RESEARCH ARTICLE

Intra- and interobserver agreement for fetal cerebral measurements in 3D-ultrasonography

Maria E. W. A. Albers^{1,2} | Erato T. I. A. Buisman³ | René S. Kahn^{2,4} |
Arie Franx¹ | N. Charlotte Onland-Moret^{5†} | Roel de Heus^{1†}

¹Division of Woman and Baby, University Medical Center Utrecht, Utrecht University, The Netherlands

²Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht University, The Netherlands

³Department of Obstetrics and Gynaecology, Jeroen Bosch Hospital, 's Hertogenbosch, The Netherlands

⁴Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, USA

⁵Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, The Netherlands

Correspondence

Roel de Heus, c/o Wilhelmina Children's Hospital (WKZ), Huispostnummer KE04.123.12, Postbus 85090, 3508 AB Utrecht, The Netherlands.
Email: R.deHeus-2@umcutrecht.nl

Funding information

Gravitation program of the Dutch Ministry of Education, Culture, and Science; Netherlands Organization for Scientific Research (NWO), Grant/Award Number: 024.001.003

Abstract

The aim of this study is to evaluate intra- and interobserver agreement for measurement of intracranial, cerebellar, and thalamic volume with the Virtual Organ Computer-aided AnaLysis (VOCAL) technique in three-dimensional ultrasound images, in comparison to two-dimensional measurements of these brain structures. Three-dimensional ultrasound images of the brains of 80 fetuses at 20–24 weeks' gestational age were obtained from YOUth, a Dutch prospective cohort study. Two observers performed offline measurement of the occipitofrontal diameter, intracranial volume, transcerebellar diameter, cerebellar volume, and thalamic width, area, and volume, independently. VOCAL was used for calculation of the volumes. The two-way random, single measures intraclass correlation coefficient (ICC) was used for analysis of agreement and Bland–Altman plots were configured. Intra- and interobserver agreement was almost perfect for occipitofrontal diameter (intra ICC 0.88, 95% CI 0.82–0.92; inter ICC 0.91, 95% CI 0.85–0.94), intracranial volume (intra ICC 0.96, 95% CI 0.91–0.98; inter ICC 0.97, 95% CI 0.96–0.98) and transcerebellar diameter (intra ICC 0.91, 95% CI 0.86–0.94; inter ICC 0.86, 95% CI 0.78–0.91). For cerebellar volume, the intraobserver agreement was almost perfect (0.85, 95% CI 0.76–0.90), whereas the interobserver agreement was substantial (0.75, 95% CI 0.44–0.88). Agreement was only moderate for thalamic measurements. Bland–Altman plots for the volume measurements are normally distributed with acceptable mean differences and 95% limits of agreement. The intra- and interobserver agreement of the measurement of intracranial and cerebellar volume with VOCAL was almost perfect. These measurements are therefore reliable, and can be used to investigate fetal brain development. Thalamic measurements are not reliable enough.

KEYWORDS

brain, cerebrum, cerebellum, fetal, intraobserver agreement, interobserver agreement, three dimensional, ultrasound

1 | INTRODUCTION

Currently, prenatal assessment of the fetal brain is mainly performed by using two-dimensional (2D) ultrasonography. In recent years, three-dimensional (3D) ultrasound has been introduced and increasingly investigated as an alternative technique (Bornstein et al., 2010b; Correa et al., 2006; Monteagudo, Timor-Tritsch, & Mayberry, 2000; Pilu et al., 2007; Salman et al., 2011; Tonni, Martins, Guimarães Filho, & Araujo

Júnior, 2015). This technique provides the opportunity to investigate prenatal brain development in more detail (Bornstein, Monteagudo, Santos, Keeler, & Timor-Tritsch, 2010a; Ioannou, Sarris, Salomon, & Papageorgiou, 2011; Pilu et al., 2006; Rizzo et al., 2012; Vinals, Muñoz, Naveas, & Giuliano, 2007).

Research in the field of Developmental Origins of Health and Disease (DOHaD) has shown that the prenatal environment of the fetus, such as maternal nutrition or stress, has influence on the risk of developing chronic adult diseases in later life, and also on the development of psychological and psychiatric problem behaviors (Barker, 1998; Gillman, 2005; Schlotz & Phillips, 2009). Likely, early

[†]N. Charlotte Onland-Moret and Roel de Heus contributed equally to this work.
Institution: Utrecht University and University Medical Center Utrecht

development of the fetal brain mediates in this link between prenatal environmental factors on the one hand and psychological and behavioral development on the other (Schlotz, Jones, Godfrey, & Phillips, 2008; Schlotz & Phillips, 2009). Therefore, it is important to elucidate how the environment of the fetus influences normal prenatal brain development and how brain development subsequently relates to neurocognitive and behavioral development.

To this purpose, a longitudinal prospective cohort study with repeated measurements during pregnancy and after birth is ongoing, with the aim to investigate how (normal) neurocognitive development mediates in the association between biological and environmental factors, and behavioral development in children. Two times during pregnancy (around 20 and around 30 weeks of gestation) 3D volume sweeps of the fetal brain are acquired with 3D ultrasound. These sweeps will be used to investigate normal structural fetal brain development in this cohort.

Structural fetal brain development can be investigated by measuring the total brain volume, grey matter volume, or the volume of brain structures, such as the cerebellum, thalami, hippocampus, or ventricles. For the measurement of volumes in 3D ultrasound images, the Virtual Organ Computer-aided AnaLysis (VOCAL) technique is a useful tool, which calculates the volume of the structure of interest after manual delimitation of the contours of the structure (Bordes, Bory, Benchaïb, Rudigoz, & Salle, 2002; Raine-Fenning et al., 2003). Concerning the brain, only methods for the measurement of intracranial volume, cerebellar volume, and thalamic volume with VOCAL have thus far been described (Araujo Júnior, Pires, Nardozza, Filho, & Moron, 2007; Benavides-Serralde et al., 2009; Rutten et al., 2009; Sotiriadis, Dimitrakopoulos, Eleftheriades, Agorastos, & Makrydimas, 2012; Zeng et al., 2015). As this technique has not been validated yet (International Society of Ultrasound in Obstetrics & Gynecology Education Committee, 2007; Ioannou et al., 2011), it is important to determine if these previously described volume measurements with VOCAL are reliable enough for research purposes, before using it in large scale. The aim of this study is therefore to analyze intra- and interobserver agreement for the measurement of intracranial, cerebellar, and thalamic volume with the VOCAL technique following previously described methods, in comparison to the reproducibility of standard 2D distance and area measurements of these structures.

2 | METHODS

2.1 | Population

The 3D ultrasound images for this study were collected as part of an ongoing longitudinal observational cohort study in the Netherlands, called YOUTH. Recruitment started in July 2015 through hospitals and midwifery practices in Utrecht and surrounding areas, and is still ongoing. The needed sample size was estimated to be at least 63 based on $\alpha = 0.05$, $\beta = 0.20$, $\rho_0 = 0.5$, and $\rho_1 = 0.7$ (Walter, Eliasziw, & Donner, 1998). As we anticipated that not all ultrasound data would be of enough quality for volume measurements, we selected the ultrasound volume data of the first 80 participants from their first visit between the

21st and 24th gestational week, to investigate intra- and interobserver agreement. This study was approved by the Medical Research Ethics Committee of the University Medical Center Utrecht and all participants provided written informed consent.

2.2 | Ultrasound imaging

3D ultrasound volume datasets were obtained transabdominally between 20 and 24 weeks gestational age by ten experienced sonographers, using a Voluson E10 (GE Healthcare, Zipf, Austria) ultrasound machine with a 2–6 MHz convex probe (RM6C). To acquire a 3D-ultrasound image, the probe sends out a sweep with sound waves at different angles. The returning echoes are processed and reconstructed in a multiplanar view of the three two-dimensional (2D) orthogonal planes. Five volume sweeps of the fetal brain were obtained from each participant; two volumes were acquired from the axial plane (transthalamic and transcerebellar), two from the coronal plane (transthalamic and transcerebellar), and one from the midsagittal plane. The angle of the sweeps was set at 65° and the maximum quality of acquisition was selected. The sweeps were stored for offline interpretation and analysis.

2.3 | Offline measurements

Two examiners (E.B. and M.A.) measured occipitofrontal diameter, intracranial volume, transcerebellar diameter, cerebellar volume, thalamic width, thalamic area, and thalamic volume offline, using the GE Medical Systems 4D View software, version 14 Ext.4 (GE Healthcare, Zipf, Austria). Virtual organ computer-aided analysis (VOCAL; GE Healthcare, Zipf, Austria) was used for volume calculations. The VOCAL software allows the observer to manually trace the outlines of a structure. When the structure has been traced once, the view is rotated around a rotation axis in six (if the rotation angle is set at 30°) to 30 steps (if the rotation angle is set at 6°). In each consecutive step, the observer manually traces the boundaries of the structure. In this study, the rotation angle was set at 30°; therefore, the examined brain structures were traced in six consecutive images. After the final trace, the VOCAL software automatically calculates the volume of the traced structure and provides a 3D image of the structure. The technical algorithm that is used by VOCAL to quantify the volume is described by Bordes et al. (2002).

The two observers measured the structures individually, blinded to each other's results. One observer (E.B.) performed the measurements twice to determine intraobserver agreement. Before performing the measurements, the observers first assessed the quality of the obtained sweeps and categorized these as inadequate, adequate, or excellent. If the image was distorted by fetal movement or if the brain structures were not visible enough for measurement due to extensive shadowing, the image was categorized as inadequate. The image was considered of adequate quality when there was some degree of shadowing but the brain structures were visible enough to perform one or more of the measurements. If there was no distortion and slight to no shadowing, the quality of the sweep was considered excellent. For each measurement, the observers selected the sweep with the best quality for that

FIGURE 1 Measurement of fetal brain structures: (a) occipitofrontal diameter, (b) intracranial volume, (c) transcerebellar diameter, (d) cerebellar volume, (e) width and area of thalami, and (f) volume of thalami. In (b, d, and f), the upper left image shows the tracing of the structure in the coronal view, the upper right image shows the sagittal view, and the bottom left image shows the axial view. The image in the bottom right shows the reconstructed 3D-image of the structure

measurement out of the set of volume sweeps (of one participant). The measurement was only performed if the observer considered the quality of the sweep at least adequate and if the structure was visible enough to perform the measurement. If there was no sweep with at least adequate quality for the measurement, that measurement was left out and registered as missing.

The occipitofrontal diameter (OFD) was measured in the axial transthalamic plane from the outer border of the frontal bone to the outer border of the occipital bone (Figure 1a) (Monteagudo & Timor-Tritsch, 2012). The same plane was selected as the reference plane for measuring intracranial volume. The outlines were defined anteriorly, posteriorly, and laterally by the inner wall of the skull and inferiorly by the floor of the skull (Figure 1b and Supporting Information, Figure 1) (Benavides-Serralde et al., 2009; Zeng et al., 2015). Transcerebellar diameter (TCD) and the cerebellar volume were measured to examine the cerebellum. The axial transcerebellar plane was used for the measurement of the TCD, from outer to outer border of the cerebellum, and as reference plane for measurement of the cerebellar volume (Monteagudo & Timor-Tritsch, 2012). To obtain the cerebellar volume, the contours of the cerebellum were traced (Figure 1c,d and Supporting Information, Figure 2) (Araujo Júnior et al., 2007). The axial transthalamic plane was selected for assessment of the thalami. The thalami can be found as darker oval structures in the middle of the plane on both sides of the midline. The width and area of both the right and the left thalamus were measured in this plane (Figure 1e). Starting in the axial transthalamic plane, the volume of each thalamus was measured by tracing the contours (Figure 1f and Supporting Information, Figure 3) (Sotiriadis et al., 2012).

All 2D measurements (OFD, TCD, thalamic width, and thalamic area) were performed once. To evaluate if the measurement of a volume is more reliable when the average of multiple measurements is used, intracranial volume, cerebellar volume, and volumes of the thalami were measured three times and the average was calculated. To test the effect of training of the observers on interobserver agreement, a volume measurement was performed again by both observers if the variance was large.

2.4 | Statistical analysis

To describe demographics of the population, the mean $\pm$ SD was calculated when the data were normally distributed, the median with interquartile range was calculated when the data were not normally distributed, and frequencies with percentage were given for nominal data. For analysis of intra- and interobserver agreement a two-way random model was used and single-measure intraclass correlation coefficients (ICCs) for absolute agreement were calculated. The two measurements of the first observer (E.B.) were compared for intraobserver agreement. For interobserver agreement the second measurements of the first observer were compared to the measurements of the second observer (M.A.). The following cutoff values were used for the quality of the agreement: An ICC value under 0.2 was considered as slight agreement; 0.21–0.40 as fair; 0.41–0.60 as moderate; 0.61–0.80 as substantial; and 0.81–1.00 as almost perfect agreement (Fleiss, 1981). Bland–Altman plots were configured to determine the bias between the two measurements and the limits of agreement (Bland & Altman, 2003). All statistical analyses were performed using IBM SPSS Statistics for Windows, version 24.0 (IBM corp. Armonk, NY).

TABLE 1 Baseline characteristics of the studied population

Variable	Mean $\pm$ SD	N (% of population)
Gestational age at examination (weeks + days)	21 + 6 $\pm$ 1 + 0	77 (100)
Maternal age (years)	33 $\pm$ 5	71 (92.2)
	Median (IQR)	
Maternal BMI	23 (20–26)	71 (92.2)
	n (%)	
Fetal position		77 (100)
Head	43 (55.8)	
Breech	29 (37.7)	
Transverse	5 (6.5)	

Note. Abbreviations: BMI = body mass index; IQR = interquartile range; SD = standard deviation.

3 | RESULTS

A total of 80 datasets of fetal brain volumes were selected for offline analysis. However, two datasets were excluded from analysis because they were from twin pregnancies and one dataset was excluded because all volume sweeps were of poor quality due to high maternal BMI. For the remaining 77 datasets, the first observer assessed the quality of the best axial transthalamic sweep as excellent in 22 sets (28.6%), as adequate in 50 sets (64.9%), and as inadequate in 5 datasets (6.5%). The best axial transcerebellar sweep was of excellent quality in 10 sets (13.0%), of adequate quality in 60 sets (77.9%), and of inadequate quality or not performed in 7 sets (9.1%). The quality of the best coronal transthalamic sweep was assessed as excellent in 7 sets (9.1%), as adequate in 42 sets (54.5%), and as inadequate or absent in 28 sets (36.4%). The best coronal transcerebellar sweep was qualified as excellent in 14 sets (18.2%), as adequate in 44 sets (57.1%), and as inadequate or absent in 19 sets (24.7%). The quality of the best midsagittal sweep was assessed as excellent in 8 sets (10.4%), as adequate in 35 sets (45.5%), and as inadequate or not performed in 34 sets (44.1%). Between the different sonographers, there was no difference in the quality of the acquired sweeps. The demographical details are described in Table 1. The mean maternal age was 33 years ($SD \pm 5$) and the mean gestational age was 21 + 6 weeks ($SD \pm 1$ week).

3.1 | Cerebrum

The ICCs of the intra- and interobserver agreement of the measurements of occipitofrontal diameter and intracranial volume are described in Table 2. Measurement of the occipitofrontal diameter was possible in all datasets by both observers with almost perfect ICCs of 0.88 (95% confidence interval (CI) 0.82–0.92) for intraobserver agreement and of 0.91 (95% CI 0.85–0.94) for interobserver agreement. The intracranial volume was measured in 73 datasets (94.8%) by both observers. The remaining four datasets were left out because one ($n = 2$) or both ($n = 2$) observers considered the quality of all sweeps in this dataset too low to perform the measurement. Agreement for this measurement

was almost perfect for both the first measurement (intra ICC 0.96, 95% CI 0.91–0.98 and inter ICC 0.86, 95% CI 0.00–0.96) and the calculated average of three measurements (intra ICC 0.96, 95% CI 0.90–0.98 and inter ICC 0.85, 95% CI 0.00–0.96). As the 95% confidence interval for interobserver agreement was extremely large, we decided to test for a training effect in the observers. Both observers independently measured the intracranial volume again, which resulted in an almost perfect interobserver agreement of 0.97 with a much smaller 95% confidence interval (0.96–0.98). Figure 2a shows the Bland–Altman plot for intraobserver agreement of intracranial volume measurement, with a mean difference of 2.15 cm³ (95% Limits of Agreement (LoA): 5.56–9.87). In Figure 2b, Bland–Altman analysis of the second time both observers measured intracranial volume shows a mean interobserver difference of 0.24 cm³ (95% LoA: 6.61–7.08).

3.2 | Cerebellum

The results of the cerebellar measurements are described in Table 3. The transverse cerebellar diameter was measured in 70 datasets by the first observer, in 75 datasets by the second observer, and in 68 datasets (88.3%) by both with almost perfect intra- and interobserver agreement (intra ICC 0.91, 95% CI 0.86–0.94 and inter ICC 0.86, 95% CI 0.78–0.91). For 65 volume datasets (84.4%) both observers measured the volume of the cerebellum, 67 datasets by observer 1 and 73 by observer 2. With an ICC of 0.85, intraobserver agreement of cerebellar volumetry is almost perfect for both the first measurement (95% CI 0.76–0.90) and the average of three measurements (95% CI 0.77–0.91). Interobserver agreement of cerebellar volumetry is substantial with an ICC of 0.74 (95% CI 0.39–0.87) for the first measurement and an ICC of 0.75 (95% CI 0.44–0.88) for the average of three measurements. The Bland–Altman analyses, presented in Figure 2c,d, show a mean intraobserver difference of 0.07 cm³ (95% LoA: 0.58–0.72) and a mean interobserver difference of 0.21 cm³ (95% LoA: 0.40–0.83) for the first measurement of the cerebellar volume.

3.3 | Thalamus

Table 4 describes the intra- and interobserver reliability of the measurements of both thalami. The width of the left thalamus was

TABLE 2 Intra- and interobserver reliability of measurement of the occipitofrontal diameter, average of three measurements of intracranial volume and of the first measurement of intracranial volume, described with intraclass correlation coefficients (ICCs), and 95% confidence intervals (95% CI)

Measurement	ICC (95% CI)	
	Intraobserver	Interobserver
Occipitofrontal diameter	0.88 (0.82–0.92)	0.91 (0.85–0.94)
Intracranial volume (average)	0.96 (0.90–0.98)	0.85 (0.00–0.96)
Intracranial volume (first)	0.96 (0.91–0.98)	0.86 (0.00–0.96)
		0.97 (0.96–0.98) ^a

^aIntracranial volume was measured again by both observers after evaluation of the method.

FIGURE 2 Bland-Altman plots of intra- and interobserver agreement of the first measurement of (a and b) the intracranial volume, (c and d) cerebellar volume, and (e and f) the volume of right thalamus. The solid line shows the mean difference, while the dashed lines show the 95% limits of agreement

measured by both observers in 68 datasets (88.3%), 72 by observer 1, and 68 by observer 2. The width of the right thalamus was measured in 64 datasets (83.1%), 71 by observer 1, and 68 by observer 2.

Agreement of the measurement of the width of the thalami was moderate with an intraobserver ICC of 0.56 (95% CI 0.37–0.70) for the left thalamus and 0.5 (95% CI 0.31–0.66) for the right thalamus, and an

TABLE 3 Intra- and interobserver reliability of measurement of the transcerebellar diameter, average of three measurements of the cerebellar volume and of the first measurement of the cerebellar volume, described with intraclass correlation coefficients (ICCs), and 95% confidence intervals (95% CI)

Measurement	ICC (95% CI)	
	Intraobserver	Interobserver
Transcerebellar diameter	0.91 (0.86–0.94)	0.86 (0.78–0.91)
Cerebellar volume (average)	0.85 (0.77–0.91)	0.74 (0.39–0.87)
Cerebellar volume (first)	0.85 (0.76–0.90)	0.75 (0.44–0.88)

interobserver ICC of 0.44 (95% CI 0.19–0.62) and 0.4 (95% CI 0.17–0.59), respectively. The area of the left thalamus was measured in 70 datasets by observer 1 and in 70 datasets by observer 2, but in only 66 datasets (85.7%) by both observers. The area of the right thalamus was measured in 70 datasets by observer 1, in 68 datasets by observer 2, and in 63 datasets (81.8%) by both. The agreement of the measurement of thalamic area was moderate with intra- and interobserver ICCs of 0.49 (95% CI 0.29–0.65) and 0.44 (95% CI 0.22–0.61) for the left thalamus and of 0.46 (95% CI 0.29–0.63) and 0.54 (95% CI 0.35–0.70) for the right thalamus. Left thalamic volume was measured by both observers in 66 datasets (85.7%), 74 by the first, and 69 by the second observer, and right thalamic volume in 59 datasets (76.6%), 67 by the first, and 65 by the second observer. The agreement of thalamic volume measurements was moderate for both the left and right thalamus and for both the first measurements and the average of three measurements, with ICCs ranging from 0.50 (95% CI 0.27–0.68) for interobserver agreement of the average of three measurements of the right thalamus to 0.58 (95% CI 0.40–0.72) for intraobserver agreement of the average of three measurements of the right thalamus. Figure 2e,f presents the Bland–Altman analyses of the first measurements of the right thalamic volume. The mean intraobserver difference was 0.03 cm³ (95% LoA: 0.26–0.32) and the mean interobserver difference was –0.05 cm³ (95% LoA: 0.38–0.27). Bland–Altman analyses of the measurements of the left thalamus show a similar distribution.

4 | DISCUSSION

The results of this study show that offline measurement of intracranial and cerebellar volume with the VOCAL technique in stored 3D

ultrasound images has good to almost perfect reproducibility, while intra- and interobserver agreement of thalamic volume measurements is moderate. The reproducibility of 3D volume measurements was equal to the reproducibility of the 2D measurements, except for the cerebellum for which interobserver agreement of the TCD was higher than interobserver agreement of the cerebellar volume. We performed volumetry three subsequent times in the selected sweep of best quality and analyzed agreement rates for the first of these three measurements as well as for the calculated average, to examine if an average of three measurements is more reliable than one measurement. Our results show that agreement for the average of three subsequent volume measurements is not higher than agreement for the first measurement. Thus, precision does not increase by multiple measurements and it is sufficient to perform volume measurements of brain structures only once, which saves a considerable amount of time.

We reported that the 95% confidence interval of interobserver agreement for the measurement of intracranial volume was initially extremely large (0.00–0.96), while the ICC was high (0.86). When both observers measured intracranial volume again, the ICC became higher with a smaller confidence interval (ICC 0.97; 95% CI 0.96–0.98), demonstrating a training effect. The reliability of the measurements by the observers increased with experience. This shows that it is pivotal that observers who aim to measure volumes of brain structures practice this technique extensively before bringing it into practice.

The ultrasound sweeps were acquired transabdominally, as the transvaginal approach is not comfortable for the pregnant women, and therefore, less feasible in observational research with healthy participants. It has been argued that the transvaginal approach is superior to the transabdominal approach. In the transvaginal approach, the transducer is placed closer to the fetal head and ultrasound waves do not have to pass through the maternal abdominal wall, uterine wall, and (anterior) placenta (Monteagudo & Timor-Tritsch, 2012; Pooh, 2012). The transvaginal approach could theoretically have provided us with sweeps of higher quality, which could have led to higher agreement rates. However, measurements were not performed if the observers considered the sweep of poor quality; nonetheless, most measurements were performed in over 80% of the volume datasets by both observers. We therefore conclude that transabdominal sweeps are of sufficient accuracy for research purposes.

We performed the repeating measurements in the same volume dataset. Consequently, some factors leading to variability in differences

TABLE 4 Intra- and interobserver reliability of measurement of the width and area of the thalami, the average of three measurements of the volume of the thalami and of the first measurement of the volume of the thalami, described with intraclass correlation coefficients (ICCs), and 95% confidence intervals (95% CI)

Measurement	ICC (95% CI)			
	Left thalamus		Right thalamus	
	Intraobserver	Interobserver	Intraobserver	Interobserver
Thalamic width	0.56 (0.37–0.70)	0.44 (0.19–0.62)	0.50 (0.31–0.66)	0.40 (0.17–0.59)
Thalamic area	0.49 (0.29–0.65)	0.44 (0.22–0.61)	0.46 (0.29–0.63)	0.54 (0.35–0.70)
Thalamic volume (average)	0.53 (0.33–0.68)	0.53 (0.31–0.69)	0.58 (0.40–0.72)	0.50 (0.27–0.68)
Thalamic volume (first)	0.52 (0.33–0.67)	0.52 (0.31–0.68)	0.57 (0.38–0.71)	0.55 (0.34–0.70)

between measurements (such as the position of the fetus and the pressure that was put on the transducer) are not accounted for in our data. This may have led to an overestimation of the reproducibility of the measurements (Martins & Natri, 2014). To minimize this effect, however, the observers independently chose one of the sweeps from each volume dataset; therefore, not all measurements were performed in the same sweep.

Our results on intra- and interobserver agreement of intracranial and cerebellar volume measurement are largely in concordance with the results described by Benavides-Serralde et al. (2009), who reported an intraobserver ICC of 0.97 (95% CI 0.95–0.98) and an interobserver ICC of 0.97 (95% CI 0.92–0.98) for intracranial volume and ICCs of 0.76 (95% CI 0.54–0.89) and 0.69 (95% CI 0.44–0.75), respectively, for cerebellar volume. However, others achieved higher rates of agreement for measurement of the cerebellar volume (intraobserver ICC 0.98, 95% CI 0.84–0.99; interobserver ICC 0.97, 95% CI 0.95–0.98), which might be explained by a much lower percentage (30.9%) of cerebellar volume measurements in their acquired volumes as compared to our study (84.4%) (Rutten et al., 2009).

The overall agreement for measurements of the thalami was rather disappointing. However, these results are in concordance with the results of previous studies that determined intra- and interobserver agreement for thalamic volumetry (Benavides-Serralde et al., 2009; Sotiriadis et al., 2012). The thalami have a very similar echogenicity to their surrounding tissue which makes it difficult to clearly distinguish them (Sotiriadis et al., 2012). This difficulty in the distinction of the thalami in ultrasound images may explain the only moderate reliability of thalamic measurements.

For this study, our aim was to investigate only the reproducibility of the volume measurements of structures for which methods and tracing boundaries were previously described, to determine whether it was feasible for us to perform these measurements reliably. Therefore, we measured the intracranial volume and not actual brain volume, and we measured cerebellar and thalamic volume and not the volumes of other brain structures, such as the hippocampus or the ventricles. Now that we know that it is feasible to measure intracranial and cerebellar volume with the VOCAL technique, it could be interesting to explore whether it is possible to measure other brain structures with VOCAL as well.

In the near future, we will measure intracranial volume and cerebellar volume in all ultrasound datasets that will be acquired in YOUth around 20 and 30 weeks gestational age, and assess whether these volumes and the amount of increase in these volumes during pregnancy are influenced by environmental factors, such as maternal smoking behavior, maternal nutrition, and maternal stress. We will also investigate if the intracranial and cerebellar volumes are related to cognitive and behavioral outcomes in children. By measuring intracranial volume and cerebellar volume prenatally in YOUth, we expect to gain detailed insight into the development of the fetal brain and its association with cognitive and behavioral development of the child, and to gain more knowledge about the prenatal factors influencing prenatal brain development.

In conclusion, intra- and interobserver agreement for intracranial volume and cerebellar volume measurement with VOCAL was almost perfect and similar to agreement for the 2D-measurements of these

structures. There is no gain in precision between three versus one volume measurements, suggesting that it is sufficient to measure volumes once. Agreement for thalamic measurements was only moderate; therefore, we do not recommend using these measurements for evaluation of fetal brain development.

ACKNOWLEDGMENTS

The YOUth cohort is part of the Consortium on Individual Development (CID), which is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003). (<https://www.uu.nl/en/research/dynamics-of-youth/youth>)

ORCID

Maria E.W.A. Albers <http://orcid.org/0000-0001-8345-7759>

REFERENCES

- Araujo Júnior, E., Pires, C. R., Nardozza, L. M. M., Filho, H. A. G., & Moron, A. F. (2007). Correlation of the fetal cerebellar volume with other fetal growth indices by three-dimensional ultrasound. *Journal of Maternal-Fetal & Neonatal Medicine*, 20(8), 581–587. <http://www.ncbi.nlm.nih.gov/pubmed/17674275>.
- Barker, D. J. P. (1998). In utero programming of chronic disease. *Clinical Science (London, England : 1979)*, 95(2), 115. <http://cs.portlandpress.com/cs/095/cs0950115.htm>.
- Benavides-Serralde, A., Hernandez-Andrade, E., Fernandez-Delgado, J., Plasencia, W., Scheier, M., Crispi, F., ... Gratacos, E. (2009). Three-dimensional sonographic calculation of the volume of intracranial structures in growth-restricted and appropriate-for-gestational age fetuses. *Ultrasound in Obstetrics & Gynecology*, 33(5), 530–537.
- Bland, J. M., & Altman, D. G. (2003). Applying the right statistics: Analyses of measurement studies. *Ultrasound in Obstetrics & Gynecology*, 22(1), 85–93.
- Bordes, A., Bory, A. M., Benchaïb, M., Rudigoz, R. C., & Salle, B. (2002). Reproducibility of transvaginal three-dimensional endometrial volume measurements with virtual organ computer-aided analysis (VOCAL) during ovarian stimulation. *Ultrasound in Obstetrics & Gynecology*, 19(1), 76–80.
- Bornstein, E., Monteagudo, A., Santos, R., Keeler, S. M., & Timor-Tritsch, I. E. (2010a). A systematic technique using 3-dimensional ultrasound provides a simple and reproducible mode to evaluate the corpus callosum. *American Journal of Obstetrics and Gynecology*, 202(2), 201.e1–201.e5. <https://doi.org/10.1016/j.ajog.2009.10.705>.
- Bornstein, E., Monteagudo, A., Santos, R., Strock, I., Tsymbal, T., Lenchner, E., & Timor-Tritsch, I. E. (2010b). Basic as well as detailed neurosonograms can be performed by offline analysis of three-dimensional fetal brain volumes. *Ultrasound in Obstetrics & Gynecology*, 36, 20–25.
- Correa, F. F., Lara, C., Bellver, J., Remohí, J., Pellicer, A., & Serra, V. (2006). Examination of the fetal brain by transabdominal three-dimensional ultrasound: Potential for routine neurosonographic studies. *Ultrasound in Obstetrics & Gynecology*, 27(5), 503–508.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed., pp. 38–46). New York, NY: John Wiley.
- Gillman, M. W. (2005). Developmental origins of health and disease. *New England Journal of Medicine*, 353(17), 1848–1850.

- International Society of Ultrasound in Obstetrics & Gynecology Education Committee. (2007). Sonographic examination of the fetal central nervous system: Guidelines for performing the "basic examination" and the "fetal neurosonogram. *Ultrasound in Obstetrics & Gynecology*, 29, 109–116.
- Ioannou, C., Sarris, I., Salomon, L. J., & Papageorghiou, A. T. (2011). A review of fetal volumetry: The need for standardization and definitions in measurement methodology. *Ultrasound in Obstetrics & Gynecology*, 38(6), 613–619.
- Martins, W., & Nastri, C. (2014). Interpreting reproducibility results for ultrasound measurements. *Ultrasound in Obstetrics & Gynecology*, 43(4), 479–480. 7
- Monteagudo, A., & Timor-Tritsch, I. E. (2012). Fetal CNS scanning—Less of a headache than you think. *Clinical Obstetrics and Gynecology*, 55(1), 249–265. <http://www.ncbi.nlm.nih.gov/pubmed/22343242>.
- Monteagudo, A., Timor-Tritsch, I. E., & Mayberry, P. (2000). Three-dimensional transvaginal neurosonography of the fetal brain: "Navigating" in the volume scan. *Ultrasound in Obstetrics & Gynecology*, 16(4), 307–313.
- Pilu, G., Ghi, T., Carletti, A., Segata, M., Perolo, A., & Rizzo, N. (2007). Three-dimensional ultrasound examination of the fetal central nervous system. *Ultrasound in Obstetrics & Gynecology*, 30(2), 233–245.
- Pilu, G., Segata, M., Ghi, T., Carletti, A., Perolo, A., Santini, D., . . . Rizzo, N. (2006). Diagnosis of midline anomalies of the fetal brain with the three-dimensional median view. *Ultrasound in Obstetrics & Gynecology*, 27(5), 522–529.
- Pooh, R. K. (2012). Normal anatomy by three-dimensional ultrasound in the second and third trimesters. *Seminars in Fetal and Neonatal Medicine*, 17(5), 269–277.
- Raine-Fenning, N. J., Clewes, J. S., Kendall, N. R., Bunkheila, A. K., Campbell, B. K., & Johnson, I. R. (2003). The interobserver reliability and validity of volume calculation from three-dimensional ultrasound datasets in the in vitro setting. *Ultrasound in Obstetrics & Gynecology*, 21(3), 283–291.
- Rizzo, G., Pietrolucci, M. E., Mammarella, S., Dijnelli, E., Bosi, C., & Arduini, D. (2012). Assessment of cerebellar vermis biometry at 18–32 weeks of gestation by three-dimensional ultrasound examination. *Journal of Maternal-Fetal & Neonatal Medicine*, 25(5), 519–522.
- Rutten, M. J., Pistorius, L. R., Mulder, E. J. H., Stoutenbeek, P., de Vries, L. S., & Visser, G. H. A. (2009). Fetal cerebellar volume and symmetry on 3-D ultrasound: Volume measurement with multiplanar and vocal techniques. *Ultrasound in Medicine & Biology*, 35(8), 1284–1289.
- Salman, M. M., Twining, P., Mousa, H., James, D., Momtaz, M., Aboulghar, M., . . . Bugg, G. J. (2011). Evaluation of offline analysis of archived three-dimensional volume datasets in the diagnosis of fetal brain abnormalities. *Ultrasound in Obstetrics & Gynecology*, 38(2), 165–169.
- Schlutz, W., Jones, A., Godfrey, K. M., & Phillips, D. I. W. (2008). Effortful control mediates associations of fetal growth with hyperactivity and behavioural problems in 7- to 9-year-old children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 49, 1228–1236.
- Schlutz, W., & Phillips, D. I. W. (2009). Fetal origins of mental health: Evidence and mechanisms. *Brain, Behavior, and Immunity*, 23(7), 905–916.
- Sotiriadis, A., Dimitrakopoulos, I., Eleftheriades, M., Agorastos, T., & Makrydimas, G. (2012). Thalamic volume measurement in normal fetuses using three-dimensional sonography. *Journal of Clinical Ultrasound*, 40(4), 207–213.
- Tonni, G., Martins, W. P., Guimarães Filho, H., & Araujo Júnior, E. (2015). Role of 3-D ultrasound in clinical obstetric practice: Evolution over 20 years. *Ultrasound in Medicine & Biology*, 41(5), 1180–1211.
- Vinals, F., Muñoz, M., Naveas, R., & Giuliano, A. (2007). Transfrontal three-dimensional visualization of midline cerebral structures. *Ultrasound in Obstetrics & Gynecology*, 30(2), 162–168.
- Walter, S. D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in Medicine*, 17(1), 101–110.
- Zeng, S., Zhou, Q. C., Zhou, J. W., Li, M., Long, C., & Peng, Q. H. (2015). Volume of intracranial structures on three-dimensional ultrasound in fetuses with congenital heart disease. *Ultrasound in Obstetrics & Gynecology*, 46(2), 174–181.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Albers MEWA, Buisman ETIA, Kahn RS, Franx A, Onland-Moret NC, de Heus R. Intra- and interobserver agreement for fetal cerebral measurements in 3D-ultrasonography. *Hum Brain Mapp*. 2018;39:3277–3284. <https://doi.org/10.1002/hbm.24076>