



# Fetal Brain Biometric Measurements on 3D Super-Resolution Reconstructed T2-Weighted MRI: An Intra- and Inter-observer Agreement Study

Marie Khawam<sup>1†</sup>, Priscille de Dumast<sup>1,2\*†</sup>, Pierre Deman<sup>1,2</sup>, Hamza Kebiri<sup>1,2</sup>, Thomas Yu<sup>2,3</sup>, Sébastien Tourbier<sup>1</sup>, Hélène Lajous<sup>1,2</sup>, Patric Hagmann<sup>1</sup>, Philippe Maeder<sup>1</sup>, Jean-Philippe Thiran<sup>1,2,3</sup>, Reto Meuli<sup>1</sup>, Vincent Dunet<sup>1</sup>, Meritxell Bach Cuadra<sup>1,2</sup> and Mériam Koob<sup>1</sup>

## OPEN ACCESS

### Edited by:

Kirsten A. Donald,  
University of Cape Town, South Africa

### Reviewed by:

Angelo Lavano,  
University of Magna Graecia, Italy  
Gemma Piella,  
Pompeu Fabra University, Spain

### \*Correspondence:

Priscille de Dumast  
priscille.guerrierdedumast@unil.ch

<sup>†</sup>These authors have contributed  
equally to this work and share  
first authorship

### Specialty section:

This article was submitted to  
Pediatric Neurology,  
a section of the journal  
Frontiers in Pediatrics

**Received:** 09 December 2020

**Accepted:** 07 June 2021

**Published:** 10 August 2021

### Citation:

Khawam M, de Dumast P, Deman P, Kebiri H, Yu T, Tourbier S, Lajous H, Hagmann P, Maeder P, Thiran J-P, Meuli R, Dunet V, Bach Cuadra M and Koob M (2021) Fetal Brain Biometric Measurements on 3D Super-Resolution Reconstructed T2-Weighted MRI: An Intra- and Inter-observer Agreement Study. *Front. Pediatr.* 9:639746. doi: 10.3389/fped.2021.639746

<sup>1</sup> Department of Radiology, Lausanne University Hospital, University of Lausanne (CHUV-UNIL), Lausanne, Switzerland,

<sup>2</sup> CIBM Center for Biomedical Imaging, Lausanne, Switzerland, <sup>3</sup> Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

We present the comparison of two-dimensional (2D) fetal brain biometry on magnetic resonance (MR) images using orthogonal 2D T2-weighted sequences (T2WSs) vs. one 3D super-resolution (SR) reconstructed volume and evaluation of the level of confidence and concordance between an experienced pediatric radiologist (obs1) and a junior radiologist (obs2). Twenty-five normal fetal brain MRI scans (18–34 weeks of gestation) including orthogonal 3-mm-thick T2WSs were analyzed retrospectively. One 3D SR volume was reconstructed per subject based on multiple series of T2WSs. The two observers performed 11 2D biometric measurements (specifying their level of confidence) on T2WS and SR volumes. Measurements were compared using the paired Wilcoxon rank sum test between observers for each dataset (T2WS and SR) and between T2WS and SR for each observer. Bland–Altman plots were used to assess the agreement between each pair of measurements. Measurements were made with low confidence in three subjects by obs1 and in 11 subjects by obs2 (mostly concerning the length of the corpus callosum on T2WS). Inter-rater intra-dataset comparisons showed no significant difference ( $p > 0.05$ ), except for brain axial biparietal diameter (BIP) on T2WS and for brain and skull coronal BIP and coronal transverse cerebellar diameter (DTC) on SR. None of them remained significant after correction for multiple comparisons. Inter-dataset intra-rater comparisons showed statistical differences in brain axial and coronal BIP for both observers, skull coronal BIP for obs1, and axial and coronal DTC for obs2. After correction for multiple comparisons, only axial brain BIP remained significantly different, but differences were small ( $2.95 \pm 1.73$  mm). SR allows similar fetal brain biometry as compared to using the conventional T2WS while improving the level of confidence in the measurements and using a single reconstructed volume.

**Keywords:** fetal, brain, biometry, magnetic resonance imaging, sequences, super-resolution

## INTRODUCTION

Biometric measurements are good markers of fetal brain maturation and growth and are a fundamental basis for the diagnosis of developmental and acquired brain abnormalities (1). Indeed, an abnormal measurement is often the first warning of disturbed fetal growth that requires further investigation. Quantifying brain development, in comparison to reference charts, is the first routine step of prenatal diagnosis on ultrasound (US) and magnetic resonance imaging (MRI). This can detect common pathologies like microcephaly, cerebellar hypoplasia, corpus callosum dysgenesis, and ventriculomegaly (2). The accuracy of biometric data is essential for the evaluation of prognosis and prediction of outcome as it may influence both prenatal and postnatal management. Indeed, it is crucial to give appropriate parental information and counseling, as termination of pregnancy may be considered in severe cases (3).

US is the first screening fetal brain imaging modality. MRI comes as a complementary imaging tool to confirm or rule out pathological US findings and to look for additional information (2). Fetal biometry performed on US may be influenced by maternal morphology, fetal position, or interference from the skull. MRI does not share these limitations and, in contrast to US, provides more accurate measurements of the fetal brain, together with better parenchymal signal and gyration analysis. Moreover, although some studies have demonstrated a good concordance between both techniques for biometry (4–6), others have shown discrepancies (7). For example, MRI has been found to be more accurate for the definitive diagnosis of microcephaly (8, 9), as preliminary US diagnoses were ruled out in 8 out of 30 cases after subsequent MRI analysis. This is valuable, as a small brain size is considered a risk of poor neurodevelopmental outcome. Similar conclusions were drawn from studies on the corpus callosum or the vermis (10, 11).

In clinical practice, fetal brain MRI biometry is performed on series of two-dimensional (2D) 3- to 4-mm-thick slices acquired in three orthogonal planes by T2-weighted sequences (T2WSs). These T2WS-based measurements are then compared to reference values (4, 12–15). In two large-scale studies, a good inter-rater agreement on T2WS was found for most brain biometric measurements (4, 16). However, a key limitation of fetal MRI is the maternal and fetal motion, which often results in inter-slice and inter-series motion as well as intensity artifacts. The acquisition plane may be oblique, thus not accurately oriented in the reference orthogonal planes, and motion may occur between acquisition planes, which makes identification of anatomical landmarks difficult and hampers subsequent measurement. Therefore, fetal brain MRI biometric analysis based on T2WS is commonly performed on low-quality images with low confidence and uncertainty. The level of confidence is another important parameter in the process of image analysis (5), as it corresponds to the “degree of certainty in the correctness of diagnosis” and is associated with relevant decisional consequences (17, 18). An imaging method where measurements would be easier to perform could help in providing accurate biometry with more confidence, particularly on suboptimal images and for less experienced radiologists.

The recent developments of advanced image processing methods based on super-resolution (SR) techniques allow the reconstruction of a 3D high-resolution motion- and inhomogeneity-free volume from T2WSs (19–26). Handling a 3D volume with isotropic spatial resolution could be of great value to conduct fetal brain biometry, as it would allow to freely navigate in any plane and to obtain SR orthogonal planes easily (27–30), as well as automated quantitative studies (31) of high-resolution 3D images. As depicted in **Figure 1A**, fetal brain SR reconstruction aims at combining orthogonal low-resolution series of thick 2D slices as acquired clinically into a single isotropic high-resolution image free of motion artifacts (the SR-reconstructed volume) that enables the visualization of fetal brain anatomy in any plane.

This study aimed at comparing 2D measurements of fetal brain biometry using three orthogonal 2D low-resolution T2WSs vs. using one single 3D SR-reconstructed volume for both supratentorial and infratentorial measurements. Three points were evaluated: the agreement between T2WS and SR measurements, the level of the observers’ confidence on both datasets, and the concordance between a junior radiologist and an experienced pediatric radiologist.

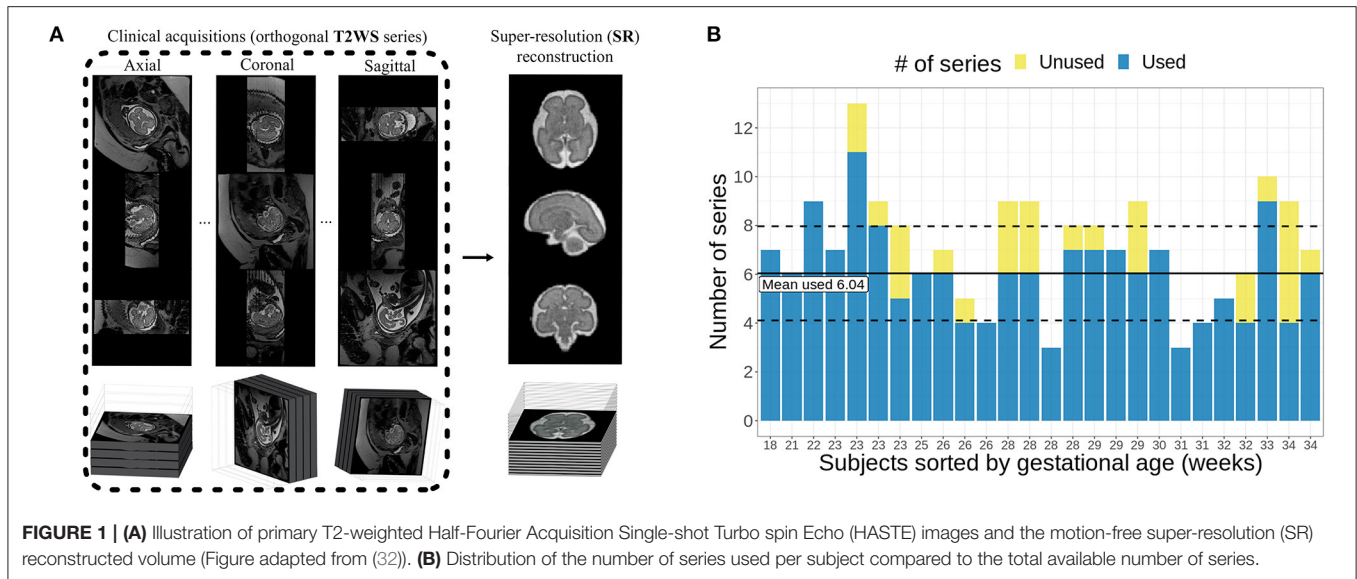
## MATERIALS AND METHODS

### Dataset Cohort

We retrospectively collected all consecutive normal fetal brain MRI exams from January 2013 to October 2018 (28 subjects in total: 15 males and 13 females) from the MRI database of our institution. All MRI scans were conducted on medical indication (ventriculomegaly, suspicion of corpus callosum or posterior fossa anomaly, microcephaly...) within 2 weeks of an expert fetal neurosonographic study and were finally considered normal. Early neonatal clinical evaluation was normal. All images were anonymized prior to further analysis. This retrospective study was part of a larger research protocol at our institution approved by the local ethics committee.

### MRI

Clinical MR images were acquired either at 1.5 T (MAGNETOM Aera, Siemens Healthcare, Erlangen, Germany) (88% of the subjects) or at 3 T (MAGNETOM Skyra<sup>fit</sup>, Siemens Healthcare, Erlangen, Germany) (12% of the subjects). The fetal brain MRI protocol included T2-weighted (T2W) Half-Fourier Acquisition Single-shot Turbo spin Echo (HASTE) sequences in the three orthogonal orientations; usually at least two acquisitions were performed in each orientation, together with axial gradient echo T1-weighted and diffusion-weighted imaging (DWI) or diffusion tensor imaging (DTI) in some cases. The coronal plane was parallel to the brain stem, and the axial plane was parallel to the corpus callosum long axis (4). We excluded twins from the study ( $n = 2$ ). At this point, 26 normal fetal brain MR images were kept for further analysis. Details on the MRI acquisition parameters can be found in **Table 1**.



**TABLE 1 |** MRI acquisition parameters of the T2-weighted Half-Fourier Acquisition Single-shot Turbo spin Echo (HASTE) sequences.

Field strength (Tesla)	Number of exams	Gestational age (weeks)	Number of series	In-plane resolution (mm)		Slice thickness (mm)		Echo time (ms)		Repetition time (ms)	
				Min-Max	Mean ± SD	Min-max	Mean ± SD	Min-Max	Mean ± SD	Min-Max	Mean ± SD
1.5T	23	18–34 27.4 ± 4.2	143	1.125–1.172	1.127 ± 0.009	2.42–4.0	3.34 ± 0.20	82–98	90 ± 1.9	832–1,200	1,185 ± 59
3T	3		14	0.547	–	3.0	–	101	–	1,000–1,100	1,090 ± 027

## Methodology

### Three-Dimensional Super-Resolution Reconstruction

In our study, 3D SR volumes were reconstructed from T2WSs within the PACS station with an in-house *syngo.via* Frontier fetal MRI prototype (33) based on the publicly available MIALSRtoolkit software (20, 34). In a nutshell, the fetal brain MRI prototype performs the following steps: fetal brain extraction, bias field correction, inter-slice motion estimation based on slice-to-volume rigid registration, and SR reconstruction based on a forward model of the image acquisition process that leads to an inverse problem that is solved *via* an efficient total variation algorithm with automatic regularization (20).

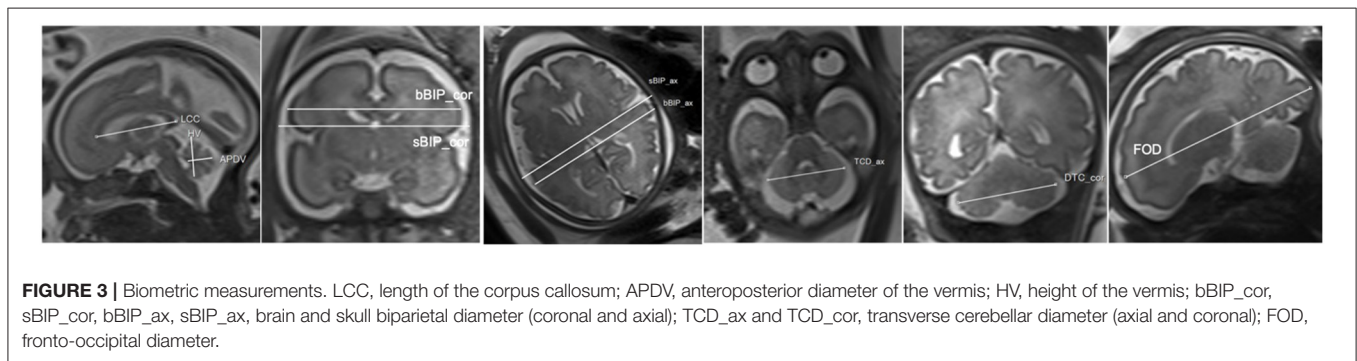
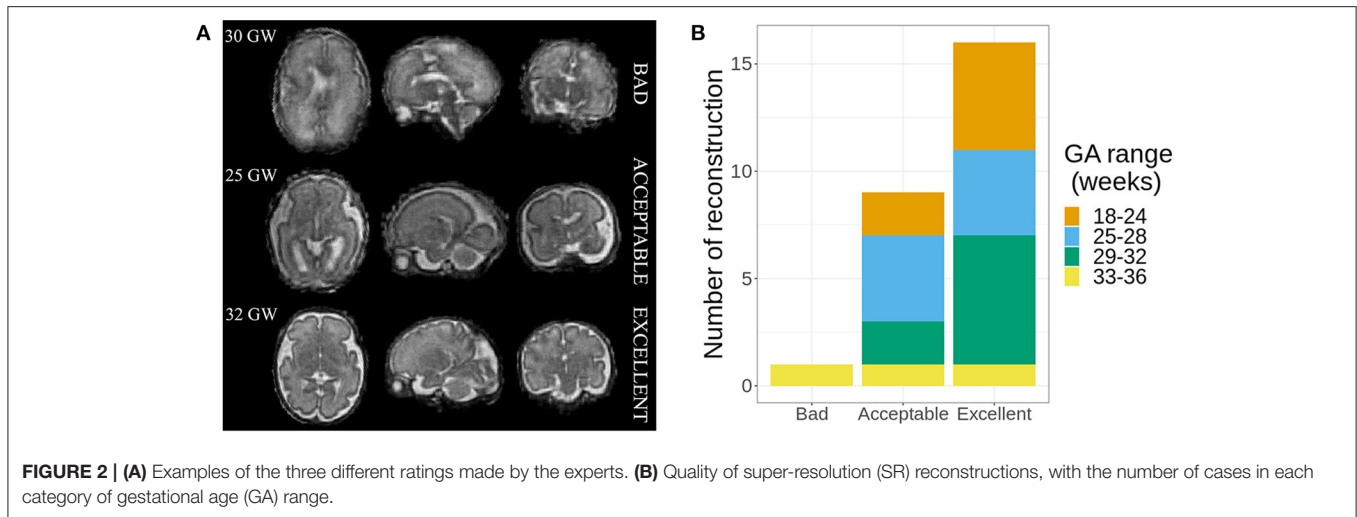
All cases were reconstructed by an engineer with 20 years of experience in medical image processing. The selection of the series used for the reconstruction was done based on visual inspection, and T2WSs that exhibited high levels of motion distortion and/or intensity signal dropout (thus, that were not exploitable for radiological reading neither) were excluded from the 3D SR reconstruction process. On average, six series were used per subject for SR reconstruction, with a range from 3 to 11 series (Figure 1B). All 3D images were reconstructed with an isotropic spatial resolution matching its input in-plane resolution (in average of around  $1.1 \times 1.1 \times 1.1 \text{ mm}^3$ ).

### Evaluation of Super-Resolution Quality

The quality of the 3D SR was assessed independently, in a blinded protocol, by three expert raters: one engineer expert in MR image analysis (rater 1, the same who performed the SR reconstructions) and two experienced pediatric radiologists (rater 2 and rater 3, with, respectively, 15 and 9 years of experience in fetal brain MRI). Reconstructed volumes were classified into one of three categories: bad (with remaining motion, very blurred, unusable for diagnosis purposes), acceptable (overall good quality with some blurring but still relevant for diagnosis purposes), or excellent (good quality without any blurring). Examples of SR-reconstructed volumes rated with the three different quality measures are shown in Figure 2A.

### Biometric Measurements

Using standard tools on the PACS station (Carestream Vue PACS®, Version 12.1.6.1005, Carestream Health Inc., NY, USA), two observers, one experienced pediatric neuroradiologist (obs1, 15 years of experience of fetal brain MRI) and one inexperienced radiologist (obs2, without any experience in fetal brain MRI) independently measured 11 fetal brain structures on both datasets (T2WS then SR) during two reading sessions separated by 3 weeks and blinded to the results of the previous measurements.



The following biometric measurements as shown in **Figure 3** were performed following previously published guidelines for fetal brain MRI biometry (2, 4, 12, 28).

The skull coronal biparietal diameter (sBIP\_cor) is defined as the greatest transversal diameter between the inner tables of parietal bones on a coronal slice through the temporal horns of the lateral ventricles. The brain coronal biparietal diameter (bBIP\_cor) was measured on the same slice.

The brain axial biparietal diameter (bBIP\_ax) is defined as the maximal brain diameter in the transverse plane through the atria. The skull axial biparietal diameter (sBIP\_ax) is defined as the inner to inner table maximal skull diameter in the transverse plane through the atria.

The corpus callosum length (LCC), height of the vermis (HV), and anteroposterior vermis diameter (APDV) were measured in the mid-sagittal plane.

The transverse cerebellar diameter was measured on a coronal slice (TCD\_cor) and on an axial slice (TCD\_ax).

The fronto-occipital diameters [right fronto-occipital diameter (rFOD) and left fronto-occipital diameter (lFOD)] were measured on a sagittal slice between the extreme points of the frontal and occipital cortices.

On T2WS, each observer independently chose the best-quality T2W series for each measurement. SR measurements were performed in orthogonal planes in multiplanar reformations

(MPRs) using the tools within the PACS system. Each observer rated the confidence of each of his measurements in both T2WS and SR datasets either as high or low.

### Statistical Analysis

Statistical analysis was conducted with R software (version 3.6.3).

#### Evaluation of Super-Resolution Quality

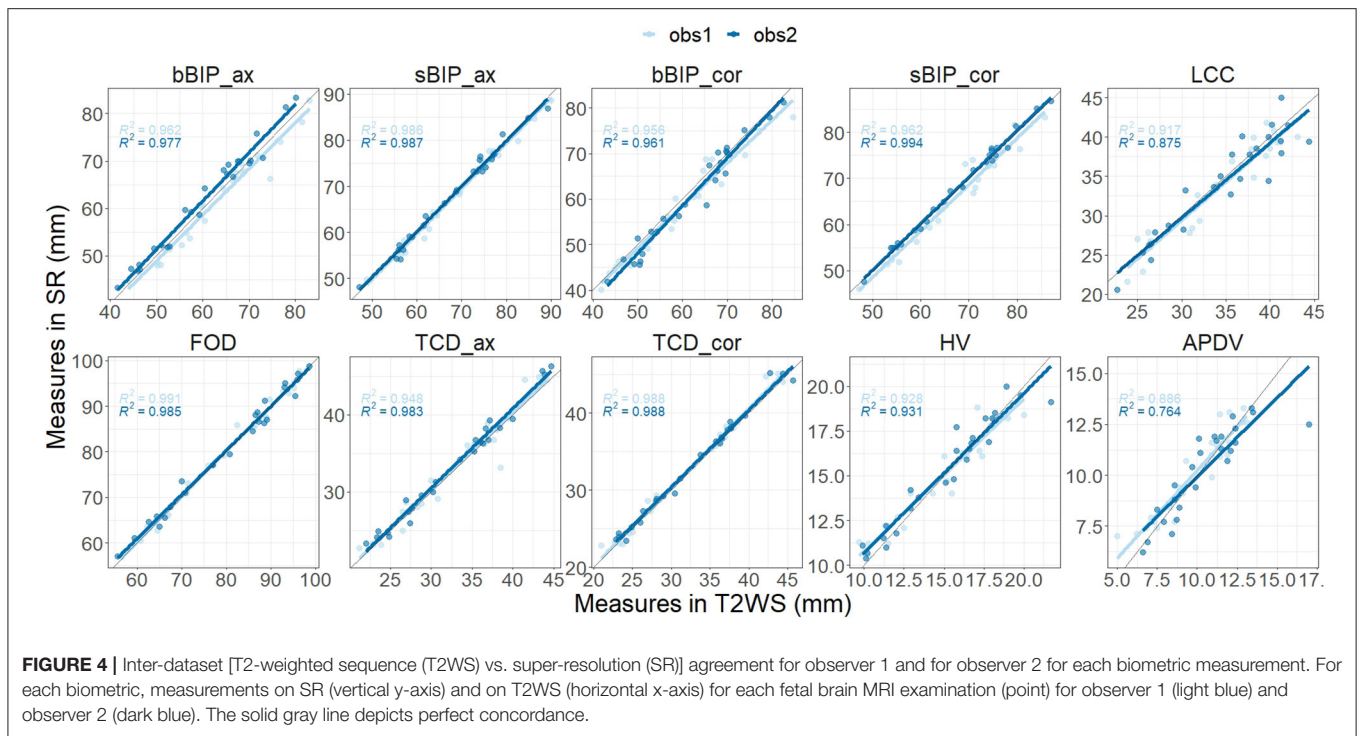
The inter-rater reliability was measured using a weighted ordinal Gwet's agreement coefficient (Gwet's AC) and interpreted according to Altman's benchmarking scale (35).

#### Level of Confidence

Chi-square test statistics were used to evaluate the dependence of the level of confidence of the raters on each dataset.

#### Biometric Measurements

The association of paired lFOD and rFOD measurements was tested using Spearman's  $\rho$  statistic (36), and the difference was tested with the paired Wilcoxon rank sum test. This analysis was performed for each pair of dataset-observer. Biometric measurements were compared statistically to determine inter-dataset and inter-observer significant difference, respectively, for each observer (obs1 and obs2) or dataset (T2WS and SR), with the paired Wilcoxon rank sum test (without and with Bonferroni multiple comparisons correction). Lin's concordance correlation



**FIGURE 4 |** Inter-dataset [T2-weighted sequence (T2WS) vs. super-resolution (SR)] agreement for observer 1 and for observer 2 for each biometric measurement. For each biometric, measurements on SR (vertical y-axis) and on T2WS (horizontal x-axis) for each fetal brain MRI examination (point) for observer 1 (light blue) and observer 2 (dark blue). The solid gray line depicts perfect concordance.

coefficient (CCC) (37, 38) and intraclass correlation coefficient (ICC) were computed for the agreement on measurements obtained on the two datasets (39). Bland–Altman plots were used to assess the agreement between the two observers for each of the two datasets and the reliability between the two datasets for each of the two observers. Agreement was rated as follows: poor, <0.5; moderate, 0.5–0.75; good, 0.75–0.9; excellent, >0.90 (35). Agreement between transverse and coronal measurements was evaluated for brain (bBIP\_ax, bBIP\_cor) and skull (sBIP\_ax, sBIP\_cor) biparietal diameters and for transverse cerebellar diameter (TCD\_ax, TCD\_cor) using ICC, error rate, and paired Wilcoxon rank sum test. The *p*-value level for statistical significance was set at 0.05.

## RESULTS

### Evaluation of Three-Dimensional Super-Resolution Reconstructions

A total of 26 cases were reconstructed based on the SR approach described above.

The estimated ordinal Gwet’s AC between the three raters was 0.85 with a standard error of 0.06. According to Altman’s benchmarking scale, this estimated coefficient was considered to be either *Good* or *Very Good* with a 0.99 probability.

As the inter-rater agreement was good, we considered the quality of a reconstruction as the averaged consensus between the three raters’ assessment. On average, experts rated one case as bad, nine as average, and 16 as excellent (**Figure 2B**). No significant differences were found regarding the quality of the SR reconstruction and the gestational age (GA) ranges. The case

rated as bad<sup>1</sup> was discarded for further analysis; thus, 25 normal fetal brain MRI (mean GA: 27.1 ± 4.2 weeks, range: 18–34 weeks) were considered for the biometric analysis.

### Confidence of Measurements

On T2WS, some measurements were made with low confidence in three out of 25 fetuses (12%) by obs1, while obs2 reported a low level of confidence in 13 out of 25 fetuses (52%). Specifically, obs1 had low confidence in FOD in two fetuses and in axial measurements in DTC, sBIP, and bBIP in another fetus. In contrast, obs2 had low confidence mostly on the LCC (in 11 fetuses) and also in FOD and axial DTC (in two fetuses). On SR, low confidence measures were made only on the LCC in two fetuses by obs2. All the remaining measurements made on SR and T2WS were rated with high confidence. Overall, the level of confidence of obs2 was dependent on the dataset used, either T2WS or SR (*p* = 0.002), with higher confidence using SR. Conversely, no significant difference was found in the confidence level for obs1 (*p* = 0.23).

### Biometric Measurements Analysis

Each observer–dataset pair showed a high correlation between right and left FOD (Spearman’s rank correlation  $\rho = 0.975, p < 0.001$ ) and no significant difference (*p* = 0.8). Hence, rFOD and lFOD were averaged. Agreement between T2WS and SR for each biometric parameter was good for both observers, with minimum Lin’s CCC estimated to be 0.86. Both Lin’s CCC and ICC were on average 0.997 for both raters (**Figure 4; Table 2**).

<sup>1</sup>Bad SR quality was primarily due to only four series were available and brain was cut in two of them.

**TABLE 2** | Lin's concordance correlation coefficient values.

	Obs1	Obs2
bBIP_ax	0.972	0.977
sBIP_ax	0.992	0.994
bBIP_cor	0.969	0.971
sBIP_cor	0.973	0.996
LCC	0.951	0.932
FOD	0.996	0.992
TCD_ax	0.972	0.987
TCD_cor	0.992	0.992
HV	0.956	0.961
APDV	0.930	0.865

Obs1, obs2, observer 1 and observer 2; LCC, length of the corpus callosum; APDV, anteroposterior diameter of the vermis; HV, height of the vermis; bBIP\_cor, sBIP\_cor, bBIP\_ax, sBIP\_ax, brain and skull biparietal diameter (coronal and axial); TCD\_ax and TCD\_cor, transverse cerebellar diameter (axial and coronal); FOD, fronto-occipital diameter.

Inter-dataset (SR vs. T2WS) intra-observer (obs1 and obs2) comparisons (**Figures 5, 6**) showed statistical differences ( $p < 0.05$ ) for brain axial and coronal BIP (bBIP\_ax and bBIP\_cor) for both observers, skull coronal BIP (sBIP\_cor) for obs1, and axial and coronal TCD for obs2 (TCD\_ax, TCD\_cor). After correction for multiple comparisons, only axial brain BIP (bBIP\_ax) remained significantly different for both observers, but differences were small ( $2.95 \pm 1.73$  mm) (**Figure 6**). Overall, the inter-dataset average error rate in our study was 3.3%. Additionally, **Supplementary Table 3** shows a high intra-rater reliability for one observer on a sub-cohort of five fetuses.

Inter-observer intra-dataset comparisons showed no significant differences ( $p > 0.05$ ), except for brain axial biparietal diameter (bBIP\_ax) on T2WS and for brain and skull coronal BIP (bBIP\_cor, sBIP\_cor) and coronal transverse cerebellar diameter (TCD\_cor) on SR. Overall, differences remained small (**Figure 7**) and independent of the measured size (**Figure 8**). After correction for multiple comparisons, none of them remained significant.

We computed the agreement between transverse and coronal measurements within each dataset and for both observers (ICC, percentage of error range, mean  $\pm$  standard deviation of percentage error). Agreement was excellent without statistical differences ( $p > 0.05$ ) for TCD on T2WS (ICC = 0.983, 0.3–16%,  $2.9 \pm 2.6\%$ ) and SR (ICC = 0.997, 0–1%,  $1 \pm 1\%$ ), for bBIP on T2WS (ICC = 0.951, 0.3–14%,  $5.2 \pm 3.5\%$ ) and SR (ICC = 0.970, 0.1–15%,  $5 \pm 3.8\%$ ), and for sBIP on T2WS (ICC = 0.972, 0–15%,  $3.4 \pm 3.2\%$ ) and SR (ICC = 0.972, 0–11%,  $3.8 \pm 3.4\%$ ). Complete results are presented in **Supplementary Tables 1, 2**.

## DISCUSSION

### Summary of Contributions

Our study showed a sound agreement for biometric measurements performed on T2WS and on SR. Thus, a good-quality SR volume is as valuable as T2WS for fetal brain biometric assessment. Our results are in line with and

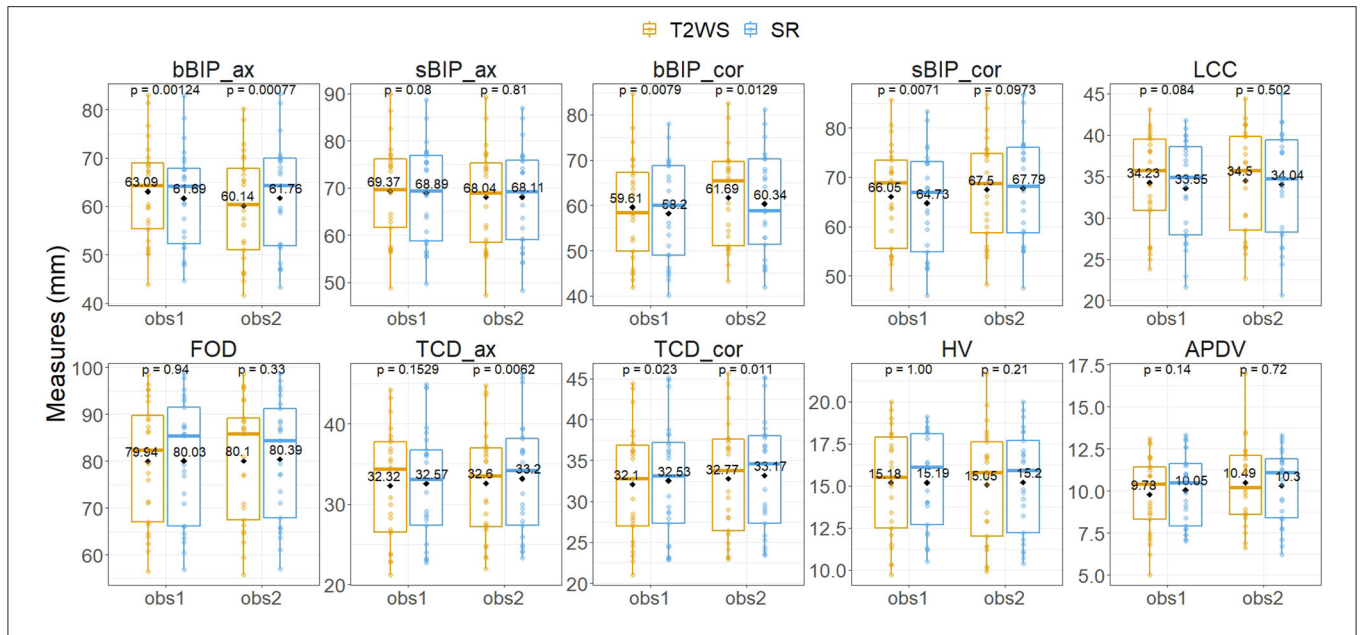
complement a previous study that has also validated the use of SR for posterior fossa (29). Our SR biometric measures are also concordant with the normative charts of the fetal brain that have been recently published (28).

The intra-observer agreement in 25 fetuses (both Lin's CCC and ICC were on average 0.997 for both observers) is in line with previous work (28) that also compared T2WS and SR measurements and reported an ICC between 0.95 and 0.99. Furthermore, the error rate is also of the same range (3.3% in average in our study as compared to 0.2–2.4%). However, let us note though that in the work by Kyriakopoulou et al. (28), the comparison of T2WS vs. SR measurements was performed in a controlled sub-cohort of 10 fetuses specifically selected with symmetrical non-rotated images.

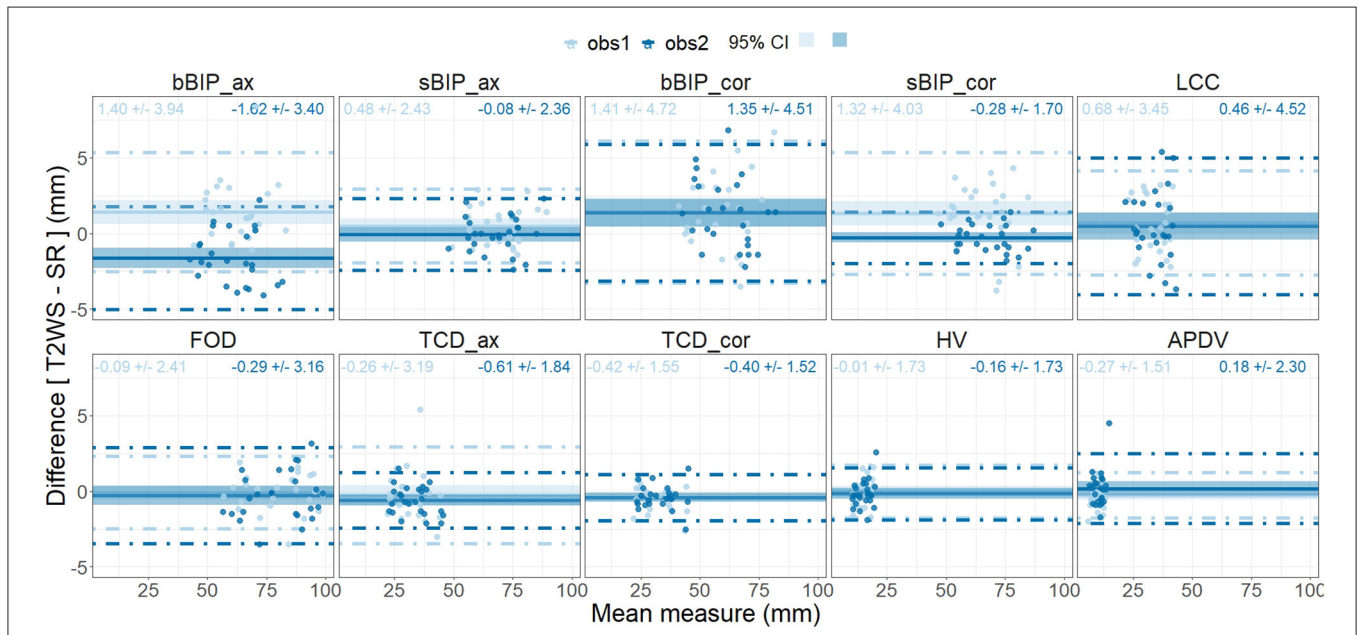
Intra-observer comparisons between T2WS and SR measurements showed significant differences only for axial brain BIP, though differences remained small ( $2.95 \pm 1.73$  mm) and acceptable in clinical practice, as age-specific reference intervals for this biometric parameter are larger (12). The discrepancies observed in our study for the axial brain BIP may be explained by a different plane used for SR and T2WS and the possibility offered by SR to change the windowing to more precisely identify anatomical landmarks. In contrast to US, there are no true standardized measurement criteria for the BIP on MRI; indeed, the chosen plane, axial or coronal, differs between authors. For instance, measurements of brain and skull BIP reference data were performed in the coronal plane through the temporal horns of the ventricles by some authors (4, 40), while it was made in the axial plane by others (28). We could question if the BIP obtained in the coronal plane really compares to the BIP obtained in the axial plane; however, our results still showed good agreement between transverse and coronal measurements for the BIP. A similar pattern was found for axial vs. coronal TCD, but with a shorter error interval for TCD in SR, similarly to another study (29) that also found superior concordance in SR rather than in T2WS measurement of TCD.

Our study also showed overall good inter-observer agreement for biometric measurements performed on T2WS and on SR, without any statistically significant difference after correction for multiple comparisons. This indicates that experience is not crucial for fetal brain biometric assessment and that SR can be used for biometry even by junior trainee radiologists without extensive expertise in fetal brain MRI. Indeed, obs2 did not report any fetal brain MRI before the study nor received special training. Obs1 showed obs2 how to perform the measurements on one subject and provided schematical guidelines similar to **Figure 3**. This is to be compared to recent work where the experience of the radiologist (number of previous fetal MRI exams reported) had an influence on the diagnostic error rate, with the less experienced radiologists having higher error rates of 11%, while experts had  $<3.8\%$  (41).

In practice, the main additional value of SR is the possibility to reorient the planes in the standard anatomical planes using MPR. In our study, the junior radiologist had more confidence in identifying and measuring the corpus callosum on SR compared to T2WS. Indeed, it may be difficult to visualize the whole



**FIGURE 5 |** Inter-dataset intra-observer comparison for each biometric measurement using the paired Wilcoxon signed rank test. Vertical y-axis, value in mm of each biometric measurement for each fetal brain exam (point) for observer 1 (obs1) and observer 2 (obs2) (horizontal x-axis) for T2-weighted sequence (T2WS; orange) and super-resolution (SR; blue). Solid horizontal lines depict the median; diamonds depict the average of measurement in mm. See **Figure 3** for measurement abbreviations.  $p < 0.05$  is considered significant.

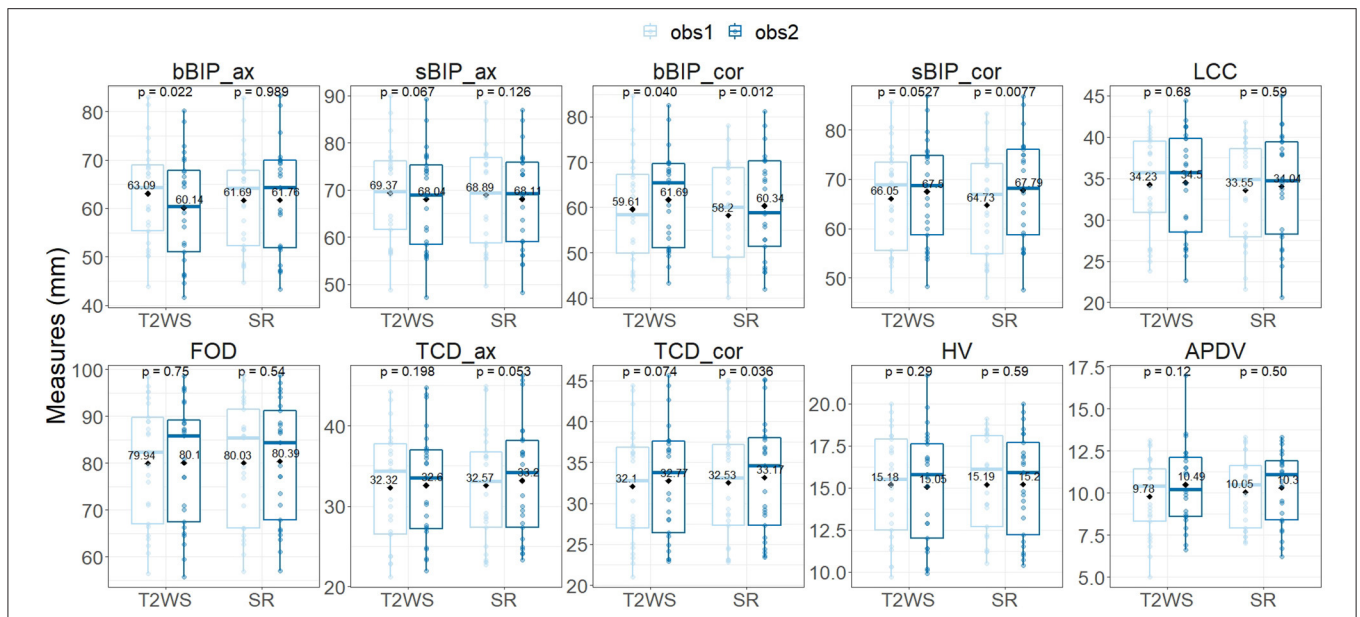


**FIGURE 6 |** Bland-Altman plots of measurement differences between datasets [inter-dataset, T2-weighted sequence (T2WS) vs. super-resolution (SR)] for observer 1 and for observer 2 (intra-observer) for each biometric measurement. Each point corresponds in the x-axis to the mean measure between datasets for each observer (light blue for observer 1, dark blue for observer 2) and on the vertical y-axis to the difference between the two measurements T2WS/SR in mm. Horizontal solid line = mean of all measurement differences. Dashed lines = 95% limits of agreement, and shadow areas correspond to the 95% confidence interval (CI).

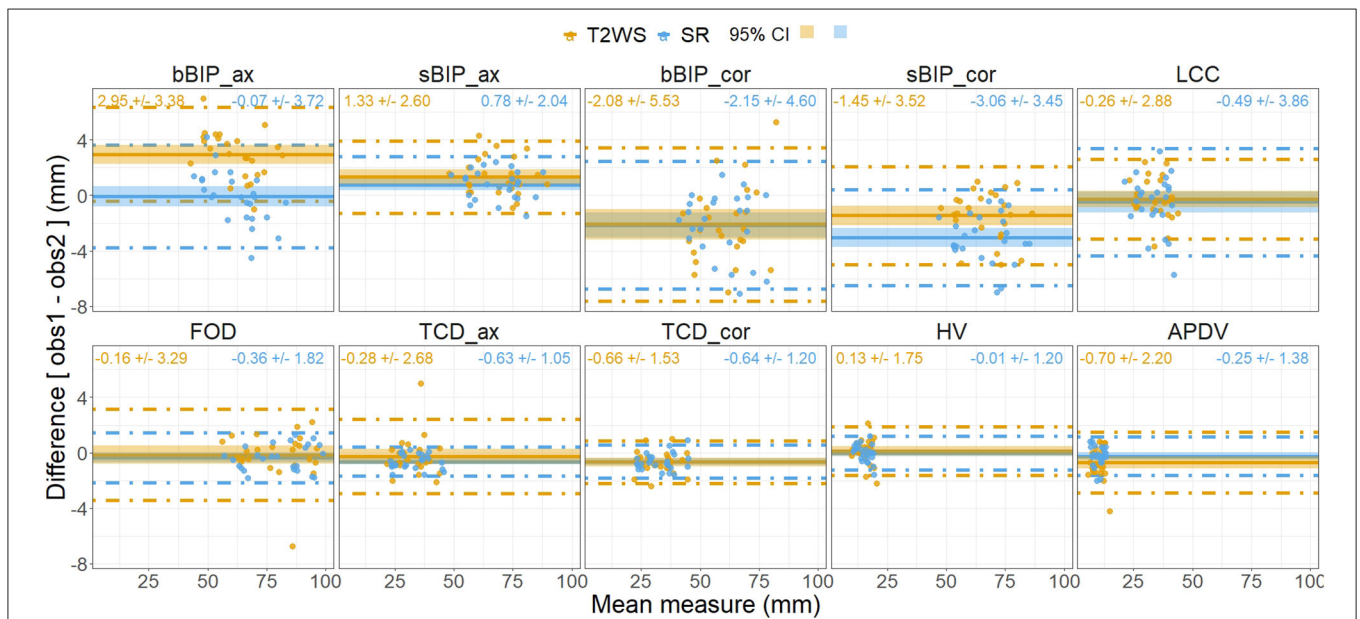
corpus callosum, particularly the rostrum, on conventional low-resolution 2D T2WS, and US is considered to have a better resolution in this case (42, 43). Our results suggest that SR makes it easier to identify the whole corpus callosum. The

potential benefit of SR in this indication should be explored on larger series.

However, the 3D SR reconstruction also has some limitations. Obviously, the quality of SR depends on the quality of the



**FIGURE 7 |** Inter-observer intra-dataset comparison for each biometric measurement using the paired Wilcoxon signed rank test. Vertical y-axis, measurement in mm for each biometric measurement for each fetal brain MRI exam (point) for T2-weighted sequence (T2WS) and super-resolution (SR) (horizontal x-axis) for observer 1 (light blue) and observer 2 (dark blue). Solid horizontal lines depict the median; diamonds depict the average of measurement in mm. See **Figure 3** for measurement abbreviations.  $p < 0.05$  is considered significant.



**FIGURE 8 |** Bland-Altman plots of measurement differences between observers (observer 1 vs. observer 2) for each dataset [intra-dataset, T2-weighted sequence (T2WS) and super-resolution (SR)] for each biometric measurement. Each point corresponds in the x-axis to the mean measure between observers and in the vertical y-axis to the difference in mm between the two observers on T2WS (orange point) and SR (blue point). Horizontal solid line = mean of all measurements for T2WS (blue) and SR (red); dashed lines = 95% limits of agreement, and shadow areas correspond to the 95% confidence interval (CI).

native T2WSs, which are frequently affected by motion. Indeed, due to the reconstruction process itself, small or narrow structures (e.g., optic chiasm or corpus callosum) can appear blurred due to partial volume effects (27). Nevertheless, in

our study, we obtained a good or excellent quality of SR images in 90% of cases, with a good visual rating concordance between three expert raters (one senior engineer and two senior radiologists). The fact that the concordance between T2WS and



SR measurements was high indicates that the process of SR reconstruction does not distort the fetal brain anatomy (29). Our success ratio in SR reconstruction is slightly higher than the one previously reported at 3T (29), where 79% of 62 cases were successfully reconstructed.

SR reconstruction requires a minimum number of orthogonal series to ensure a good reconstruction quality (44, 45). In our study, six series were used on average (range of 3–11) per subject's reconstruction with an average computing time of 1 h. Previous works reported a similar number of series and processing time [from 4 to 15, in average eight series, 1–20 h in Gholipour et al. (27); and eight series and 40 min in Kyriakopoulou et al. (28)]. The processing time, the need for user interaction for the series selection or the refinement of brain masks, and the lack of integration into the clinical environment preclude the use of SR volumes for fetal brain exploration in daily practice. However, current developments aim at automatizing and accelerating SR reconstruction (24, 25, 34) to facilitate its integration as a supporting tool in clinical routine and consequently also to be adopted by non-engineer users.

The clinical adoption of SR-reconstructed fetal MRI is thus at its earliest stage (46–49). In our center, SR-reconstructed MRI is used only for certain cases, in particular to confirm or reject suspected cortical abnormalities on 2D T2WS. In our opinion, for now, SR could be complementary to but cannot replace low-resolution 2D T2W planes for brain parenchyma analysis, as voxel intensities in SR volumes have undergone many changes. Indeed, beyond the promising value of SR fetal brain biometry, the intensity contrast of SR images still has to be evaluated for its diagnostic value in comparison to native T2WS and eventually improved. Let us recall that voxel intensities in the reconstructed SR volume are computed from the original voxel values within the multiple series throughout many image processing steps; therefore, they are not directly generated from the MRI scanner, and their interpretation as pathological features has to be done with caution.

## Strengths and Limitations

Our study is the first to compare both supratentorial and infratentorial measurements with SR, while only posterior fossa measurements were made in Pier et al. (29); this latter study included only normal volunteers, while our cases were taken from clinical workflow. We also compared SR measurements between a junior and a senior radiologist to evaluate the influence of experience on biometric results. We have shown that even a radiologist without expertise in fetal brain MRI can perform accurate fetal brain biometry—thanks to SR. In contrast, multiple raters who received extensive training developed normative charts for 2D T2WS and 3D SR fetal brain measurements (28). In their study, the agreement analysis was performed between measurements on native T2WS and SR in a sub-cohort of only 10 fetuses, with the same rater, and on selected symmetrical non-rotated images. Our study on 25 fetuses comparing two raters provides additional insight into the value of SR techniques for biometry of fetal brain MRI.

A limitation of our study is the small-size cohort, and so the few subjects available per gestational week. For this reason, we cannot draw conclusions about gestational age or sex influence on measurement quality. Finally, the reproducibility of our measurements were tested only with one observer on a sub-cohort of five fetuses.

## CONCLUSION

This study demonstrates that SR is a valid, reliable, and simple method for biometric measurements. SR measurements are concordant with T2WS measurements, even when conducted by non-expert radiologists. Some biometric measurements like the biparietal diameter show small statistically significant differences, which can be explained by poorly defined and standardized measurement criteria. As soon as full automatization of SR is available in the clinical environment, the use of 3D SR could initially complement conventional T2WSs by faster providing the reference planes and further facilitating biometric measurements. Afterward, SR could provide a new standard of measurements in real orthogonal planes more than mimicking ill-defined or incorrect 2D planes. Future studies on pathological cases will enable the evaluation of other potential benefits of SR in clinical practice.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: the ethical approval for the use of these data did not include public release. Requests to access these datasets should be directed to Meritxell Bach Cuadra, meritxell.bachcuadra@unil.ch.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Commission cantonale (VD) d'éthique de la recherche sur l'être humain (CER-VD). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

MKh, PDu, VD, MB, and MKo contributed to the conception and design of the study and wrote sections of the manuscript and the first draft of the manuscript. PH, PM, RM, and VD acquired the images and evaluated the subjects used in this study. MKh and MKo performed biometric measurements. PDu and VD performed statistical analysis. MB performed the image reconstruction. VD, MB, and MKo evaluated the images. PDe, ST, and MB contributed to the software development used in this study. HL, HK, TY, and J-PT contributed to the analysis and interpretation of the results. All authors contributed to manuscript revision and read and approved the submitted version.

## FUNDING

This work was supported by the Swiss National Science Foundation (project 205321-182602) and the Hasler Foundation (project 17029). We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), Ecole polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE) and Geneva University

## REFERENCES

- Guibaud L, Lacalm A. Diagnostic imaging tools to elucidate decreased cephalic biometry and fetal microcephaly: a systematic analysis of the central nervous system: editorial. *Ultrasound Obstet Gynecol.* (2016) 48:16–25. doi: 10.1002/uog.15926
- Prayer D, Malinger G, Brugger PC, Cassady C, De Catte L, De Keersmaecker B, et al. ISUOG practice guidelines: performance of fetal magnetic resonance imaging. *Ultrasound Obstet Gynecol.* (2017) 49:671–80. doi: 10.1002/uog.17412
- Hern WM. Fetal diagnostic indications for second and third trimester outpatient pregnancy termination: fetal indications in late outpatient abortion. *Prenatal Diagn.* (2014) 34:438–44. doi: 10.1002/pd.4324
- Tilea B, Alberti C, Adamsbaum C, Armoogum P, Oury JF, Cabrol D, et al. Cerebral biometry in fetal magnetic resonance imaging: new reference data. *Ultrasound Obstet Gynecol.* (2009) 33:173–81. doi: 10.1002/uog.6276
- Griffiths PD, Bradburn M, Campbell MJ, Connolly DJA, Cooper CL, Jarvis D, et al. Change in diagnostic confidence brought about by using in utero MRI for fetal structural brain pathology: analysis of the MERIDIAN cohort. *Clin Radiol.* (2017) 72:451–7. doi: 10.1016/j.crad.2017.01.009
- Gafner M, Fried S, Gosher N, Jeddah D, Sade EK, Barzilay E, et al. Fetal brain biometry: is there an agreement among ultrasound, MRI and the measurements at birth? *Eur J Radiol.* (2020) 133:109369. doi: 10.1016/j.ejrad.2020.109369
- Katorza E, Bertucci E, Perlman S, Tschini S, Ber R, Gilboa Y, et al. Development of the fetal vermis: new biometry reference data and comparison of 3 diagnostic modalities-3d ultrasound, 2d ultrasound, and MR imaging. *Am J Neuroradiol.* (2016) 37:1359–66. doi: 10.3174/ajnr.A4725
- Griffiths PD, Mooney C, Bradburn M, Jarvis D. Should we perform in utero MRI on a fetus at increased risk of a brain abnormality if ultrasonography is normal or shows non-specific findings? *Clin Radiol.* (2018) 73:123–34. doi: 10.1016/j.crad.2017.09.007
- Yaniv G, Katorza E, Tshmaister Abitbol V, Eisenkraft A, Bercovitz R, Bader S, et al. Discrepancy in fetal head biometry between ultrasound and MRI in suspected microcephalic fetuses. *Acta Radiol.* (2017) 58:1519–27. doi: 10.1177/0284185117698865
- Griffiths PD, Brackley K, Bradburn M, Connolly DJA, Gawne-Cain ML, Kilby MD, et al. Anatomical subgroup analysis of the MERIDIAN cohort: posterior fossa abnormalities. *Ultrasound Obstet Gynecol.* (2017) 50:745–52. doi: 10.1002/uog.17485
- Griffiths PD, Brackley K, Bradburn M, Connolly DJA, Gawne-Cain ML, Griffiths DI, et al. Anatomical subgroup analysis of the MERIDIAN cohort: failed commissuration. *Ultrasound Obstet Gynecol.* (2017) 50:753–60. doi: 10.1002/uog.17502
- Garel C, Chantrel E, Sebag G. *Le développement du cerveau foetal: atlas IRM et biométrie.* Montpellier: Sauramps médical (2000).
- Mckinnon K, Kendall GS, Tann CJ, Dyet L, Sokolska M, Baruteau KP, et al. Biometric assessments of the posterior fossa by fetal MRI: a systematic review. *Prenatal Diagn.* (2021) 41:258–70. doi: 10.1002/pd.5874
- Cai S, Zhang G, Zhang H, Wang J. Normative linear and volumetric biometric measurements of fetal brain development in magnetic resonance imaging. *Childs Nerv Syst.* (2020) 36:2997–3005. doi: 10.1007/s00381-020-04633-3
- Hospitals (HUG). TY was supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie project TRABIT (agreement No. 765148).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fped.2021.639746/full#supplementary-material>

- measurements: super-resolution motion-corrected MRI of the fetal posterior fossa. *J Neuroimaging*. (2016) 26:539–44. doi: 10.1111/jon.12342
30. Velasco-Annis C, Gholipour A, Afacan O, Prabhu SP, Estroff JA, Warfield SK. Normative biometrics for fetal ocular growth using volumetric MRI reconstruction: normative biometrics for fetal ocular growth. *Prenat Diagn*. (2015) 35:400–8. doi: 10.1002/pd.4558
  31. Makropoulos A, Counsell SJ, Rueckert D. A review on automatic fetal and neonatal brain MRI segmentation. *Neuroimage*. (2018) 170:231–48. doi: 10.1016/j.neuroimage.2017.06.074
  32. Tourbier S. Novel image processing methods for improved fetal brain MRI. (2016). [https://serval.unil.ch/en/notice/serval:BIB\\_92AC4B1CA9EE](https://serval.unil.ch/en/notice/serval:BIB_92AC4B1CA9EE)
  33. Deman P, Tourbier S, Meuli R, Cuadra MB. *meribach/mevislabFetalMRI: MEVISLAB MIAL Super-Resolution Reconstruction of Fetal Brain MRI v1.0*. Zenodo. (2020). Available online at: <https://zenodo.org/record/3878563> (accessed June 5, 2020).
  34. Tourbier S. *sebastientourbier/mialsuperresolutiontoolkit*. (2020). Available online at: <https://github.com/sebastientourbier/mialsuperresolutiontoolkit> (accessed March 19, 2019).
  35. Gwet KL. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters; [a handbook for researchers, practitioners, teachers & students]*. 4th ed. Gaithersburg, MD: Advanced Analytics, LLC (2014). 410 p.
  36. Hollander M, Wolfe DA, Chicken E. *Nonparametric Statistical Methods*. 3rd edition. Hoboken, New Jersey: John Wiley & Sons, Inc. (2014). 819 p.
  37. Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. (1989) 45:255. doi: 10.2307/2532051
  38. Lin LI-K. A note on the concordance correlation coefficient. *Biometrics*. (2000) 56:324–5.
  39. Morgan CJ, Aban I. Methods for evaluating the agreement between diagnostic tests. *J Nuclear Cardiol*. (2016) 23:511–3. doi: 10.1007/s12350-015-0175-7
  40. Garel C. Fetal cerebral biometry: normal parenchymal findings and ventricular size. *Eur Radiol*. (2005) 15:809–13. doi: 10.1007/s00330-004-2610-z
  41. Batty R, Gawne-Cain ML, Mooney C, Mandefield L, Bradburn M, Mason G, et al. Analysis of errors made on in utero MR studies of the foetal brain in the MERIDIAN study. *Eur Radiol*. (2019) 29:195–201. doi: 10.1007/s00330-018-5508-x
  42. Cardenas AM, Whitehead MT, Bulas DI. Fetal neuroimaging update. *Semin Pediatr Neurol*. (2020) 33:100801. doi: 10.1016/j.spen.2020.100801
  43. Garel C, Cassart M. *Imagerie du fœtus au nouveau-né*, Lavoisier, Paris. (2016). p. 104–99.
  44. Lin Z, Shum H-Y. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *IEEE Trans Pattern Anal Mach Intelligence*. (2004) 26:83–97. doi: 10.1109/TPAMI.2004.1261081
  45. Rousseau F, Kim K, Studholme C, Koob M, Dietemann JL. On super-resolution for fetal brain MRI. *Med Image Comput Comput Assist Interv*. (2010) 13 (Pt 2):355–62. doi: 10.1007/978-3-642-15745-5\_44
  46. Im K, Guimaraes A, Kim Y, Cottrill E, Gagoski B, Rollins C, et al. Quantitative folding pattern analysis of early primary sulci in human fetuses with brain abnormalities. *Am J Neuroradiol*. (2017) 38:1449–55. doi: 10.3174/ajnr.A5217
  47. Ortinau CM, Rollins CK, Gholipour A, Yun HJ, Marshall M, Gagoski B, et al. Early-Emerging sulcal patterns are atypical in fetuses with congenital heart disease. *Cerebral Cortex*. (2019) 29:3605–16. doi: 10.1093/cercor/bhy235
  48. Benkarim OM, Hahner N, Piella G, Gratacos E, González Ballester MA, Eixarch E, et al. Cortical folding alterations in fetuses with isolated non-severe ventriculomegaly. *Neuroimage Clin*. (2018) 18:103–14. doi: 10.1016/j.nicl.2018.01.006
  49. Hahner N, Benkarim OM, Aertsen M, Perez-Cruz M, Piella G, Sanroma G, et al. Global and regional changes in cortical development assessed by MRI in fetuses with isolated nonsevere ventriculomegaly correlate with neonatal neurobehavior. *Am J Neuroradiol*. (2019) 40:1567–74. doi: 10.3174/ajnr.A6165
- Conflict of Interest:** 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.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2021 Khawam, de Dumast, Deman, Kebiri, Yu, Tourbier, Lajous, Hagmann, Maeder, Thiran, Meuli, Dunet, Bach Cuadra and Koob. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.