Citation: Skelton, E. ORCID: 0000-0003-0132-7948, Matthew, J., Li, Y., Khanal, B., Cerrolaza Martinez, J. J., Toussaint, N., Gupta, C., Knight, C., Kainz, B., Hajnal, J. V. and Rutherford, M. (2020). Towards automated extraction of 2D standard fetal head planes from 3D ultrasound acquisitions: A clinical evaluation and quality assessment comparison. Radiography, doi: 10.1016/j.radi.2020.11.006

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/25421/

Link to published version: http://dx.doi.org/10.1016/j.radi.2020.11.006

Copyright and reuse: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.
Towards automated extraction of 2D standard fetal head planes from 3D ultrasound acquisitions: A clinical evaluation and quality assessment comparison

E. Skelton a, b, *, J. Matthew a, b, Y. Li c, B. Khanal b, J.J. Cerrolaza Martinez c, N. Toussaint b, C. Gupta a, C. Knight a, d, B. Kainz b, c, J.V. Hajnal a, b, M. Rutherford a, d

a Perinatal Imaging Department, King’s College London, UK
b School of Biomedical Engineering and Imaging Sciences, King’s College London, UK
c Department of Computing, Imperial College London, UK
d Guy’s & St Thomas’ NHS Foundation Trust, UK

Article info
Article history:
Received 25 September 2020
Received in revised form
5 November 2020
Accepted 9 November 2020
Available online xxx

Keywords:
Clinical evaluation
Fetal imaging
Quality assessment
Ultrasound

Introduction and literature review

The fetal head and brain is examined during 18+0-20+6 fetal anomaly ultrasound (US) examinations to assess growth and development of the skull and intracranial structures.1 Transventricular (TV) and trans cerebellar (TC) views are routinely assessed in the basic screening examination. These two-dimensional (2D) planes allow identification of intracranial landmarks and acquisition of specific biometric measurements which, if absent or outside expected reference ranges, may be indicative of an anomaly.2 Occasionally, it is not possible to obtain these planes, resulting in an incomplete or suboptimal assessment.3

* Corresponding author. St Thomas’ Hospital, 4th Floor, North Wing, Westminster Bridge Road, London, SE1 7EH, UK.
E-mail address: emily.skelton@kcl.ac.uk (E. Skelton).

https://doi.org/10.1016/j.radi.2020.11.006
1076-8174/© 2020 The College of Radiographers. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Advances in three-dimensional US (3D-US) and multi-planar reconstruction can complement conventional 2D-US and overcome some of its limitations. The 3D-transabdominal “single-shot” technique can be used to acquire a reproducible fetal head US-volume with reconstructed images enabling detailed retrospective review by clinicians when multiple datasets are acquired from different insonation angles (e.g. transverse, sagittal, coronal). Biometric measurements from 3D-images have also demonstrated good correlation with 2D-methods.

Visualisation of intracranial structures using 3D-US can be superior to 2D-US as reviewers are able to manipulate the image planes for optimisation, although this process can be time-consuming, requiring additional training in 3D-techniques and experience with different manufacturer’s platforms, therefore is not feasible in clinics where real-time assessment is required. A deep learning approach to automatically-extract required standard planes from 3D-volumes could overcome these barriers, and facilitate the accessible use of 3D-techniques within US screening clinics.

Robust clinical evaluation of DL tools is essential to build on reported technical accuracy metrics as part of the clinical trans-formation process. However, for fetal US, quality assessment (QA) is laborious and subjective because of variation in screening programmes and a lack of agreement on quantitative assessment criteria. Building on the work of Li et al. this study aims to clinically evaluate the quality of automatically extracted standard TV and TC fetal head planes from 3D-US volumes in comparison to standard planes that are manually-acquired from 2D-US and operator-selected from 3D-US.

**Methods**

Data was acquired between 2016 and 2019 as part of the Intelligent Fetal Imaging and Diagnosis (iFIND) project (NRES reference number = 14/LO/1086 and 07/H0707/105). Participants gave informed written consent.

Inclusion criteria were: completed 18-26 clinical fetal anomaly scan and consent to fetal research imaging. Scans were undertaken by four operators (3 research sonographers, 1 obstetrician) (TF/JM/CK/ES) using a Philips EpiQ (Philips Healthcare, Best, Netherlands) US system with an X6-1 MHz matrix transducer to acquire the following image planes:

1. 2D-TV
2. 2D-TC
3. 3D-TV
4. 3D-TC.

3D-volumes were acquired from standard head planes (angle of insonation at 90° to mid-line echoes), using an acquisition sweep angle of 90° to ensure complete coverage of the fetal cranium.

**Algorithm development**

Using open-source software (MITK workbench 2016.11), 303 TV and 248 TC-planes were manually annotated by 3 observers (2 research sonographers, 1 medical student) (ES/JM/CG) to provide training data for tool development. For consistency, observers received in-person training by JM prior to annotating. The annotation process required the observers to identify 14 cranial landmarks (Fig. 1).

The tool is implemented using the M4+: Iterative Transformation Network (ITN) approach. ITN utilises a convolutional neural network to learn the mapping between a 2D-image and the rigid transformation required to move that plane towards the location and orientation of the standard plane in the 3D-volume.

The dataset of manually annotated 3D-fetal head volumes was split into training and test sets (Fig. 2). There are fewer TC-plane annotations because of lack of available image data. Two separate models were trained for TV and TC-planes respectively.

The ITN model was implemented in Tensorflow running on a machine with Intel Xeon CPU E5-1630 at 3.70 GHz and one NVIDIA Titan Xp 12 GB GPU. Accuracy with the ITN model was evaluated quantitatively using: 1) distance between the plane centres (dx) and 2) rotation angle between the planes (θ). On the test dataset, the ITN approach achieved a detection accuracy of dx=(3.68 ± 1.69) mm, θ=12.5 ± 6.1° for TV-planes and dx=(3.69 ± 1.75) mm, θ=(12.9 ± 6.9)° for TC-planes. These results show the ITN model can accurately predict both the location and orientation of the planes.

Using a randomisation tool developed in MATLAB (The MathWorks, USA), a selection of the remaining planes were presented for retrospective quality assessment (Fig. 2).

**Quality assessment**

For quality assessment (QA), a software tool (USQA) and a quality scoring system was developed, devised from peer-reviewed studies and guidance by the NHS Fetal Anomaly Screening Programme. Two observers (>20-years combined ultrasound experience) (ES/JM) performing QA received prior formal USQA training from a software engineer (NT) using unrelated images. To minimise recall bias for the observers, a break of 4-weeks was included between annotation and QA. The observers were blinded to the plane’s modality and instructed to perform a binary assessment of predefined criteria (Table 1) derived from existing literature for each image, including a final assessment of overall adequacy. Not all criteria required a “pass” for the image to be considered adequate overall: some technical factors (e.g. hemisphere asymmetry) may not be sufficiently detrimental to render the image inadequate, but criteria related to visualisation of key anatomical landmarks for the standard plane (e.g. cavum septum pellucidum, CSP) are essential. For intra-observer agreement, both observers were presented with a subset (n = 25) of randomly selected images for re-review after 2-weeks. Primary outcome measures were:

1. Overall adequacy of image quality from TV and TC-planes between three modalities: 2D manually-acquired (2D-MA), 3D operator-selected (3D-OS) and 3D-automatically-extracted (3D-DL)
2. Assessment of specific image features contributing to inadequate images
3. Assessment of intra-observer variation agreement of quality scoring.

**Statistical analysis**

Data was analysed using SPSS (version 24, SPSS Inc, USA). Non-parametric statistical analysis was undertaken because of the categorical nature of the dataset. McNemar’s test was used to determine any differences between the number of adequate planes per modality for each observer. Where there was inadequate image quality for 3D-planes, a sub-analysis was performed to identify the criteria contributing to lower quality. Inter-observer agreement strength was assessed using Cohen’s Kappa, and the percentage intra-observer agreement was calculated (p < 0.05 was used to determine statistical significance).
Results

Participant demographics

Images from 91-TV and 73-TC cases were available for review. There were 29-TV and 19-TC 2D-MA images that observers agreed were of overall adequate quality. These were selected as the gold-standard for comparative analysis against their corresponding image from 3D-OS and 3D-DL. Mean maternal age at consent was 33 years (range 24–40). Mean maternal BMI was 27 kg/m² (range 20.05–45.6 kg/m²). Mean gestational age (GA) at the time of scan was 26 completed weeks (range 20¹⁵⁻³⁻³⁺ weeks). There were 3 cases of fetal cardiac anomaly: 2 right aortic arch and 1 suspected coarctation of the aorta. Two additional cases had placenta praevia. No cases had any identifiable structural head or brain anomalies.
Quality assessment of TV-planes (Table 2)

Of 29 TV-planes, observer-1 rated 62% of 3D-OS TV-planes as adequate. Observer-2 rated 72% of 3D-OS TV-planes as adequate. Of the planes that observer-1 rated inadequate (n = 11), observer-2 gave the same rating in 7 (64%). For 3D-DL TV-planes, observer-1 rated 69% as adequate, and observer-2 rated 66% as adequate. Of the 3D-DL TV-planes that observer-1 rated inadequate (n = 9), observer-2 gave the same rating in 7 (78%). There were 3 cases (10%) where observers agreed that both the corresponding 3D-OS and 3D-DL TV-planes were inadequate.

Quality assessment of TC-planes (Table 3)

Of 19 TC-planes, observer-1 rated 58% of corresponding 3D-OS planes as adequate. Observer-2 rated 32% as adequate. Of the 3D-DL TC planes that observer-1 rated inadequate (n = 12), observer-2 gave the same rating in 9 (75%). There were 5 cases (26%) where observers agreed that both the 3D-OS and 3D-DL TC-planes were inadequate.

Sub-analysis of inadequate 3D planes

For inadequate 3D-OS TV-planes (n = 7), the image quality criterion with the highest proportion of agreed failures was visualisation of the CSP (71%). For 3D-OS TC-planes (n = 7), this was poor visualisation of the cerebellar hemispheres (86%).

Figure 2. Study flowchart illustrating allocation of image data.

Table 1

| TV plane | TC plane |
|----------|----------|
| Symmetrical appearance of right-left hemispheres | Symmetrical appearance of right-left hemispheres |
| CSP well visualised | CSP well visualised |
| Posterior ventricle/choroid plexus well visualised | Equal size cerebellar hemispheres (borders not obscured by posterior fossa) |
| No cerebellum visualised | Occipital bone and nuchal skin fold visible |
| Oval skull shape > round skull shape | Well visualised intra-cranial structures |
| Well visualised intra-cranial structures | Overall adequate view |
ventricle (86%) (Fig. 3). For failed 3D-DL TC-planes (n = 9), this was poor visualisation of the cerebellar hemispheres (80%) (Fig. 4).

**Observer agreement of image adequacy**

Observers agreed 59% of 3D-OS and 3D-DL TV-planes were of overall adequate quality. Of these, there were 10 cases where both 3D-TV-planes were of overall adequate quality. Observers agreed 24% of 3D-OS and 10% of 3D-DL TC-planes included were of overall adequate quality. There was only one case where both 3D-TC-planes were overall adequate quality.

Cohen’s kappa was calculated to assess the strength of the inter-observer agreement in the overall adequacy of the image achieved beyond chance as per McHugh (Table 4).

Inter-observer agreement for the overall adequacy of TC-planes was weak for 3D-OS (k = 0.486). No inter-observer agreement in the overall adequacy of 3D-DL TC-planes was demonstrated. The power of these findings is uncertain because of the low number of cases.

Due to small intra-observer case numbers, it was not possible to statistically analyse the strength of intra-observer agreement, therefore, the percentage agreement was calculated.

Eleven TV-planes were included for intra-observer re-review: 2D-MA (n = 5), 3D-OS (n = 3) and 3D-DL (n = 4). Observer-1 had the highest agreement with 8 (72%) of TV-planes rated the same in both review periods. Observer-2 rated 6 (54%) of the included TV-planes as the same in both reviews.

The TC-planes included for intra-observer re-review were: 2D-MA (n = 1), 3D-OS (n = 3) and 3D-DL (n = 1). There were no agreed images for either observer.

### Table 2
Observer assessment of overall TV-plane quality.

|                     | 2D manually-acquired and 3D operator-selected both adequate | 2D manually-acquired and 3D-operator selected both inadequate | Total |
|---------------------|-----------------------------------------------------------|-------------------------------------------------------------|-------|
| Observer 1          | 18                                                        | 11                                                          | 29    |
| Observer 2          | 21                                                        | 8                                                           | 29    |
| Total               | 39                                                        | 19                                                          | 58    |

### Table 3
Observer assessment of overall TC-plane quality.

|                     | 2D manually-acquired and 3D operator-selected both adequate | 2D manually-acquired and 3D-operator selected both inadequate | Total |
|---------------------|-----------------------------------------------------------|-------------------------------------------------------------|-------|
| Observer 1          | 11                                                        | 8                                                           | 19    |
| Observer 2          | 6                                                         | 13                                                          | 19    |
| Total               | 17                                                        | 21                                                          | 38    |

**Figure 3.** 3D-DL automatically-extracted TV-plane rated as overall inadequate image quality due to poor visualisation of the posterior horns of the lateral ventricle (solid white arrow to demonstrate poorly visualised posterior horn, star indicates cavum septum pellucidum).

**Figure 4.** 3D-DL automatically-extracted TC-plane rated as overall inadequate image quality due to poor visualisation of the cerebellar hemispheres (solid white arrow to demonstrate non-visualisation of cerebellar hemispheres in posterior fossa, star indicates cavum septum pellucidum).
Comparison of imaging modalities per observer

McNemar’s tests found significant differences in the overall adequacy of the 2D-MA planes compared to 3D as rated by both observers, however no significant difference between 3D-images was demonstrated ($p < 0.005$) (Table 5).

Discussion

This evaluation suggests that a 3D-DL algorithm can automatically extract standard planes from fetal head volumes of comparable quality to 3D-image planes selected by an operator from the same volume. No significant difference in image quality was demonstrated between 3D-modalities, although compared to corresponding 2D-planes, there was a significant reduction in quality. This infers image quality limitations are related to acquired 3D-volumes used to train and test the algorithm. Quality of the 3D-volume and resultant planes may be limited by fetal head position within the maternal pelvis, and strong ossification of cranial bones causing shadowing artefact and decreasing visibility of intra-cranial structures.

This is particularly relevant to TC-planes which had fewer adequate images in this study, likely due to the difficulty in visualising the cerebellum through obscuration of the posterior fossa. GA distribution of this dataset was wider than a typical antenatal ultrasound screening clinic where there may be a lack of expertise. Acquiring 3D-volumes may be faster than current practice which requires the operator to navigate to the region of interest and obtain a specific 2D-plane. Benacerraf et al., reported the acquisition times for a routine 20-week fetal anomaly scan were halved using 3D-approaches, on average taking 6.6 min compared to 13.9 min for 2D. This could ease physical demands for US scanning.

Advantages of 3D-fetal sonography in improved image quality/anomaly detection and acquisition time have been reported in published literature, although without further validation

### Table 4
Cohen’s kappa for inter-observer agreement of quality scoring per standard plane and modality.

| Overall adequate view | Inter-observer agreement (Observer 1 and Observer 2) |
|-----------------------|-----------------------------------------------------|
| 3D operator-selected | 3D-DL automatically-extracted                      |
| TV plane              | Moderate ($k = 0.613, p = 0.001$)                   |
| TC plane              | Weak ($k = 0.486, p = 0.026$)                      |

### Table 5
Comparison of imaging modalities and overall adequacy of image per observer (McNemar’s test). $p < 0.05$ was used to determine statistical significance.

| Standard plane | 2D manually-acquired | 3D operator-selected |
|----------------|----------------------|----------------------|
| TV             | Fail                 | Pass                 |
| Pass           | 11                   | 18                   |
| McNemar’s p    | 0.001                |                      |
| TC             | Fail                 | Pass                 |
| Pass           | 8                    | 11                   |
| McNemar’s p    | 0.008                |                      |

| Standard plane | 3D automatically-extracted |
|----------------|---------------------------|
| TV             | Fail                      |
| Pass           | 9                         |
| McNemar’s p    | 0.004                     |
| TC             | Pass                      |
| Pass           | 12                        |
| McNemar’s p    | 0.000                     |
of 3D-neurosonography, it is only recommended that volumetric interpretation is used to compliment conventional 2D-assessment.26 Whilst outside of the scope for this study, future work may further develop the tool towards automatic extraction of additional planes (sagittal and coronal) required for complete fetal neurosonographic examination.27,30

High-quality fetal US images are essential to optimise visualisation, improve evaluation of anatomical structures and avoid litigation in cases of undetected anomalies.21 Image QA is subjective,22 and whilst attempts have been made to standardise evaluation, this is usually associated with biometrics.15,16 The USQA tool uses pre-defined criteria to help reduce subjectivity, and demonstrated moderate inter-observer agreement in assessment of overall image quality for 3D TV-planes. Image quality checklists may help guide observers undertaking image quality analysis; however, may not always reduce variation between observers’ evaluations (e.g. weak inter-observer agreement for 3D TC-planes). This may be related to the small sample size, and/or the subjectivity of TC-plane assessment at later gestations (the TC-plane is infrequently imaged beyond 20-weeks unless a posterior fossa anomaly is suspected).15 Salomon et al., found that reviewers assessing fetal head image quality disagreed in over one third of cases, and the same reviewer could give varying judgements in up to 25% of cases,13 emphasising the subjectivity of review processes and the difficulties of proposing objective approaches. Further evaluation of the USQA software at earlier GAs using multiple reviewers is required to validate this approach to QA.

The accuracy of standard plane localisation using the ITN model was assessed quantitatively prior to image quality evaluation.9 A comparison of image quality scores with quantitative measurements was not conducted, however, even with a large rotation angle between the planes (90°), the resultant images are often still visually similar (e.g. the image boundary region may be rotated significantly but the central brain region and key structures still remain intact and visually comparable). The algorithm may be more sensitive to alterations in the distance between the planes (dx) which could result in non-visualisation of key structures (thus failing to meet corresponding image quality criteria). These considerations highlight the differences between technical and clinical evaluation of DL-tools, and development in this area may help to further align the results.

Conclusions

The clinical value of automatic-plane extraction tools for fetal head screening is not established. This study demonstrates that standard planes can be automatically-extracted from 3D-fetal head volumes to a similar quality of an operator-selected 3D-plane, however, quality of the 3D-volume from which the plane is extracted remains a limitation. Further work should focus on refining the algorithm using datasets from cohorts at lower gestations to improve this.

Author’s statement

All authors have made a significant contribution to the design, data collection and analysis and production of this manuscript.

Data statement

The data sets collected and analysed during this study are available from the corresponding author on request.

Funding

This work was supported by the Wellcome Trust Council IEH Award [102431] for the Intelligent Fetal Imaging and Diagnosis project (www.ifindproject.com) and the Wellcome/EPSRC Centre for Medical Engineering [WT 203148/Z/16/Z]. The authors acknowledge support from the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London. The views expressed are those of the authors and not necessarily those of the NIHS, the NIHR or the Department of Health. The study has been granted NHS R&D and ethics approval, NRES ref no = 14/LO/1086 and 07/H0707/105.

Conflict of interest statement

None declared.

Acknowledgements

The authors would like to thank Tara Fletcher for participating in image data collection.

References

1. Public Health England. Fetal anomaly screening programme standards 2018. 2018 [Online] Available at: https://www.gov.uk/government/publications/fetal-anomaly-screening-programme-standards/standards-valid-from-1-april-2018. [Accessed 23 September 2020].
2. Harada A, Miyashita S, Nagai R, Makino S, Murotsuki J. Prenatal sonographic findings and prognosis of craniosynostosis diagnosed during the fetal and neonatal periods. Congenit Anom (Kyoto) 2019;59:132–41. https://doi.org/10.1111/cga.12308.
3. Endres LK, Cohen L. Reliability and validity of three-dimensional fetal brain volumes. J Ultrasound Med 2001;20:1265–9. https://doi.org/10.7863/jum.2001.20.12.1265.
4. Salman MM, Twinning P, Mousa H, James D, Montzax M, Aboughar M, et al. Evaluation of offline analysis of archived three-dimensional volume datasets in the diagnosis of fetal brain abnormalities. Ultrasound Obstet Gynecol 2011;38:165–9. https://doi.org/10.1002/uog.9821.
5. Fratelli N, Taddei F, Perfumo F, Franceschetti L, Farina G, Frusa T. Interobserver reproducibility of transabdominal 3-dimensional sonography of the fetal brain. J Ultrasound Med 2009;28:1009–13. https://doi.org/10.7863/jum.2009.28.8.1009.
6. Borstein E, Monteagudo A, Santos R, Strock I, Tyszmbal T, Lenchen E, et al. Basic as well as detailed neonrosonograms can be performed by offline analysis of three-dimensional fetal brain volumes. Ultrasound Obstet Gynecol 2009;36:20–5. https://doi.org/10.1002/uog.7527.
7. Muňoz H, Pedeza D, Astudillo J, Yamamoto M, Valenti P, Carrillo J, et al. OP03.01: neurosonographic fetal assessment with the “single shot” technique in transabdominal 3-D multiplanar ultrasound during routine 20-24 weeks’ gestation scan. Ultrasound Obstet Gynecol 2008;32:524–31. https://doi.org/10.1002/uog.5234.
8. D’Ast A, Paramasivam G, Bashier SN, Whithy E, Tahir Z, Lees C. How to obtain diagnostic planes of the fetal central nervous system using three-dimensional ultrasound and a context-preserving rendering technology. Am J Obstet Gynecol 2019;220:215–29. https://doi.org/10.1016/j.ajog.2019.06.010.00051.4080.51709.93.
9. Li Y, Khanal B, Hou B, Alansary A, Cerrolaza J, Sinclair M, et al. Standard plane detection in 3D fetal ultrasound using an iterative transformation network. In: Lecture Notes in computer science (including subseries lecture Notes in artificial Intelligence and lecture Notes in bioinformatics) 11070 LNCS. Springer Verlag; 2018. p. 392–400. https://doi.org/10.1007/978-3-030-09928-1_45.
10. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2018;17. https://doi.org/10.1186/s12916-019-1426-2.
11. Urasen MT, Peters I, Kraan-van der Est M, Reijerink-Verheij J, Knaphel M, Cohen-Overbeek P. An audit of second-trimester fetal anomaly scans based on a novel image-scoring method in the southwest region of The Netherlands. J Ultrasound Med 2017;36:1171–9. https://doi.org/10.7863/ultra.16.0605S.
12. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: a system for large-scale machine learning. 12th Symp. Oper. Syst. Des. Implement. 2016; 655–83 [Online] Available at: http://arxiv.org/abs/1605.08605. [Accessed 23 September 2020].
13. Salomon LJ, Bernard JP, Duyme M, Doris B, Mas N, Ville Y. Feasibility and reproducibility of an image-scoring method for quality control of fetal
biometry in the second trimester. Ultrasound Obstet Gynecol 2005;27:34–40.

14. Salomon LJ, Winer N, Bernard JP, Ville Y. A score-based method for quality control of fetal images at routine second-trimester ultrasound examination. Prenat Diagn 2008;28:822–7. https://doi.org/10.1002/pd.2016.

15. Sairam S, Awadh AMA, Cook K, Papageorghiou AT, Carvalho JS. Impact of audit of routine second-trimester cardiac images using a novel image-scoring method. Ultrasound Obstet Gynecol 2009;33:545–51. https://doi.org/10.1002/uog.6323.

16. Jaudi S, Tezenas Du Montcel S, Fries N, Nizare J, Halley Desfontaines V, Dommergues M. Online evaluation of fetal second-trimester four-chamber view images: a comparison of six evaluation methods. Ultrasound Obstet Gynecol 2011;38:185–90. 0.10097/01.uog.000514080.51709.93.

17. Yaqub M, Kelly R, Stobart H, Napolitano R, Noble AJ, Papageorghiou AT. Quality-improvement program for ultrasound-based fetal anatomy screening using large-scale clinical audit. Ultrasound Obstet Gynecol 2019;54:239–45. https://doi.org/10.1002/uog.20144.

18. McHugh M. Interrater reliability: the kappa statistic. Biochem Med 2012;22:276–82.

19. Rizzo G, Capponi A, Pietrolucci ME, Capece A, Aiello E, Mammarella S, et al. An algorithm based on Omniview technology to reconstruct sagittal and coronal planes of the fetal brain from volume datasets acquired by three-dimensional ultrasound. Ultrasound Obstet Gynecol 2011;38:158–64. https://doi.org/10.1002/uog.8959.

20. Lloyd DFA, Pushparajah K, Simpson JM, von Amerom JFP, van Poppel MPM, Schultz A, et al. Three-dimensional visualisation of the fetal heart using prenatal MRI with motion-corrected slice-volume registration: a prospective, single-centre cohort study. Lancet 2019;393:1619–27.

21. Yaqub M, Rueda S, Kopuri A, Melo P, Papageorghiou AT, Sullivan PB, et al. Plane localization in 3-D fetal neurosonography for longitudinal analysis of th developing brain. IEEE J Biomed Heal Informatics. 2016;20:1–9. https://doi.org/10.1109/JBHI.2015.2435651.

22. Gonçalves LF, Lee W, Espinoza J, Romero R. Three- and 4-dimensional ultrasound in obstetric practice does it help? J Ultrasound Med 2005;24:1599–624. https://doi.org/10.7863/jum.2005.24.12.1599.

23. Abuhamad AZ. Standardization of 3-dimensional volumes in obstetric sonography: a required step for training and automation. J Ultrasound Med 2005;24:397–401. https://doi.org/10.7863/jum.2005.24.4.397.

24. The Society and College of Radiographers. Ultrasound workforce UK census. 2019. 2019 [Online] Available at: https://www.sor.org/sites/default/files/document-versions/ultrasound_workforce_uk_census_2019.pdf. [Accessed 3 November 2020].

25. Roy-Lacroix ME, Moretti F, Ferraro ZM, Brosseau L, Clancy J, Fung-Kee-Fung KA. A comparison of standard two-dimensional ultrasound to three-dimensional volume sonography for routine second-trimester fetal imaging. J Perinatol 2017;37:380–6. https://doi.org/10.1038/jp.2016.212.

26. Yaqub M, Cuinnet R, Napolitano R, Roundhill D, Papageorghiou AT, Ardron R, et al. Volumetric segmentation of key fetal brain structures in 3D ultrasound. In: Lecture Notes in computer science (including subseries Lecture Notes in artificial Intelligence and lecture Notes in bioinformatics) 8184 LNCS. Springer Verlag; 2013. p. 25–32. https://doi.org/10.1007/978-3-319-02267-3_4.

27. Correa FJ, Lara C, Belyer J, Remohi J, Pellicer A, Serra V. Examination of the fetal brain by transabdominal three-dimensional ultrasound: potential for routine neurosonographic studies. Ultrasound Obstet Gynecol 2006;27:503–8. https://doi.org/10.1002/uog.2756.

28. Benacerfar BR, Shipp TD, Bromley B. How sonographic tomography will change the face of obstetric sonography: a pilot study. J Ultrasound Med 2005;24:37–8. https://doi.org/10.7863/jum.2005.24.3.371.

29. Zhou SK, Sofka M, Zhang J, Good S, Comaniciu D. Automatic detection and measurement of structures in fetal head ultrasound volumes using sequential estimation and integrated detection network (IDN). IEEE Trans Med Imag 2014;33. https://doi.org/10.1109/TMI.2014.2301936.

30. Paladini D, Malinger M, Moteagudo A, Pilu G, Timor-Tritsch I, Toi A. Sonographic examination of the fetal central nervous system: guidelines for performing the ‘basic examination’ and the ‘fetal neurosonogram’. Ultrasound Obstet Gynecol 2007;29:109–16. https://doi.org/10.1002/uog.3909.

31. Malik Y, Jackson D. Litigation in obstetrical sonography. Obstet Gynecol 2017;129:5165. https://doi.org/10.1097/AOG.0000514080.51709.93.

32. Ville Y. ‘Ceci n’est pas une échographie’: a plea for quality assessment in prenatal ultrasound. Ultrasound Obstet Gynecol 2008;31:1–5. https://doi.org/10.1002/uog.5248.

33. Sarris I, Ionnou C, Dighie M, Oberto M, Qingqing W, Shah J, et al. Standardization of fetal ultrasound biometry measurements: improving the quality and consistency of measurements. Ultrasound Obstet Gynecol 2011;38:581–7. https://doi.org/10.1002/uog.8997.