Automated Craniofacial Biometry with 3D T2w Fetal MRI

Jacqueline Matthew[1,2,5*†], Alena Uus[1,5†], Alexia Egloff Collado[1,2], Aysha Luis[1,2], Sophie Arulkumaran[2], Abi Fukami-Gartner[1], Vanessa Kyriakopoulou[1], Daniel Cromb[1], Robert Wright[1], Kathleen Colford[5], Maria Deprez[1], Jana Hutter[3,5], Jonathan O'Muircheartaigh[5], Christina Malamateniou[4], Reza Razavi[1], Lisa Story[2,5], Jo Hajnal[1,5], Mary A. Rutherford[2,5]

[1*]School of Biomedical Engineering and Imaging Sciences, King's College London, St Thomas' Hospital, London, UK.
[2]Guy's and St Thomas' NHS Foundation Trust, London, UK.
[3]Smart Imaging Lab, Radiological Institute, University Hospital Erlangen, Erlangen, Germany.
[4] Division of Midwifery and Radiography, City University of London, London, UK.
[5]Department of Early Life Imaging, King's College London, St Thomas' Hospital, London, UK.

*Corresponding author(s). E-mail(s): jacqueline.matthew@kcl.ac.uk;
†These authors contributed equally to this work.

Abstract

Objectives: Craniofacial phenotype-genotype correlations prenatally remain subjective, and detailed evaluation with 3D ultrasound is challenging and time-consuming. Recent methods for automating MRI fetal brain biometry have shown technical feasibility. Using 3D motion-corrected, slice-to-volume reconstructed (SVR) fetal MRI, we developed an automated landmark propagation pipeline specifically for a large set of cranial and facial measurements.
Methods: A literature review and expert consensus identified 31 craniofacial biometrics for fetal MRI. A 3D spatiotemporal SVR fetal MRI atlas was used to define anatomical landmark points and as a template for subject registration, auto-labelling and finally landmark-based biometric calculation. Visual inspections of subject-level landmark labels, as well as multivariate and univariate

analyses, were performed in 108 healthy control fetuses and 24 fetuses with Down syndrome, T21, (gestational age, GA, 29-36 weeks), to assess differences between groups and identify meaningful landmarks in T21. Reliability and reproducibility were assessed in 10 random datasets by four observers.
Results: Automated labels were produced for all 132 subjects, with a 0.03% major error rate. Significant differences in biometrics were found between T21 and control groups (MANOVA, $p<0.001$). There were significant differences with large effect sizes for seven measurements (ANOVA, $p<0.001$), including the anterior base of skull length and maxillary length. Manual measurements took 25-35 minutes per case, while automated extraction took approximately 5 minutes. Bland-Altman plots showed a relative mean bias and limits of agreement within the ranges seen for manual observers except for the mandibular width, which had an agreement threshold for manual observers of good (ICC = 0.904, 95% CI = 0.677-0.975) versus moderate for automated plus manual observers (ICC = 0.635, 95%, CI = 0.176-0.888). Extended GA growth charts (19-39 weeks) based on 280 control fetuses were produced for all craniofacial biometrics and made publicly available for future research.
Conclusion: We present the first automated atlas-based protocol using 3D motion-corrected MRI for fetal craniofacial biometrics and, using this method, accurately characterise differences in craniofacial morphology in Downs syndrome fetuses, with findings consistent with the known phenotype. Future work should include studies to improve measurement reliability, larger clinical cohort studies, technical development of craniofacial optimised MRI sequences and 3D deep learning for anatomical landmarking. Future development may reduce clinical reporting time, improve accuracy, increase understanding of prenatal development and enhance personalised prenatal care with accurate genotype-phenotype correlations.

Keywords: Fetal MRI, Slice-to-volume reconstruction, craniofacial development, automated biometry

1 Introduction

Comprehensive prenatal characterisation of craniofacial development remains a challenge for obstetric ultrasound due to limitations caused by fetal position, artefacts, and technical difficulties in the 2D and 3D domain. Currently, clinical imaging techniques broadly rely on subjective assessment of facial features, and in high-risk cases, quantitative linear measurements, facial angles, and indexes, have been proposed for use with prenatal ultrasound [1–4]. Prenatal detection of face, ear and neck anomalies are low during universal second-trimester ultrasound screening in pregnancy. Indeed, a recent systematic review found them to have the lowest detection rate of 46 organ groups assessed, with a sensitivity of 32.3% [5]. Prenatal detection of facial anomalies, even in high-risk groups referred for a targeted specialist ultrasound, are even more likely to remain undetected if isolated, or, if presenting with an additional non-genetic body structural anomaly [6]. The fetal MRI craniofacial clinical literature usually describes subjective imaging assessments, and are often clinical reviews or case studies, however,

its complementary role to ultrasound is often highlighted. For example, a recent historical cohort study of 45 patients referred to fetal MRI for a wide range of suspected craniofacial malformations at an anomaly level, e.g. cleft lip and palate, craniosynostosis, hyper/hypotelorism, ear structure anomalies, facial cysts and masses, found MRI was more likely to make a confident diagnosis and less likely to over-diagnose when compared to ultrasound [7].

1.1 Imaging craniofacial development in-utero

Imaging the craniofacial complex prenatally requires expert image acquisition, precise 2D image planes, or, 3D data that can be aligned to the region of interest (ROI). A quantitative assessment, for a prenatal phenotypic characterisation, is thus time-consuming and subject to observer variation. International guidelines for routine mid-trimester ultrasound, performed at approximately 20 weeks gestational age, GA, suggests the facial examination should be limited to a qualitative assessment of the upper lip, orbits, and an optional examination of the mid-sagittal facial profile and nasal bone [8].

The increased use of 3D data in fetal ultrasound and the expanding applications of structural fetal MRI have resulted in the feasibility of using extended biometrics methods to better characterise and/or diagnosis subtle craniofacial dysmorphology [9, 10]. In an expert consensus paper, Merz et al (2012) suggested a targeted craniofacial examination to include 3D ultrasound, with multiplanar and aligned 2D views to allow the biometric assessment of the nasal bone, NB, frontomaxillar facial angle, FMA, inferior facial angle, IFA, orbital diameters, OD, interorbital distance, IOD, and outer orbital distance, (or bi-orbital distance, BOD), maxilla width, MxW, and mandibular width, MdW (all in addition to standard head biometry i.e. head circumference, HC, occipitofrontal diameter, OFD, and biparietal diameter, BPD) [1]. However, few antenatal imaging studies have sought to comprehensively assess multiple craniofacial biometric profiles, often focussing on fetal estimated weight parameters (i.e. HC, OFD, BPD, in addition to abdominal circumference, and femur length), orbital and mandibular regions [3, 11–15].

1.2 Fetal MRI for craniofacial assessment

Toren et al in (2020), confirmed the feasibility of multiple manually extracted fetal MRI craniofacial biometrics [16]. The authors reviewed the use of eight fetal MRI 2D measurements related to the mandible and nasal cavity, these included the, previously cited measurements, IFA, BPD, and IOD, and four new measurements, the mandibular anterior-posterior diameter, mandibular vertebral length, maximum nasal length, septal height, and septal length. The authors highlighted the additional need for structural radiological biomarkers to characterise fetal facial development and, importantly, noted that 70% (843 MRI scans) had to be excluded from their final sample due to motion artefact that degraded the image quality or the absence of a true orthogonal plane to produce the required measurement.

Except for the established cranial vault measurements (HC, OFD, BPD), there are limited examples of fetal MRI craniofacial reference ranges. Noteable anatomical areas

examined are the orbits [17–19], and includes our previous work on the automated extraction of fetal 2D orbial biometry from 3D volumes [20]; the mandible [21] and more recently maxillary sinuses [22]. A reason for slower development of the MRI craniofacial literature may be because 3D fetal MRI has focused on brain development and assessment. Indeed, the first step would be to accurately define MRI landmarks for any new measurements and to ensure that image quality enables the accurate localisation of rarely assessed structures in-vivo. Furthermore, there is also a known lack of clear consensus on formalisation of fetal MRI biometry protocols, nomogram model formulas and measurement techniques for MRI between different clinical centres [23]. T2-weighted MRI is considered the optimal choice of image contrast for fetal structural assessment due to the faster acquisition times and good fluid tissue differentiation. The effect of field strength (1.5 Tesla compared to 3 Tesla, T) on brain biometry has been shown to produce small absolute differences for some measurements. This is likely due to the increased spatial resolution at 3T and resulting in larger discrepancies particularly for smaller structures [24]. In addition, whilst brain and facial anatomical detail are diagnostically acceptable at both 1.5T and 3T, however image quality may be be better at 3T or vary according to acquisition parameters [25]. There are, of course additional sources of error due to clinician training and experience, reporting software and environment, maternal breathing and fetal motion leading to imprecise acquisition planes and calliper placement that result in increased intra- and inter-observed variability.

1.3 Automation of biometry for fetal MRI

Motion correction methods, based on 3D slice-to-volume registration (SVR) [26], partially resolves these challenges since the 3D reconstructed images can be reoriented to any plane. 3D-derived biometric measurements are reportedly comparable with 2D slice-wise biometry [27, 28]. Yet, there may be considerable operator variability when placing landmarks in a 3D volume due to the requirement to choose the correct plane for the measurement and then define the anatomical landmarks in 3D space.

Theoretically, in addition to being faster, automation of biometry should also allow reproducible biometric measurements. Recently, there have been several proposed methods for automated fetal MRI biometry with deep learning brain measurements such as biparietal and transverse cerebellar diameters and atrial diameters [29, 30] in 2D slices and ocular measurements using registration and deep learning [20, 31] in 3D motion-corrected images. However, outside of the cranial vault, there have been no reported automated solutions for craniofacial measurements for fetal MRI.

1.4 Contributions

In this study, we formalise the first landmark-based protocol for craniofacial biometry for 3D T2w fetal head MRI in the atlas space and develop the first automated pipeline for extraction of 31 craniofacial biometry measurements based on label propagation. The performance of the pipeline is extensively evaluated with respect to traditional direct measurements by expert observers as well as the analysis of common errors and the effects of MRI image quality and field strength. Next, the utility of the proposed

biometry protocol is assessed by quantitative comparison of 108 normal control and 24 T21 subjects, characterised by well-known craniofacial dysmorphology prenatally. In addition, we generated normal craniofacial biometry growth charts from 280 control subjects from 19 to 38 weeks GA range.

2 Methods

2.1 Fetal MRI datasets and preprocessing

Participants were scanned between 2014 and 2024 at a single site (St. Thomas' Hospital, London, UK) and all maternal participants gave written informed consent for the use of data acquired under one of eight ethically approved MRI research studies. The datasets were acquired under different research studies: The Placental Imaging Project (PIP, REC 14/LO/1169)[1]; the Intelligent Fetal Imaging and Diagnosis (iFIND, REC 14/LO/1806); the quantification of fetal growth and development with MRI study (fetal MRI, REC 07/H0707/105)[2]; the fetal CMR service at Evelina London Children's Hospital (REC 07/H0707/105); the developing human connectome project (dHCP, REC 14/LO/1169); the early brain imaging in Down syndrome study (eBiDS, REC 19/LO/0667); the Individualised risk prediction of adverse neonatal outcome in pregnancies that deliver preterm using advanced MRI techniques and machine learning study (PRESTO: REC 21/SS/0082); and the Cardiac and Placental Imaging in Pregnancy project (CARP; REC 08/LO/1958). The inclusion criteria for case selection included: singleton pregnancy, fetal MRI stacks with full ROI coverage, acceptable quality whole head SVR output. The normal control cohort included 314 cases without reported fetal or maternal anomalies with moderate to excellent image quality from four different acquisition protocols, from 19 to 39 weeks GA. The T21 cohort was curated primarily based on the availability of datasets with research consent and acceptable 3D head SVR reconstruction quality. In total, we selected 24 T21 cases from 3 different acquisition protocols and 29 - 36 weeks GA range.

2.1.1 MRI acquisition protocols

The included datasets were acquired with different MRI acquisition protocols depending on the recruiting study:

- 4 DS and 34 healthy control subjects scanned on a 1.5T Philips Ingenia MRI system using 28-channel torso coil with TE=80ms, 1.25x1.25mm resolution, 2.5mm slice thickness, -1.25mm gap and 9-11 stacks (iFIND, FCMR studies);
- 17 DS and 106 healthy control subjects were scanned on 3T Philips Achieva MRI system using a 32-channel cardiac coil with TE=180ms, 1.25 x 1.25mm resolution, 2.5mm slice thickness, -1.5mm gap and 5-6 stacks (PIP, PRESTO, eBIDs studies);
- 3 DS and 130 healthy control subjects were scanned on 3T Philips Achieva MRI system with a 32-channel cardiac coil using a dedicated dHCP fetal acquisition protocol with TE=250ms, 1.1 x 1.1mm resolution, 2.2mm slice thickness, -1.1mm gap and 6 stacks (dHCP, fetal MRI studies).

[1] PiP project: https://placentaimagingproject.org/project/
[2] iFIND project: https://www.ifindproject.com/

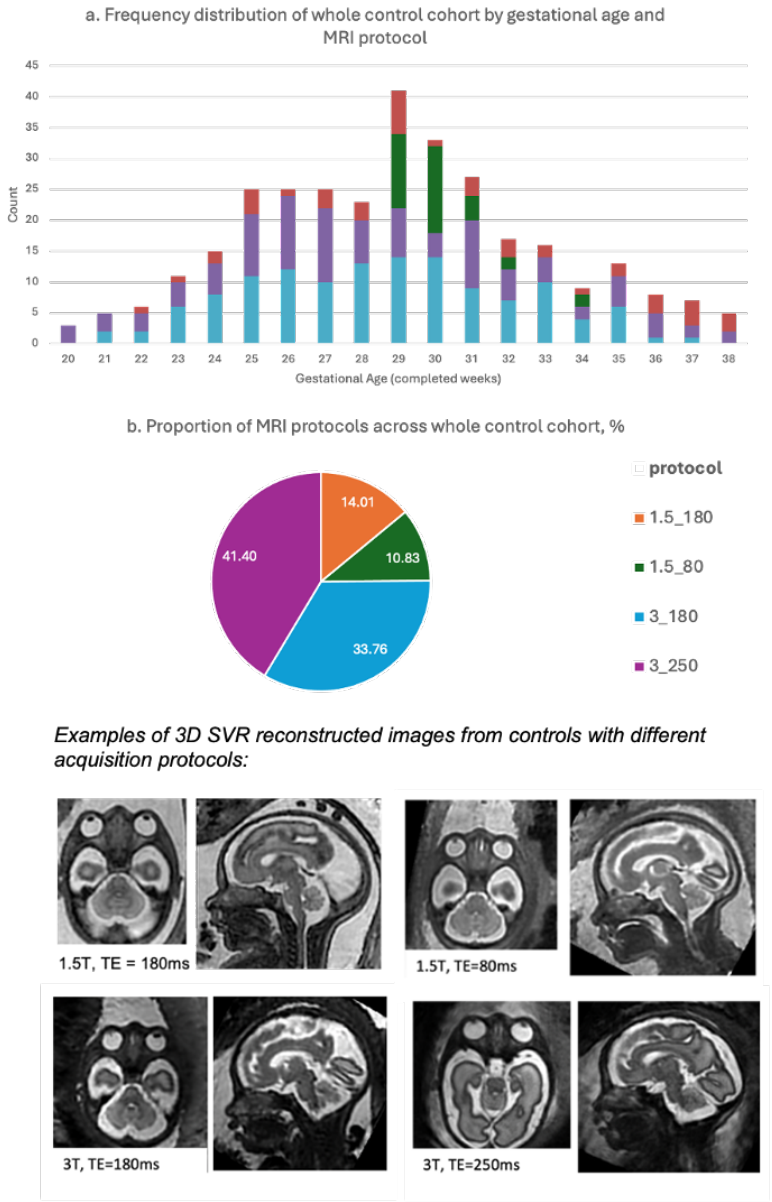


Fig. 1: a. GA distribution of control subject datasets in the study per MRI protocol; b. proportional distribution of MRI protocols in the whole control group (MRI protocol (field strength/TE) = 1.5T/ 180ms; 1.5T/ 80ms; 3T/ 180ms; or, 3T/ 250ms); and examples of a 3D SVR fetal head reconstructions at different acquisition parameters.

- 44 healthy control subjects scanned on a 1.5T Philips Ingenia MRI system using 28-channel torso coil with TE=180ms, 1.25x1.25mm resolution, 2.5mm slice thickness, -1.25mm gap and 9-11 stacks (PIP, CARP studies).

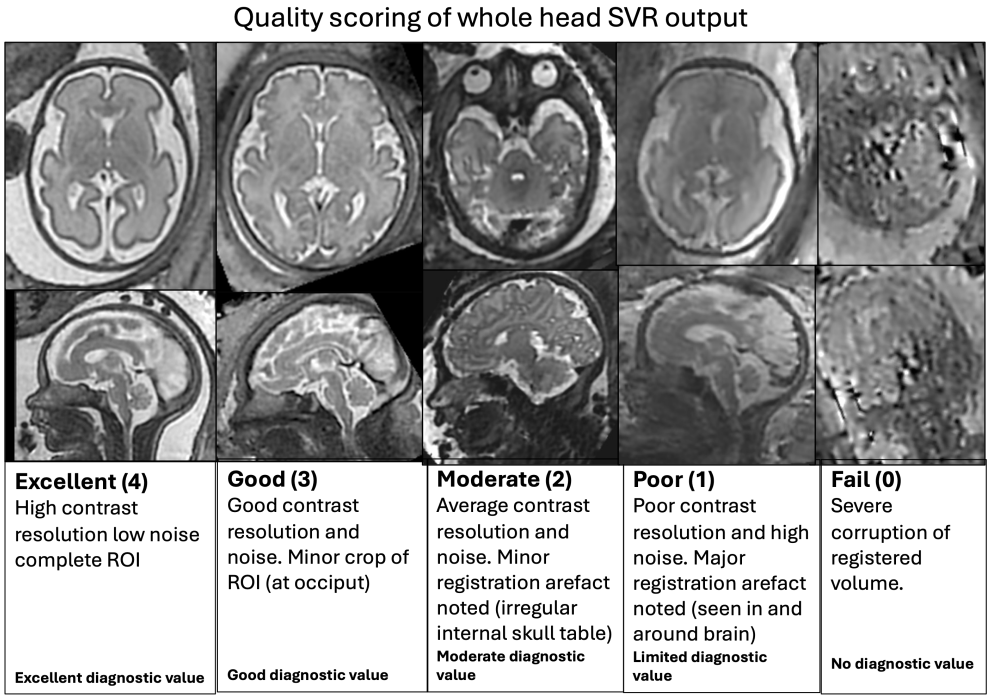


Fig. 2: Image examples of SVR head image quality scoring for inclusion in the dataset, scores of 3 or 4 were considered adequate quality for inclusion in the study.

2.1.2 3D SVR head reconstruction

All datasets were reconstructed for the whole head using two different automated SVR methods: the dedicated SVR pipeline developed for dHCP project [32] and the optimised automated version [33] of the classical 3D SVR reconstruction method [34] in SVRTK package[3] [4] [35]. The reconstructed 3D head images have 0.8 mm isotropic resolution and are reoriented to the standard radiological space (see examples in Fig. 1). In order to account for the small dimensions of some of the biometrics, we applied additional resampling of 0.5mm isotropic resolution to all 3D reconstructions prior to landmark label propagation.

The 3D whole head MRI image quality scoring protocol is shown in Fig. 2 similar to that proposed in our previous qualitative 3D MRI assessments [10]. An image score of

[3]SVRTK toolbox: https://github.com/SVRTK/SVRTK
[4]Automated SVRTK-based analysis tools: https://github.com/SVRTK/auto-proc-svrtk

'good' or 'excellent' was given when the brain could be visualised with no, or minimal, image or reconstruction artefacts, in addition, the mid and lower facial region and the facial profile should be included within the image volume.

2.2 Formalisation of 3D MRI craniofacial biometry protocol

Following an extensive literature search (JM), two 3-hour consensus workshops (JM, MR, AEC, AL, SA) were held which included an image review of proposed 3D landmarking points. A set of 31 biometric measurements (35 points) were agreed as relevant to clinical craniofacial assessment and feasible/reliable in terms of landmark visibility in 3D fetal MRI and included distance, angular and area measurements of the deep viscerocanial and cranial vault regions, see Fig. 3 for a visual representation of the 3D point cloud.

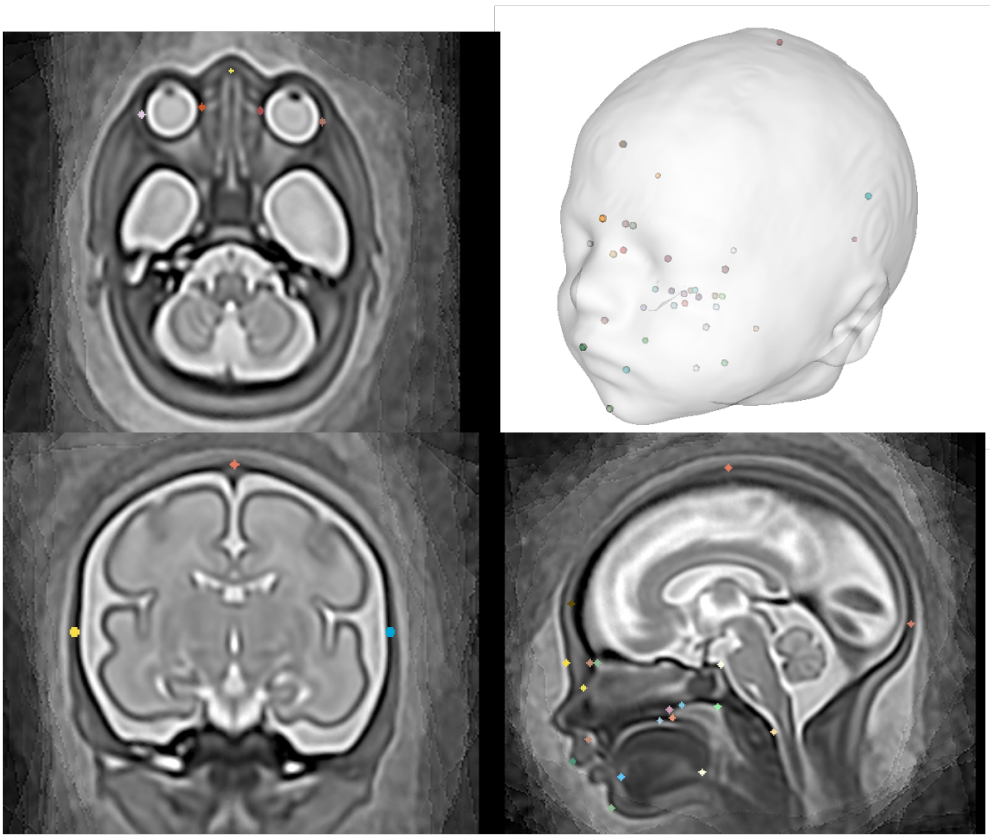
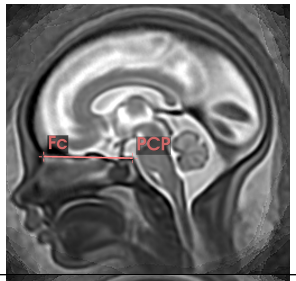
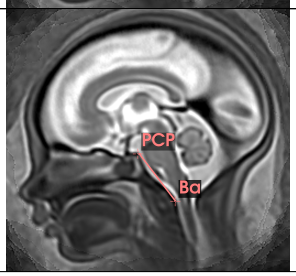


Fig. 3: Visual representation of 3D landmarks placed within a 3D population-averaged MRI atlas (31 weeks GA)

The formalised biometry protocol, with abbreviations used in this work, is summarised in Tab. 1 along with citations of publications that directly relate to the

proposed measurement technique and/or the related reference charts [16–18, 21, 27, 36–48]. Additionally, the wider literature supported the rationale for landmark definition and included: prenatal craniofacial biometry and anatomical MRI literature [49, 50]; 2nd and 3rd trimester prenatal US measurement literature [51–53]; ex-vivo/post-mortem anatomical studies [54, 55]; and, neonatal, paediatric or adult clinical imaging literature, where relevant [22, 56–59]. The label definition and location of anatomical points of interest were agreed upon in the consensus workshop, resulting in 35 points related to the measurements and a further 15 anatomical points of interest that may be of future interest (the latter not investigated in this work). These points are summarised in supplementary Tab. B1 and B2. The points were manually placed in 3D space on to a population-average atlas of the volumetric whole fetal head [60] using a 5mm 3D isotropic spherical 'paintbrush' with the ITK-SNAP software[5]; a visual overview of the points can be seen in Fig. 3. The atlas template, extended 3D segmentation file (50 anatomical points) and the landmarking protocol are publicly available online at the KCL CDB data repository[6].

Table 1: Formalised measurement definitions for the proposed landmark-based craniofacial biometry protocol with 3D T2w fetal MRI, with measurement and landmark abbreviations.

Measurement name (abbreviation)	Fetal MRI description (related reference[s])	Illustration
1. Anterior base of skull length (ABSL)	Distance of line projected from the foramen caecum to the posterior clinoid process (Fc-PCP) [36, 61].	
2. Posterior base of skull length (PBSL)	Distance of line projected from the posterior clinoid process to the anterior border of the foramen magnum, (PCP-Ba) [36, 61].	

[5]ITK-SNAP tool: http://www.itksnap.org/pmwiki/pmwiki.php
[6]KCL CDB fetal MRI head atlas repository: https://gin.g-node.org/kcl_cdb/craniofacial_fetal_mri_atlas

Table 1: Continued: Formalised measurement definitions for the proposed landmark-based craniofacial biometry protocol with 3D T2w fetal MRI, with measurement and landmark abbreviations.

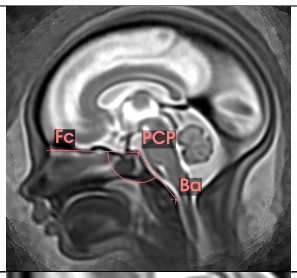
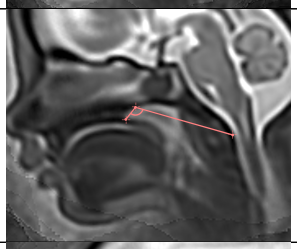
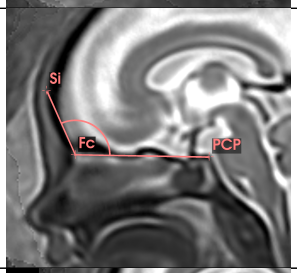
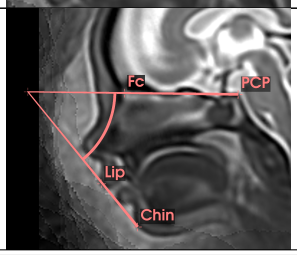
Measurement (abbreviation)	Fetal MRI description (related reference[s])	Illustration
3. Internal base of skull angle (CBA1)	Angle formed between the projected lines from the Foramen caecum to the posterior clinoid process to the anterior border of the foramen magnum (Fc-PCP-Ba)[36, 38–40].	
4. External base of skull angle (CBA2)	Angle formed between the projected lines from the posterior nasal spine to the hormion to the anterior border of the foramen magnum (PNS-H-Ba) [36, 38, 39].	
5. Fronto Maxillary Angle (FMA)	Fronto Measured on the sagittal plane and defined as the angle between the forehead and the ABSL (Si-Fc-PCP) [41]	
6. Inferior Facial Angle (IFA)	Measured on the sagittal plane and defined as the angle between the ABSL (PCP-Fc) and the line between the midline soft tissue of the mandible (chin) and upper lip (lip) [21](modified)	

Table 1: Continued: Formalised measurement definitions for the proposed landmark-based craniofacial biometry protocol with 3D T2w fetal MRI, with measurement and landmark abbreviations.

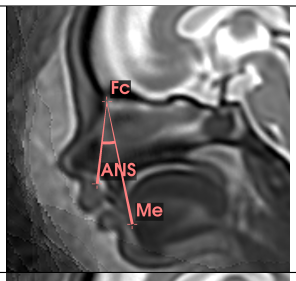
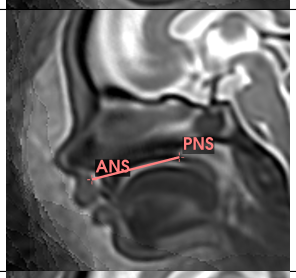
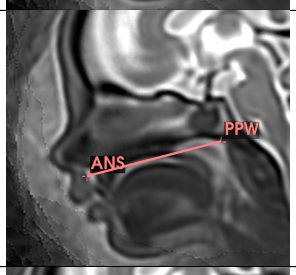
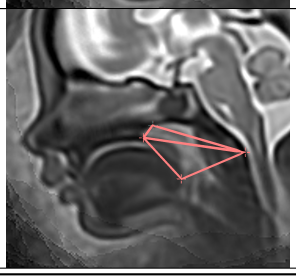
Measurement (abbreviation)	Fetal MRI description (related reference[s])	Illustration
7. Maxillary nasion mandibular Angle (MNMA)	Measured on the sagittal plane and defined as the angle between the alveolar ridge, foramen caecum and symphysis mentum (ANS-Fc-Me) [2]	
8. Hard palate length (HPL)	Distance between the alveolar ridge and the posterior nasal spine (ANS-PNS) [38, 39, 44].	
9. Velopharangeal length (VPL)	Linear distance of a projected line from the alveolar ridge, through the posterior nasal spine to the posterior pharangeal wall (ANS-PNS-PPW) [43].	
10. Nasopharyngeal area (NASO) and 11. Oropharangeal area (ORO)	In the mid-sagittal plane, NASO = area of an enclosed triangle formed by the landmarks of the Posterior nasal spine, Hormion point and the basion (PNS-H-Ba). ORO = area of an enclosed triangle formed by the landmarks of the posterior nasal spine, basion and posterior border of the tongue (PNS-Ba-TP) [38, 39].	

Table 1: Continued: Formalised measurement definitions for the proposed landmark-based craniofacial biometry protocol with 3D T2w fetal MRI, with measurement and landmark abbreviations.

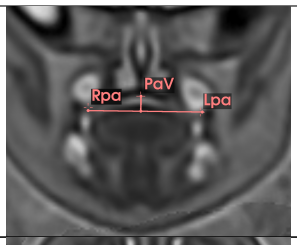
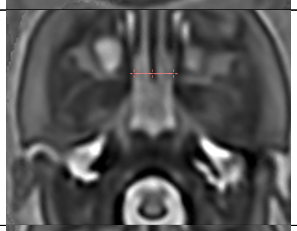
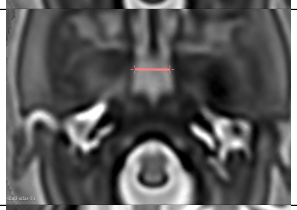
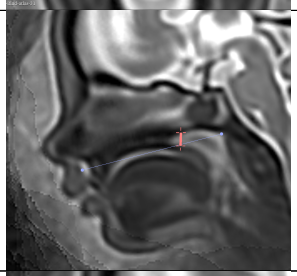
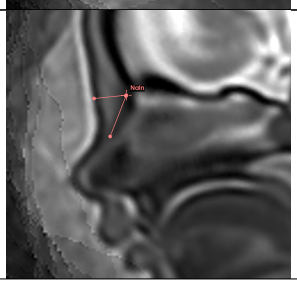
Measurement (abbreviation)	Fetal MRI description (related reference[s])	Illustration
12. Palate width (PaW) and 13. Palate height (PaH)	In the mid-coronal plane aligned orthogonally to the face, PaW = the distance between the inferior margin of the most posterior maxillary toothbud visualised at its widest point (RtPa-LtPa). PaH = the distance from the palate vault, i.e. inferior border of the hard palate, to the mid-PaW line (PV-PaW), [43] .	
14. and 15. Right and left choanal width (RtCho/LtCho)	In axial plane, the linear distance from the medial pteygoid process on the lateral border of the choanae to the midline of the vomer posteriorly (Rt/LtMPP - Vo), [45].	
16. Nasopharyngeal width (NPW)	In axial plane, the linear distance of the lateral walls of the anterior nasopharynx level with the hormion point superiorly and at the widest point (LtCho-RtCho) [45].	
17. Choanal height (ChoH)	In Sagittal plane, Distance of line projected from the hormion point to a perpendicular intersect with anterior to posterior hard palate (H-HPL) [16, 45].	
18. Nasal Bone (NB) and 19. Pre-nasal Thickness (PnTh)	In Sagittal plane, NB = distance from the nasion (outer bony border, NaIn) to the tip of the nasal bone (NBt). PnTh = distance from the NaIn to the outer soft tissue border of the nasion (NaO) [41, 42, 62].	

Table 1: Continued: Formalised measurement definitions for the proposed landmark-based craniofacial biometry protocol with 3D T2w fetal MRI, with measurement and landmark abbreviations.

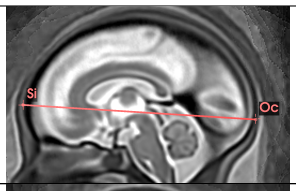
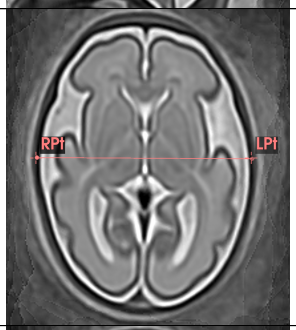
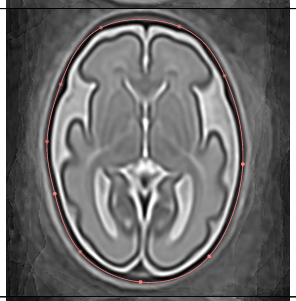
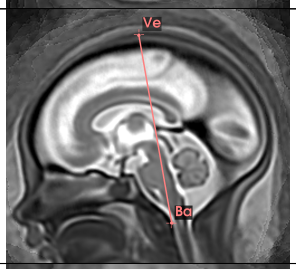
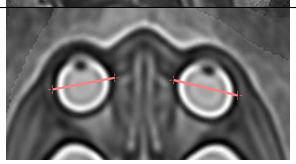
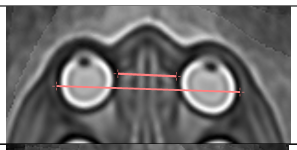
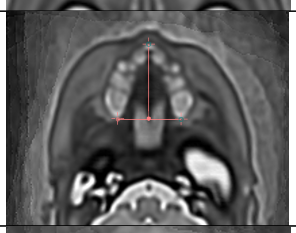
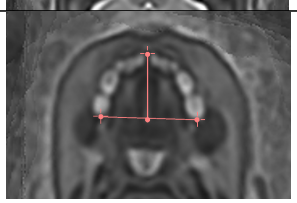
Measurement (abbreviation)	Fetal MRI description (related reference[s])	Illustration
20. Occipitofrontal Diameter (OFD)	In mid-sagittal plane, distance between the most anterior point of frontal bone to the posterior-most point of the occiput (Si-Oc) [27]	
21. Biparietal diameter (BPD)	In axial plane, distance between the widest lateral point of the parietal bone at the level of the OFD (RtPt-LtPt) [27]	
22. Head circumference (HC)	Outer circumference of the skull, calculated from OFD and BPD using an ellipse formula, an established approach in ultrasound clinical practice [27, 46]	
23. Maximum cranial height, (MCh),	In sagittal plane, distance between the vertex of the skull and the basion (Ve-Ba) [47]	
24.-25. Right and Left occular diameter (OD)	In axial plane, distance between the widest lateral borders of each orbit to include the sclera of the eye. Note, in MRI the term 'occular' is used to refer to the globe, rather than the bony orbital margin, that is measured [17, 31]	

Table 1: Continued: Formalised measurement definitions for the proposed landmark-based craniofacial biometry protocol with 3D T2w fetal MRI, with measurement and landmark abbreviations.

Measurement (abbreviation)	Fetal MRI description (related reference[s])	Illustration
26. Interoccular diameter (IOD) 27. Bioccular diameter (BOD)	In axial plane, IOD = distance between the widest medial points of both orbits. BOD = distance between widest lateral points between both orbits [17, 31]	
28. Maxillary width, (MxW) 29. Maxillary length (MxL)	Aligned orthgonally to the face, MxW = the distance between the posterior-most margins of the posterior maxillary toothbuds visualised at their widest point axially. MxL = distance alveolar ridge (ANS) to the midpoint of the MxW line [48]	
30. Mandibular width, (MdW) 31. Mandibular length (MdL)	Aligned orthgonally to the mandible, MdW = the distance between the posterior-most margins of the posterior mandibular toothbuds as they emerge from the masseter muscle, in axial plane. MdL = distance the outer border of the symphysis menti at its mid-point to the midpoint of the MdW line [21]	

2.3 Automated 3D craniofacial biometry pipeline

The proposed pipeline for automated biometry is outlined in Fig. 4. Firstly, the defined 3D labels from the atlas were propagated to the subject space via registration. We employed classical affine + non-rigid free form deformation registration [63] with local normalised cross-correlation similarity metric with a 6 mm control point spacing implemented in MIRTK[7]. This registration-based approach was feasible since the 3D head images were reoriented in the standard atlas radiological space apriori after SVR reconstruction. The registration parameter were optimised for this particular task. Next, the landmark labels from the atlas were transformed to the subject space using output transformations. Label propagation was followed by computation of the defined biometry parameters based on the landmark coordinates including distances, angles and areas. Next, the landmark coordinates were computed as centre-points of propagated landmark labels in 3D world space. Lastly, this was followed by the calculation of the defined biometry measurements (Tab. 1) using the formulas given below:

[7]MIRTK package: https://github.com/BioMedIA/MIRTK

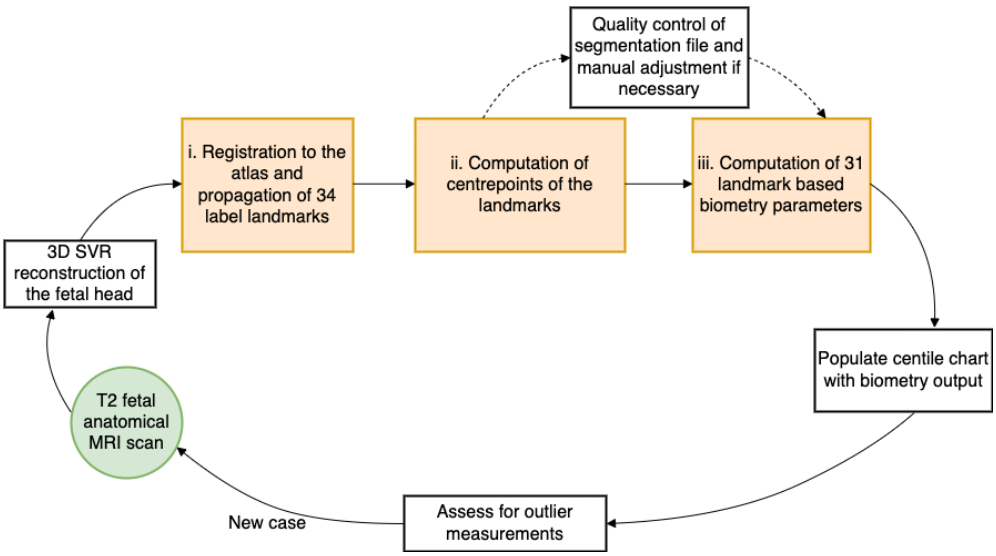


Fig. 4: Proposed pipeline for atlas-based 3D craniofacial biometry for fetal MRI (orange boxes)

- the distance d between two landmarks $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ in 3D space:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{1}$$

- the angle θ between two lines defined by landmarks $P_1(x_1, y_1, z_1)$, $P_2(x_2, y_2, z_2)$, and $Q_1(x_3, y_3, z_3)$, $Q_2(x_4, y_4, z_4)$ in 3D space:

$$\theta = \arccos\left(\frac{(x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3) + (z_2 - z_1)(z_4 - z_3)}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}\sqrt{(x_4 - x_3)^2 + (y_4 - y_3)^2 + (z_4 - z_3)^2}}\right) \tag{2}$$

- the distance d between a landmark $P(x_3, y_3, z_3)$ and the center point of a line defined by landmarks $Q_1(x_1, y_1, z_1)$ and $Q_2(x_2, y_2, z_2)$ in 3D space:

$$d = \sqrt{\frac{(x_1 + x_2 - 2x_3)^2 + (y_1 + y_2 - 2y_3)^2 + (z_1 + z_2 - 2z_3)^2}{4}} \tag{3}$$

- the area A of a triangle from three landmarks $P_1(x_1, y_1, z_1)$, $P_2(x_2, y_2, z_2)$, and $P_3(x_3, y_3, z_3)$ in 3D space:

$$A = \frac{1}{2}\|\mathbf{x}_1(x_2 - x_1, y_2 - y_1, z_2 - z_1) \times \mathbf{x}_1(x_3 - x_1, y_3 - y_1, z_3 - z_1)\| \tag{4}$$

- the length of an ellipse L is calculated by a formula using two perpendicular lines (the OFD and BPD), each defined by landmarks $P_1(x_1, y_1, z_1)$, $P_2(x_2, y_2, z_2)$, and

$Q_1(x_3, y_3, z_3)$, $Q_2(x_4, y_4, z_4)$ in 3D space:

$$L = 1.62\times\left(\sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}+\sqrt{(x_4-x_3)^2+(y_4-y_3)^2+(z_4-z_3)^2}\right) \tag{5}$$

The implemented function for landmark-based biometry calculations *craniofacial-biometry* is publicly available as a part of SVRTK package.

2.4 Evaluation of the proposed biometry protocol

With the pipeline outputs of the automated label segementation files and 31 biometrics per subject, we performed an extensive evaluation of the feasibility of the proposed protocol and pipeline on normal and T21 cohorts from different acquisition protocols. This included qualitative assessment of landmark localisation in 132 datasets followed by extensive quantitative evaluation based on comparison with manual measurements in 10 cases.

2.4.1 Qualitative visual evaluation of landmarks

A single operator visually inspected all automated 3D points for 132 subjects (later used in normal and T21 comparison). The whole head SVR volume and landmarks for each subject were loaded into Slicer 3D in turn to detect any major errors in the landmark placement. 35 landmarks per subject related to biometry were inspected and scored as 'minimal or no' error, or as having a 'major' error (i.e. requiring a manual change of more than a few millimeters, degrees or mm^3). Any measurement that was documented as an outlier (based on the distribution of cohort), had a detailed inspection and qualitative comments on the suitability of the related landmark placement. During this assessment each case was rescored for image quality, blinded to the intial image score during data curation and including an additional criteria to those presented in Fig. 2 so that *'reasonable image contrast to differentiate deep naso-oral soft tissue structures'* was also considered in the qualitative assessment.

2.4.2 Measurement Validation, intraobserver, interoberver, intramethod and intermethod

We performed assessment of intra-, inter-observer and intermethod variability in a subset of 10 random cases (from the control and T21 cohorts) given in Tab. 2. Intraobserver repeatability was assessed by repeating measurements on the subset using a similar method to the label propagation pipeline [64]. That is, the point related to the biometric was placed using ITKsnap software, with a 5mm 3D sphere, and then the biometry was indirectly calculated automatically from the centroid centre points. The assessment was repeated after a washout period of 2 weeks, to reduce reviewer bias.

To understand the expected variability within clinical raters, the interobserver reproducibility was assessed with three fetal radiology experts. After importing the SVR volume into MITK workbench, the use of coupled cross-hairs in the x-y-z planes allowed for fine and precise adjustment of image planes within the 3D volume, required for the direct calliper measurements.

Table 2: Demographics of the 10 cases selected for quantitative evaluation of the proposed biometry protocol and pipeline.

ID	GA	Group	Sex	TE	Field Strength	HeadSVR quality
1	29.86	T21	Male	180ms	3.0T	Good
2	35.43	T21	Male	180ms	3.0T	Moderate
3	32.71	T21	Male	250ms	3.0T	Moderate
4	33.71	T21	Female	80ms	1.5T	Good
5	32.71	T21	Female	80ms	1.5T	Excellent
6	31.86	Control	Male	80ms	1.5T	Moderate
7	32.86	Control	Female	80ms	1.5T	Good
8	29.43	Control	Female	80ms	1.5T	Moderate
9	30.00	Control	Male	180ms	3.0T	Moderate
10	35.14	Control	Female	250ms	3.0T	Moderate

The absolute and relative differences were calculated for intra-observer, inter-observer and the automated measurements compared with the most experienced fetal radiologist, considered to be the ground truth. Bland Altman plots were constructed to visualise the variability of the automated and manual observers and compared to the most expert observer. An intraclass correlation coefficient (ICC) (two-way random effects model with absolute agreement) was performed to statistically assess the reliability of both systematic and random error to include; intraobserver (single observer, repeated indirect measurements), interobserver (3 expert raters, single direct measurements), and intermethod agreement (4 human observers and automated measures, single measurements). The ICC interpretation threshold values were reported as: <0.50 = poor; 0.50 -0.75 = moderate; 0.75 - 0.90 = good; and, >0.90 = excellent as defined by Koo and Li (2016) [65]. A Cronbach's Alpha test was applied to assess internal validity for each measurement, as a high internal error within the subset will reduce the power and therefore validity of the ICC [66]. A threshold of 0.7 was set as a minimum value to interpret the ICC value safely.

All human observers were asked to rate their diagnostic confidence of measuring on a binary scale (confident/not confident) for all the measurements or landmarks, where relevant. Due to the small sample size the results were presented descriptively.

2.5 Comparison of normal and abnormal cohorts

Next, in order to assess the clinical utility of the proposed biometry pipeline for automated analysis of a large number of craniofacial biometrics, we ran a multivariate analysis of covariance (MANOVA) to assess the impact of scanner field strength with GA as a covariate (control group only). A comparison with 24 T21 cases and 108 GA-matched normal control cases, including three different acquisition protocols, was tested with a univariate analysis (ANOVA) using robust standard errors to account for multiple comparisons [67]. Posthoc power calculations were included to assess the risk of type 2 error and magnitude of effect size (using, partial estimated squares, η_p^2) [68]. In addition, the centile normal range charts were used to assess the proportion of T21 cases falling above the 95% or below the 5% for gestation for the most relevant biometrics identified by the ANOVA.

To ensure the test assumptions were met for the analysis of variance, normality was evaluated using Q-Q plots and Shapiro-Wilk tests, multicollinearity was assessed with a Pearson's correlation coefficient, tabulated correlation matrix, and a variance inflation factor (VIF) test was performed to detect severe multicollinearity (i.e. VIF >10). Lastly, linearity was visually assessed with scatterplots of the variable against GA and lastly, homogeneity of variance was assessed with a Levene's Test, [69]. Parameter estimates using robust standard errors [67] and using the HC3 method [70] are reported for the analysis of variance to account for any violations in test assumptions.

2.6 Normal craniofacial biometry growth charts

Next, the automated label propagation for all suitable control cases from 19 to 38 weeks GA was performed to create extended GA nomograms for the proposed biometry protocol. The 3D points were visually inspected and corrected if necessary, where required. After extraction of all biometric measurements, the 5th, 50th, and 95th centiles for the automated biometry results were calculated based on the widely used statistical approach for growth chart construction [71] similar to fetal MRI brain charts described by Kyriakopoulou et al. (2017) [27]. Normal range plots with centile trendlines and quadratic formula (or linear formula where relevant) formula were then produced.

2.7 Image Processing and Statistical Analysis Software

All image processing steps, including image reconstruction, reorientation and landmark propagation, were implemented using SVRTK and MIRTK packages.

All image review software used was open source and compatible with nifti and/or dicom format 3D image volumes. Image landmark labelling for the atlas template and for the intraobserver measurements was performed in ITKsnap. Interobserver measurements were performed in MITK workbench. The review of labels for data quality was performed in 3D slicer [72].

Data was analysed in Excel (Microsoft Excel for Mac, Version 16.85, 2024; descriptive statistics, plots), SPSS Statistics (IBM corp, version 29.0.2.0 (20), 2023; growth curves, ICC, ANOVA, MANOVA), and, RStudio (R version 4.3.3 (2024-02-29, data visualisation and outlier assessment).

3 Results

3.1 Evaluation of reliability of the proposed biometry protocol

3.1.1 Visual assessment of the proposed automated method

The automated pipeline produced an output of labels, centre-points and biometrics for all cases. The quality of the label placement is fundamental to the centre-point extraction and therefore the subsequent biometric calculation, thus a detailed assessment of image quality and appropriateness of landmark placement is described below:

Visual inspection and outlier assessment of all subjects The 132 subjects used in evaluation the had paired SVR and landmark data available for visual assessment, which took approximately 5 minutes per case. Of the 4620 total landmarks assessed, n = 15, 0.3%, were deemed as requiring major editing/unsuccessful by a single operator. The 15 unsuccessful labels were all in SVR images of moderate quality and no single subject had more than one major label error. Qualitative comments were collected about the landmark placement, with the most comments made for the lip (n = 3), posterior tongue (n = 3) and chin (n = 2). The limitations in image quality described were related to low contrast resolution and noise in the lower facial region as well as compression of superficial soft tissue by external structures e.g. placenta or maternal uterine wall. Poor visibility was also described in the naso-oropharyngeal area, which received multiple mentions (n = 17) largely due to fluid motion artefact, especially in the choanal region. Fig. 5 gives an overview of the SVR QA results stratified by field strength over gestational age and image examples of major landmark errors.

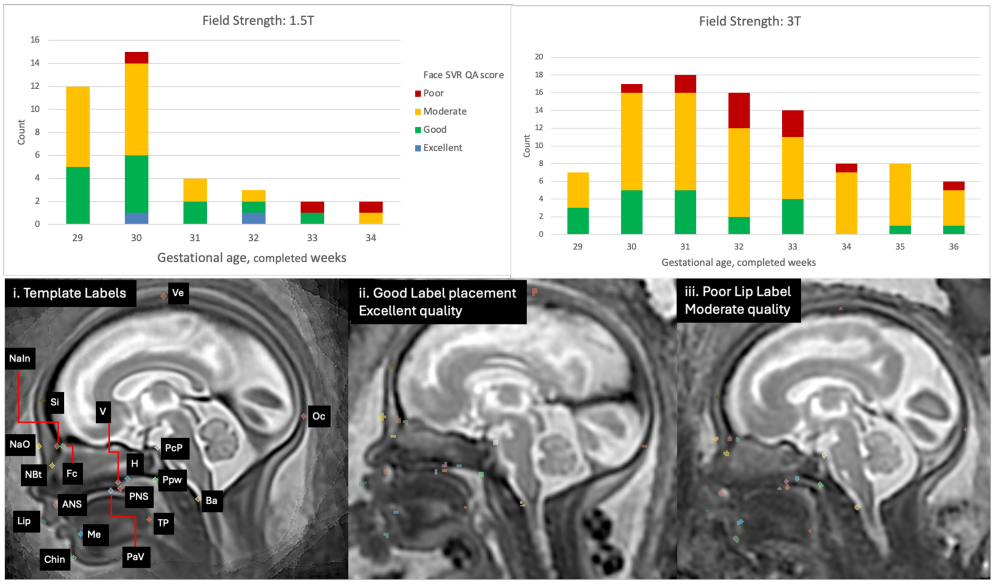
Outlier measurements in the control and T21 groups were defined by assessing the measurement distributions, see boxplots in Fig. 6. 57 measurements were outliers from 32/132 unique subjects, 4 (12.5%) of the 32 outlier subjects were scanned at 1.5T and the remaining at 3T MRI field strength. The image quality of the outlier subjects ranged from poor to excellent, with a moderate score having the highest frequency. Only one measurement, the internal cranial base angle, in a 33 week old fetus scanned at 3T, was an 'extreme' outlier, i.e., it had values above the 3rd quartile and also was more than 3 times the interquartile range. The labels related to the angle measurement appeared correctly placed on visual inspection.

3.1.2 Automated and human observer agreement and reliability

The observer variability was assessed to understand if the measurement error found for the automated biometrics were within the limits expected for expert observers. 10 cases were randomly selected for review and are presented in Tab. 2, and were balanced interms of fetal sex, MRI field strength and were not selected for quality apriori. The repeated measures (for intraobserver agreement and reliability) were performed by a single observer, a clinical researcher with 15 years fetal imaging experience. The reproducibility measures (for interobserver agreement and reliability) were performed by 3 consultant fetal MRI neuroradiologists, with experience ranging from ****XX to xx years**** RADIOLOGISTS TO ADD.

For the assessment of variation of the automated measurements, the most experienced consultant radiologist was considered the ground truth, to which the absolute and relative difference were compared and presented with its mean bias and limits of agreement in a selection of Bland Altman (BA) plots, see Fig. 7 (absolute and relative difference tables for all variables can be found in supplementary Fig. A1). Intermethod agreement and reliability included all observations except the first set of intraobserver measures. An acceptable relative mean bias of less than $+/-10\%$ was seen for 22/31 automated measurements. Despite this, all automated measurements had mean error within the range seen for human observers except for the mandibular width (mean absolute difference; automated = 5.30mm (14.62%) versus manual

Fig. 5: Upper row: Face SVR quality assessment results stratified by MRI field strength (1.5T/3T) and gestational age. Lower Row: i. automated landmarks with labels, ii. Example: Excellent quality SVR with good landmark placement, iii. Example: Moderate quality SVR (poor contrast resolution lower face), with poor lip label placement due to limited boundary definition adjacent to maternal tissue. (Label key: Ve=vertex; Oc=occiput; Si=sinciput; NaIn=inner nasion; NaO=outer nasion; Fc=foramen caecum; NBt=nasal bone tip; v=vomer; H=hormion; PcP=posterior clinoid process; Ppw=posterior pharyngeal wall; Pav=palate vault; Ba=basion; ANS=anterior nasal spine; PNS; posterior nasal spine; Tp=posterior tongue; Me=bony mentum; Lip=lip; Chin=chin)



(range for 3 observers) = -0.99 - 1.28mm (-2.63 - 4.55%)). The automated random error in all 31 measurements was within the limits of agreement for that seen in the manual observers, however, nine measures had an unacceptably high random error of more than $+/-20\%$ (ChoL, PnTh, CBA2, MNMA, IFA, PAH, ChoH, NASO and ORO) [73].

Intermethod reliability (i.e. for all measurements/methods including the automated) was excellent for the cranial measurements of; the anterior and posterior base of skull; the occipitofrontal diameter; biparietal diameter; maximum cranial height; internal cranial base angle; head circumference and, bi-occular diameter, (ICC range = 0.915-0.986, 95% CI range, 0.770-0.995). The intermethod reliability threshold was not below that of the interobserver reliability (expert radiologists) in any assessed variables except for the mandibular width where the ICC score changed from good (ICC = 0.904, 95% CI = 0.677-0.975) to moderate (ICC = 0.635, 95%, CI = 0.176-0.888). The ICC values were interpreted only when good internal validity of the measurement

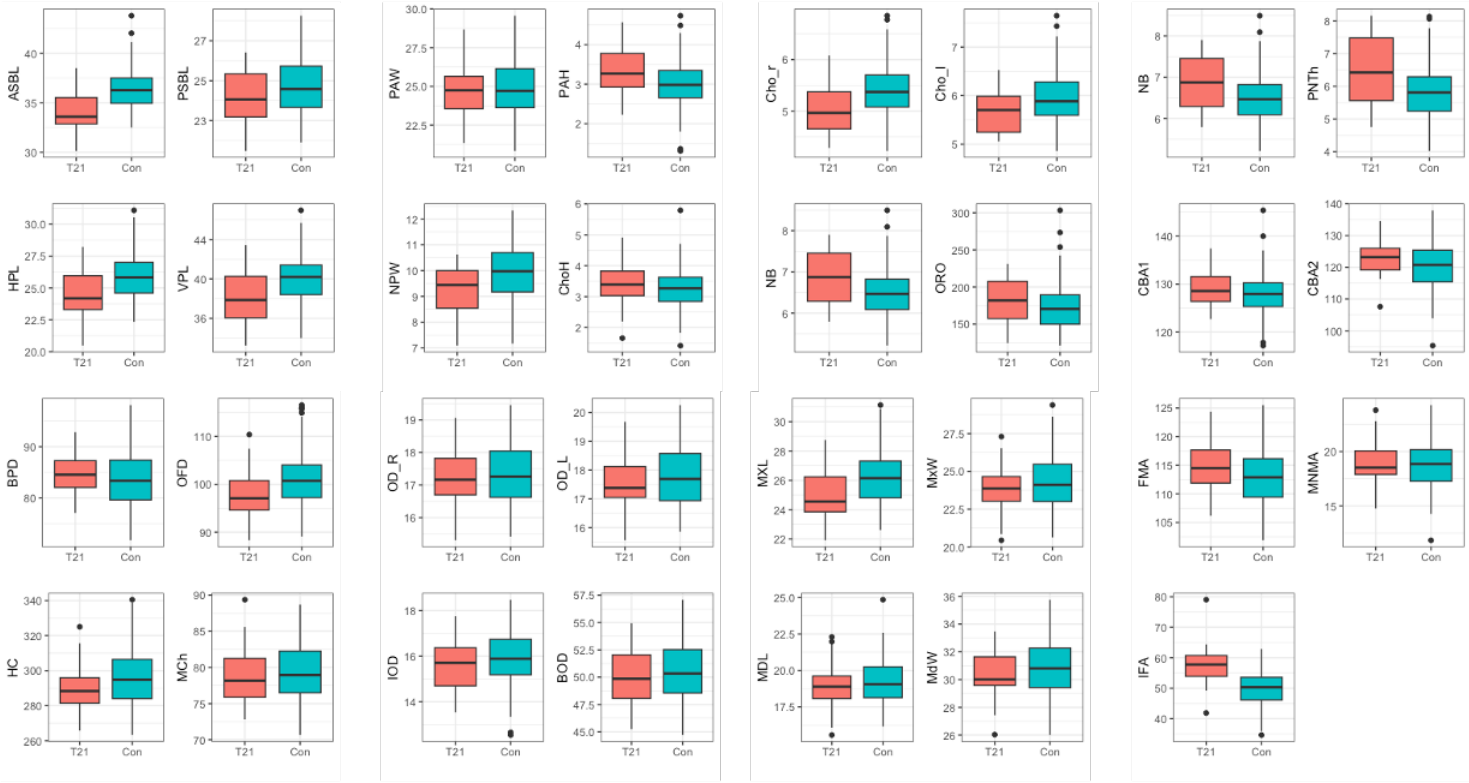


Fig. 6: Boxplots of measurements stratified by healthy control (blue) and T21 groups (red). Central tendency is represented by the black horizontal line, and the upper and lower limits of the box represent the interquartile range (2nd to 3rd quartile). The black dots represent outliers

was present, i.e. Cronbach's Alpha >0.70, violation of the ICC test assumptions was seen in 12/31 measurements. Full ICC results can be found in supplementary Fig. A2.

3.1.3 Human observer confidence

Subjective scoring of diagnostic confidence found that observers were least confident with measurements in the choanal and nasopharyngeal area. The intraobserver repeated measures included failed landmark placement for the nasopharyngeal space, i.e. choanal and pharyngeal width measurements and the nasal bone tip, see Fig. 8a and the average confidence scores and the variation across the three radiologist measurements are presented as boxplots in Fig. 8b. Of note, a total of 4/10 cases could not have an accurate measurement of the nasal bone performed. For the measurements obtained, the confidence scoring demonstrated that there was minimal/no variability in confidence scores, with biometrics generally scored as 'confident'. The exception was for the choanae width measurements (left, right and total width, i.e. NPW) and for cases 7 and 10 where multiple measures were more difficult. Both cases were at different scanner field strengths (1.5T/3T) and GAs (32.38 and 35.14 weeks), and image quality was subjectively scored as moderate.

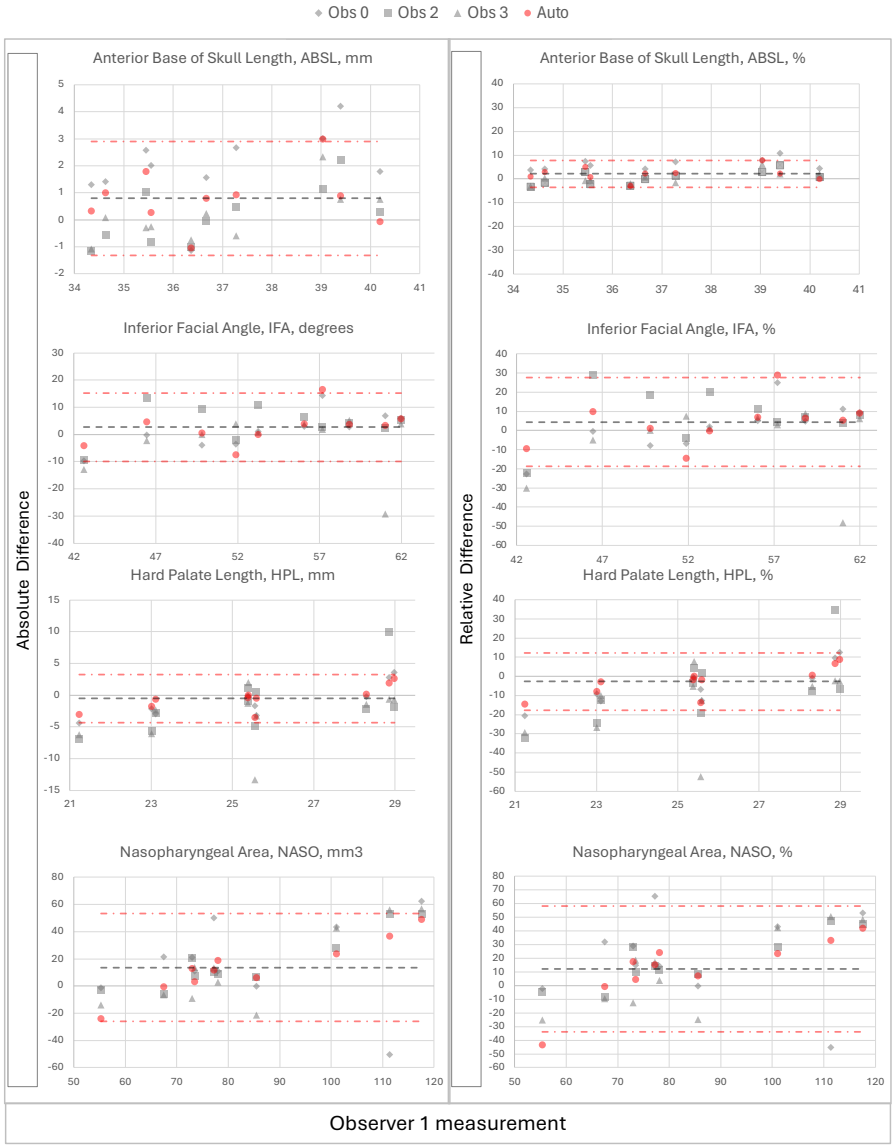
3.2 Comparison of normal control and T21 cohorts

In order to understand the clinical utility of the proposed automated pipeline, and its ability to assess differences in craniofacial development, an analysis of variance was performed on the dataset.

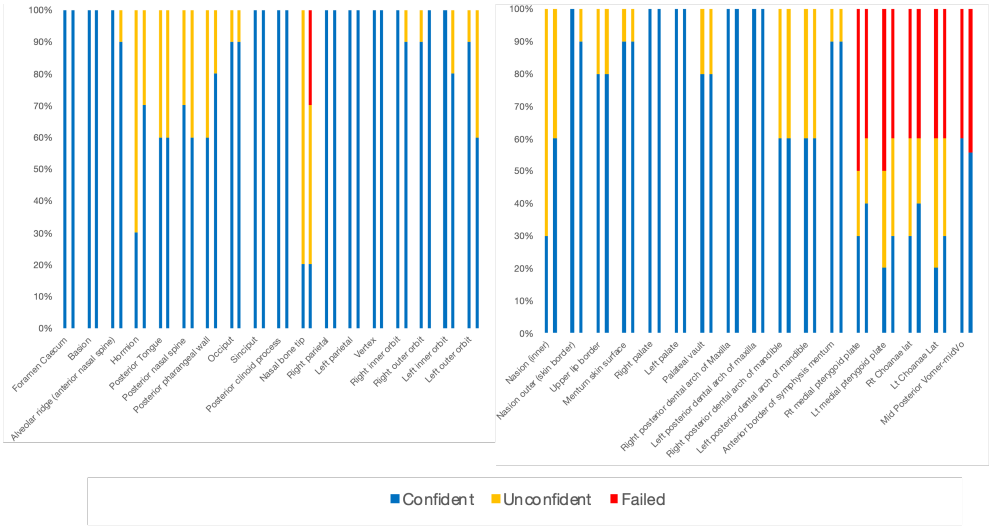
3.2.1 Demographics and baseline characteristics

The final retrospective sample contained 108 control and 24 T21 subjects. The mean GA for the control group was 31 weeks and 6 days (range 29 weeks and 0 days to 36 and 0 days) and for the T21 group the mean GA was 32 weeks and 3 days (range 29 and 6 days to 35 and 5 days). The distribution of MRI protocols used in the healthy control group differed from the T21 group (1.5 Tesla and echotime of 80ms in 34 (31.5%) and 4 (16.7%) respectively; 3 Tesla and echotime of 180ms in 25 (23.1%) and 17 (70.1%) respectively; and, 3 Tesla and echotime of 250ms in 49 (45.4%) and 3 (12.5%) respectively). Most datasets in the T21 cohort were performed at 3 Tesla at 180ms echotime whereas the datasets were more evenly spread in the control group, with most scans performed at 3 Tesla and 250ms. The ratio of female and male fetuses were evenly split between the cohort groups (healthy controls = 53 (49.1%)/ 54 (50.0%), and, T21 = 10 (41.7%)/10 (41.7%), the remaining fetuses had unknown sex at the time of the scan). Fig. 9 shows the spread of fetal sex and protocols across the gestational age, with the control group having a peak at 30 weeks GA and the smaller sample of T21 fetuses being more evenly spread. The maternal ethnicity across the control group was largely of White European origin (79.6%), with 9.3% of participants reporting as having an Asian background, 3.7% as having a Black background, and the remaining participants of other or unknown ethnicity, see Tab. B4.

Fig. 7: Bland Altman plots of absolute and relative difference for observer 0, grey diamond (using manual landmark-based indirect measurement method), observers 2 and 3, grey square and triangle respectively (using direct manual measurement method), and automated biometry, red circle, all compared to expert observer 1 (direct manual method) for a selection of biometrics (ABSL, IFA, HPL, and NASO). Grey dash=mean difference for automated biometry and red dash=upper and lower limits of agreement for automated method.



(a) Confidence in 3D landmark placement (n=35) between first and second repeated observation



(b) Boxplots of average confidence for 3 raters for obtaining measurements (scale: 0 = failed, 1 = unconfident, 2 = Confident). * Represents case number outliers, see 2. Note, small boxes at point 2 on the y-axis represents no variation.

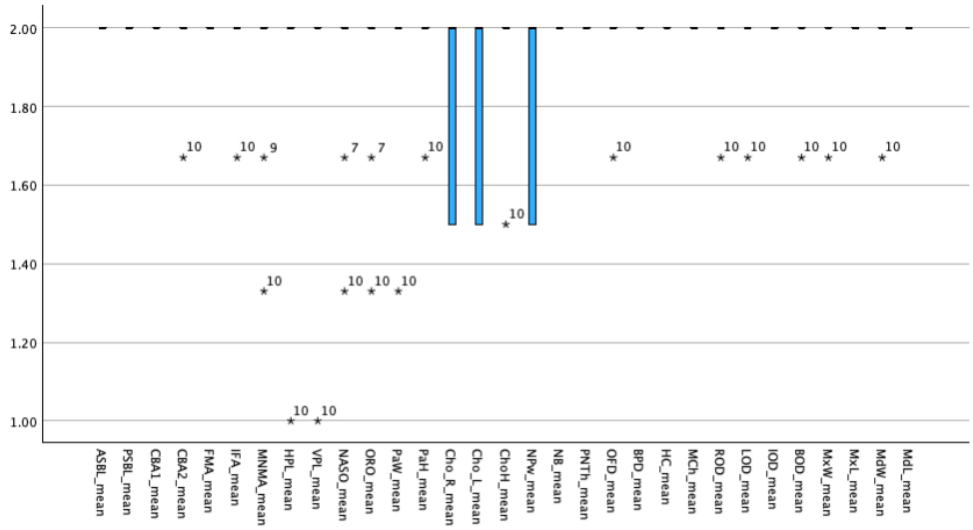


Fig. 8: Human Observer Diagnostic Confidence of Landmark placement and measurement accuracy. a. Observer 0 (repeated measures). b. Average of observers 1 to 3 (each performing single independent measurements

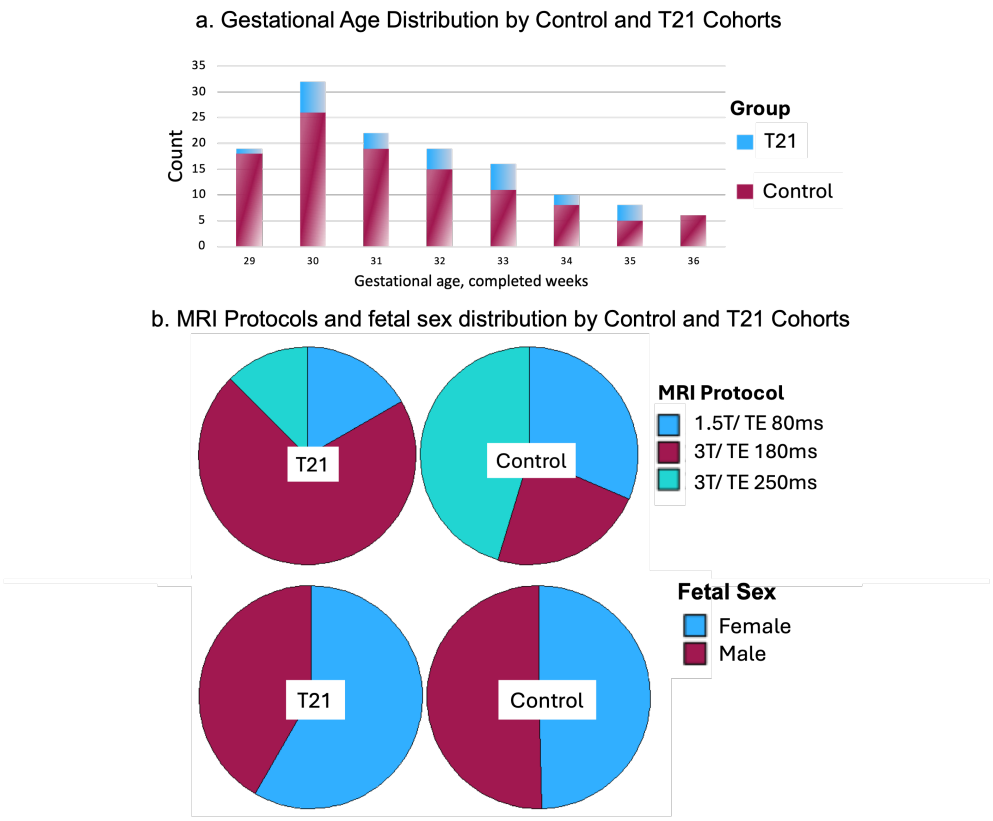


Fig. 9: GA-matched fetal MRI datasets used in the control (104) and T21 (24) comparison a. GA distribution b. MRI protocols and fetal sex

3.2.2 Summary statistics of the dataset

The mean measurements and standard deviations were similar between the gestational age-matched T21 and healthy control groups, see supplementary Tab. B3, however, there were relatively larger mean differences between the T21 and control groups for; the occipitofrontal diameter (97.87mm and 101.05mm, respectively); inferior facial angle (55.16 and 48.97^0, respectively); and, the nasopharyngeal and oropharyngeal areas (58.42 and 63.88mm^3 and 181.41 and 173.97mm^3, respectively). Box plots for the measurements indicate these differences graphically and it was noted that outlier cases were noted predominantly in the control group (see Fig. 6).

3.2.3 Growth Chart Utility and Biometric Variability Assessment

Biometric growth charts based on the age-matched control cohort were constructed for each automated biometric measurement with a calculated best-fit line and standard deviation regression equations (presented in the supplementary Tab. B5). Using the centile ranges to assess the T21 subjects falling outside the normal ranges, the 5th

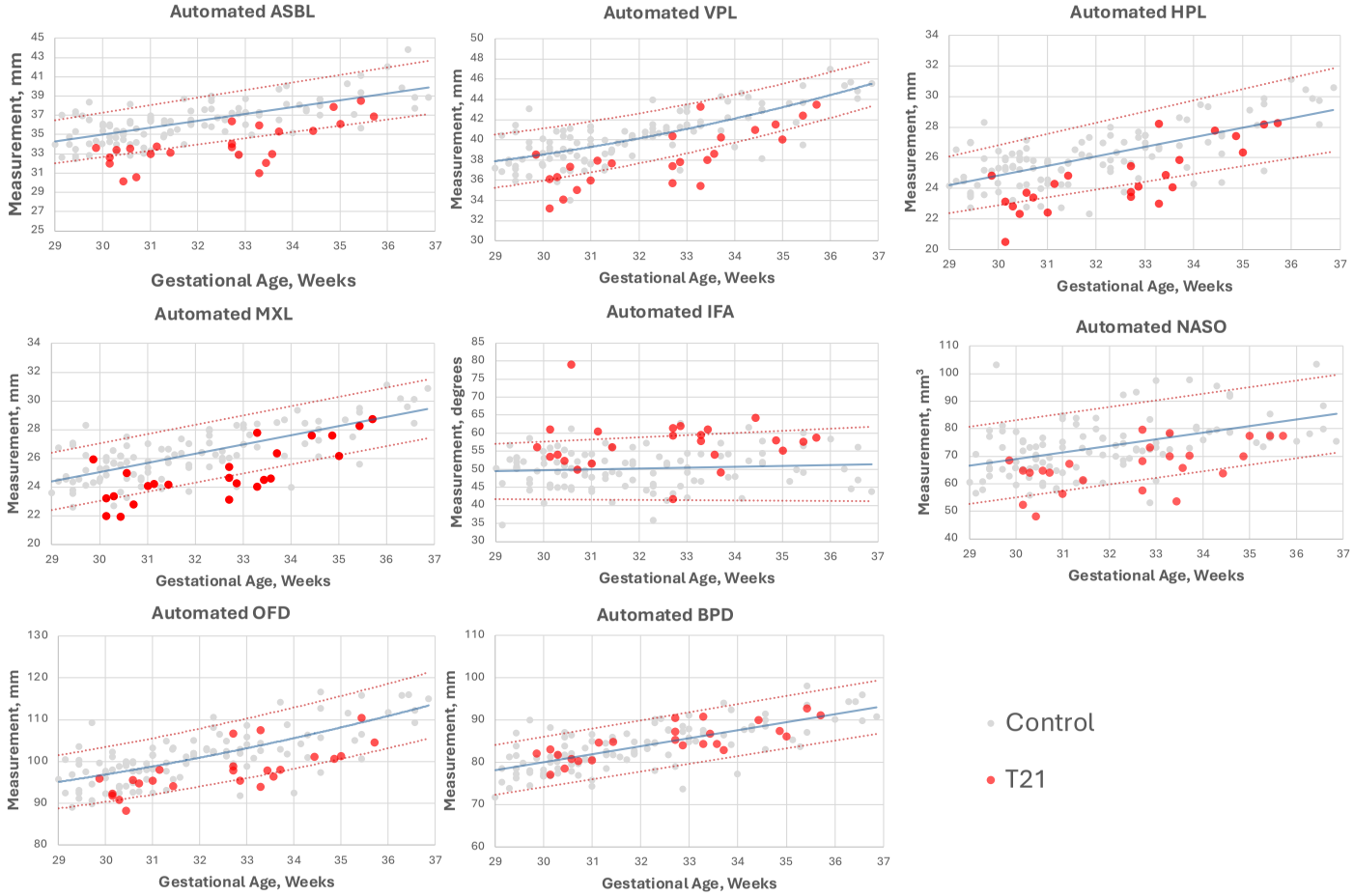


Fig. 10: Growth charts for seven craniofacial biometrics* with statistically significant differences from control group and a large effect size, and biparietal diameter, BPD, with no significant differences between groups. (*ASBL, anterior skull base length; VPL, velopharyngeal length; HPL, hard palate length; MXL, maxillary length, IFA, inferior facial angle; NASO, nasopharyngeal area; and, occipitofontal diameter.)

to the 95th percentile were considered the threshold thus, by definition, only 10% of the healthy control cohort would be expected to fall outside of this range. 17/24 of the T21 fetal subjects had measurements that were out of range for at least one of the most meaningful 7 biometrics, i.e. the ASBL, VPL, HPL, OFD, IFA, MXL, and NASO (i.e. 71% true positive rate, TP), see Fig. 10. When considering all biometrics 23/24 cases had at least one biometrics falling outside of normal range (96% TP), with 21/24 cases having at least two measurements falling outside of the normal range (88% TP), see Fig. 11.

Biometric variation in the control population for subgroups of MRI field strength and GA

A MANOVA was conducted to examine the effects of MRI field strength (1.5T/3T) on the combination of 31 dependent continuous variables, i.e. the craniofacial measurements, with gestational age in completed weeks included as a covariate.

Variation with Gestational age: Evaluating the control group, the overall biometrics varied with GA, p <0.001, $\eta_p^2 = 0.866$, observed power = 1.00 and on univariate analysis there was no statistically significant measurement variation seen with GA for the FMA, IFA, CBA1 or ChoH (p = 0.943, 0.432, 0.639, and 0.827 respectively).

Variation with field strength: Within the control group there was a statistically significant difference in overall biometric variation between 1.5T and 3T MRI scans (with GA as a covariate), $p = 0.006$, $\eta_p^2 = 0.537$, observed power = 0.99. ****check correction for multiple variables****

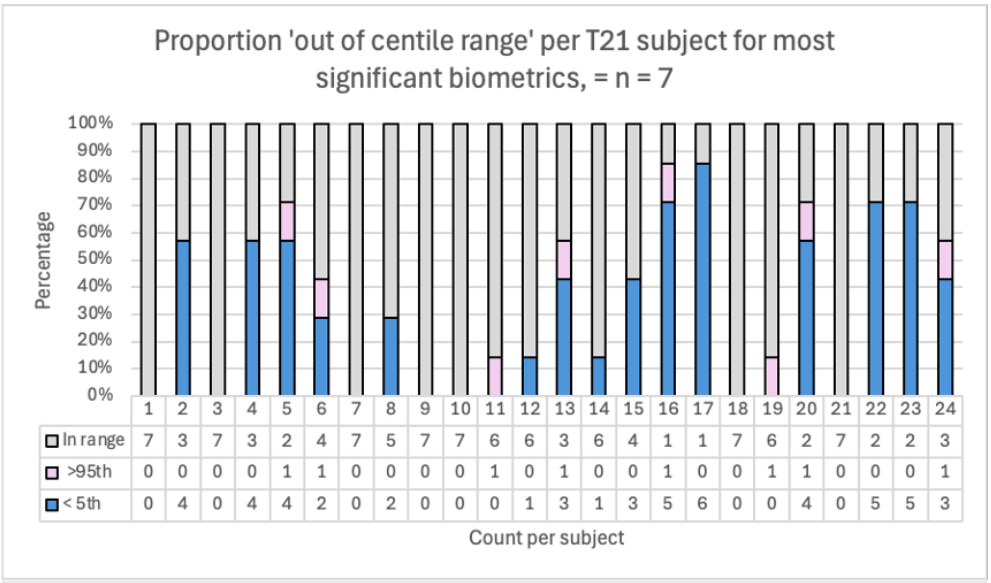
Overall biometric variation between control and T21 groups

A two-way MANOVA was performed to examine whether the dependent variables (the craniofacial measurements) differed by disease status i.e. control group or confirmed T21. The combined dependent variables of the main effect variable (disease status) showed statistically significant differences between groups $p < 0.001$, $\eta_p^2 = 0.710$, observed power = 1.000.

Individual biometric variation between control and T21 groups

The ANOVA indicated that 7/31 variables; ASBL, HPL, VPL, OFD; IFA, MXL, and, NASO, were statistically significantly different (p <0.05) and had large effect sizes ($\eta_p^2 = 13.6$ and 30.4%), with the ASBL having the largest effect, see supplementary Tab. B6. 13 additional variables were statistically significant (PSBL, ChoL, ChoR, NPW, PnTh, ODL, ODR, IOD, BOD, MXW, MDL, MDL, and, HC) however had relatively small effect sizes of between 3.4 and 11.7%.

Sensitivity analysis: All non-significant variables (11/31) had a power of less than < 80% (PAH, PAW, CHOH, BPD, MCh, CBA1, CBA2, FMA, MNMA, ORO and interestingly NB). There were four variables reaching statistical significance despite a power below 80%; ODR; IOD; MXW; and, MDL; all had small effect sizes of less than 5.5%. These are therefore variables that would benefit from a larger sample to meaningfully assess statistical differences in control and T21 populations with MRI.



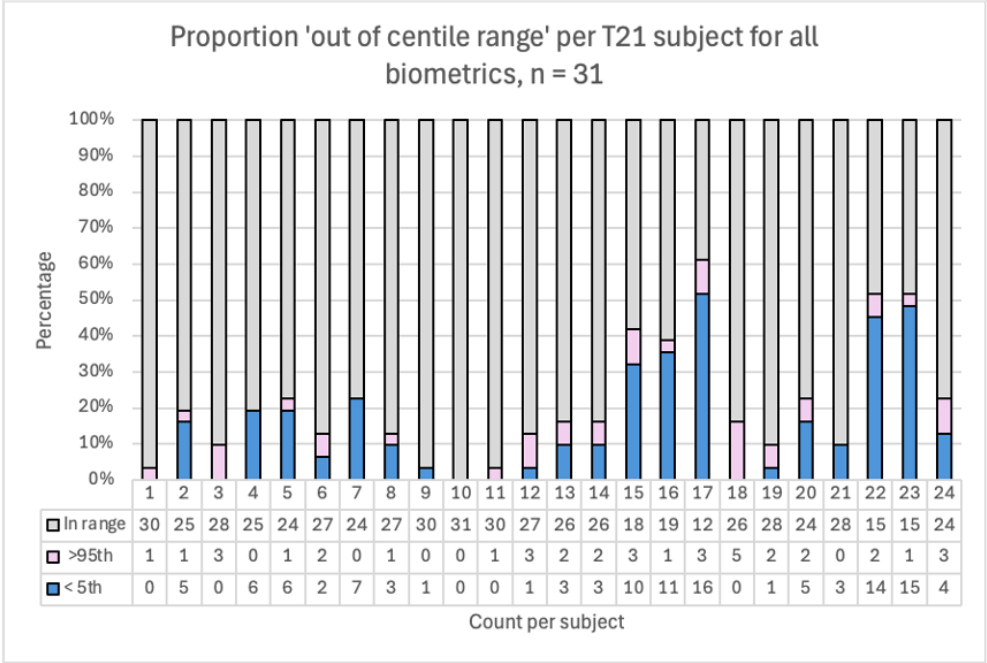


Fig. 11: Chart demostrating the proportion of biometrics that fell outside of normal range (5th - 95th percentile) for each T21 subject, upper chart: for most significant biometrics and lower chart all biometrics.

3.3 2nd and 3rd trimester normal growth charts for craniofacial biometry

Following the analysis of utility of the proposed craniofacial biometry protocol for the assessment of abnormal cases we generated extended nomograms with a mixture of 3 different acquisition protocols. Datasets were included to cover the 2nd and 3rd trimester of pregnancy (>20 weeks GA), and with an even distribution of protocols across the GA range. Automated biometric data from 280 subjects were included, with a GA range of 19.67 to 38.62, mean 27.79 weeks. 84.3% of the sample (n = 236) was at 3T field strength and 15.7% (n = 44) was performed at 1.5T.

Linear growth patterns were seen in 5/31 measurements, IFA, FMA, CBA1, CBA2 and the NB, with the remaining variables having a quadratic growth pattern. These findings contrast with the third trimester only growth charts in section 3.2 where 21/31 measurements had a linear growth pattern. In the extended GA range growth charts most variables varied with GA, however, the IFA and FMA appeared relatively stable across GA which contrasts with the four measures that were stable in the third trimester, i.e. IFA, FMA, CBA1, and ChoH. A selection of plots with the mean bestfit regression equations are shown in Fig. 12 and the full range of plots, along with dataset demographics and sample charateristics, are publicly available for research purposes online at the 3D fetal craniofacial atlas KCL repository [8].

4 Discussion

This work proposed the first comprehensive craniofacial biometry protocol and automated measurement pipeline for 3D T2w fetal MRI. We began with an extensive literature review which supported the selection of 31 biometric measurements, 29 being unique (i.e. non-bilateral), to characterise craniofacial development. The anatomical landmarks, related to the biometrics, were defined within a 3D population-averaged fetal head MRI atlas space and included 35 landmark points and corresponding mathematical 3D vector formulae for the measures. Next, we developed a pipeline for automated biometry, based on the registration of the atlas to the individual subject space and label propagation. The biometry protocol and automated pipeline were evaluated on a T21 cohort and a GA-matched healthy control cohort with a mixture of MRI acquisition protocols. This included a qualitative assessment of 132 cases by visual inspection of landmark placement, which saw a low error rate of 0.03%. The quantitative comparison included an analysis of differences between groups and an assessment of intermethod variability comparing four expert observers (manual method) and automated outputs.

This is the largest study of multiple prenatal craniofacial measurements in fetuses with confirmed T21, and whilst the MANOVA comparison of biometrics revealed statistically significant differences in many measurements (19/31), only 7 had strong effect sizes (ASBL, HPL, VPL, OFD, MXL and NASO ($p<0.001$)). These variables are all consistent with a shorter anteroposterior length of the skull and/or a smaller maxillary area, related to mid-face hypoplasia, and the findings are consistent with

[8] 3D fetal head MRI atlas repository: https://gin.g-node.org/kcl_cdb/craniofacial_fetal_mri_atlas

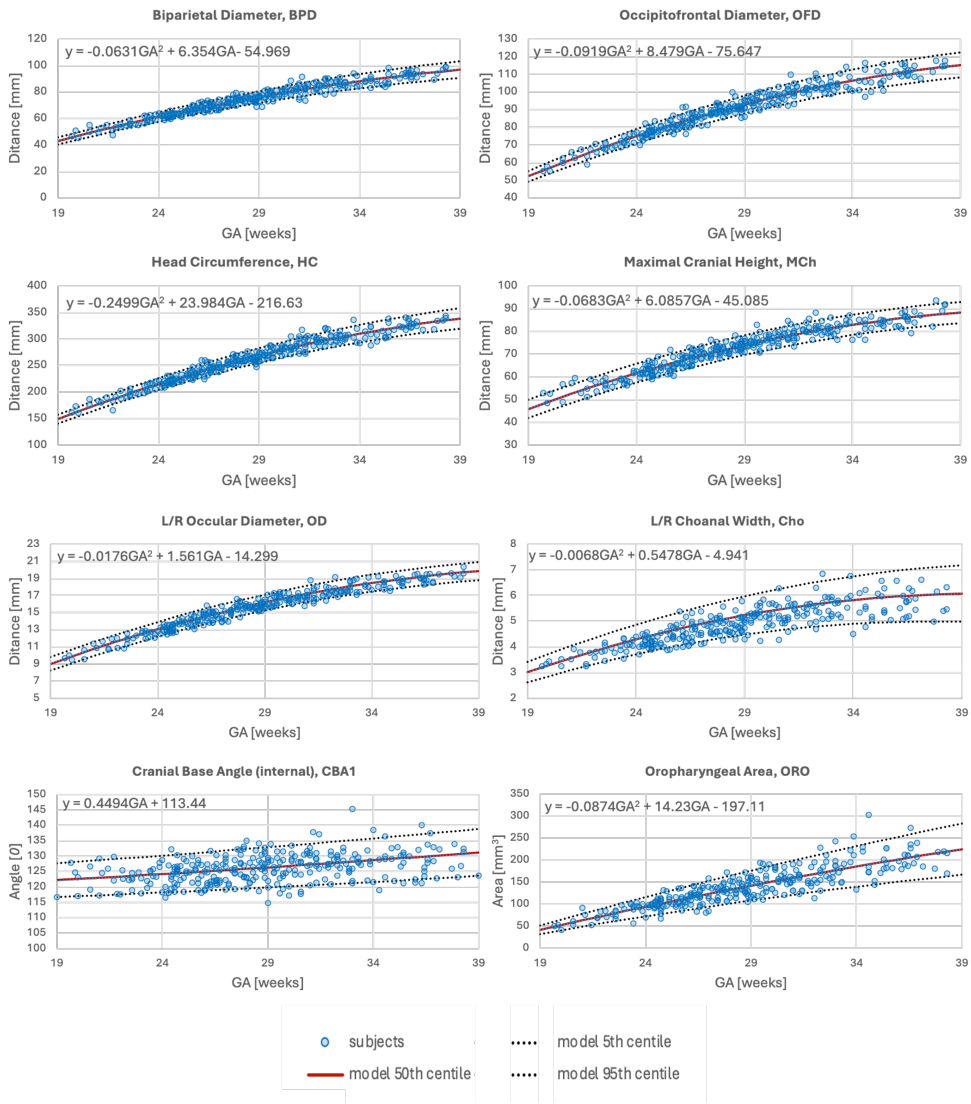


Fig. 12: Selected growth charts for craniofacial biometry from 280 normal controls (blue circles) during 2nd and 3rd trimesters. The quadratic or linear regression equation for the 50th centile bestfit line is included on the chart (y = the mean measurement of the variable under investigation, and GA is the selected gestational age).

ultrasound and postmortem findings in T21 fetal and neonatal cohorts [59, 74, 75]. Interestingly the NB measurement was not found to be significantly associated with T21 in our study, which is in stark contrast to ultrasound nasal bone measurement literature [12, 42]. This result is most likely due to the differing contrasts between the two imaging modalities, with MRI T2 weighted sequences having a far poorer contrast

resolution of bone compared to adjacent soft tissue, affecting visibility and precision of measurements. There has been some promise of 'black bone' imaging and, more recently, zero TE MRI sequences development to examine fetal bone, however, there are limited investigations related to craniofacial applications in fetal life [76–79]. The centiles charts related to the 7 biometrics of interest in the ANOVA model gave a true positive rate (sensitivity) of 71% that increased to 91% when all 31 variables were included. In clinical practice, the assessment of the true negative rate (specificity) is also important and not assessed here. It is highly likely that including more variables with lower effect sizes, would increase sensitivity but also increase false positives and thus reduce the specificity.

Our results indicated that the automated 3D landmark-based measurements were within the variance range of expert manual calliper placement for measurements. Systematic differences were only notable for the mandibular width and this is likely related to the manual measuring of an angular structure which requires precise slice alignment in a 3D volume of non-standard image plane, which adds complexity compared to traditional 2D measurements. It is important to consider that without a reliable ground truth measurement, the automated method produces the same results for the same datasets, if repeated, however, there are always small differences for an individuals' repeated measures and therefore automated biometry will reduce human random and systematic error caused by differences in training, experience, environment, or fatigue. Furthermore, one of the advantages of the proposed automated pipeline is that even with manual verification (and minor refinement of landmarks) it allows significantly faster and more consistent biometry, approximately 5 minutes, in comparison to the classical manual approach that can take 25-35 minutes per case for all measurements.

Lastly, the publicly available 2nd and 3rd trimester normal biometry growth charts, based on 280 normal control subjects from 19 to 39 weeks GA, could be used as a reference for future research studies. This is a first step towards standardisation of automated 3D MRI fetal craniofacial biometry for quantitative analysis that could potentially allow efficient assessment of large cohorts.

Limitations and future work

In this paper, we focus on the variability and reliability of fetal MRI measurements of our new baseline protocol with a first evaluation of the feasibility in a T21 cohort. Despite, yielding novel and comprehensive results in this group, a deeper analysis will be required for its application in clinical practice.

We conducted a detailed evaluation of autobiometry performance in the 3rd trimester, however further optimisation 3D landmark localisation based on deep learning, rather than atlas registration, should improve the reliability of the results. This is important to consider as the proposed method cannot exclude a landmark based on low confidence as a human observer would. Incorporation of surface-derived information could also ensure that the "upper/lower limit" measurements (e.g., skull OFD) will correspond to true anatomical values [30]. Automated quality control, QC, of 3D T2w image volumes and confidence in landmarks placement could be achieved based on deep learning classification which would require a rigorous definition of a QC protocol.

Martins et.al (2014), suggests cut off values for the limits of agreement (i.e. random errors) in measurement variability studies and their suitability for research or clinical use as +/- <5% to 10% is considered good or very good for clinical precision. However, any variability of 20-50% may still be useful for research but should be used with caution in clinical practice [73]. In our case there were large random errors observed in the manual measurements of more than 20% for nine variables and therefore any future work should exhibit caution and investigate error reduction.

Additional sources of error include the MRI scanner field strength which may impact spatial resolution, contrast resolution and the presence of artefacts. In our data, we found statistically significant differences in variability between 1.5T and 3T autobiometry. Image quality is very likely a factor that resulted in this difference with the image contrast of the lower face noted qualitatively to be a factor of variability, especially in the choanal and nasopharyngeal regions. Our analysis controlled for GA, however there are other factors that may have also contributed to differences in the automated biometry including the impact of fetal sex or ethnicity which may all influence natural biological variation. Furthermore, taking into account the wide range of values that naturally occur at a single GA stage, a thorough assessment of the impact of biological parameter (sex, ethnicity), parental characteristics and normalisation of results to individual fetal anatomical size via measurement indexes require further investigation.

Our constructed growth charts spanning the 2nd and 3rd trimester of pregnancy also offers opportunities for detailed characterisation of fetal craniofacial features in future clinical studies. This, together with investigation of the growth trends in longitudinal datasets could potentially help with development of patient-specific approach for evaluation of high risk cases. Evaluating the clinical utility of faciliting diagnoses particularly in syndromic cases is a route of future investigation. A wider range of confirmed craniofacial structural anomalies or genetic and syndromic cases with a known craniofacial phenotypes could help to understand which selection biometrics are reliable for characterisation. Assessing the facial phenotype in the newborn can be made by expert geneticists and supportive imaging results that support selection of additional diagnostic tests could be an outcome. This might also require an extension of the proposed biometry protocol with additional measurements related to the extended landmarking protocol we have provided publicly. Indeed, our MRI 3D volume landmarking protocol offers opportunities to expand on the choice of traditional biometry, to further characterise the complex craniofacial region with 3D landmark based geometric morphometry studies or statistical shape modelling to assist diagnostic prediction, as an emerging method in clinical practice (ref). ****

A clinical dilemma when counselling parents includes assessing the severity of expression of the T21 phenotype, thus with advances in early volumetric imaging assessment of the fetal and neonatal brain and further understanding of the morphological covariation with craniofacial characteristics, future research could begin to explore these questions [80].

5 Conclusions

This study presents the first comprehensive craniofacial biometry protocol and automated measurement pipeline for 3D T2w fetal MRI, identifying 31 key biometric measurements, and defining corresponding landmarks in a 3D fetal head MRI atlas. An automated biometry pipeline was developed and validated with a T21 cohort and GA-matched controls, showing comparable accuracy to manual methods. Significant differences in craniofacial measurements were found in T21 fetuses, particularly highly sensitive biometrics indicating mid-face hypoplasia, a feature consistent with the T21 phenotype. The automated method reduces human error and speeds up measurements significantly, although further optimisation, especially using deep learning, is needed for clinical practice. Future work may include evaluation across earlier GA ranges and addressing potential sources of error like MRI image quality and craniofacial specific sequences at different field strengths. In addition, a focus on reducing measurement variability and exploring additional diagnostic biometrics, with growth charts and longitudinal datasets will offer a potential for personalised evaluation in high-risk cases.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

J.M. and A.U. contributed equally to this work. J.M. performed literature review and formalised the biometry protocol, analysed the datasets, performed all evaluation experiments, interpreted the results and prepared the manuscript. A.U. reconstructed 3D SVR images, implemented automated biometry pipeline, processed all datasets and prepared all technical descriptions. A.E.C., A.L and S.A. participated in evaluation experiments and formalisation of the biometry protocol. A.F.G. participated in processing and analysis of T21 cohort cases. V.K. contributed to formalisation of the biometry protocol. D.C. contributed to analysis of the datasets. R.W. contributed to formalisation of the craniofacial analysis pipeline. M.D. supervised various parts of the project. C.M. provided supervision and manuscript review. J.H. provided fetal MRI datasets. J.O.M. contributed to formalisation of the craniofacial analysis pipeline. C.M. supervised various parts of the project. L.S. provided fetal MRI datasets. J.V.H. provided fetal MRI datasets and supervised various parts of the project. M.A.R. provided fetal MRI dataset, contributed to analysis of the datasets and supervised the project. All authors reviewed the final version of the manuscript.

Funding

This work was supported by the NIHR clinical doctoral research fellowship to JM (NIHR300555), the European Research Council under the European Union's Seventh Framework Programme ([FP7/ 20072013]/ERC grant agreement no. 319456)

for the dHCP project; the Wellcome Trust and EPSRC IEH award [102431] for the iFIND project and the Wellcome/EPSRC Centre for Medical Engineering at King's College London [WT 203148/Z/16/Z]; the NIH (Human Placenta Project [grant 1U01HD087202-01]) for the PIP study; NIHR Advanced Fellowship awarded to Lisa Story [NIHR30166]; the Medical Research Council grant [MR/X010007/1]; and the NIHR Clinical Research Facility at Guy's and St Thomas' and by the National Institute for Health Research Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London.

Acknowledgements

We thank everyone who was involved in the acquisition and analysis of the datasets at the Department of Perinatal Imaging and Health at King's College London. We thank all participating mothers and families.

The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Data availability statement

Fetal MRI dataset related to the dHCP project are available here https://biomedia.github.io/dHCP-release-notes/. Additional fetal MRI datasets used for this study are not publicly available due to ethics regulations. The proposed biometry protocol defined in the 3D T2w atlas space and craniofacial growth charts are publicly available online at KCL CDB fetal MRI atlas repository: https://gin.g-node.org/kcl_cdb/craniofacial_fetal_mri_atlas.

References

[1] Merz, E., Abramovicz, J., Baba, K., Blaas, H.G.K., Deng, J., Gindes, L., Lee, W., Platt, L., Pretorius, D., Schild, R., Sladkevicius, P., Timor-Tritsch, I.: 3D imaging of the fetal face - Recommendationsfrom the International 3D Focus Group. Ultraschall in der Medizin **33**(2), 175–182 (2012) https://doi.org/10.1055/s-0031-1299378

[2] Jong-Pleij, E.A.P.D., Ribbert, L.S.M., Manten, G.T.R., Tromp, E., Bilardo, C.M.: Maxilla-nasion-mandible angle: A new method to assess profile anomalies in pregnancy. Ultrasound in Obstetrics and Gynecology **37**, 562–569 (2011) https://doi.org/10.1002/uog.7768

[3] Rotten, D., Levaillant, J.M., Martinez, H., pointe, H.D.L., Vicaut, : The fetal mandible: a 2d and 3d sonographic approach to the diagnosis of retrognathia and micrognathia. Ultrasound in Obstetrics Gynecology (2002) https://doi.org/10.1046/j.0960-7692.2001.00622.x

[4] Roelfsema, N.M., Hop, W.C.J., Wladimiroff, J.W.: Three-dimensional sonographic determination of normal fetal mandibular and maxillary size during the

second half of pregnancy. Ultrasound in Obstetrics and Gynecology **28**, 950–957 (2006) https://doi.org/10.1002/uog.3866

[5] Buijtendijk, M.F.J., Bet, B.B., Leeflang, M.M.G., Shah, H., Reuvekamp, T., Goring, T., Docter, D., Timmerman, M.G.M.M., Dawood, Y., Lugthart, M.A., Berends, B., Limpens, J., Pajkrt, E., Hoff, M.J.B., Bakker, B.S.: Diagnostic accuracy of ultrasound screening for fetal structural abnormalities during the first and second trimester of pregnancy in low-risk and unselected populations. Cochrane Database of Systematic Reviews **2024** (2024) https://doi.org/10.1002/14651858.CD014715.pub2

[6] Lamanna, B., Dellino, M., Cascardi, E., Rooke-Ley, M., Vinciguerra, M., Cazzato, G., Malvasi, A., Vitagliano, A., Nicolì, P., Cosola, M.D., Ballini, A., Cicinelli, E., Vimercati, A.: Efficacy of systematic early-second-trimester ultrasound screening for facial anomalies: A comparison between prenatal ultrasound and postmortem findings. Journal of Clinical Medicine **12** (2023) https://doi.org/10.3390/jcm12165365

[7] Zemet, R., Amdur-Zilberfarb, I., Shapira, M., Ziv-Baran, T., Hoffmann, C., Kassif, E., Katorza, E.: Prenatal diagnosis of congenital head, face, and neck malformations—Is complementary fetal MRI of value? Prenatal Diagnosis **40**(1), 142–150 (2020) https://doi.org/10.1002/pd.5593

[8] Salomon, L.J., Alfirevic, Z., Berghella, V., Bilardo, C.M., Chalouhi, G.E., Da Silva Costa, F., Hernandez-Andrade, E., Malinger, G., Munoz, H., Paladini, D., Prefumo, F., Sotiriadis, A., Toi, A., Lee, W.: ISUOG Practice Guidelines (updated): performance of the routine mid-trimester fetal ultrasound scan. Ultrasound in Obstetrics Gynecology **59**(6), 840–856 (2022) https://doi.org/10.1002/UOG.24888

[9] Clark, A.E., Biffi, B., Sivera, R., Dall'asta, A., Fessey, L., Wong, T.L., Paramasivam, G., Dunaway, D., Schievano, S., Lees, C.C.: Developing and testing an algorithm for automatic segmentation of the fetal face from three-dimensional ultrasound images: Automatic segmentation of the fetal face. Royal Society Open Science **7** (2020) https://doi.org/10.1098/rsos.201342

[10] Matthew, J., Uus, A., De Souza, L., Wright, R., Fukami-Gartner, A., Priego, G., Saija, C., Deprez, M., Collado, A.E., Hutter, J., Story, L., Malamateniou, C., Rhode, K., Hajnal, J., Rutherford, M.A.: Craniofacial phenotyping with fetal MRI: a feasibility study of 3D visualisation, segmentation, surface-rendered and physical models. BMC Medical Imaging **24**(1), 52 (2024) https://doi.org/10.1186/s12880-024-01230-7

[11] Salomon, L.J., Alfirevic, Z., Da Silva Costa, F., Deter, R.L., Figueras, F., Ghi, T., Glanc, P., Khalil, A., Lee, W., Napolitano, R., Papageorghiou, A., Sotiriadis, A., Stirnemann, J., Toi, A., Yeo, G.: ISUOG Practice Guidelines: ultrasound assessment of fetal biometry and growth. Ultrasound in Obstetrics & Gynecology

53(6), 715–723 (2019) https://doi.org/10.1002/uog.20272

[12] Moreno-Cid, M., Rubio-Lorente, A., Rodríguez, M.J., Bueno-Pacheco, G., Tenías, J.M., Román-Ortiz, C., Arias, Á.: Systematic review and meta-analysis of performance of second-trimester nasal bone assessment in detection of fetuses with Down syndrome. Ultrasound in Obstetrics & Gynecology **43**(3), 247–253 (2014) https://doi.org/10.1002/uog.13228

[13] Cang, Z., Cui, J., Pei, J., Wang, Z., Du, Y., Mu, S., Dou, W., Fan, X., Zhang, X., Li, Y.: Prenatal diagnosis of micrognathia: a systematic review. Frontiers in Pediatrics **11** (2023) https://doi.org/10.3389/fped.2023.1161421

[14] Rotten, D., Levaillant, J.M.: Two- and three-dimensional sonographic assessment of the fetal face. 2. Analysis of cleft lip, alveolus and palate. Ultrasound in Obstetrics & Gynecology **24**(4), 402–411 (2004) https://doi.org/10.1002/uog.1718

[15] Goldstein, I., Tamir, A., Zimmer, E.Z., Itskovitz-Eldor, J.: Growth of the fetal orbit and lens in normal pregnancies. Ultrasound in Obstetrics and Gynecology **12**, 175–179 (1998) https://doi.org/10.1046/j.1469-0705.1998.12030175.x

[16] Toren, A., Spevac, S., Hoffman, C., Katorza, E.: What does the normal fetal face look like? MR imaging of the developing mandible and nasal cavity. European Journal of Radiology **126**, 108937 (2020) https://doi.org/10.1016/j.ejrad.2020.108937

[17] Robinson, A.J., Blaser, S., Toi, A., Chitayat, D., Pantazi, S., Keating, S., Viero, S., Ryan, G.: MRI of the fetal eyes: morphologic and biometric assessment for abnormal development with ultrasonographic and clinicopathologic correlation. Pediatric Radiology **38**(9), 971–981 (2008) https://doi.org/10.1007/s00247-008-0929-3

[18] Velasco-Annis, C., Gholipour, A., Afacan, O., Prabhu, S.P., Estroff, J.A., Warfield, S.K.: Normative biometrics for fetal ocular growth using volumetric MRI reconstruction. Prenatal Diagnosis **35**(4), 400–408 (2015) https://doi.org/10.1002/pd.4558

[19] Li, X.B., Kasprian, G., Hodge, J.C., Jiang, X.L., Bettelheim, D., Brugger, P.C., Prayer, D.: Fetal ocular measurements by MRI. Prenatal Diagnosis **30**(11), 1064–1071 (2010) https://doi.org/10.1002/pd.2612

[20] Uus, A., Matthew, J., Grigorescu, I., Jupp, S., Grande, L.C., Price, A., Hughes, E., Patkee, P., Kyriakopoulou, V., Wright, R., Roberts, T., Hutter, J., Pietsch, M., Hajnal, J.V., Edwards, A.D., Rutherford, M.A., Deprez, M.: Spatio-temporal atlas of normal fetal craniofacial feature development and cnn-based ocular biometry for motion-corrected fetal mri. In: PIPPI MICCAI Workshop, pp. 168–178 (2021). https://doi.org/10.1007/978-3-030-87735-4_16

[21] Nemec, U., Nemec, S.F., Brugger, P.C., Weber, M., Bartsch, B., Bettelheim, D., Gruber, M., Prayer, D.: Normal mandibular growth and diagnosis of micrognathia at prenatal MRI. Prenatal Diagnosis **35**(2), 108–116 (2015) https://doi.org/10.1002/pd.4496

[22] Ozcan, U.A., Yildiz, M.E., Ulus, S., Turk, A., Erzen, C.: Magnetic resonance imaging evaluation of fetal maxillary sinuses. The Journal of craniofacial surgery **25**, 363–366 (2014) https://doi.org/10.1097/SCS.0000000000000625

[23] Mascio, D.D., Khalil, A., Rizzo, G., Kasprian, G., Caulo, M., Manganaro, L., Odibo, A.O., Flacco, M.E., Giancotti, A., Buca, D., Liberati, M., Timor-Tritsch, I.E., D'Antonio, F.: Reference ranges for fetal brain structures using magnetic resonance imaging: systematic review (2022). https://doi.org/10.1002/uog.23762

[24] Shrot, S., Hadi, E., Barash, Y., Hoffmann, C.: Effect of magnet strength on fetal brain biometry — a single-center retrospective mri-based cohort study. Neuroradiology **65**, 1517–1525 (2023) https://doi.org/10.1007/s00234-023-03193-y

[25] Priego, G., Barrowman, N., Barrowman, N., Hurteau-Miller, J., Miller, E.: Does 3T Fetal MRI Improve Image Resolution of Normal Brain Structures between 20 and 24 Weeks' Gestational Age? American Journal of Neuroradiology **38**(8), 1636–1642 (2017) https://doi.org/10.3174/ajnr.a5251

[26] Uus, A.U., Collado, A.E., Roberts, T.A., Hajnal, J.V., Rutherford, M.A., Deprez, M.: Retrospective motion correction in foetal mri for clinical applications: existing methods, applications and integration into clinical practice. The British Journal of Radiology (2022) https://doi.org/10.1259/bjr.20220071

[27] Kyriakopoulou, V., Vatansever, D., Davidson, A., Patkee, P., Elkommos, S., Chew, A., Martinez-Biarge, M., Hagberg, B., Damodaram, M., Allsop, J., Fox, M., Hajnal, J.V., Rutherford, M.A.: Normative biometry of the fetal brain using magnetic resonance imaging. Brain Structure and Function **222**, 2295–2307 (2017) https://doi.org/10.1007/s00429-016-1342-6

[28] Khawam, M., Dumast, P., Deman, P., Kebiri, H., Yu, T., Tourbier, S., Lajous, H., Hagmann, P., Maeder, P., Thiran, J.P., Meuli, R., Dunet, V., Cuadra, M.B., Koob, M.: Fetal brain biometric measurements on 3d super-resolution reconstructed t2-weighted mri: An intra- and inter-observer agreement study. Frontiers in Pediatrics **9** (2021) https://doi.org/10.3389/fped.2021.639746

[29] Avisdris, N., Yehuda, B., Ben-Zvi, O., Link-Sourani, D., Ben-Sira, L., Miller, E., Zharkov, E., Bashat, D.B., Joskowicz, L.: Automatic linear measurements of the fetal brain on mri with deep neural networks. International Journal of Computer Assisted Radiology and Surgery **16**, 1481–1492 (2021) https://doi.org/10.1007/s11548-021-02436-8

[30] She, J., Huang, H., Ye, Z., Huang, W., Sun, Y., Liu, C., Yang, W., Wang, J.,

Ye, P., Zhang, L., Ning, G.: Automatic biometry of fetal brain mris using deep and machine learning techniques. Scientific Reports **13** (2023) https://doi.org/10.1038/s41598-023-43867-4

[31] Velasco-Annis, C., Gholipour, A., Afacan, O., Prabhu, S.P., Estroff, J.A., Warfield, S.K.: Normative biometrics for fetal ocular growth using volumetric mri reconstruction. Prenatal Diagnosis **35**, 400–408 (2015) https://doi.org/10.1002/pd.4558

[32] Cordero-grande, L., Price, A.N., Hughes, E.J., Wright, R., Rutherford, M.A., Hajnal, J.V.: Automating fetal brain reconstruction using distance regression learning, pp. 3–6 (2019). https://archive.ismrm.org/2019/4779.html

[33] Uus, A.U., Silva, S.N., Verdera, J.A., Payette, K., Hall, M., Colford, K., Luis, A., Sousa, H.S., Ning, Z., Roberts, T., McElroy, S., Deprez, M., Hajnal, J.V., Rutherford, M.A., Story, L., Hutter, J.: Scanner-based real-time 3d brain+body slice-to-volume reconstruction for t2-weighted 0.55t low field fetal mri. medRxiv, 2024–042224306177 (2024) https://doi.org/10.1101/2024.04.22.24306177

[34] Kuklisova-Murgasova, M., Quaghebeur, G., Rutherford, M.A., Hajnal, J.V., Schnabel, J.A.: Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. Medical Image Analysis **16**(8), 1550 (2012) https://doi.org/10.1016/J.MEDIA.2012.07.004

[35] Uus, A., Zhang, T., Jackson, L.H., Roberts, T.A., Rutherford, M.A., Hajnal, J.V., Deprez, M.: Deformable Slice-to-Volume Registration for Motion Correction of Fetal Body and Placenta MRI. IEEE Transactions on Medical Imaging, 1–1 (2020) https://doi.org/10.1109/tmi.2020.2974844

[36] Roelfsema, N.M., Grijseels, E.W.M., Hop, W.C.J., Wladimiroff, J.W.: Three-dimensional sonography of prenatal skull base development. Ultrasound in Obstetrics Gynecology (2007) https://doi.org/10.1002/uog.3946

[37] Nemzek, W.R., Brodie, H.A., Hecht, S.T., Chong, B.W., Babcook, C.J., Seibert, J.A.: Mr, ct, and plain film imaging of the developing skull base in fetal specimens. AJNR: American Journal of Neuroradiology **21**, 1699 (2000)

[38] Begnoni, G., Serrao, G., Musto, F., Pellegrini, G., Triulzi, F.M., Dellavia, C.: Craniofacial structures' development in prenatal period: An MRI study. Orthodontics Craniofacial Research **21**(2), 96–103 (2018) https://doi.org/10.1111/ocr.12222

[39] Jeffery, N.: Cranial base angulation and growth of the human fetal pharynx. Anatomical Record - Part A Discoveries in Molecular, Cellular, and Evolutionary Biology **284**(1), 491–499 (2005) https://doi.org/10.1002/ar.a.20183

[40] Hirunpat, S., Wimolsiri, N., Sanghan, N.: Normal value of skull base angle using

the modified magnetic resonance imaging technique in thai population. Journal of Oral Health and Craniofacial Science **2**, 017–021 (2017) https://doi.org/10.29328/JOURNAL.JOHCS.1001006

[41] Vos, F.I., De Jong-Pleij, E.A., Ribbert, L.S., Tromp, E.: Three-dimensional ultrasound imaging and measurement of nasal bone length, prenasal thickness and frontomaxillary facial angle in normal second- and third-trimester fetuses. Ultrasound in obstetrics gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology **39**(6), 636–641 (2012) https://doi.org/10.1002/uog.10058

[42] De Jong-Pleij, E.A.P., Vos, F.I., Ribbert, L.S.M., Pistorius, L.R., Tromp, E., Bilardo, C.M.: Prenasal thickness-to-nasal bone length ratio: a strong and simple second- and third-trimester marker for trisomy 21. Ultrasound in Obstetrics Gynecology **39**(2), 185–190 (2012) https://doi.org/10.1002/uog.9047

[43] Perry, J.L., Kollara, L., Kuehn, D.P., Sutton, B.P., Fang, X.: Examining age, sex, and race characteristics of velopharyngeal structures in 4- to 9-year old children using magnetic resonance imaging. The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association **55**(1), 21–34 (2018) https://doi.org/10.1177/1055665617718549

[44] Jeffery, N., Spoor, F.: Ossification and Midline Shape Changes of the Human Fetal Cranial Base. American Journal of Physical Anthropology **123**(1), 78–90 (2004) https://doi.org/10.1002/ajpa.10292

[45] Tan, H.B., Rathnayake, C., Selvam, S., Berkowitz, R.G.: Magnetic resonance imaging evaluation of choanal atresia in newborns. Australian Journal of Otolaryngology **5**(0) (2022)

[46] Loughna, P., Chitty, L., Evans, T., Chudleigh, T.: Fetal size and dating: charts recommended for clinical obstetric practice. Ultrasound **17**, 161–167 (2009) https://doi.org/10.1179/174313409X448543

[47] Leibovitz, Z., Shiran, C., Haratz, K., Tamarkin, M., Gindes, L., Schreiber, L., Malinger, G., Ben-Sira, L., Lev, D., Shapiro, I., Bakry, H., Weizman, B., Zreik, A., Kidron, D., Egenburg, S., Arad, A., Lerman-Sagie, T.: Application of a novel prenatal vertical cranial biometric measurement can improve accuracy of microcephaly diagnosis in utero. Ultrasound in Obstetrics Gynecology **47**, 593–599 (2016) https://doi.org/10.1002/UOG.15886

[48] Mailath-Pokorny, M., Klein, K., Worda, C., Weber, M., Brugger, P.C., Czerny, C., Nemec, U., Prayer, D.: Maxillary dental arch biometry: assessment with fetal mr imaging. Prenatal Diagnosis **32**, 530–535 (2012) https://doi.org/10.1002/pd.3851

[49] Garel, C., Vande Perre, S., Guilbaud, L., Soupre, V., Blondiaux, E., Pointe, H.: Contribution of computed tomography and magnetic resonance imaging in the

analysis of fetal craniofacial malformations. Pediatric radiology **51**(10), 1917–1928 (2021) https://doi.org/10.1007/S00247-021-05033-8

[50] Kunzendorf, B., Diogo, M.C., Covini, D.I., Weber, M., Gruber, G.M., Zeilhofer, H.-F., Berg, B.-I., Prayer, D.: Comparison of the visibility of fetal tooth buds on 1.5 and 3 tesla mri. Journal of Clinical Medicine **9**, 3424 (2020) https://doi.org/10.3390/jcm9113424

[51] Geerts, L., Meyer, R., Nolan, H., Nel, L., Nel, D.G., Brink, L., Odendaal, H.: Reference standards for facial measurements in early third trimester South African fetuses, and the effect of maternal and fetal characteristics. Journal of Maternal-Fetal and Neonatal Medicine **35**(25), 8434–8442 (2022)

[52] Degani, S., Leibovitz, Z., Shapiro, I., Gonen, R., Ohel, G.: Ultrasound evaluation of the fetal skull base throughout pregnancy. Ultrasound in Obstetrics and Gynecology **19**, 461–466 (2002) https://doi.org/10.1046/j.1469-0705.2002.00639.x

[53] Grzonkowska, M., Baumgart, M., Badura, M., Wiśniewski, M., Lisiecki, J., Szpinda, M.: Quantitative anatomy of primary ossification centres of the lateral and basilar parts of the occipital bone in the human foetus. Folia Morphologica (Poland) **80**(4), 895–903 (2021) https://doi.org/10.5603/FM.A2021.0115

[54] Radlanski, R.J., Heikinheimo, K.: Cephalometric assessment of human fetal head specimens. Journal of orofacial orthopedics = Fortschritte der Kieferorthopadie : Organ/official journal Deutsche Gesellschaft fur Kieferorthopadie **74**, 332–348 (2013)

[55] Lesieur, E., Delteil, C., Gorincour, G., Panuel, M., Chaumoitre, K., Saliba, B., Bretelle, F., Piercecchi-Marti, M.D., Adalian, P.: Fetal facial bone growth: Post-mortem ct analysis. Morphologie **107**, 116–126 (2023) https://doi.org/10.1016/J.MORPHO.2022.05.006

[56] Raut, A.A., Naphade, P.S., Chawla, A.: Imaging of skull base: Pictorial essay. The Indian Journal of Radiology Imaging **22**, 305 (2012) https://doi.org/10.4103/0971-3026.111485

[57] Siddiqui, A., Connor, S.E.J.: Imaging of developmental skull base abnormalities. Neuroimaging Clinics of North America **31**, 621–647 (2021) https://doi.org/10.1016/J.NIC.2021.06.004

[58] Delteil, C., Mopin, C., Tuchtan, L., Carballeira Alvarez, A., Piercecchi, M.-D., Adalian, P.: Ontogeny of cranial base during the first two years of life. Morphologie **106**(353), 98–106 (2022) https://doi.org/10.1016/j.morpho.2021.03.006

[59] Vicente, A., Bravo-González, L.A., López-Romero, A., Muñoz, C.S., Sánchez-Meca, J.: Craniofacial morphology in down syndrome: a systematic review and

meta-analysis. Scientific Reports 2020 10:1 **10**(1), 1–14 (2020) https://doi.org/10.1038/s41598-020-76984-5

[60] Uus, A., Matthew, J., Grigorescu, I., Jupp, S., Grande, L.C., Price, A., Hughes, E., Patkee, P., Kyriakopoulou, V., Wright, R., Roberts, T., Hutter, J., Pietsch, M., Hajnal, J.V., Edwards, A.D., Rutherford, M.A., Deprez, M.: Spatio-Temporal Atlas of Normal Fetal Craniofacial Feature Development and CNN-Based Ocular Biometry for Motion-Corrected Fetal MRI. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **12959 LNCS**, 168–178 (2021)

[61] Jeffery, N.: A high-resolution MRI study of linear growth of the human fetal skull base. Neuroradiology **44**(4), 358–366 (2002) https://doi.org/10.1007/s00234-001-0753-z

[62] Sonek, J.D., McKenna, D., Webb, D., Croom, C., Nicolaides, K.: Nasal bone length throughout gestation: Normal ranges based on 3537 fetal ultrasound measurements. Ultrasound in Obstetrics and Gynecology **21**, 152–155 (2003) https://doi.org/10.1002/uog.41

[63] Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: Application to breast mr images. IEEE Transactions on Medical Imaging **18**(8), 712–721 (1999) https://doi.org/10.1109/42.796284

[64] Bartlett, J.W., Frost, C.: Reliability, repeatability and reproducibility: Analysis of measurement errors in continuous variables (2008). https://doi.org/10.1002/uog.5256 . http://doi.wiley.com/10.1002/uog.5256

[65] Koo, T.K., Li, M.Y.: A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine **15**(2), 155–163 (2016) https://doi.org/10.1016/j.jcm.2016.02.012

[66] Tavakol, M., Dennick, R.: Making sense of Cronbach's alpha. Int J Med Educ **2**, 53–55 (2011) https://doi.org/10.5116/ijme.4dfb.8dfd

[67] Wilcox, R.R., Rousselet, G.A.: An Updated Guide to Robust Statistical Methods in Neuroscience. Current Protocols **3**(3), 719 (2023) https://doi.org/10.1002/cpz1.719

[68] Olejnik, S., Algina, J.: Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs. American Psychological Association, Olejnik, Stephen: U Georgia, Coll of Education, Dept of Educational Psychology, 325 Aderhold Hall, Athens, GA, US, 30602-7143, olejnik@coe.uga.edu (2003). https://doi.org/10.1037/1082-989X.8.4.434

[69] Bittner, A.: Analysis-of-variance (ANOVA) Assumptions Review: Normality, Variance Equality, and Independence, pp. 28–33 (2022). https://doi.org/10.47461/isoes.2022_bittner

[70] Hayes, A.F., Cai, L.: Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. Behavior Research Methods **39**(4), 709–722 (2007) https://doi.org/10.3758/BF03192961

[71] Royston, P., Wright, E.M.: How to construct 'normal ranges' for fetal variables. Ultrasound in Obstetrics & Gynecology **11**(1), 30–38 (1998) https://doi.org/10.1046/j.1469-0705.1998.11010030.x

[72] Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J.V., Pieper, S., Kikinis, R.: 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magnetic Resonance Imaging **30**(9), 1323–1341 (2012) https://doi.org/10.1016/J.MRI.2012.05.001

[73] Martins, W.P., Nastri, C.O.: Interpreting reproducibility results for ultrasound measurements (2014). https://doi.org/10.1002/uog.13320

[74] Ettema, A.M., Wenghoefer, M., Hansmann, M., Carels, C.E.L., Borstlap, W.A., Bergé, S.J.: Prenatal Diagnosis of Craniomaxillofacial Malformations: A Characterization of Phenotypes in Trisomies 13, 18, and 21 by Ultrasound and Pathology. The Cleft Palate-Craniofacial Journal **47**(2), 189–196 (2010) https://doi.org/10.1597/08-285_1

[75] Guihard-Costa, A.M., Khung, S., Delbecque, K., Ménez, F., Delezoide, A.L.: Biometry of Face and Brain in Fetuses with Trisomy 21. Pediatric Research 2006 59:1 **59**(1), 33–38 (2006) https://doi.org/10.1203/01.pdr.0000190580.88391.9a

[76] Matsubara, Y., Higaki, T., Tani, C., Kamioka, S., Harada, K., Aoyama, H., Nakamura, Y., Akita, T., Awai, K.: Demonstration of human fetal bone morphology with mr imaging: A preliminary study. Magnetic Resonance in Medical Sciences **19**(4), 310–317 (2020) https://doi.org/10.2463/mrms.mp.2019-0105

[77] Lu, A., Gorny, K.R., Ho, M.-L.: Zero TE MRI for Craniofacial Bone Imaging. American Journal of Neuroradiology **40**(9), 1562–1566 (2019) https://doi.org/10.3174/ajnr.A6175

[78] Kamona, N., Jones, B.C., Lee, H., Song, H.K., Rajapakse, C.S., Wagner, C.S., Bartlett, S.P., Wehrli, F.W.: Cranial bone imaging using ultrashort echo-time bone-selective MRI as an alternative to gradient-echo based "black-bone" techniques. Magnetic Resonance Materials in Physics, Biology and Medicine **37**(1), 83–92 (2024) https://doi.org/10.1007/s10334-023-01125-8

[79] Robinson, A.J., Blaser, S., Blaser, S., Vladimirov, A., Drossman, D., Chitayat, D., Ryan, G., Ryan, G.: Foetal "black bone" MRI: utility in assessment of the foetal spine. British Journal of Radiology (2015) https://doi.org/10.1259/bjr.20140496

[80] Fukami-Gartner, A., Baburamani, A.A., Dimitrova, R., Patkee, P.A., Ojinaga-Alfageme, O., Bonthrone, A.F., Cromb, D., Uus, A.U., Counsell, S.J., Hajnal, J.V., O'Muircheartaigh, J., Rutherford, M.A.: Comprehensive volumetric phenotyping of the neonatal brain in Down syndrome. Cerebral Cortex **33**(14), 8921–8941 (2023) https://doi.org/10.1093/cercor/bhad171

Appendix A Figures

Absolute difference	Observer 0 *					Observer 2					Observer 3					Automated				
ID Variable	N_Obs	Mean	Lower_CI	Upper_CI	abs LoA range (+/- mean)	N_Obs	Mean	Lower_CI	Upper_CI	abs LoA range (+/- mean)	N_Obs	Mean	Lower_CI	Upper_CI	abs LoA range (+/- mean)	N_Obs	Mean	Lower_CI	Upper_CI	abs LoA range (+/- mean)
1 ASBL	10	1.94	-0.80	4.69	2.75	10	0.15	-1.99	2.30	2.14	10	0.12	-1.81	2.06	1.93	10	0.79	-1.32	2.91	2.12
2 PSBL	10	0.11	-2.51	2.74	2.63	10	-1.05	-4.76	2.67	3.72	10	-0.04	-3.18	3.10	3.14	10	-0.27	-2.81	2.27	2.54
3 HPL	10	-0.88	-5.91	4.15	5.03	10	-1.38	-10.67	7.91	9.29	10	-3.37	-11.82	5.08	8.45	10	-0.55	-4.32	3.23	3.77
4 VPL	10	0.02	-5.08	5.12	5.10	10	0.81	-8.79	10.42	9.61	10	-1.73	-8.33	4.88	6.61	10	-2.08	-5.03	0.88	2.95
5 PAW	10	-0.53	-4.13	3.07	3.60	10	-2.03	-5.74	1.68	3.71	10	-2.64	-5.68	0.40	3.04	10	-0.13	-4.06	3.79	3.92
6 Cho_r	6	0.34	-6.48	7.17	6.83	7	-0.11	-1.63	1.42	1.53	6	1.15	-1.89	4.18	3.03	7	0.16	-0.80	1.11	0.95
7 Cho_l	6	-0.01	-7.61	7.58	7.59	7	-0.17	-1.67	1.33	1.50	6	0.91	-1.87	3.70	2.79	7	-0.20	-1.53	1.13	1.33
8 NPW	6	0.23	-3.50	3.97	3.73	7	0.12	-2.74	2.99	2.87	6	2.09	-4.19	8.36	6.28	7	1.66	-0.64	3.96	2.30
9 NB	7	0.67	-4.20	5.55	4.88	8	2.58	-0.71	5.86	3.29	9	0.10	-3.88	4.07	3.98	9	1.72	-0.26	3.69	1.97
10 PNTh	10	0.38	-3.70	4.46	4.08	10	-0.56	-5.59	4.48	5.03	10	-0.31	-4.33	3.72	4.03	10	0.12	-3.08	3.32	3.20
11 OFD	10	0.30	-4.50	5.10	4.80	10	-1.18	-6.68	4.31	5.50	10	0.25	-7.05	7.55	7.30	10	0.44	-4.70	5.57	5.13
12 BPD	10	1.29	-6.27	8.85	7.56	10	-0.81	-7.76	6.15	6.95	10	-0.61	-7.75	6.52	7.14	10	1.96	-4.44	8.36	6.40
13 MCh	10	1.81	-3.90	7.51	5.71	10	1.14	-3.68	5.95	4.81	10	0.74	-5.26	6.74	6.00	10	-0.33	-5.92	5.26	5.59
14 OD_R	10	-0.70	-2.95	1.54	2.24	10	-0.90	-2.58	0.77	1.67	10	-1.00	-3.26	1.27	2.26	10	0.21	-1.60	2.03	1.81
15 OD_L	10	-0.48	-1.83	0.87	1.35	10	-0.71	-2.95	1.54	2.25	10	-1.11	-2.61	0.39	1.50	10	0.11	-1.01	1.24	1.12
16 IOD	10	0.45	-2.00	2.89	2.44	10	-1.23	-4.43	1.98	3.20	10	0.63	-1.87	3.13	2.50	10	0.14	-2.13	2.40	2.27
17 BOD	10	-0.34	-3.44	2.75	3.09	10	-0.71	-3.74	2.33	3.04	10	-1.06	-4.27	2.15	3.21	10	0.51	-2.05	3.08	2.56
18 MxW	10	1.28	-0.91	3.47	2.19	10	0.72	-3.25	4.70	3.98	10	-0.99	-5.95	3.97	4.96	10	2.05	-0.20	4.29	2.24
19 MdW	10	1.67	-1.87	5.21	3.54	10	0.34	-1.11	1.79	1.45	10	-0.89	-3.41	1.64	2.52	10	5.30	1.58	9.01	3.71
20 CBA1	10	0.60	-4.49	5.70	5.09	10	2.08	-5.43	9.58	7.51	10	-0.03	-7.39	7.33	7.36	10	0.39	-6.89	7.66	7.28
21 CBA2	10	-7.44	-48.97	34.09	41.53	10	-1.41	-42.85	40.04	41.44	10	8.32	-18.69	35.33	27.01	10	-9.93	-30.10	10.24	20.17
22 FMA	10	0.78	-12.79	14.35	13.57	10	2.64	-9.01	14.30	11.66	10	-5.65	-22.02	10.72	16.37	10	0.40	-12.46	13.26	12.86
23 MNMA	10	-10.78	-20.54	-1.01	9.76	10	-3.61	-12.23	5.00	8.61	10	-10.35	-23.02	2.33	12.67	10	-7.73	-16.16	0.71	8.43
24 IFA	10	1.65	-11.41	14.70	13.06	10	4.27	-8.68	17.21	12.95	10	-2.53	-23.60	18.54	21.07	10	2.66	-9.91	15.26	12.59
25 PAH	10	-1.37	-4.07	1.32	2.70	10	-1.49	-4.09	1.12	2.60	10	-0.87	-5.08	3.33	4.21	10	-0.15	-2.40	2.10	2.25
26 ChoH	6	1.32	-0.59	3.23	1.91	6	1.00	-1.15	3.14	2.15	6	-0.53	-2.92	1.86	2.39	6	0.74	-1.39	2.87	2.13
27 MXL	10	-1.00	-6.33	4.34	5.34	10	0.30	-3.28	3.89	3.58	10	2.47	-8.71	13.64	11.18	10	-0.76	-3.98	2.45	3.22
28 MDL	10	2.85	-2.43	8.13	5.28	10	0.33	-3.26	3.92	3.59	10	1.04	-6.31	8.40	7.35	10	4.21	0.21	8.22	4.01
29 NASO	10	16.79	-45.71	79.30	62.50	10	17.93	-22.98	58.85	40.91	10	13.29	-43.36	69.93	56.64	10	13.67	-26.00	53.35	39.67
30 ORO	10	-8.35	-68.43	51.73	60.08	10	-69.47	-175.44	36.50	105.97	10	-34.70	-119.64	50.25	84.95	10	-47.18	-108.70	14.33	61.51
31 HC	10	1.65	-15.09	18.39	16.74	10	-6.15	-22.44	10.14	16.29	10	2.67	-17.82	23.17	20.50	10	2.94	-15.94	21.82	18.88

Relative Difference	Observer 0 *					Observer 2					Observer 3					Automated				
ID Variable	N_Obs	Mean	Lower_CI	Upper_CI	rel LoA range (+/- mean)	N_Obs	Mean	Lower_CI	Upper_CI	rel LoA range (+/- mean)	N_Obs	Mean	Lower_CI	Upper_CI	rel LoA range (+/- mean)	N_Obs	Mean	Lower_CI	Upper_CI	rel LoA range (+/- mean)
1 ASBL	10	5.20	-1.94	12.34	7.14	10	0.32	-5.43	6.07	5.75	10	0.24	-4.86	5.34	5.10	10	2.13	-3.46	7.72	5.59
2 PSBL	10	0.36	-9.87	10.58	10.22	10	-4.16	-18.69	10.38	14.54	10	-0.19	-13.61	13.23	13.42	10	-1.36	-11.63	8.91	10.27
3 HPL	10	-4.30	-24.46	15.87	20.16	10	-6.58	-42.75	29.60	36.18	10	-13.91	-48.32	20.50	34.41	10	-2.70	-17.71	12.31	15.01
4 VPL	10	-0.21	-13.16	12.75	12.95	10	1.76	-22.87	26.39	24.63	10	-4.71	-21.98	12.56	17.27	10	-5.58	-13.48	2.32	7.90
5 PAW	10	-2.75	-19.79	14.28	17.04	10	-8.69	-24.70	7.32	16.01	10	-11.18	-25.98	3.62	14.80	10	-1.26	-18.81	16.29	17.55
6 Cho_r	6	9.50	-4.04	23.03	13.53	7	-2.76	-30.54	25.02	27.78	6	19.99	-32.33	72.31	52.32	7	2.18	-15.43	19.78	17.61
7 Cho_l	6	0.27	-17.90	18.43	18.16	7	-3.74	-29.71	22.22	25.96	6	16.30	-31.80	64.39	48.10	7	-4.47	-27.69	18.74	23.22
8 NPW	6	2.21	-30.95	35.38	33.17	7	0.86	-23.75	25.46	24.60	6	17.82	-34.58	70.22	52.40	7	14.20	-4.58	32.97	18.78
9 NB	7	8.21	-51.13	67.54	59.34	8	29.36	-6.32	65.04	35.68	9	-0.35	-50.25	49.56	49.90	9	19.69	-0.06	39.44	19.75
10 PNTh	10	-0.69	-67.52	66.13	66.83	10	-18.60	-138.39	101.19	119.79	10	-12.10	-90.88	66.68	78.78	10	-2.88	-59.90	54.14	57.02
11 OFD	10	0.24	-4.60	5.07	4.84	10	-1.16	-6.58	4.26	5.42	10	0.32	-6.65	7.29	6.97	10	0.36	-4.77	5.49	5.13
12 BPD	10	1.48	-7.16	10.11	8.64	10	-0.93	-8.90	7.04	7.97	10	-0.73	-8.98	7.52	8.25	10	2.24	-5.04	9.52	7.28
13 MCh	10	2.23	-4.62	9.09	6.86	10	1.43	-4.32	7.18	5.75	10	0.99	-6.24	8.21	7.23	10	-0.47	-7.36	6.42	6.89
14 OD_R	10	-4.12	-17.23	8.98	13.11	10	-5.34	-15.77	5.09	10.43	10	-5.82	-19.01	7.37	13.19	10	1.01	-9.43	11.46	10.45
15 OD_L	10	-2.76	-10.67	5.15	7.91	10	-3.90	-16.01	8.20	12.11	10	-6.30	-14.83	2.23	8.53	10	0.56	-5.78	6.90	6.34
16 IOD	10	2.52	-11.81	16.84	14.32	10	-7.90	-27.23	11.42	19.33	10	3.56	-11.96	19.07	15.51	10	0.46	-13.08	13.99	13.54
17 BOD	10	-0.65	-6.53	5.22	5.88	10	-1.38	-7.28	4.53	5.91	10	-2.05	-8.24	4.13	6.18	10	0.98	-4.02	5.99	5.01
18 MxW	10	4.81	-3.29	12.91	8.10	10	2.59	-12.26	17.45	14.86	10	-3.82	-23.17	15.52	19.34	10	7.67	-0.42	15.76	8.09
19 MdW	10	4.55	-5.27	14.37	9.82	10	0.93	-3.14	4.99	4.06	10	-2.63	-9.97	4.70	7.34	10	14.62	5.02	24.22	9.60
20 CBA1	10	0.41	-3.55	4.38	3.96	10	1.61	-4.20	7.42	5.81	10	-0.08	-5.78	5.63	5.71	10	0.27	-5.37	5.91	5.64
21 CBA2	10	-7.61	-46.82	31.61	39.21	10	-2.64	-40.13	34.85	37.49	10	6.36	-17.86	30.57	24.22	10	-9.89	-30.61	10.84	20.72
22 FMA	10	0.43	-11.48	12.34	11.91	10	2.25	-8.03	12.53	10.28	10	-5.24	-20.49	10.01	15.25	10	0.06	-11.74	11.87	11.81
23 MNMA	10	-108.17	-266.13	49.79	157.96	10	-41.68	-172.55	89.20	130.88	10	-106.79	-275.52	61.94	168.73	10	-81.55	-224.31	61.21	142.76
24 IFA	10	1.98	-23.19	27.16	25.17	10	7.69	-20.03	35.41	27.72	10	-4.95	-42.06	32.17	37.11	10	4.38	-18.75	27.52	23.13
25 PAH	10	-52.80	-163.65	58.05	110.85	10	-56.26	-175.35	62.83	119.09	10	-47.42	-224.12	129.27	176.69	10	-18.30	-144.68	108.09	126.38
26 ChoH	6	33.23	-7.10	73.56	40.33	6	20.17	-35.33	75.68	55.50	6	-15.54	-71.05	39.97	55.51	6	10.91	-62.52	84.35	73.43
27 MXL	10	-4.69	-24.68	15.29	19.99	10	1.07	-13.58	15.72	14.65	10	9.50	-35.23	54.24	44.73	10	-3.46	-16.83	9.90	13.36
28 MDL	10	11.62	-10.17	33.41	21.79	10	0.83	-14.45	16.11	15.28	10	3.31	-27.34	33.97	30.66	10	17.31	4.32	30.30	12.99
29 NASO	10	20.20	-42.15	82.55	62.35	10	17.76	-19.50	55.03	37.26	10	10.62	-46.51	67.75	57.13	10	12.14	-33.73	58.01	45.87
30 ORO	10	-10.05	-73.24	53.14	63.19	10	-62.32	-213.87	89.23	151.55	10	-28.78	-107.80	50.24	79.02	10	-39.77	-110.85	31.32	71.08
31 HC	10	0.55	-5.10	6.21	5.65	10	-2.05	-7.49	3.39	5.44	10	0.93	-5.86	7.72	6.79	10	0.98	-5.29	7.25	6.27

Mean within +/- 10% or LoA within +/- 20%
Automated variation higher human observers

Fig. A1: Results of absolute (upper table) and relative (lower table) differences with observer 0 (manual landmark based indirect measurement), observer 2 and 3 (direct manual measurement) and automated results compared to a single expert radiologist (direct manual measurement. Green represents relative mean difference or <10% or relative limits of agreement (+/− mean) of <20%. Orange represents automated measurements with more variation than any one manual observer.

*average measures			Intraobserver					Interobserver					Intermethod			ICC threshold criteria: <0.50 = poor; 0.50 -0.75 = moderate; 0.75 - 0.90 = good; >0.90 = excellent (Koo and Li, 2016)		
				95% Confidence					95% Confidence					95% Confidence				
Variable	n	Cronbach's Alpha	Intraclass Correlation*	Lower Bound	Upper Bound	n	Cronbach's Alpha	Intraclass Correlation*	Lower Bound	Upper Bound	n	Cronbach's Alpha	Intraclass Correlation*	Lower Bound	Upper Bound	Intrasobserver (indirect measurement	Interobserver (direct measurement	Intermethod (direct, indirect, automated
ASBL	10	0.952	**0.952**	0.818	0.988	10	0.936	**0.942**	0.829	0.984	10	0.963	**0.926**	0.770	0.980	Excellent	Excellent	Excellent
PSBL	10	0.747	**0.754**	0.033	0.938	10	0.889	**0.879**	0.659	0.967	10	0.918	**0.915**	0.793	0.976	Good	Good	Excellent
HPL	10	0.795	**0.811**	0.190	0.954	10	0.434	**0.392**	-0.463	0.822	10	0.640	**0.603**	0.099	0.882	Good	n/a	n/a
VPL	10	0.529	**0.431**	-0.476	0.836	10	0.418	**0.397**	-0.569	0.829	10	0.588	**0.554**	-0.009	0.867	n/a	n/a	n/a
PAW	10	0.973	**0.975**	0.901	0.994	10	0.899	**0.791**	0.256	0.947	10	0.926	**0.856**	0.596	0.960	Excellent	Good	Good
Cho_r	5	0.741	**0.769**	-1.788	0.977	6	0.431	**0.275**	-0.386	0.840	5	0.834	**0.660**	0.130	0.953	Good	n/a	Moderate
Cho_l	5	0.750	**0.429**	-0.257	0.909	6	0.627	**0.396**	-0.206	0.870	5	0.712	**0.464**	-0.054	0.911	Good	n/a	Poor
NPW	6	0.959	**0.964**	0.759	0.995	6	-0.104	**-0.029**	-0.326	0.608	5	-1.419	**-0.292**	-0.616	0.566	Excellent	n/a	n/a
NB	7	0.417	**0.429**	-2.607	0.903	8	-0.578	**-0.254**	-0.964	0.553	5	0.458	**0.448**	-0.864	0.934	n/a	n/a	n/a
PNTh	10	0.560	**0.521**	-0.522	0.873	10	0.345	**0.361**	-1.043	0.833	10	0.507	**0.514**	-0.223	0.863	n/a	n/a	n/a
OFD	10	0.981	**0.981**	0.927	0.995	10	0.974	**0.974**	0.926	0.993	10	0.982	**0.982**	0.956	0.995	Excellent	Excellent	Excellent
BPD	10	0.994	**0.987**	0.765	0.998	10	0.933	**0.937**	0.818	0.983	10	0.923	**0.916**	0.795	0.976	Excellent	Excellent	Excellent
MCh	10	0.959	**0.927**	0.446	0.984	10	0.958	**0.958**	0.882	0.989	10	0.971	**0.968**	0.921	0.991	Excellent	Excellent	Excellent
OD_R	10	0.943	**0.926**	0.661	0.982	10	0.773	**0.691**	0.180	0.912	10	0.895	**0.837**	0.584	0.953	Excellent	Moderate	Good
OD_L	10	0.866	**0.774**	-0.058	0.947	10	0.846	**0.793**	0.403	0.943	10	0.916	**0.878**	0.687	0.966	Good	Good	Good
IOD	10	0.909	**0.861**	0.284	0.968	10	0.685	**0.556**	-0.033	0.866	10	0.824	**0.750**	0.413	0.926	Good	n/a	Moderate
BOD	10	0.955	**0.943**	0.743	0.986	10	0.941	**0.930**	0.793	0.981	10	0.960	**0.950**	0.873	0.986	Excellent	Excellent	Excellent
MxW	10	0.936	**0.942**	0.762	0.986	10	0.428	**0.411**	-0.559	0.835	10	0.796	**0.727**	0.372	0.919	Excellent	n/a	Moderate
MdW	10	0.823	**0.833**	0.323	0.959	10	0.934	**0.904**	0.677	0.975	10	0.901	**0.635**	0.176	0.888	Good	Good	Moderate
CBA1	10	0.555	**0.581**	-0.965	0.899	10	0.915	**0.907**	0.735	0.974	10	0.934	**0.934**	0.839	0.981	n/a	Excellent	Excellent
CBA2	10	0.855	**0.745**	-0.139	0.822	10	-0.332	**-0.300**	-2.319	0.627	10	0.154	**0.130**	-0.755	0.722	Moderate	n/a	n/a
FMA	10	0.607	**0.602**	-0.474	0.940	10	0.777	**0.692**	0.181	0.913	10	0.817	**0.777**	0.475	0.935	n/a	Moderate	Good
MNMA	10	0.612	**0.619**	-0.514	0.905	10	-0.557	**-0.176**	-0.581	0.423	10	0.197	**0.090**	-0.246	0.574	n/a	n/a	n/a
IFA	10	0.863	**0.872**	0.479	0.968	10	0.730	**0.701**	0.201	0.916	10	0.760	**0.747**	0.395	0.927	Good	Moderate	Good
PAH	10	0.890	**0.827**	0.138	0.960	10	0.082	**0.067**	-1.011	0.709	10	0.044	**0.036**	-0.880	0.685	Good	n/a	n/a
ChoH	10	0.440	**0.445**	-1.226	0.862	6	0.465	**0.375**	-0.565	0.886	6	0.617	**0.505**	-0.139	0.907	n/a	n/a	n/a
MXL	10	0.555	**0.556**	-0.704	0.883	10	0.543	**0.528**	-0.274	0.869	10	0.501	**0.471**	-0.195	0.843	n/a	n/a	n/a
MDL	10	0.654	**0.647**	-0.299	0.910	10	0.442	**0.448**	-0.634	0.852	10	0.674	**0.515**	0.057	0.839	n/a	n/a	n/a
NASO	10	0.847	**0.833**	0.384	0.958	10	-0.219	**-0.168**	-1.421	0.625	10	0.316	**0.309**	-0.647	0.800	Good	n/a	n/a
ORO	10	0.671	**0.675**	-0.265	0.919	10	0.601	**0.431**	-0.146	0.813	10	0.745	**0.618**	0.187	0.880	n/a	n/a	Moderate
HC	10	0.988	**0.979**	0.839	0.996	10	0.968	**0.953**	0.831	0.988	10	0.964	**0.959**	0.899	0.988	Excellent	Excellent	Excellent

Fig. A2: Table of Intraclass Correlation Coefficient (ICC) results. Green is and excellent or good agreement, yellow a moderate agreement, red is poor agreement (n.b. ICC threshold criteria interpreted only if Cronbach's Alpha is ¿0.70).

Appendix B Tables

Table B1: 3D Anatomical Landmarks (*Landmark not used for biometry extraction)

Label number	Anatomical point	Abbr.	Description
1	Foramen Caecum	Fc	The midline point marking the pit between the fetal crista galli and the endocranial wall of the frontal bone
2	Basion	Ba	The midline point on the anterior margin of the foramen magnum
3	Alveolar ridge (anterior nasal spine)	ANS	The midline central point on the tip of the alveolar ridge (also synonymous with anterior nasal spine, ANS)
4	Hormion	H	The posterior-most midline point on the junction between the ventral surface of the sphenoid and the vomeral root (also, point located at the intersection between the perpendicular line to S-Ba from PNS and the cranial base)
5	Posterior Tongue	TP	Posterior most aspect of the tongue in the midline
6	Posterior nasal spine	PNS	The midline point of the superior surface of the hard palate where the horizontal aspect of the palatine bone meets the posterior most aspect of the vomer bone.
7	Posterior pharangeal wall	PPW	The posterior point of the pharanyx on a linear line contiguous with the hard palate length (ANS-PNS)
8	Occiput	Oc	Posterior most skull border in axial plane at level of thalami
9*	Cervical axis	C1	The midline point on the first vertebral body anterior to the basion point
10	Sinciput	Si	Anterior most skull border in axial plane at level of thalami

Continued on next page

Table B1 – Continued from previous page

Label number	Anatomical point	Abbr.	Description
11	Posterior clinoid process	PCP	The midline point on the posteior raised tuberculum sella of the body of the sphenoid
12	Nasal bone tip	NB	Inferior most tip of the nasal bone
13*	Nose tip	NBT	Tip of the superficial surface of nasal soft tissue
14*	Tongue root	TR	Most inferior posterior muscular part of the tongue (estimated site of hyoid bone)
15	Right parietal	Pt_r	Right lateral most skull border in axial plane at level of SI-Oc plane (OFD/BPD)
16	Left parietal	Pt_l	Left lateral most skull border in axial plane at level of Si-Oc plane (OFD/BPD)
17	Vertex	Ve	Most superior skull border in a midline sagittal plane (at level of cerebral peduncles when oriented to brain orthgonal planes in coronal)
18	Right palate	pa_r	Right mid-inferior border of posterior most maxillary molar in coronal plane (when oriented for coronal facial views)
19	Left palate	pa_l	Left mid-inferior border of posterior most maxillary molar in coronal plane (when oriented for coronal facial views)
20	Palateal vault	PaV	Midline point in coronal plane at level of posterior most maxillary molar at the inferior border of the hard palate
21	Right posterior dental arch of Maxilla	Mx_r	Posterior point of last right molar tooth socket at the level of the widest point of the maxillary dental arch
22	Left posterior dental arch of maxilla	Mx_l	Posterior point of last left molar tooth socket at the level of the widest point of the maxillary dental arch

Continued on next page

Table B1 – Continued from previous page

Label number	Anatomical point	Abbr.	Description
23	Right posterior dental arch of mandible	man_r	Posterior point of last right molar tooth socket at the level of the widest point of the mandibular dental arch (at the level of the masseter on its entrance point to the mandible)
24	Left posterior dental arch of mandible	man_l	Posterior point of last left molar tooth socket at the level of the widest point of the mandibular dental arch (at the level of the masseter on its entrance point to the mandible)
25	Anterior border of symphysis mentum	Me	Midline point at the level of the symphysis mentis of the mandible
26*	Right Superior external ear	hel_r	Right most superior point of the helix (external ear)
27*	Right inferior external ear	lob_r	Right most interior point of the auricular lobule (external ear)
28*	Right external auditory meatus	EAM_r	Midpoint of the right external auditory meatus (between the tragus and the concha)
29*	Left Superior external ear	hel_l	Left most superior point of the helix (external ear)
30*	Left inferior external ear	lob-l	Left most interior point of the auricular lobule (external ear)
31*	Left external auditory meatus	EAM_l	Midpoint of the left external auditory meatus (between the tragus and the concha)
32	Right inner orbit	RO_in	Right inner border of the orbital globe (i.e. medial outer edge of sclera) at widest point in axial view
33	Right outer orbit	RO_o	Right outer border of the orbital globe (i.e. lateral outer edge of sclera) at widest point in axial view
34	Left inner orbit	LO_in	Left inner border of the orbital globe (i.e. medial outer edge of sclera) at widest point in axial view
35	Left outer orbit	LO_o	Left outer border of the orbital globe (i.e. lateral outer edge of sclera) at widest point in axial view

Continued on next page

Table B1 – Continued from previous page

Label number	Anatomical point	Abbr.	Description
36	Nasion (inner)	NaIn	Inner border of nasion
37	Nasion outer (skin border)	NaO	Outer skin border of nasion
38*	Nuchal fold inner	NF_in	Nuchal inner border (medial subcutaneous layer)
39*	Nuchal fold outer (skin border)	NF_o	Nuchal outer border (posterior border subcutaneous layer)
40	Upper lip border	Lip	Skin border of the upper lip midline
41	Mentum skin surface	Chin	Skin border of the chin midline
42*	opisthion	Op	Midline posterior border of the foramen magnum
43*	Right Lateral Foramen Magnum	FM_r	Right lateral border of the foramen magnum
44*	Left lateral Foramen Magnum	FM_l	Left lateral inner border of the foramen Magnum
45*	Tongue tip	TT	Most anterior part of tongue in the midline
46	Rt medial pterygoid plate	MPP_r	Point on the right posterior lateral border of the nasal labrinth at a level of the vomer
47	Lt medial pterygoid plate	MPP_l	Point on the left posterior lateral border of the nasal labrinth at a level of the vomer
48	Rt Choanae lat	Cho_r	Right posterior lateral border of the anterior nasopharyneal space, posterior to the vomer
49	Lt Choanae Lat	Cho_l	Left posterior lateral border of the anterior nasopharyneal space, posterior to the vomer
50	Mid Posterior Vomer-midVo	Vo	Point in the midline of the vomer level with the right and left medial pterygoid plate

Table B2: Formalised measurement definitions for landmark-based craniofacial biometry protocol with 3D T2w fetal MRI.

Anatomical group	Measure Number	Measurement name	Abbreviation	Landmark points used
1. Base of skull	1	Anterior skull base length	ASBL	Fc (1), PCP (11)
	2	Posterior skull base length	PSBL	PCP (11), Ba (2)
	3	Internal cranial Base angle (°)	CBA1	Fc(1), PCP (11), Ba (2)
	4	External cranial base angle (°)	CBA2	PNS (6), H (4), Ba (2)
2. Facial angles	5	Fronto maxillary angle	FMA	Si (10), Fc (1), PCP (11)
	6	Inferior facial angle	IFA	PCP (11), Fc (1), Lip (40), Chin (41)
	7	Maxillary nasion mandibular angle	MNMA	ANS (3), Fc (1), Me (25)
3. Oropharangeal	8	Hard palate length	HPL	ANS (3), PNS (6)
	9	Velopharangeal length	VPL	ANS (3), PNS (6), PPW (7)
	10	Nasopharyngeal area (mm2)	NASO	PNS (6), H (4), Ba (2)
	11	Oropharyngeal area (mm2)	ORO	PNS (6), Ba (2), TP (5)
	12	Palatal width	PaW	Pa (Rt/Lt) (18,19)
	13	Palatal height	PaH	Pav (20), RtPa, Lt Pa (18, 19)
4. Nasal	14-15	Choanae width (rt/Lt)	Cho_R/Cho_L	MPP (Rt/Lt) (46, 47), Vo (50)
	16	Choanae height (mm)	ChoH	H (4), ANS (3), PNS (6)
	17	Nasopharynx width	NPw	Cho (Lt/Rt) (48, 49)
	18	Nasal bone	NB	NaIn (36), NB (12)
	19	Prenasal thickness	PNTh	NaIn (36), NaO (37)
	20	Occipital frontal diameter	OFD	Si (10), Oc (8)
5. Cranial vault	21	Bi parietal diameter	BPD	Pt (Rt/Lt) (15/16)
	22	Head circumference	HC	SI (10), Oc (8), RtPt (15), LtPt (16)
	23	Maximum cranial height	MCh	Ve (17), Ba (2)
	24-25	Orbital distance	ROD/LOD	RO/LO_in (33/35), RO/LO_o (34/36)
6. Orbits	26	Interoccular distance	IOD	RO_in (33), LO_i (35)
	27	Biocular distance	BOD	RO/LO_o (34/36)
7. Maxilla	28	Maxillary width	MxW	RtMx (21), LtMx (22)
	29	Maxillary length	MxL	ANS (3), RtMx (21), LtMx (22)
8. Mandible	30	Mandibular Width	MdW	RtMd (23), LtMd (24)
	31	Mandibular length	MdL	Me (25), RtMd (23), LtMd (24)

Table B3: Table of number of subjects (n), mean measurement, and standard deviation (SD) - stratified by healthy control and T21 groups

		T21 (mean GA: 32.48)			Control (mean GA: 31.88)		
Biometry ID	Biometry name, (unit)	n	Mean	SD	n	Mean	SD
1	ASBL, (mm)	24	34.45	2.20	108	36.76	2.13
2	PSBL, (mm)	24	24.10	1.40	108	24.77	1.45
3	HPL, (mm)	24	24.69	2.10	108	26.02	1.85
4	VPL, (mm)	24	38.21	2.82	108	40.22	2.53
5	PAW, (mm)	24	24.76	1.71	108	24.88	1.72
6	Cho_r, (mm)	24	5.04	0.46	108	5.40	0.50
7	Cho_l, (mm)	24	5.68	0.45	108	5.94	0.53
8	NPW, (mm)	24	9.29	0.95	108	9.87	1.08
9	NB, (mm)	24	7.92	0.74	108	7.41	0.69
10	PNTh, (mm)	24	7.81	1.21	108	7.14	0.88
11	OFD, (mm)	24	97.82	5.41	108	101.05	6.34
12	BPD, (mm)	24	84.43	4.17	108	83.38	5.43
13	MCh, (mm)	24	79.80	4.13	108	80.36	4.26
14	OD_R, (mm)	24	17.18	0.88	108	17.35	0.98
15	OD_L, (mm)	24	17.53	0.99	108	17.80	1.07
16	IOD, (mm)	24	15.48	1.14	108	15.91	1.18
17	BOD, (mm)	24	49.95	2.64	108	50.64	2.77
18	MxW, (mm)	24	23.90	1.62	108	24.25	1.67
19	MdW, (mm)	24	30.42	1.86	108	30.89	2.26
20	CBA1, (^{0})	24	131.20	3.76	108	129.27	4.31
21	CBA2, (^{0})	24	116.89	5.72	108	115.39	7.50
22	FMA, (^{0})	24	116.28	4.56	108	114.04	4.99
23	MNMA, (^{0})	24	17.66	2.19	108	17.91	2.27
24	IFA, (^{0})	24	55.16	5.22	108	48.97	4.95
25	PAH, (mm)	24	3.28	0.63	108	3.00	0.58
26	ChoH, (mm)	24	3.30	0.54	108	3.22	0.59
27	MXL, (mm)	24	24.99	1.95	108	26.23	1.80
28	MDL, (mm)	24	19.02	1.68	108	19.27	1.53
29	NASO, (mm^3)	24	58.42	7.18	108	63.88	9.37
30	ORO, (mm^3)	24	181.41	32.81	108	173.97	33.06
31	HC, (mm)	24	289.62	14.46	108	296.01	18.20

Table B4: Summary of included datasets for the 3rd trimester T21 and age-matched control cohorts, by GA, MRI protocol, fetal sex and ethnicity

		T21	**Control**	**Total count all, n**
Gestational age	Range	29.86 - 35.71	29.00 - 36.00	-
	Mean (SD)	32.48 (1.85)	31.88 (2.01)	-
MRI Protocol, n (%)	1.5T / 80ms	4 (16.6)	34 (31.5)	38
	3T / 180ms	17 (70.8)	25 (23.1)	42
	3T / 250ms	3 (0.1)	49 (45.4)	52
	MRI protocol total, n (%)	**24(100)**	**108 (100)**	132
Fetal Sex, n (%)	Female	10 (41.7)	53 (49.1)	63
	Male	10 (41.7)	54 (50.0)	64
	Unknown	4 (5.4)	1 (0.9)	5
	Fetal sex total,n (%)	**24 (100)**	**108 (100)**	132
Ethnicity, n (%)	Asian	1 (4.2)	10 (9.3)	11
	Black	1 (4.2)	4 (3.7)	5
	Other	0 (0)	3 (2.8)	3
	Unknown	21 (87.5)	5 (4.6)	26
	White	1 (4.2)	86 (79.6)	87
	Ethnicity total, n (%)	**24 (100)**	**108 (100)**	132

Variable	Bestfit line equation	SD equation	Mean bestfit method
ASBL	0.7191 *GA + 13.340	0.039 *GA + 0.221	Linear
PSBL	28.312 + -0.6771 * GA + 0.018 *GA*GA	0.060 * GA + -0.898	Quadratic
HPL	0.6207 * GA + 6.230	0.068 * GA + -0.843	Linear
VPL	58.7882 + -2.0538 * GA + 0.0460 * GA*GA	-0.031 * GA + 2.508	Quadratic
PAW	0.6259 * GA + 4.925	0.042 * GA + -0.178	Linear
Cho_r	0.1199 * GA + 1.575	0.019* GA + -0.173	Linear
Cho_l	0.1438 * GA + 1.359	0.049 * GA + -1.150	Linear
NPW	0.1517 * GA + 5.0364	-0.022 * GA + 1.71	Linear
NB	35.307 + -1.940 * GA + 0.032 * GA*GA	0.008 * GA + 0.188	Quadratic
PNTh	0.211 * GA + -0.926	-0.013* GA + 1.085	Linear
OFD	113.735 + -2.971 * GA + 0.0803 * GA*GA	0.118 * GA + 0.432	Quadratic
BPD	1.885 * GA + 23.520	0.037 * GA + 2.485	Linear
MCh	1.410 * GA + 34.461	0.119 * GA + -1.101	Linear
OD_R	0.3607 * GA + 5.856	0.011 * GA + 0.242	Linear
OD_L	0.393 * GA + 5.281	0.037 * GA + -0.501	Linear
IOD	0.2642 * GA + 7.485	0.023 * GA + 0.255	Linear
BOD	1.049 * GA + 17.203	0.037 * GA + 0.435	Linear
MxW	0.609 * GA + 4.849	0.081 * GA + -1.470	Linear
MdW	-19.0189+ 2.345 * GA + -0.024 * GA*GA	0.002 * GA + 1.572	Quadratic
CBA1	0.230 * GA + 120.529	0.266 * GA + -4.4712	Linear
CBA2	1.301 * GA + 79.057	0.266 * GA + -4.471	Linear
FMA	0.091 * GA + 109.872	0.140* GA + 0.693	Linear
MNMA	0.159 * GA + 13.787	0.0158 * GA + 1.795	Linear
IFA	0.253 * GA + 42.129	0.206 * GA + -1.307	Linear
PAH	0.067 * GA + 0.858	0.004 * GA + 0.392	Linear
ChoH	0.0129 * GA + 2.880	0.023* GA + -0.126	Linear
MXL	55.528 + -2.427 * GA + 0.047 * GA*GA	0.005 * GA + 1.058	Quadratic
MDL	33.245 + -1.336* GA + 0.028 * GA*GA	0.0321 * GA + 0.136	Quadratic
NASO	2.376 * GA + -2.227	0.0142 * GA + 8.092	Linear
ORO	9.157 * GA + -117.920	1.535 * GA + -21.413	Linear
HC	322.684 + -7.907 * GA + 0.221 * GA*GA	-0.145 * GA + 16.694	Quadratic

Table B5: Best linefit and Standard Deviation (SD) regression equations and method, for each biometric growth chart. GA = Gestational age. Note: all SD regression formulae have linear bestfit method.

Dependent Variable	Test parameter with covariate	B	Robust Std. Error	t	Sig.	95% C.I. Lower Bound	95% C.I. Upper Bound	Partial Eta Squared	Noncent. Parame-ter	Observed Power
Anterior Skull Base Length, ASBL	T21 with GA*	-31.631	6.334	-4.994	0.000	-44.163	-19.099	0.162	4.994	0.999
Hard Palate Length, HPL	T21 with GA*	-1.682	0.329	-5.107	0.000	-2.334	-1.030	0.168	5.107	0.999
Velopharyngeal Length, VPL	T21 with GA*	-2.536	0.438	-5.784	0.000	-3.403	-1.668	0.206	5.784	1.000
Occipitofrontal Diameter, OFD	T21 with GA*	-4.389	0.908	-4.835	0.000	-6.185	-2.593	0.153	4.835	0.998
Biparietal Diameter, BPD	T21 with GA§	0.297	0.677	0.438	0.662§	-1.043	1.636	0.001	0.438	0.072
Inferior Facial Angle, IFA	T21 with GA*	7.042	1.563	4.505	0.000	3.949	10.135	0.136	4.505	0.994
Maxillary Length, MXL	T21 with GA*	-1.608	0.308	-5.217	0.000	-2.217	-0.998	0.174	5.217	0.999
Nasopharyngeal Area, NASO	T21 with GA*	-8.483	1.777	-4.774	0.000	-11.999	-4.968	0.150	4.774	0.997

* = Large effect size
§ = non-significant result and at risk of Type II error
GA = Gestational Age

Table B6: ANOVA results for statistically significant biometry with largest effect size in T21 cohort (ABSL, HPL, VPL, OFD, IFA, NASO) and BPD as a non-significant results. Results corrected with robust standard errors.