks' gestation as the prior shape for each structure. The warping process involves performing a continuous diffeomorphic deformation that resamples the initial binary mask to its corresponding structure within each brain scan. This constrained deformation provides direct topological guarantees by explicitly discouraging the introduction of gaps on the surfaces of the target anatomies.

TEDS-Net was used to directly segment the 10 cortical and subcortical regions delineated on the atlases: CoP, WM, DGM, CB, Th, LV, ChP, FH, BS, and CSP. On the unseen $\mathcal{D}_{testing}$, it achieved an average Dice overlap of 0.78 ± 0.03 across these structures, with per-structure scores provided in Supplementary Fig. 9. Visual inspection of the test dataset's segmentation masks further confirmed that no manual refinement was necessary.

Lobar parcellations (FL, TL, PL, OL, IL) were generated by warping the 22 weeks' gestation atlas-based parcellation templates to each individual scan using the continuous diffeomorphic deformation learnt in TEDS-Net. As these lobar labels were derived from propagated templates rather than independently hand-delineated ground truth, per-lobe Dice coefficients are not

ARTICLE IN PRESS

reported. Dice metrics shown in Supplementary Fig. 9 therefore refer only to the manually delineated anatomical structures.

To ensure segmentation accuracy and exclude hallucination artifacts, we used a two-tier QC strategy: 1) manual review of a representative 2% test subset (≈70 scans) spanning gestational ages and image qualities, and 2) validation on an independent dataset of 90 fetal brain volumes with paired same-day MRI, where US-derived labels closely matched MRI-derived segmentations[54]. Together, these confirm anatomically reliable outputs.

Another challenge in US brain image analysis is that the increasingly calcifying skull bones and their convex shape create an acoustic cavity that, when interacting with the sound waves, cause shadows and reverberation artifacts. These limit the visualisation of both cerebral hemispheres in the brain volumes. As such, TEDS-Net yielded segmentation masks only for the visible hemisphere (distal to the US probe), shown in Table 1. To achieve this, distal hemisphere was used as an input into TEDS-Net, to ensure the correct side was labelled. Hemisphere labelling was performed using a deep learning classifier which achieved an accuracy of 99.9% on the held-out test set. In our dataset, more right hemispheres were imaged at sufficient quality than left (Supplementary Table 10), consistent with previous reports that fetuses more often adopt a rightward head position at this gestational age range[76]. Consequently, each 3D US volume produced one value per brain measure.

Image Derived Phenotypes (IDPs)

From the trained segmentation models, masks and volumes were computed for the entire FGLS population. For example, once the CoP voxels had been segmented from each 3D scan, the global size measures, i.e. volume (V) and surface area (SA), were computed from each segmentation mask generated by TEDS-Net, $\boldsymbol{Y}_i \in \mathbb{R}^3$, from image i. CoPV was calculated by summing the voxels ($\boldsymbol{v}$) labelled as CoP (i.e. CoPV $= \alpha \sum_{\boldsymbol{v} \in S} \boldsymbol{v}$, where $S = \{\boldsymbol{v} : \boldsymbol{Y}_i = 1\}$ and α is the voxel resolution (i.e. isotropic, 1 voxel = 0.6 mm^3)). To facilitate the computation of SA, the CoP segmentation mask was converted into a triangular mesh $\mathcal{M}_i$ using the Lewiner marching cube algorithm (f_L), which, crucially, also preserves the topology of the underlying structure [77]. CoPA was estimated by summing the areas of all triangular faces of the mesh.

ARTICLE IN PRESS

CoPD was defined as the closest distance between the skull boundary and each position within the CoP. To measure the distance, we computed the Euclidean distance transform of the TBV label and recorded the distances at each position within the CoP label. Mathematically this can be written as:

$$\mathbf{CoPD}_i = \text{dist_transf}(\mathbf{TBV}_i) \times \mathbf{CoP}_i \tag{1}$$

$$\text{dist_transf}(\mathbf{TBV}_i) = \min_{p \in \mathbf{TBV}_i} \sqrt{(q_x - p_x)^2 + \left(q_y - p_y\right)^2 + (q_z - p_z)^2} \tag{2}$$

where q represents the locations of background voxels and p are the locations within TBV.

CoPT was measured at each point along the cortical surface by computing the orthogonal projection from the surface to the nearest boundary point along the local gradient vector, inspired by previous voxel-wise techniques for estimating CoPT in MRI[78,79]. This voxel-wise distance transform approach achieves the same goal as MRI-based cortical thickness estimation but uses a different implementation since surface-based potential field models commonly applied in MRI cannot be used reliably in ultrasound. To validate these measures, we confirmed that CoPT estimates were within the 1.5-3.0 mm range reported in MRI atlases for 18-27 weeks' gestation[16,55] and showed consistent age-related trends across the cohort.

To generate growth standards for brain regions of functional relevance, the CoPV and average CoPT and CoPD was found at each parcellation lobe. To approximate SFD, instead of measuring the average depth in the insula, with incorporates both the inner and outer boundaries, we measured the median depth in the insula.

MRI Atlas Comparison

To validate the structural labelling, the atlas labels were compared to five publicly available, labelled MRI atlases, spanning a similar gestational age range [8,15,17,18,80]. An overview of each atlas used is provided in Supplementary Table 11. As different structures are labelled across the atlases, ranging between 4 and 124 labels, a common set of labels was created by combing the

ARTICLE IN PRESS

available structures (described in Supplementary Table 12 and shown in Fig. 4a). In US images, the FH, LV and ChP have clear tissue boundaries; however, the distinction between structures is lost in MR images resulting in this ventricular system (VS) being segmented as one structure. Thus, to compare the VS volume across the five atlases, the US-derived FH, LW and ChP labels were combined across both hemispheres and the volume of each structure computed.

Fetal Brain Maturation Index

To compare biological and gestational ages from the IDPs, we largely followed the methodology presented by Smith et al. [30] for the prediction of Brain Age Delta from adult brain MRI data from the UK Biobank. All subsequent procedures used three-fold cross-validation, split at the subject level, with 1,938 subjects used for training and validation and 968 held-out subjects used for testing.

We used the 28 extracted IDPs to construct a feature matrix $\boldsymbol{X}$. The primary goal was to predict gestational age from this feature matrix, denoted as $\hat{y} = f(\boldsymbol{X})$, where f represents the predictive model. Each IDP was normalised to have zero mean and unit variance [81]. Subsequently, confounding variables were regressed out from the IDPs, following the approach detailed by Alfaro-Almagro *et al.* (2021) [82]. In our dataset, the set of identified confounds included sex, study site, and visible hemisphere. Notably, age-dependent confounds could not be removed due to their intrinsic relationship with the target variable.

Each confound was numerically encoded and then demeaned. These confounds collectively regressed out from the feature matrix $\boldsymbol{X}$ using a linear model without an intercept (as all confound variables were demeaned):

$$\widehat{\boldsymbol{X}} = \boldsymbol{X} - \boldsymbol{V}\boldsymbol{V}^{+}\boldsymbol{X} \qquad (\mathbf{3})$$

Here, $\boldsymbol{V}$ is the matrix of demeaned confounds, and $\boldsymbol{V}^{+}$ denotes the Moore-Penrose pseudo-inverse of $\boldsymbol{V}$. This regression-based unconfounding was performed simultaneously for all confound variables together.

ARTICLE IN PRESS

To quantify the influence of each individual confound, we calculated the percentage variance explained ($\%VE$) and the percentage unique variance explained ($\%UVE$). Specifically, for a given confound $\boldsymbol{V}_1$, we define the predicted component of $\boldsymbol{X}$ explained solely by this confound (if acting alone) as:

$$\widehat{\boldsymbol{X}}_1 = \boldsymbol{V}_1 \boldsymbol{V}_1^{+} \boldsymbol{X} \tag{4}$$

where $\boldsymbol{V} = [\boldsymbol{V}_1\ \boldsymbol{V}_n]$ is the matrix of all confounds and $\boldsymbol{V}_n$ are the remaining confounds. The percentage variance explained by $\boldsymbol{V}_1$ for each IDP (column j in $\boldsymbol{X}$) was then computed as:

$$\%VE_j = 100\ \times \frac{\widehat{\boldsymbol{X}}_{1,j}{}^{T} \widehat{\boldsymbol{X}}_{1,j}}{\boldsymbol{X}_j^T \boldsymbol{X}_j} \tag{5}$$

Similarly, the percentage unique variance explained ($\%UVE$) for each IDP (column j), representing the variance explained by confound $\boldsymbol{V}_1$ not already accounted for by other confounds $\boldsymbol{V}_n$, was computed as:

$$\%UVE_j = 100\ \times \frac{\boldsymbol{X}_j^T \boldsymbol{X}_j - \widehat{\boldsymbol{X}}_{n,j}{}^{T} \widehat{\boldsymbol{X}}_{n,j}}{\boldsymbol{X}_j^T \boldsymbol{X}_j} \tag{6}$$

where $\widehat{\boldsymbol{X}}_{n,j} = \boldsymbol{V}_n \boldsymbol{V}_n^{+} \boldsymbol{X}_j$ represents the adjusted component of feature j explained only by the remaining confounds $\boldsymbol{V}_n$. Although each confound individually accounted for a small portion of the variance in $\boldsymbol{X}$ (sex: 0.485%, hemisphere: 1.71%, site: 1.28%), all were removed to avoid arbitrary thresholding in confound selection.

The fully normalised and deconfounded feature matrix $\widehat{\boldsymbol{X}}$ served as input to a Random Forest model [83] to regress gestational age, implemented using the Scikit-learn implementation [84]. The model comprised 100 estimators with a maximum feature ratio of 0.15.

ARTICLE IN PRESS

ARTICLE IN PRESS

To identify the most informative IDPs contributing to the fetal brain maturational index, we examined feature importance measures derived from the Random Forest model. This included the mean and standard deviation of impurity decreases within each tree. Additionally, we used SHAP (Shapley Additive exPlanation) values [85] to assess the impact of each feature on individual predictions, thereby evaluating the consistency of feature importance across the gestational age range.

Statistical Analysis

To investigate the relationship between specific brain volumes (TBV and all 15 brain regions) and gestational age, regional brain growth was modelled as a function of gestational age. To account for the increasing inter-subject variability with gestational age, the mean and standard deviation (SD) were modelled separately using fractional polynomial regression within the GAMLSS framework in R. This approach also allows for modelling of higher-order distributional parameters, such as skewness and kurtosis, thereby accommodating non-Gaussian distributions and providing robust estimations of centiles.

Based on our previous publications and WHO recommendations for comparing the similarities in growth across populations [19,20,65] we used variance component analysis to calculate the percentage of total variance due to between-study site differences for all brain measures. Variance components were calculated by nonparametric generalised linear mixed models (using the R lme4), which included as covariates sex and gestational age at assessment. These variables were included as fixed effects with study site as random effect. The method was chosen because it assumes a discrete instead of a Gaussian distribution of the random effects and showed better fit than Maximum Likelihood and Restricted Maximum Likelihood models according to the lower Akaike information and Bayesian information criteria.

In addition, for each brain measure, SSDs were calculated as the difference between the mean from a given site and the mean of all sites together, expressed as a proportion of the SD of the pooled data across sites for each measure. As pre-specified in the INTERGROWTH-21st Project protocol and recommended by WHO [20], SSD values in the range of ± 0.5 units of the pooled SD were taken as adequate to determine whether the data from all sites could be pooled.

ARTICLE IN PRESS

ARTICLE IN PRESS

Infant follow-up

Across all sites, standardised clinical care and feeding practices were implemented using the INTERGROWTH-21st Neonatal Group protocols (www.intergrowth21.org.uk). Exclusive breastfeeding up to 6 months and appropriate nutritional support for infants born preterm were promoted during and after pregnancy, as recommended [86]. Detailed information was obtained from the mother at age 1 and 2 about the infant's health, severe morbidities, hospitalisations, length of breastfeeding, timing of the introduction of solid food, age at weaning, feeding practices and food intake, using specially produced forms (www.intergrowth21.org.uk). The proportion of infants receiving breast milk, and vitamin and mineral supplements, and those following a special diet were estimated at age 1 and 2. Similarly, at age 1 and 2, the infant's weight, length and head circumference were measured following WHO protocols, and their age- and sex-specific z-scores and centiles were compared to the WHO Child Growth Standards [20]. These anthropometric measures, as indicators of general nutrition at age 2, are strongly predictive of later attained height, development and human capital [44].

For the neurodevelopmental evaluation, children were scheduled for assessment at 2 years of age in five of the eight original sites: Nagpur, Nairobi, Oxford, Pelotas and Turin using tools specifically developed or selected for this purpose [87]. The sites in Beijing, Muscat and Seattle did not participate because of logistical and administrative reasons, delays in the start of the study and/or staff availability.

We assessed neurodevelopment using the INTERGROWTH-21st Neurodevelopmental Assessment (INTER-NDA; www.inter-nda.com), an international, psychometrically valid and reliable, standardised tool, targeted at children aged 22-30 months, which measures multiple dimensions of early development using a combination of directly administered, concurrently observed and caregiver reported items [87]. It was designed to be implemented by non-specialists across multinational settings, and includes a reduced number of culture-specific items comprising six domains measuring cognition, language, fine and gross motor skills, and positive and negative behaviour, in an assessment time of 15 min on average.

ARTICLE IN PRESS

ARTICLE IN PRESS

The INTER-NDA has been validated against the Bayley Scales of Infant Development III edition [88]. Based on established guidelines [89], it showed 'good' agreement with interclass coefficient correlations across domains ranging between 0.75 and 0.88. Its norms are the first international standards of early child development, constructed according to the prescriptive WHO approach using data from five of the eight original INTERGROWTH-21st study sites [12]. To date, over 40,000 children in 26 countries have been assessed using the INTER-NDA.

Vision was assessed using the Cardiff Visual Acuity and Contrast Sensitivity tests for binocular vision [13]. These are indicative of the integrity of the central visual pathway, and as directly observed neurodevelopmental markers, are unlikely to be affected by cultural influences and co-occurring disturbances in cognitive, hearing and language skills.

Motor development was assessed against four WHO milestones that are less likely to be affected by recall bias: sitting without support, hands-and-knees crawling, standing alone, and walking alone [90]. Trained staff collected the data on a form with pictures of the relevant child positions and corresponding definitions. Parents were asked to report the age in months and weeks when they "first observed" or "never observed" the milestones. We assessed the age (in months) at which WHO gross motor milestones were first achieved.

All INTER-NDA assessors were trained and standardised centrally. All assessors were subject to a protocol adherence and reliability assessment following training; only those with protocol adherence scores in excess of 90% and inter-rater reliability of >0.8 conducted assessments. The administration of the above tests was supported by a tablet-based data collection and management system. Field staff were unaware of the INTER-NDA domain and total scores for individual children and sites. Data were uploaded onto secure servers as soon as each assessment was completed and compared to the international normative values established by the INTERGROWTH-21st Project [12].

DATA AVAILABILITY

The labeled ultrasound atlases will be made publicly available on https://intergrowth21.com/research/brain-atlas-project. The derived growth curves are

ARTICLE IN PRESS

available in the Supplementary materials. Owing to the data still being under analysis for the principal and secondary objectives of the study protocol, anonymized image data will be made available on reasonable request for academic use only and within the limitations of the informed consent. Requests must be made to the corresponding author or to the INTERGROWTH-21st Consortium at intergrowth@wrh.ox.ac.uk. Full conditions of access are available in the INTERGROWTH-21st study protocol at https://intergrowth21.com/research/brain-atlas-project. Every request will be reviewed by the INTERGROWTH-21st Consortium Executive Committee with due promptness. After approval, the researcher will need to sign a data access agreement with the INTERGROWTH-21st Consortium.

CODE AVAILABILITY

The complete data analysis pipeline is available here: https://oxford-omni-lab-org.github.io/OMNI_ultrasound/NormativeGrowthTrajec/index.html

All ultrasound processing was performed on Python 3.11, with the deep-learning implementation using Pytorch 2.1.0. Statistical analysis was performed in R-Studio (version 2024.04.2+764) using GAMLSS (version 5.4-22).

ARTICLE IN PRESS

REFERENCES

1 Bethlehem, R. A. I. *et al.* Brain charts for the human lifespan. *Nature* **604**, 525-533 (2022). https://doi.org/10.1038/s41586-022-04554-y

2 Studholme, C., Kroenke, C. D. & Dighe, M. Motion corrected MRI differentiates male and female human brain growth trajectories from mid-gestation. *Nat Commun* **11**, 3038 (2020). https://doi.org/10.1038/s41467-020-16763-y

3 de Onis, M. & Habicht, J. P. Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee. *The American journal of clinical nutrition* **64**, 650-658 (1996).

4 Sun, L. *et al.* Human lifespan changes in the brain's functional connectome. *Nat Neurosci* **28**, 891-901 (2025). https://doi.org/10.1038/s41593-025-01907-4

5 Namburete, A. I. L. *et al.* Normative spatiotemporal fetal brain maturation with satisfactory development at 2 years. *Nature* **623**, 106-114 (2023). https://doi.org/10.1038/s41586-023-06630-3

ARTICLE IN PRESS

6 Wyburd, M. K., Jenkinson, M. & Namburete, A. I. in *Annual Conference on Medical Image Understanding and Analysis.* 56-68 (Springer).

7 Hesse, L. S. *et al.* Subcortical segmentation of the fetal brain in 3D ultrasound using deep learning. *NeuroImage* **254**, 119117 (2022).

8 Uus, A. U. *et al.* BOUNTI: Brain vOlumetry and aUtomated parcellatioN for 3D feTal MRI. *elife* **12:RP88818** (2023). https://doi.org/https://doi.org/10.7554/eLife.88818.1

9 Wyburd, M. K., Dinsdale, N. K., Namburete, A. I. & Jenkinson, M. in *International Conference on Medical Image Computing and Computer-Assisted Intervention.* 250-260 (Springer).

10 Wyburd, M. K., Dinsdale, N. K., Jenkinson, M. & Namburete, A. I. L. Anatomically plausible segmentations: Explicitly preserving topology through prior deformations. *Med Image Anal* **97**, 103222 (2024). https://doi.org/10.1016/j.media.2024.103222

11 Villar, J. *et al.* Association between fetal abdominal growth trajectories, maternal metabolite signatures early in pregnancy, and childhood growth and adiposity: prospective observational multinational INTERBIO-21st fetal study. *Lancet Diabetes Endocrinol* **10**, 710-719 (2022). https://doi.org/10.1016/S2213-8587(22)00215-7

12 Fernandes, M. *et al.* INTERGROWTH-21st Project international INTER-NDA standards for child development at 2 years of age: an international prospective population-based study. *BMJ Open* **10:e035258** (2020). https://doi.org/10.1136/bmjopen-2019-035258

13 Adoh, T. O., Woodhouse, J. M. & Oduwaiye, K. A. The Cardiff Test: a new visual acuity test for toddlers and children with intellectual impairment. A preliminary report. *Optom Vis Sci* **69**, 427-432 (1992).

14 Striedter, G. F., Srinivasan, S. & Monuki, E. S. Cortical folding: when, where, how, and why? *Annu Rev Neurosci* **38**, 291-307 (2015). https://doi.org/10.1146/annurev-neuro-071714-034128

15 Gholipour, A. *et al.* A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth. *Sci Rep* **7**, 476 (2017). https://doi.org/10.1038/s41598-017-00525-w

16 Xu, X. *et al.* Spatiotemporal atlas of the fetal brain depicts cortical developmental gradient. *Journal of Neuroscience* **42**, 9435-9449 (2022).

17 Wu, J. *et al.* Age-specific structural fetal brain atlases construction and cortical development quantification for chinese population. *Neuroimage* **241**, 118412 (2021).

18 Serag, A. *et al.* Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. *Neuroimage* **59**, 2255-2265 (2012).

19 Villar, J. *et al.* The likeness of fetal growth and newborn size across non-isolated populations in the INTERGROWTH-21st Project: the Fetal Growth Longitudinal Study and Newborn Cross-Sectional Study. *Lancet Diabetes Endocrinol* **2**, 781-792 (2014). https://doi.org/10.1016/S2213-8587(14)70121-4

20 de Onis, M., Garza, C., Onyango, A. W. & Martorell, R. WHO Child Growth Standards. *Acta Paediatr Suppl* **450**, 1-101 (2006).

21 Pistorius, L. R. *et al.* Grade and symmetry of normal fetal cortical development: a longitudinal two- and three-dimensional ultrasound study. *Ultrasound Obstet Gynecol* **36**, 700-708 (2010). https://doi.org/10.1002/uog.7705

ARTICLE IN PRESS

ARTICLE IN PRESS

22 Afif, A., Bouvier, R., Buenerd, A., Trouillas, J. & Mertens, P. Development of the human fetal insular cortex: study of the gyration from 13 to 28 gestational weeks. *Brain Struct Funct* **212**, 335-346 (2007). https://doi.org/10.1007/s00429-007-0161-1

23 Chi, J. G., Dooling, E. C. & Gilles, F. H. Gyral development of the human brain. *Ann Neurol* **1**, 86-93 (1977). https://doi.org/10.1002/ana.410010109

24 Vasung, L. *et al.* Spatiotemporal Differences in the Regional Cortical Plate and Subplate Volume Growth during Fetal Development. *Cereb Cortex* **30**, 4438-4453 (2020). https://doi.org/10.1093/cercor/bhaa033

25 Huttenlocher, P. R. & Dabholkar, A. S. Regional differences in synaptogenesis in human cerebral cortex. *J Comp Neurol* **387**, 167-178 (1997). https://doi.org/10.1002/(sici)1096-9861(19971020)387:2<167::aid-cne1>3.0.co;2-z

26 Innocenti, G. M. & Price, D. J. Exuberance in the development of cortical networks. *Nat Rev Neurosci* **6**, 955-965 (2005). https://doi.org/10.1038/nrn1790

27 Rakic, P., Bourgeois, J. P., Eckenhoff, M. F., Zecevic, N. & Goldman-Rakic, P. S. Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. *Science* **232**, 232-235 (1986). https://doi.org/10.1126/science.3952506

28 Kostovic, I. & Rakic, P. Developmental history of the transient subplate zone in the visual and somatosensory cortex of the macaque monkey and human brain. *J Comp Neurol* **297**, 441-470 (1990). https://doi.org/10.1002/cne.902970309

29 Xu, X. *et al.* Spatiotemporal Atlas of the Fetal Brain Depicts Cortical Developmental Gradient. *J Neurosci* **42**, 9435-9449 (2022). https://doi.org/10.1523/JNEUROSCI.1285-22.2022

30 Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E. & Miller, K. L. Estimation of brain age delta from brain imaging. *Neuroimage* **200**, 528-539 (2019).

31 Cole, J. H. & Franke, K. Predicting age using neuroimaging: innovative brain ageing biomarkers. *Trends in neurosciences* **40**, 681-690 (2017).

32 Millar, P. R. *et al.* Predicting brain age from functional connectivity in symptomatic and preclinical Alzheimer disease. *Neuroimage* **256**, 119228 (2022).

33 Fung, R. *et al.* Achieving accurate estimates of fetal gestational age and personalised predictions of fetal growth based on data from an international prospective cohort study: a population-based machine learning study. *Lancet Digit Health* **2**, e368-e375 (2020). https://doi.org/10.1016/S2589-7500(20)30131-X

34 Villar, J. *et al.* Fetal cranial growth trajectories are associated with growth and neurodevelopment at 2 years of age: INTERBIO-21st Fetal Study. *Nat Med* **27**, 647-652 (2021). https://doi.org/10.1038/s41591-021-01280-2

35 Xue, J. *et al.* Ultrasound assessment of insular development in adequate-for-gestational-age fetuses and fetuses with early-onset fetal growth restriction using 3D-ICRV technology. *Front Med (Lausanne)* **11**, 1393115 (2024). https://doi.org/10.3389/fmed.2024.1393115

36 Egana-Ugrinovic, G., Sanz-Cortes, M., Figueras, F., Couve-Perez, C. & Gratacos, E. Fetal MRI insular cortical morphometry and its association with neurobehavior in late-onset small-for-gestational-age fetuses. *Ultrasound Obstet Gynecol* **44**, 322-329 (2014). https://doi.org/10.1002/uog.13360

ARTICLE IN PRESS

ARTICLE IN PRESS

37 Stiles, J. & Jernigan, T. L. The Basics of Brain Development. *Neuropsychology review* **20**, 327-348 (2010). https://doi.org/10.1007/s11065-010-9148-4

38 Sun, T. & Hevner, R. F. Growth and folding of the mammalian cerebral cortex: from molecules to malformations. *Nature reviews. Neuroscience* **15**, 217-232 (2014). https://doi.org/10.1038/nrn3707

39 Zilles, K., Armstrong, E., Moser, K. H., Schleicher, A. & Stephan, H. Gyrification in the Cerebral Cortex of Primates. *Brain, behavior and evolution* **34**, 143-150 (1989). https://doi.org/10.1159/000116500

40 Cohen-Sacher, B., Lerman-Sagie, T., Lev, D. & Malinger, G. Sonographic developmental milestones of the fetal cerebral cortex: a longitudinal study. *Ultrasound in obstetrics & gynecology* **27**, 494-502 (2006). https://doi.org/10.1002/uog.2757

41 Chi, J. G., Dooling, E. C. & Gilles, F. H. Gyral development of the human brain. *Annals of neurology* **1**, 86-93 (1977). https://doi.org/10.1002/ana.410010109

42 Daniels, M. C. & Adair, L. S. Growth in young Filipino children predicts schooling trajectories through high school. *The Journal of nutrition* **134**, 1439-1446 (2004).

43 Feinstein, L. Inequality in the early cognitive development of British children in the 1970 cohort. *Economica* **70**, 73-97 (2003).

44 Victora, C. G. *et al.* Maternal and child undernutrition: consequences for adult health and human capital. *The lancet* **371**, 340-357 (2008).

45 Mendez, M. A. & Adair, L. S. Severity and timing of stunting in the first two years of life affect performance on cognitive tests in late childhood. *The Journal of nutrition* **129**, 1555-1562 (1999).

46 Glewwe, P., Jacoby, H. G. & King, E. M. Early childhood nutrition and academic achievement: a longitudinal analysis. *Journal of public economics* **81**, 345-368 (2001).

47 Sadhwani, A. *et al.* Fetal Brain Volume Predicts Neurodevelopment in Congenital Heart Disease. *Circulation (New York, N.Y.)* **145**, 1108-1119 (2022). https://doi.org/10.1161/CIRCULATIONAHA.121.056305

48 Ren, J.-Y. *et al.* Quantification of intracranial structures volume in fetuses using 3-D volumetric MRI: normal values at 19 to 37 weeks' gestation. *Frontiers in Neuroscience* **16**, 886083 (2022).

49 Roelfsema, N. M., Hop, W. C., Boito, S. M. & Wladimiroff, J. W. Three-dimensional sonographic measurement of normal fetal brain volume during the second half of pregnancy. *American journal of obstetrics and gynecology* **190**, 275-280 (2004).

50 Babucci, G., Rosen, K., Cappuccini, B. & Clerici, G. 3D evaluation of fetal brain structures: reference values and growth curves. *The Journal of Maternal-Fetal & Neonatal Medicine* **34**, 3546-3551 (2021).

51 Villar, J. *et al.* The satisfactory growth and development at 2 years of age of the INTERGROWTH-21st Fetal Growth Standards cohort support its appropriateness for constructing international standards. *Am J Obstet Gynecol* **218**, S841-S854 e842 (2018). https://doi.org/10.1016/j.ajog.2017.11.564

52 Villar, J. *et al.* Neurodevelopmental milestones and associated behaviours are similar among healthy children across diverse geographical locations. *Nat Commun* **10**, 511 (2019). https://doi.org/10.1038/s41467-018-07983-4

ARTICLE IN PRESS

ARTICLE IN PRESS

53 Studholme, C., Kroenke, C. D. & Dighe, M. Motion corrected MRI differentiates male and female human brain growth trajectories from mid-gestation. *Nature communications* **11**, 3038 (2020).

54 Wyburd, M. K. *et al.* Cross-Modality Comparison of Fetal Brain Phenotypes: Insights From Short-Interval Second-Trimester MRI and Ultrasound Imaging. *Human Brain Mapping* **46**, e70349 (2025).

55 Corbett-Detig, J. *et al.* 3D global and regional patterns of human fetal subplate growth determined in utero. *Brain Structure and Function* **215**, 255-263 (2011).

56 Xu, F. *et al.* Morphological development trajectory and structural covariance network of the human fetal cortical plate during the early second trimester. *Cerebral Cortex* **31**, 4794-4807 (2021).

57 Clarkson, M. J. *et al.* A comparison of voxel and surface based cortical thickness estimation methods. *Neuroimage* **57**, 856-865 (2011).

58 Jones, S. E., Buchbinder, B. R. & Aharon, I. Three-dimensional mapping of cortical thickness using Laplace's equation. *Human brain mapping* **11**, 12-32 (2000).

59 Moser, F., Huang, R., Consortium, I.-., Papiez, B. W. & Namburete, A. I. L. BEAN: Brain Extraction and Alignment Network for 3D Fetal Neurosonography. *Neuroimage* **258**, 119341 (2022). https://doi.org/10.1016/j.neuroimage.2022.119341

60 Villar, J. *et al.* The objectives, design and implementation of the INTERGROWTH-21st Project. *BJOG: An International Journal of Obstetrics & Gynaecology* **120**, 9-26 (2013).

61 Graham, G. N. Why your ZIP code matters more than your genetic code: promoting healthy outcomes from mother to child. *Breastfeeding Medicine* **11**, 396-397 (2016).

62 Noble, K. G. *et al.* Family income, parental education and brain structure in children and adolescents. *Nature neuroscience* **18**, 773-778 (2015).

63 Heckman, J. J. The economics, technology, and neuroscience of human capability formation. *Proceedings of the national Academy of Sciences* **104**, 13250-13255 (2007).

64 Rushton, J. P. & Rushton, E. W. Brain size, IQ, and racial-group differences: Evidence from musculoskeletal traits. *Intelligence* **31**, 139-155 (2003).

65 Papageorghiou, A. T. *et al.* International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. *Lancet* **384**, 869-879 (2014). https://doi.org/10.1016/S0140-6736(14)61490-2

66 Robinson, H. P. & Fleming, J. E. A critical evaluation of sonar "crown-rump length" measurements. *Br J Obstet Gynaecol* **82**, 702-710 (1975).

67 Ioannou, C., Sarris, I., Hoch, L., Salomon, L. & Papageorghiou, A. Standardisation of crown-rump length measurement. *BJOG : an international journal of obstetrics and gynaecology* **120 Suppl 2**, 38-41 (2013). https://doi.org/10.1111/1471-0528.12056

68 Sarris, I. *et al.* Fetal biometry: how well can offline measurements from three-dimensional volumes substitute real-time two-dimensional measurements? *Ultrasound Obstet Gynecol* **42**, 560-570 (2013). https://doi.org/10.1002/uog.12410

69 Rodriguez-Sibaja, M. J. *et al.* Fetal cerebellar growth and Sylvian fissure maturation: international standards from Fetal Growth Longitudinal Study of INTERGROWTH-21st Project. *Ultrasound Obstet Gynecol* **57**, 614-623 (2021). https://doi.org/10.1002/uog.22017

ARTICLE IN PRESS

ARTICLE IN PRESS

70 Napolitano, R. *et al.* International standards for fetal brain structures based on serial ultrasound measurements from Fetal Growth Longitudinal Study of INTERGROWTH-21st Project. *Ultrasound Obstet Gynecol* **56**, 359-370 (2020). https://doi.org/10.1002/uog.21990

71 Khalili, N. *et al.* Automatic brain tissue segmentation in fetal MRI using convolutional neural networks. *Magnetic resonance imaging* **64**, 77-89 (2019).

72 Payette, K. *et al.* Fetal brain tissue annotation and segmentation challenge results. *Medical image analysis* **88**, 102833 (2023).

73 Bayer, S. A. & Altman, J. *The Human Brain during the Second Trimester 96–to 150–mm Crown-Rump Lengths: Atlas of Human Central Nervous System Development, Volume 8*. (CRC Press, 2023).

74 Yushkevich, P. A. *et al.* User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **31**, 1116-1128 (2006). https://doi.org/10.1016/j.neuroimage.2006.01.015

75 Ronneberger, O., Fischer, P. & Brox, T. in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18.* 234-241 (Springer).

76 Ververs, I. A., de Vries, J. I., van Geijn, H. P. & Hopkins, B. Prenatal head position from 12-38 weeks. I. Developmental aspects. *Early Hum Dev* **39**, 83-91 (1994). https://doi.org/10.1016/0378-3782(94)90157-0

77 Lewiner, T., Lopes, H., Wilson Vieira, A. & Tavares, G. Efficient implementation of marching cubes cases with topological guarantees. *J Graph Tools* **8**, 1-15 (2003).

78 Miller, M. I., Massie, A. B., Ratnanather, J. T., Botteron, K. N. & Csernansky, J. G. Bayesian construction of geometrically based cortical thickness metrics. *NeuroImage* **12**, 676-687 (2000).

79 Hutton, C., De Vita, E., Ashburner, J., Deichmann, R. & Turner, R. Voxel-based cortical thickness measurements in MRI. *Neuroimage* **40**, 1701-1710 (2008).

80 Uus, A. *et al.* Multi-channel spatio-temporal mri atlas of the normal fetal brain development from the developing human connectome project. *G-Node doi* **10**, 2023 (2023).

81 Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature neuroscience* **19**, 1523-1536 (2016).

82 Alfaro-Almagro, F. *et al.* Confound modelling in UK Biobank brain imaging. *NeuroImage* **224**, 117002 (2021).

83 Breiman, L. Random forests. *Machine learning* **45**, 5-32 (2001).

84 Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **12**, 2825-2830 (2011).

85 Lundberg, S. M. *et al.* Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* **2**, 749-760 (2018). https://doi.org/10.1038/s41551-018-0304-0

86 Cheikh Ismail, L. *et al.* Preterm feeding recommendations are achievable in large-scale research studies. *BMC Nutr* **2** (2016).

87 Fernandes, M. *et al.* The INTERGROWTH-21st Project Neurodevelopment Package: A novel method for the multi-dimensional assessment of neurodevelopment in pre-school

ARTICLE IN PRESS

ARTICLE IN PRESS

age children. *PloS one* **9**, e113360 (2014). https://doi.org/10.1371/journal.pone.0113360

88 Murray, E. *et al.* Evaluation of the INTERGROWTH-21st Neurodevelopment Assessment (INTER-NDA) in 2 year-old children. *PloS one* **13**, e0193406 (2018). https://doi.org/10.1371/journal.pone.0193406

89 Koo, T. K. & Li, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine* **15**, 155-163 (2016). https://doi.org/10.1016/j.jcm.2016.02.012

90 Group, W. H. O. M. G. R. S. WHO Motor Development Study: windows of achievement for six gross motor development milestones. *Acta Paediatr Suppl* **450**, 86-95 (2006).

ACKNOWLEDGEMENTS

M.K.W and A.I.L.N are supported by the Bill and Melinda Gates Foundation. M.K.W was also supported by the EPSRC Doctoral Prize Scheme [grant number EP/W524311/1]. M.F. is supported by a Clinical Research Training Fellowship (MR/V029169/2) funded by the Medical Research Council, UK.

We would like to thank the Health Authorities in Pelotas, Brazil; Beijing, China; Nagpur, India; Turin, Italy; Nairobi, Kenya; Muscat, Oman; Oxford, UK and Seattle, USA, who facilitated the project by allowing participation of the study sites. We are grateful to Philips Medical Systems for the ultrasound equipment and technical assistance. We also thank MedSciNet U.K. Ltd for the website development and maintenance and support of the online data management system. We thank the participating hospitals, parents and infants who participated in the studies and the more than 200 members of the research teams who made this project possible.

Full acknowledgment appears at https://intergrowth21.com/research/intergrowth-21st

Luke Conaboy, Stephan Uphoff, Albert Alonso, Valentin Bacher, Stephan Uphoff, and the OMNI Lab provided invaluable help and advice.

AUTHOR CONTRIBUTIONS

M.K.W, A.I.L.N, N.K.D, S.H.K and J.V designed the study with input from M.F., A.T.P. and A.S.

M.F., M.C., L.C.I., Y.J., M.G., W.Q., E.B., M.P. and F.C.B. oversaw data collection. The data were curated by M.K.W. and A.I.L.N. A.I.L.N., M.K.W., N.K.D., R.B.G. and E.O.O. analysed the data.

ARTICLE IN PRESS

ARTICLE IN PRESS

A.I.L.N., M.K.W., N.K.D., and L.S.H. designed and implemented the image analysis algorithms, with input from M.J., S.S., and T.E.N. L.C.I., A.T.P., and A.W. coordinated the study. Z.M. assisted with the interpretation of the results. A.I.L.N., S.H.K. and M.K.W wrote the initial drafts of the paper; JV expanded the initial draft and made it consistent with the INTERGROWTH 21ST conceptual frame; all other authors including A.J.N and Z.B, contributed and revised it critically for important intellectual content. All authors approved the final version for publication.

COMPETING INTERESTS

J.A.N. and A.T.P. are Senior Scientific Advisers of Intelligent Ultrasound but the company has no financial or intellectual property links with the research described in this paper and the described work is entirely based on their academic work. We declare that all the other authors have no competing interests as defined by Nature Research, or other interests that might be perceived to influence the results and/or discussion reported in this paper. A.T.P is supported by the Oxford Partnership Comprehensive Biomedical Research Centre with funding from the NIHR Biomedical Research Centre (BRC) funding scheme. The views expressed herein are those of the authors and not necessarily those of the NHS, the NIHR the Department of Health or any of the other funders. The remaining authors declare no competing interests.

ARTICLE IN PRESS

TABLES

Table 1: An overview of the acronyms describing the 28 image derived phenotypes (IDPs). The volume of each region and parcellation was assessed, with the mean thickness and depth also measured for the cortical parcellations. Additional measures included cortical plate (CoP) SA and Sylvian fissure depth. TB and CSP was labelled across the entire brain, whereas, the remaining regions were only segmented on the distal hemisphere. Full list of the 28 IDPs: Total brain volume (TBV), cortical plate volume (CoPV), white matter volume (WMV), deep grey matter volume (DGMV), cerebellum volume (CBV), thalamus volume (ThV), lateral posterior ventricle horns volume (LVV), choroid plexus volume (ChPV), frontal horns volume (FHV), brainstem volume (BSV), cavum septum volume (CSPV), cortical plate surface area (CoPSA), Sylvian fissure depth (SFD), frontal lobe volume (FLV), temporal lobe volume (TLV), parietal lobe volume (PLV), occipital lobe volume (OLV), insular lobe volume (ILV), frontal lobe depth (FLD), temporal lobe depth (TLD), parietal lobe depth (PLD), occipital lobe depth (OLD), insular lobe depth (ILD), frontal lobe thickness (FLT), temporal lobe thickness (TLT), parietal lobe thickness (PLT), occipital lobe thickness (OLT) and insular lobe thickness (ILT).

11 regions of interest		
Both	TB	Total Brain
	CSP	Cavum septum
Distal hemisphere only	CoP	Cortical plate
	WM	White matter
	DGM	Deep grey matter
	CB	Cerebellum
	ChP	Choroid plexus
	LV	Lateral posterior ventricle horns
	FH	Frontal horns
	BS	Brain stem
	Th	Thalamus
5 cortical parcellations		
Distal hemisphere only	**FL**	Frontal Lobe
	PL	Parietal lobe
	OL	Occipital lobe
	TL	Temporal lobe
	IL	Insular lobe
Measurement type		
	V	Volume
	SA	Surface area
	D	Depth
	T	Thickness
	SFD	Sylvian Fissure depth

Table 2: Formulae for total brain volume. Equations for estimating the mean and SD for total brain volume (TBV); the remaining 27 equations are in Suppl Table 8. For SD, the reported equation is on the log scale and must be exponentiated to obtain the standard deviation.

	Number included scans		**Regression Equation**
TBV	4196	Mean	$1.954510 + 0.018205xGA^3 + -0.178633xGA^2$
		SD	$-0.702623 + 0.150265xGA$

ARTICLE IN PRESS

FIGURE CAPTIONS

Fig. 1. Summary statistics of fetuses from the INTERGROWTH-21st Fetal Growth Longitudinal Study (FGLS) included in the present study. (a) Flowchart summarising the inclusion criteria, number of fetuses, and number of 3D ultrasound scans remaining at each step. (b) Lasagna plot of the number of scanning sessions (visits) per fetus in the FGLS dataset. Of the 2906 fetuses included in the analyses of brain regions between 18 and 27 weeks' gestation, 49.5% (n=1433) had at least two scans. (c) Map showing the contribution of each study site to the total dataset, designed by Freepik http://www.freepik/.

Fig. 2. Proposed tissue parcellation. *Proposed tissue parcellation protocol on the fetal brain ultrasound atlas at 22 weeks' gestation (Namburete, 2023), shown in (a) the coronal view, (b) in 3D and (c) in 3D and encompassed by the total brain (TB). The structures labeled are cortical plate (CoP), white matter (WM), deep grey matter (DGM), cerebellum (CB), thalamus (Th), lateral posterior ventricle horns (LV), choroid plexus (ChP), frontal horns (FH), brainstem (BS) and cavum septum (CSP). A full description of the image-derived phenotypes is shown in Table1. It should be noted that, although both the left and right hemispheres were labelled on each atlas, only the distal hemisphere was labelled on individual scans, with the exception of TB and CSP.*

ARTICLE IN PRESS

Fig.3. Measures of the fetal brain cortex. (a) The cortical lobes' parcellation for an individual fetal brain shown overlaid on the scan (top row) and as a 3D reconstruction (bottom row). The same is displayed for (b) cortical depth and (c) cortical thickness. The cortical depth measures the distance from the skull to each position within the cortex. The 3D projection shows the white matter cortical boundary. Cortical thickness measures the shortest distance between the pial membrane and white matter boundary, the 3D projection is on the pial membrane. These maps provide complementary visualisation of fetal cortical maturation, integrating morphological and microstructural developmental trajectories. The cortical parcellation map

ARTICLE IN PRESS

are for: frontal lobe (FL), temporal lobe (TL), parietal lobe (PL), occipital lobe (OL) and insular lobe (IL). A full description of the image-derived phenotypes is shown in Table 1.

Fig. 4. Comparison between the atlas labels on the ultrasound (US) and five magnetic resonance imaging (MRI) atlases. (a) The correspondence between MRI-atlas labels and our US-atlas tissue labels and (b) lobar parcellations at 24 weeks' gestation, using frontal (red), parietal (green), occipital (yellow), insula (purple) and temporal (blue) lobes. (c) The volume measures of each atlas-labelled structure across both hemispheres for each atlas. Each atlas spans a one-week period, e.g., the 26th-week atlas spans 26^{+0days} to 26^{+6days}. The MRI labels show the mean (o) with the error bars showing the standard deviation across the datasets. The number in brackets under each week of gestation shows how many MRI atlases were used for that structure. For instance, three MRI atlases had deep grey matter labelled at 22 weeks' gestation. The MRI atlases included are: CRL (Gholipour et al. 2017 [15]), KCL (Uus et al. 2023 [80]), Serag (Serag et al. 2012 [18]), CHN (Xu et al. 2022 [16]) and FBA (Wu et al. 2021 [17]).

Fig. 5. The effect of study site on fetal brain trajectories. (a) Re-generating the fetal trajectories excluding one site at a time (b) cross-site variance (c) standardised site differences (SSD), all values given in Supplementary Table 5. All 28 image derived phenotypes (IDPs) are included: total brain volume (TBV), cortical plate volume (CoPV), white matter volume (WMV), deep grey matter volume (DGMV), cerebellum volume (CBV), thalamus volume (ThV), lateral posterior ventricle horns volume (LVV), choroid plexus volume (ChPV), frontal horns volume (FHV), brainstem volume (BSV), cavum septum volume (CSPV), cortical plate surface area (CoPSA), Sylvian fissure depth (SFD), frontal lobe volume (FLV), temporal lobe volume (TLV), parietal lobe volume (PLV), occipital lobe volume (OLV), insular lobe volume (ILV), frontal lobe depth (FLD), temporal lobe depth (TLD), parietal lobe depth (PLD), occipital lobe depth (OLD), insular lobe depth (ILD), frontal lobe thickness (FLT), temporal lobe thickness (TLT), parietal lobe thickness (PLT), occipital lobe thickness (OLT) and insular lobe thickness (ILT).

ARTICLE IN PRESS

ARTICLE IN PRESS

Fig. 6. Volumetric growth trajectories with fitted percentiles lines at the 3rd, 50th and 97th centiles. Growth charts showing total brain volume (TBV), cortical plate volume (CoPV), white matter volume (WMV), deep grey matter volume (DGMV), cerebellum volume (CBV), thalamus volume (ThV), lateral posterior ventricle horns volume (LVV), choroid plexus volume (ChPV), frontal horns volume (FHV), brainstem volume (BSV), cavum septum volume (CSPV), cortical plate surface area (CoPSA) and Sylvian fissure depth (SFD). The equation for the mean is written on TBV, for the remaining equations see Supplementary Table 8. All IDPs except TBV and CSPV, measure only the distal hemisphere. The number of subjects used to create each trajectory is given in Supplementary Table 8.

Fig. 7. Cortical growth trajectories with fitted percentiles lines at the 3rd, 50th and 97th centiles. These trajectories show frontal lobe volume (FLV), temporal lobe volume (TLV), parietal lobe volume (PLV), occipital lobe volume (OLV) and insular lobe volume (ILV), frontal lobe depth (FLD), temporal lobe depth (TLD), parietal lobe depth (PLD), occipital lobe depth (OLD) and insular lobe depth (ILD), frontal lobe thickness (FLT), temporal lobe thickness (TLT), parietal lobe thickness (PLT), occipital lobe thickness (OLT) and insular lobe thickness (ILT). Each of these IDPs is measured from the distal hemisphere. The number of subjects used to create each trajectory is given in Supplementary Table 8.

ARTICLE IN PRESS

Fig. 8 Relative cortical lobe growth and ratio. (a) Relative parietal lobe volume (rPLV) and relative insular lobe volume (rILV) from the distal hemisphere. Both volumes are relative to total cortical plate volume (CoPV). Left panel: progressive decline in rPLV with advancing gestational age, consistent with the relative volumetric redistribution of the developing CoP. Right panel: increase in rILV reflecting the accelerated maturation of the insular region. (b) Insular-Parietal lobes Volume Ratio (rILV/rPLV) from 18 to 27 weeks' gestation. The rILV/rPLV ratio shows the evolving temporal asynchrony between these cortical regions. Longitudinal growth of 10 randomly subjects is shown in pink. The blue dash lines show the 3rd, 50th and 97th

ARTICLE IN PRESS

centiles, with individual values in the background. Relative growth for the remaining lobes is shown in Supp Fig. 2.

ARTICLE IN PRESS

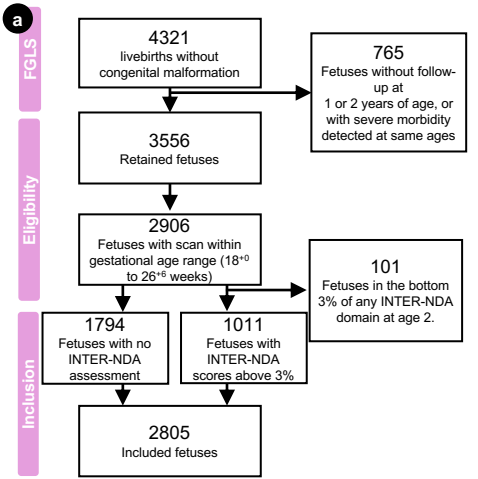
ARTICLE IN PRESS

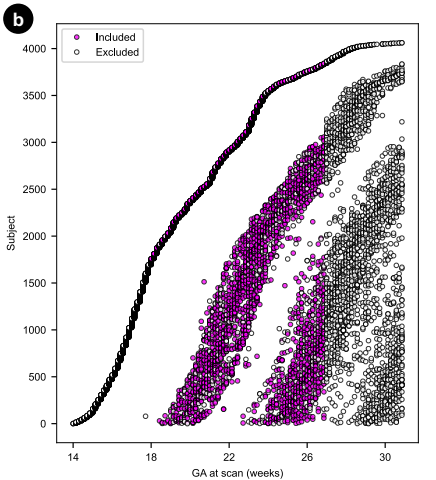
Editor's Summary: This study mapped normative growth trajectories of 28 fetal brain phenotypes derived from 4,205 3D ultrasound scans collected across 7 international sites. The low variance between sites reinforces the principle that the brain develops similarly when environmental constraints are minimal.

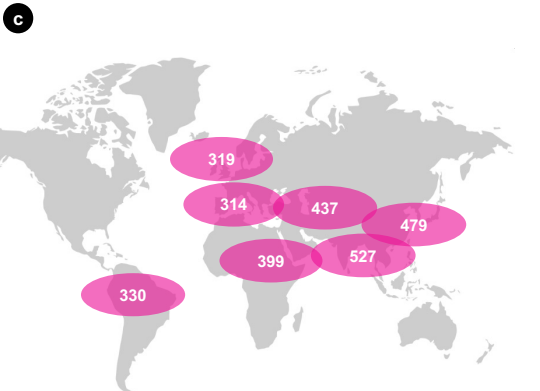
Peer review information: *Nature Communications* thanks Dan Wu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

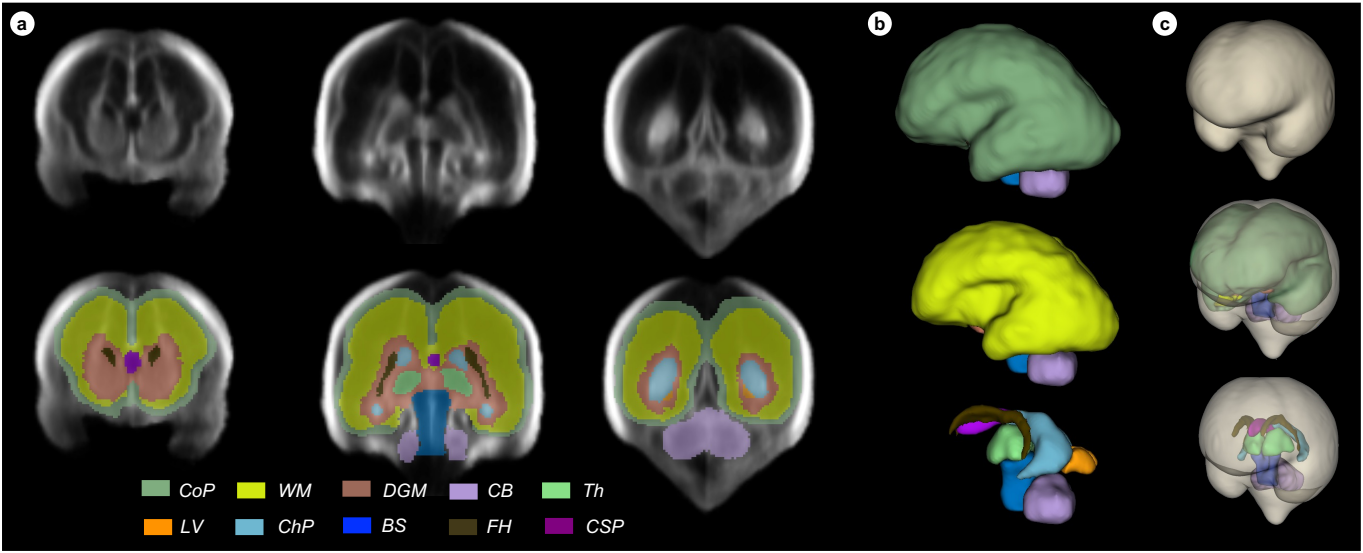
ARTICLE IN PRESS

ARTICLE IN PRESS

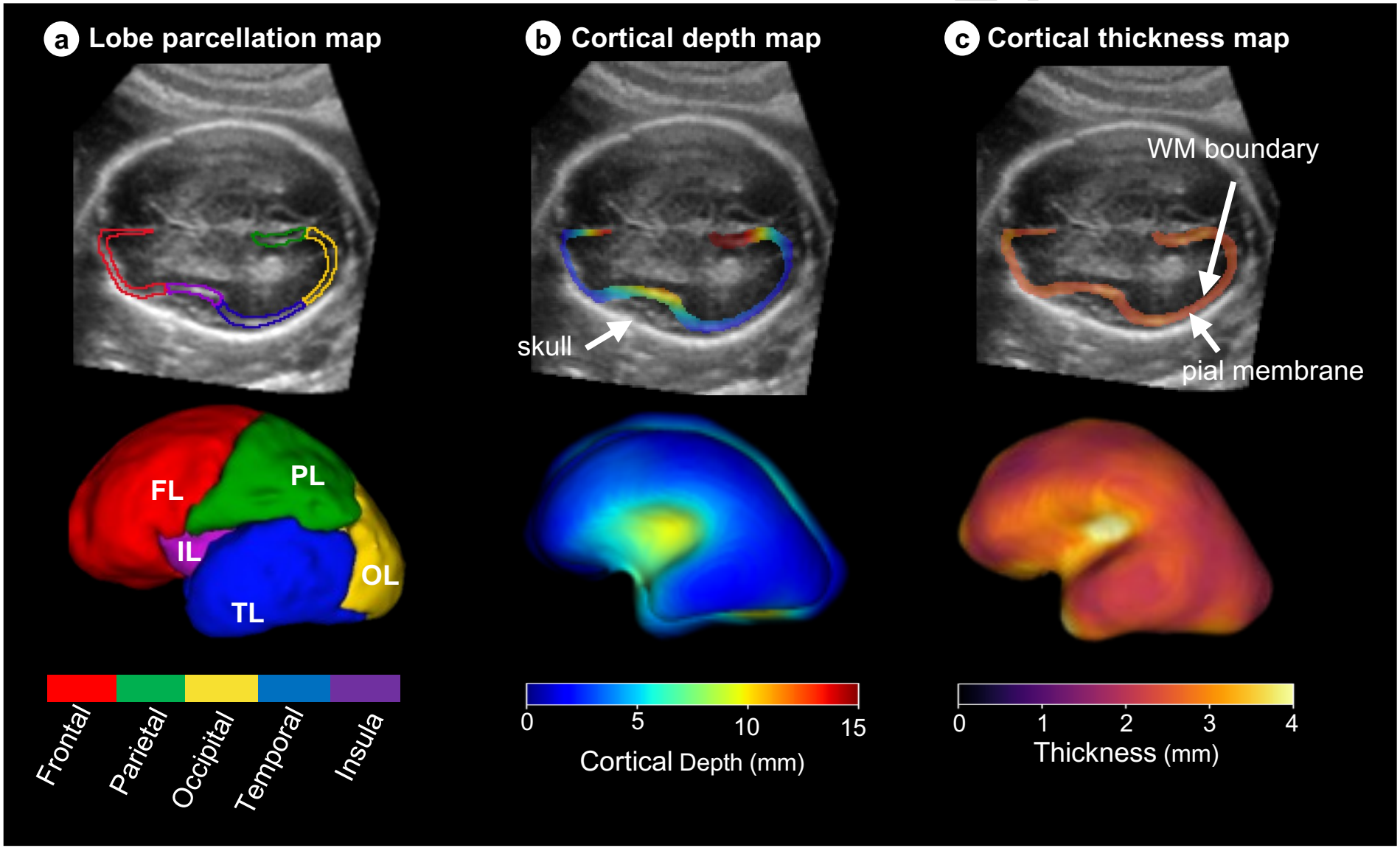


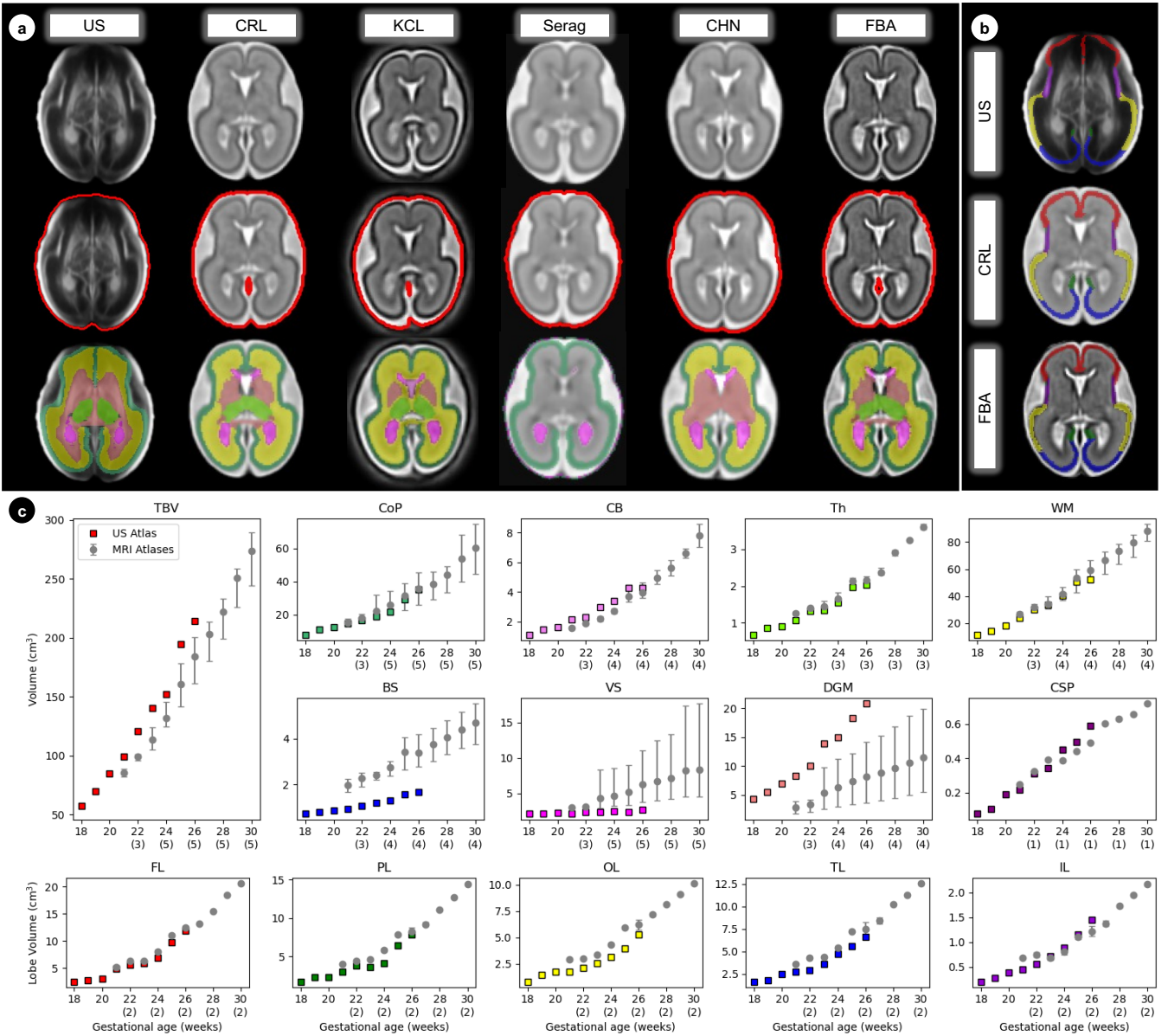


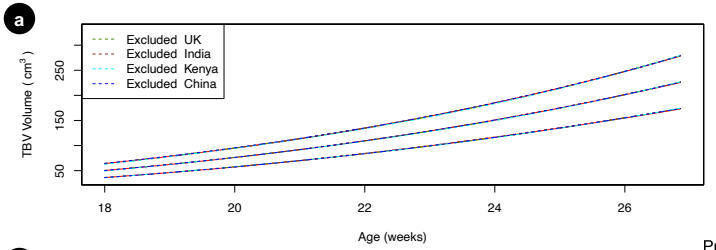


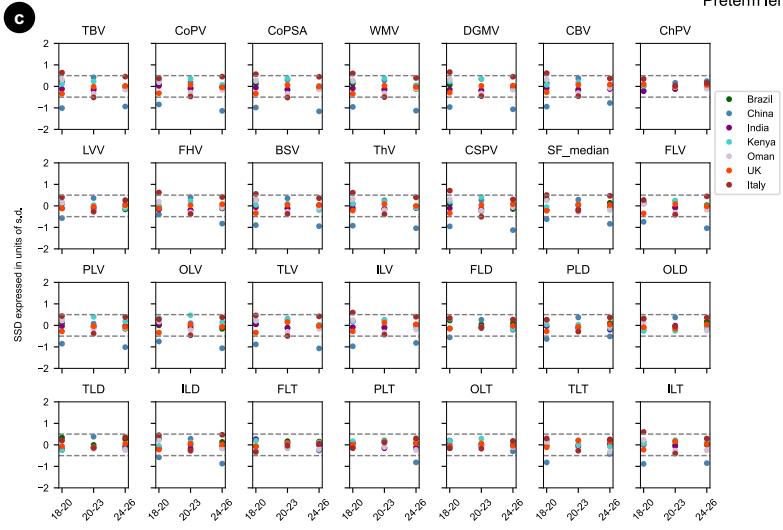

a
b
c
CoP
WM
DGM
CB
Th
LV
ChP
BS
FH
CSP

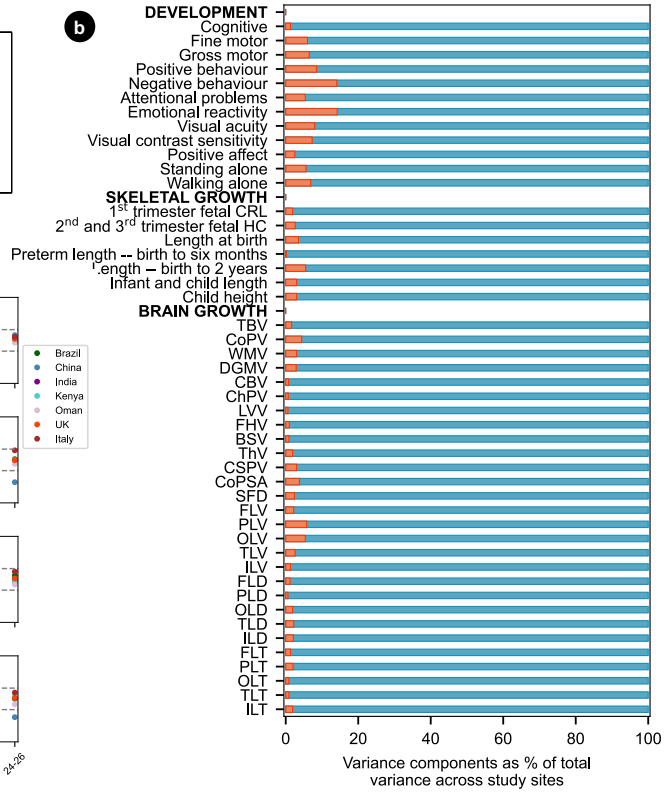
ARTICLE IN PRESS



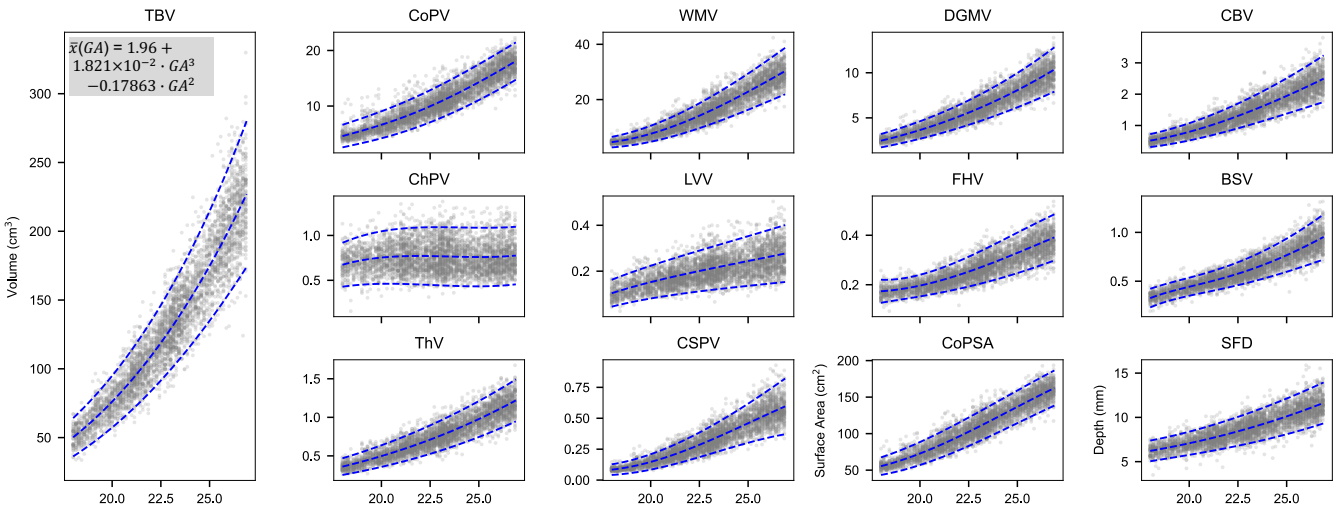

a
US
CRL
KCL
Serag
CHN
FBA
b
US
CRL
FBA
c
TBV
CoP
CB
Th
WM
BS
VS
DGM
CSP
FL
PL
OL
TL
IL
US Atlas
MRI Atlases
Volume (cm³)
Lobe Volume (cm³)
Gestational age (weeks)

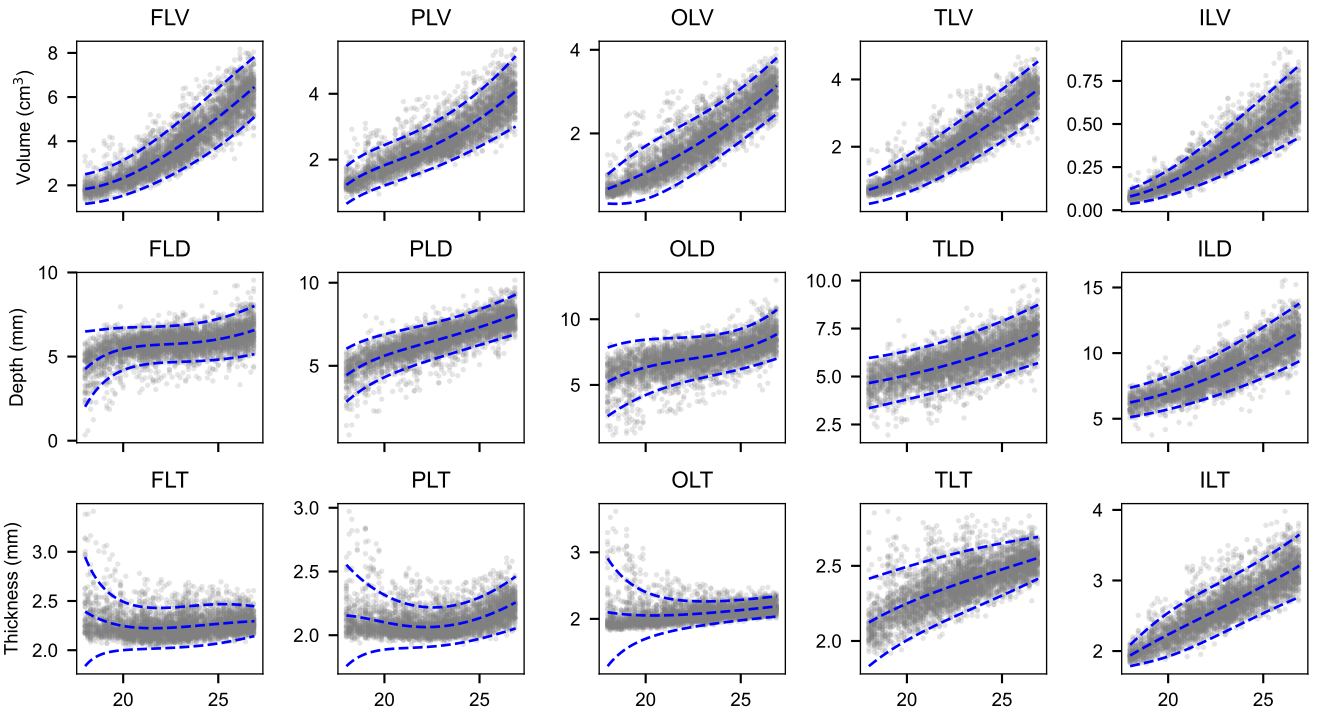

a
Excluded UK
Excluded India
Excluded Kenya
Excluded China
TBV Volume (cm3)
50
150
250
18
20
22
24
26
Age (weeks)

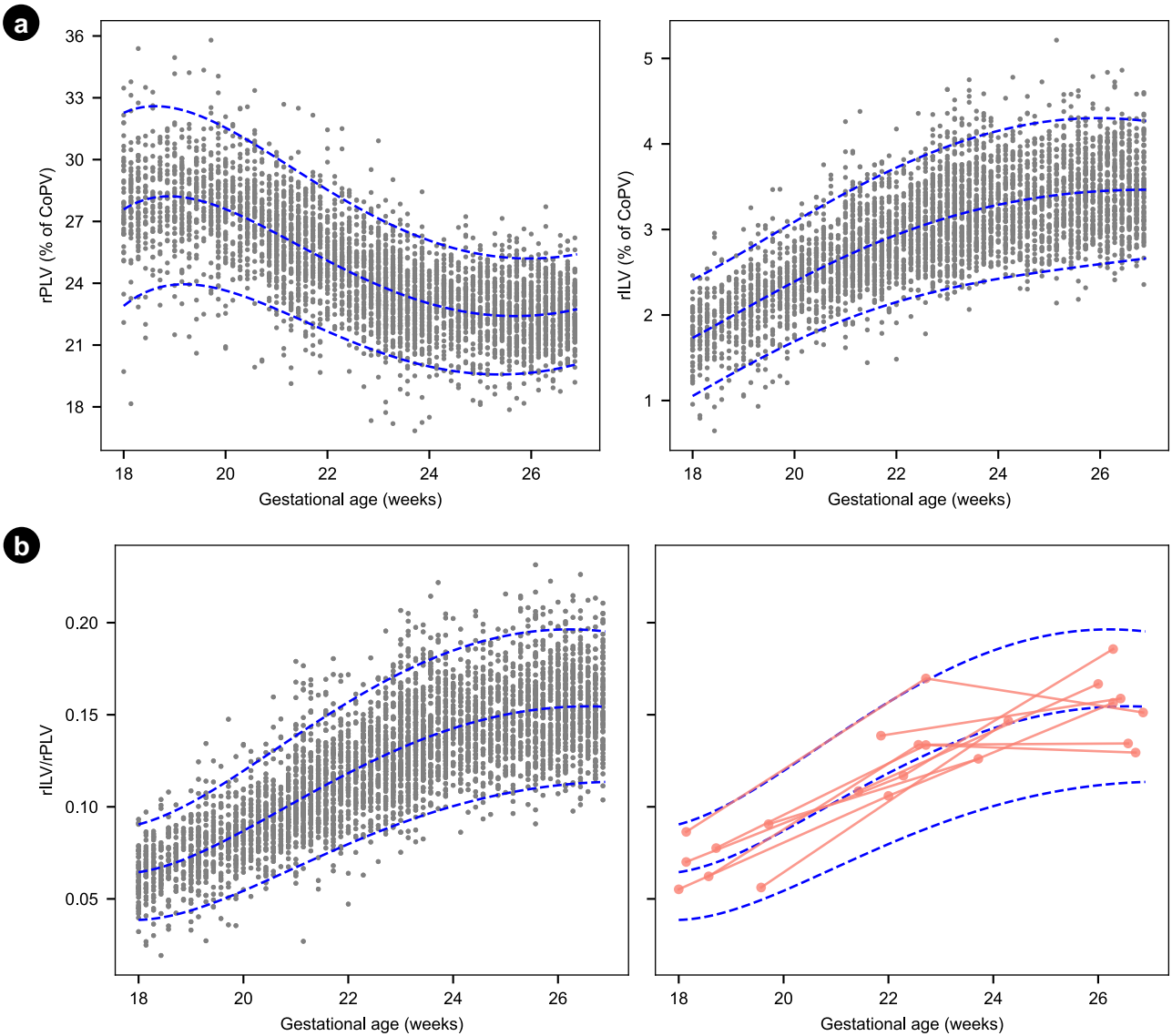

c
TBV
CoPV
CoPSA
WMV
DGMV
CBV
ChPV
LVV
FHV
BSV
ThV
CSPV
SF_median
FLV
PLV
OLV
TLV
ILV
FLD
PLD
OLD
TLD
ILD
FLT
PLT
OLT
TLT
ILT
Brazil
China
India
Kenya
Oman
UK
Italy
SSD expressed in units of s.d.
18-20
20-23
24-26


b
DEVELOPMENT
Cognitive
Fine motor
Gross motor
Positive behaviour
Negative behaviour
Attentional problems
Emotional reactivity
Visual acuity
Visual contrast sensitivity
Positive affect
Standing alone
Walking alone
SKELETAL GROWTH
1st trimester fetal CRL
2nd and 3rd trimester fetal HC
Length at birth
Preterm length -- birth to six months
Length -- birth to 2 years
Infant and child length
Child height
BRAIN GROWTH
TBV
CoPV
WMV
DGMV
CBV
ChPV
LVV
FHV
BSV
ThV
CSPV
CoPSA
SFD
FLV
PLV
OLV
TLV
ILV
FLD
PLD
OLD
TLD
ILD
FLT
PLT
OLT
TLT
ILT
0
20
40
60
80
100
Variance components as % of total variance across study sites

ARTICLE IN PRESS




FLV
PLV
OLV
TLV
ILV
Volume (cm^3)
FLD
PLD
OLD
TLD
ILD
Depth (mm)
FLT
PLT
OLT
TLT
ILT
Thickness (mm)
Gestational age (weeks)


a
rPLV (% of CoPV)
36
33
30
27
24
21
18
18
20
22
24
26
Gestational age (weeks)
rILV (% of CoPV)
5
4
3
2
1
18
20
22
24
26
Gestational age (weeks)
b
rILV/rPLV
0.20
0.15
0.10
0.05
18
20
22
24
26
Gestational age (weeks)
18
20
22
24
26
Gestational age (weeks)