








# Quality assessment of fetal middle cerebral and umbilical artery Doppler images using an objective scale within an international randomized controlled trial

M. RIAL-CRESTELO<sup>1</sup>, J. MORALES-ROSELLÓ<sup>2</sup>, E. HERNÁNDEZ-ANDRADE<sup>3</sup>, F. PREFUMO<sup>4</sup> , D. OROS<sup>5</sup> , D. CAFFICI<sup>6</sup>, A. SOTIRIADIS<sup>7</sup> , E. ZOHAV<sup>8</sup>, R. CRUZ-MARTINEZ<sup>9</sup> , M. PARRA-CORDERO<sup>10</sup>, M. LUBUSKY<sup>11</sup>, M. KACEROVSKY<sup>12</sup> and F. FIGUERAS<sup>1</sup> 

<sup>1</sup>Fetal Medicine Research Center, BCNatal - Barcelona Center for Maternal-Fetal and Neonatal Medicine (Hospital Clinic and Hospital Sant Joan de Déu), Institut Clínic de Ginecología, Obstetricia y Neonatología, IDIBAPS, Obstetrics and Gynecology, University of Barcelona, Barcelona, Spain; <sup>2</sup>Department of Obstetrics and Gynecology, Hospital Universitario y Politécnico la Fe, Valencia, Spain; <sup>3</sup>Department of Obstetrics and Gynecology, Division of Maternal Fetal Medicine, Wayne State University, Hutzel Women's Hospital, Perinatology Research Branch, NICHD/NIH/DHHS, Detroit, MI, USA; <sup>4</sup>Department of Obstetrics and Gynecology, University of Brescia, Brescia, Italy; <sup>5</sup>Aragon Institute for Health Research (IIS Aragón), Obstetrics Department, Hospital Clínico Universitario Zaragoza, Zaragoza, Spain; <sup>6</sup>Sociedad Argentina de Ultrasonografía en Medicina y Biología, Buenos Aires, Argentina; <sup>7</sup>2nd Department of Obstetrics and Gynecology, Faculty of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece; <sup>8</sup>Tel Aviv Sourasky Medical Center and Sackler School of Medicine, Lis Maternity and Women's Hospital University, Tel Aviv University, Tel Aviv, Israel; <sup>9</sup>Department of Fetal Medicine and Surgery, Children's and Women's Specialty Hospital of Querétaro, Querétaro, Mexico; <sup>10</sup>Fetal Medicine Unit, Obstetrics and Gynecology, University of Chile Hospital, Santiago de Chile, Chile; <sup>11</sup>Department of Obstetrics and Gynecology, Palacky University Olomouc, Faculty of Medicine and Dentistry, University Hospital Olomouc, Olomouc, Czech Republic; <sup>12</sup>Charles University in Prague, Faculty of Medicine Hradec Kralove, Department of Obstetrics and Gynecology, University Hospital Hradec Kralove, Prague, Czech Republic

**KEYWORDS:** Doppler; middle cerebral artery; quality control; randomized multicenter trial; reliability; umbilical artery

## CONTRIBUTION

*What are the novel findings of this work?*

The quality of Doppler ultrasound images of the fetal middle cerebral artery and umbilical artery can be evaluated reliably using an objective scale.

*What are the clinical implications of this work?*

Auditing of Doppler ultrasound images of the middle cerebral artery and umbilical artery has a key role in quality assessment in both research and clinical practice.

## ABSTRACT

**Objectives** To determine the quality of Doppler images of the fetal middle cerebral artery (MCA) and umbilical artery (UA) using an objective scale, and to determine the reliability of this scale, within a multicenter randomized controlled trial (Revealed versus concealed criteria for placental insufficiency in unselected obstetric population in late pregnancy (Ratio37)).

**Methods** The Ratio37 trial is an ongoing randomized, open-label, multicenter controlled study of women with a low-risk pregnancy recruited at 20 weeks. Doppler measurements of the fetal MCA and UA were performed

at 37 weeks. Twenty patients from each of the six participating centers were selected randomly, with two images evaluated per patient (one each for the MCA and UA). The quality of a total of 240 images was evaluated by six experts, scored on an objective scale of six items. Inter- and intrarater reliability was assessed using the Fleiss-modified kappa statistic for ordinal scales.

**Results** On average, 89.2% of MCA images and 85.0% of UA images were rated as being of perfect (score of 6) or almost perfect (score of 5) quality. Kappa values for intrarater reliability of quality assessment were 0.90 (95% CI, 0.88–0.92) and 0.90 (95% CI, 0.88–0.93) for the MCA and UA, respectively. The corresponding inter-rater reliability values were 0.85 (95% CI, 0.81–0.89) and 0.84 (95% CI, 0.80–0.89), respectively.

**Conclusion** The quality of MCA and UA Doppler ultrasound images can be evaluated reliably using an objective scale. Over 85% of images, which were obtained by operators from a broad range of clinical practices within a multicenter study, were rated as being of perfect or almost perfect quality. Intra- and inter-rater reliability of quality assessment was very good. Copyright © 2019 ISUOG. Published by John Wiley & Sons Ltd.

Correspondence to: Dr F. Figueras, Fetal Medicine Research Center, BCNatal - Barcelona Center for Maternal-Fetal and Neonatal Medicine (Hospital Clinic and Hospital Sant Joan de Déu), Institut Clínic de ginecología, Obstetricia y Neonatología, IDIBAPS, Obstetrics and Gynecology, University of Barcelona, Barcelona 08028, Spain (e-mail: ffiguera@clinic.ub.es)

Accepted: 28 May 2019

## INTRODUCTION

In the field of ultrasound, there are quality control systems aimed at ensuring acceptable levels of reliability among operators and consistency of measurements<sup>1–4</sup>. This issue is particularly important for the performance of multicenter studies because of their inherent heterogeneity<sup>5</sup>. Reliability and consistency are known to improve when using quality control systems<sup>6</sup> in research settings, but this has not been proven extensively in routine clinical practice.

Doppler evaluation of fetal vessels plays a key role in managing high-risk pregnancies, mainly those with fetal growth restriction and pre-eclampsia, with proven impact in improving perinatal outcome<sup>7</sup>. More recently, in low-risk or unselected pregnancies it has been suggested as potentially useful near term, when most cases of still-birth remain undetected<sup>8</sup>. With the potential advent of Doppler evaluation as a screening test, concerns regarding its reliability have become a key issue, since both false positives and false negatives due to poor image acquisition or measurement may have detrimental consequences.

A quality control system for fetal middle cerebral artery (MCA) Doppler assessment was published by Ruiz-Martinez *et al.*<sup>9</sup> and subsequently adapted for the umbilical (UA) and uterine arteries<sup>9</sup>. The system consists of an objective scale of six items, each of which counts for 1 point. This objective assessment showed higher agreement between raters than did subjective evaluation.

The aim of this study was to determine the quality of Doppler images of the MCA and UA using this objective scale, and to determine the reliability of the scale, within a multicenter randomized controlled trial ('Revealed versus concealed criteria for placental insufficiency in unselected obstetric population in late pregnancy' (Ratio37))<sup>10</sup>, in which a broad range of settings, women and operators were involved.

## METHODS

The Ratio37 trial is an ongoing randomized, open-label, multicenter ( $n=6$ ), controlled study recruiting women with a low-risk pregnancy at 20 weeks of gestation, with Doppler measurements of the UA and fetal MCA performed at 37 weeks<sup>10</sup>. The Doppler images are stored systematically in Digital Imaging and Communications in Medicine (DICOM). The standardized technique agreed

by the researchers to obtain Doppler measurements was in accordance with the recommendations of the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG)<sup>11</sup>. For this study, a random procedure was performed stratified by center to select images for quality assessment. Women were selected randomly by creating a random seed. In all cases, Doppler measurements of the MCA and UA had been obtained.

The quality of each image was scored according to the six-point scoring system<sup>9</sup>, in which 1 point was awarded for each of following items: (1) anatomical site, (2) magnification, (3) angle of insonation, (4) image clarity, (5) sweep-speed adjustment and (6) velocity scale and baseline adjustment. Table 1 details the quality criteria for each vessel. Images with a score of 4–6 were defined as good quality whereas those with a score of 0–3 were defined as poor quality. A total of six expert raters were selected for this study, based either on being a coauthor of the ISUOG guidelines (D.C., E.H.A. and F.P.) or on their research experience in the field (J.M.R., D.O. and A.S.). For assessment of inter-rater reliability, three raters scored independently each of the images, blinded to center. For assessment of intrarater reliability, the remaining three raters re-evaluated half of the images 3 months after the first evaluation, so as to avoid memory bias.

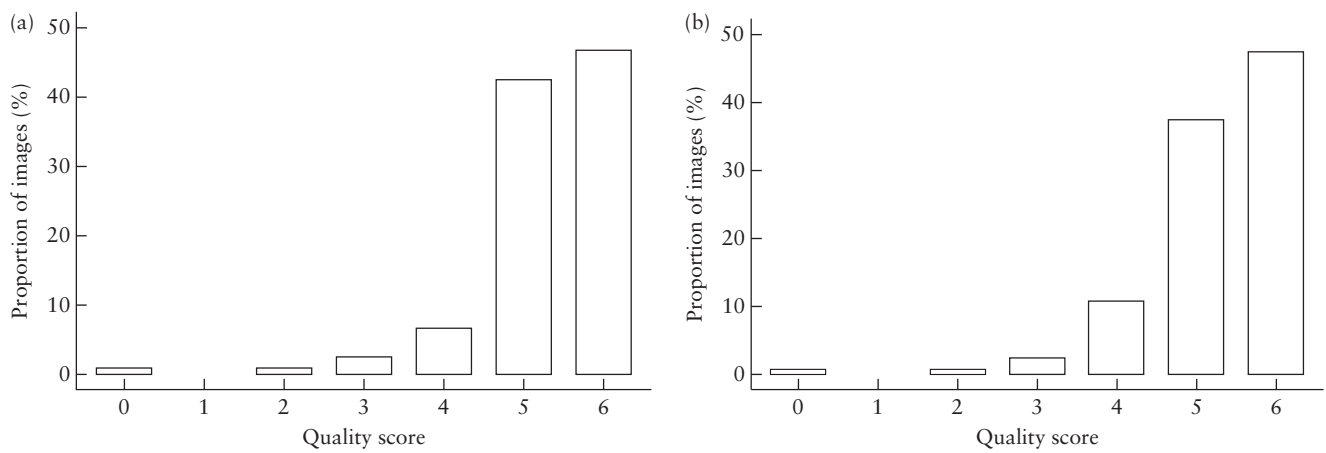
## Statistical analysis

For sample size estimation, a Monte Carlo simulation procedure of 10 000 replicates was performed for expected squared weighted kappa values of 0.7, 0.8 and 0.9 for three raters and expected frequencies of six ordinal categories of 0.01, 0.02, 0.03, 0.05, 0.1 and 0.7<sup>12</sup>. A sample size of 120 images would allow estimation of the quadratic-weighted kappa value with 10% precision for an alpha value of 0.05 (20 cases for each of the six participating centers).

To assess the reliability (i.e. the degree of agreement between measurements within (intrarater) and between (inter-rater) raters) of quality assessment, the Fleiss-modified kappa statistic for ordinal scales with quadratic weighting was calculated using the procedure implemented in the R package 'raters'<sup>12</sup>, which avoids kappa paradoxes<sup>13</sup>. Confidence intervals were calculated by bootstrapping 10 000 replicates. Three matrices are involved: the matrix of observed scores; the matrix of expected scores based on chance agreement; and the

**Table 1** Description of criteria for quality assessment of umbilical artery (UA) and fetal middle cerebral artery (MCA) Doppler images

| Criterion                              | Description  |
|--|--|
| Anatomical site                        | Identification of circle of Willis, and pulsed-wave Doppler gate placed at proximal third of MCA<br>Insonation of free loop of cord for UA |
| Magnification                          | Vessel tract occupies at least 50% of screen   |
| Angle of insonation                    | Angle of insonation between vessel tract and Doppler beam < 15° for MCA and < 30° for UA   |
| Image clarity                          | Clear waveform, without artifacts, and accurate trace  |
| Sweep-speed adjustment                 | Sweep speed: 3–10 waveforms visualized   |
| Velocity scale and baseline adjustment | Waveforms occupy 75% of pulsed Doppler y-axis  |



**Figure 1** Distribution of quality scores for Doppler images of fetal middle cerebral artery (a) and umbilical artery (b).

weight matrix. Weight matrix cells located on the diagonal (upper-left to bottom-right) represent agreement and thus contain zero. Off-diagonal cells contain weights indicating the seriousness of the disagreement<sup>14</sup>. Kappa values were interpreted as follows: 0.4–0.6, moderate reliability; > 0.6–0.8, good reliability; > 0.8, very good reliability<sup>15</sup>.

The degree of agreement was further examined using limits of agreement and Bland–Altman analysis<sup>16</sup>, which allows calculation of the range in which 95% of the disagreement between observations is likely to occur and is defined as the mean difference (systematic error)  $\pm$  1.6 SD (random error).

Statistical analyses were performed using SPSS 13.1 (SPSS Inc., Chicago, IL, USA); R version 2.15.1 (R Foundation for Statistical Computing, Vienna, Austria) with the packages ggplot2 version 3.1.0 and raters 2.0.1; and GraphPad Prism version 6.0.1 (GraphPad Software Inc., La Jolla, CA, USA).

## RESULTS

Twenty women were selected randomly from each of the six participating centers, giving a total of 240 images (120 each for the MCA and UA) for quality assessment.

The distribution of average image quality score among the three raters who participated in the inter-rater reliability analysis is shown in Figure 1. On average, 89.2% of MCA images and 85.0% of UA images were rated as being of perfect (score of 6) or almost perfect (score of 5) quality. The proportion of images of each vessel meeting each quality criterion, overall and according to rater, is shown in Table 2 and Figure 2, respectively. Of note, velocity scale and baseline adjustment was the only item that did not meet the standard criterion of quality in more than 20% of the images for both the MCA and the UA.

Figure 3 shows the intra- and inter-rater kappa values with 95% CI. Kappa values for intrarater reliability of quality assessment were 0.90 (95% CI, 0.88–0.92) and 0.90 (95% CI, 0.88–0.93) for the MCA and UA,

**Table 2** Proportion of Doppler images of fetal middle cerebral artery (MCA) and umbilical artery (UA) meeting quality criteria as defined in Table 1, according to rater

| Criterion       | Rater 1<br>(n = 120) | Rater 2<br>(n = 120) | Rater 3<br>(n = 120) |
|-----------------|----------------------|----------------------|----------------------|
| <b>MCA</b>      |                      |                      |                      |
| Anatomical site | 115 (95.8)           | 118 (98.3)           | 118 (98.3)           |
| Magnification   | 112 (93.3)           | 86 (71.7)            | 116 (96.7)           |
| Angle           | 106 (88.3)           | 110 (91.7)           | 115 (95.8)           |
| Waveform        | 103 (85.8)           | 107 (89.2)           | 83 (69.2)            |
| Sweep speed     | 111 (92.5)           | 111 (92.5)           | 118 (98.3)           |
| Scale           | 115 (95.8)           | 62 (51.7)            | 92 (76.7)            |
| <b>UA</b>       |                      |                      |                      |
| Anatomical site | 109 (90.8)           | 113 (94.2)           | 117 (97.5)           |
| Magnification   | 99 (82.5)            | 88 (73.3)            | 116 (96.7)           |
| Angle           | 96 (80.0)            | 102 (85.0)           | 116 (96.7)           |
| Waveform        | 106 (88.3)           | 104 (86.7)           | 81 (67.5)            |
| Sweep speed     | 108 (90.0)           | 110 (91.7)           | 119 (99.2)           |
| Scale           | 112 (93.3)           | 71 (59.2)            | 100 (83.3)           |

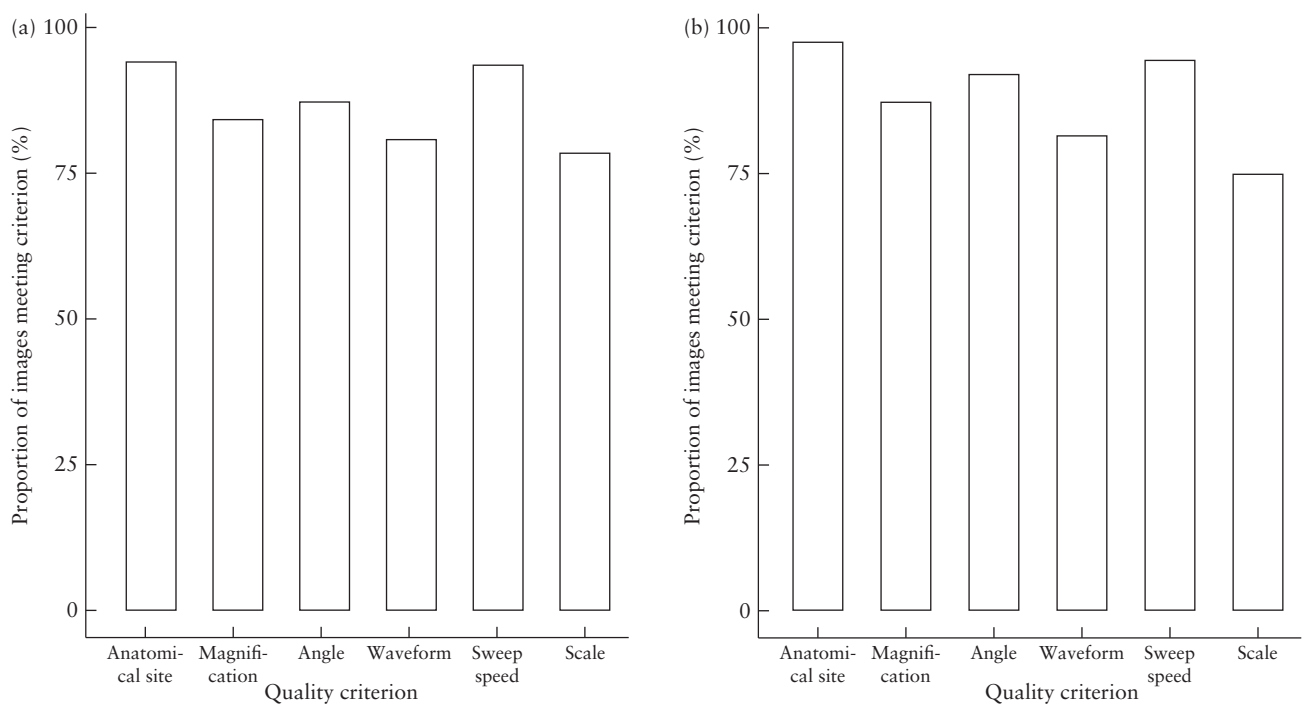
Data are given as *n* (%).

respectively. The corresponding inter-rater values were 0.85 (95% CI, 0.81–0.89) and 0.84 (95% CI, 0.80–0.89), respectively. Of note, for both vessels, inter- and intrarater reliability was found to be above 0.8, corresponding to very good agreement.

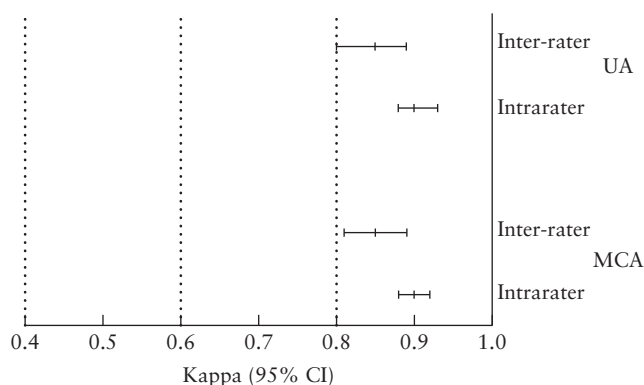
On average, the systematic and random errors of the differences in image quality score between raters (in absolute values) were 0.36 and 0.99 for MCA images and 0.32 and 0.96 for UA images, respectively. Figure S1 shows the paired Bland–Altman graphs.

## DISCUSSION

The implementation of an objective scale to evaluate the quality of ultrasound images is useful, especially in large multicenter trials, in which heterogeneity is a major issue. In this study, we have demonstrated that the scoring system is feasible and has almost perfect inter- and intrarater reliability. The Ratio37 study was designed as a pragmatic trial<sup>17</sup>, because the



**Figure 2** Proportion of Doppler images of fetal middle cerebral artery (a) and umbilical artery (b) meeting each quality criterion as defined in Table 1.



**Figure 3** Inter- and intrarater agreement in quality score for Doppler images of fetal middle cerebral artery (MCA) and umbilical artery (UA). Kappa cut-offs: 0.4–0.6, moderate reliability; > 0.6–0.8, good reliability; > 0.8, very good reliability.

main objective is to test the real-world effectiveness of the intervention (Doppler evaluation at 37 weeks) in broad patient groups from different practices and among different practitioners. Under this design, quality assurance is crucial to determine to what extent the findings are attributable to implementation issues. This control system allowed us to detect centers in which image quality was below the average standard and target them for improvement.

The vast majority of published studies on the reliability of ultrasound measurements have limitations in terms of design, reporting or interpretation<sup>18</sup>. Regarding the reliability of Doppler measurements of pulsatility index, we described previously the impact on inter-rater

reliability of different sampling sites in both the UA and the MCA. The reliability of MCA Doppler measurement was also assessed by Salvi *et al.*<sup>19</sup>. These studies were well designed and highlighted the importance of adherence to methodological recommendations to obtain reliable measurements. ISUOG set quality criteria for Doppler evaluation, aiming to improve its accuracy and reproducibility<sup>11</sup>. This quality can be assessed subjectively by judging a Doppler acquisition as acceptable or not. However, objective quality scoring would be more informative, particularly in a research setting in which interpretation of the results could be highly influenced by the quality of the measurements. A scoring system for MCA Doppler quality was developed by Ruiz-Martinez *et al.*<sup>9</sup> and later adapted for UA and uterine artery Doppler<sup>4</sup>. This objective assessment showed higher inter-rater reliability and agreement than did subjective evaluation. Intra-rater reliability was not reported. Of note, this study was performed in an ideal research setting, in which only one ultrasound machine model was used and only by certified sonographers, which confers good internal validity.

The strengths of our study are that it allowed objective assessment of image quality. All images were obtained by sonographers from a broad range of clinical practices, thus ensuring external validity of our findings, and were evaluated by experts in Doppler, some of whom were contributors to the ISUOG guidelines, which ensures internal validity of the study. A limitation of the study is that we did not evaluate all images from all centers, but only a small proportion of them. As women were selected randomly, we believe that the images are a good

representation of the overall population and that the effect of this limitation is therefore minimal. It should be stressed that this study did not aim to assess the reliability or agreement of Doppler measurements, but rather the agreement of a scoring system to evaluate the quality of the Doppler images. For the former objective, several measurements in the same waveform or repeated measurements in different waveforms would have been needed.

In conclusion, the quality of MCA and UA Doppler ultrasound images can be evaluated reliably using an objective scale. The quality scores of images obtained within a multicenter study by a broad range of operators were very high in over 85% of images. Intra- and inter-rater reliability were very good.

## REFERENCES

- Herman A, Maymon R, Dreazen E, Caspi E, Bukovsky I, Weinraub Z. Nuchal translucency audit: a novel image-scoring method. *Ultrasound Obstet Gynecol* 1998; 12: 398–403.
- 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* 2006; 27: 34–40.
- Wanyonyi SZ, Napolitano R, Ohuma EO, Salomon LJ, Papageorgiou AT. Image-scoring system for crown-rump length measurement. *Ultrasound Obstet Gynecol* 2014; 44: 649–654.
- Molloholli M, Napolitano R, Ohuma EO, Ash S, Wanyonyi SZ, Cavallaro A, Giudicepietro A, Barros F, Carvalho M, Norris S, Min AM, Zainab G, Papageorgiou AT. Image-scoring system for umbilical and uterine artery pulsed-wave Doppler ultrasound measurement. *Ultrasound Obstet Gynecol* 2019; 53: 251–255.
- Cavallaro A, Ash ST, Napolitano R, Wanyonyi S, Ohuma EO, Molloholli M, Sande J, Sarris I, Ioannou C, Norris T, Donadono V, Carvalho M, Purwar M, Barros FC, Jaffer, YA, Bertino E, Pang R, Gravett MG, Salomon LJ, Noble JA, Altman DG, Papageorgiou AT. Quality control of ultrasound for fetal biometry: results from the INTERGROWTH-21<sup>st</sup> Project. *Ultrasound Obstet Gynecol* 2018; 52: 332–339.
- Sarris I, Ioannou C, Ohuma E, Altman D, Hoch L, Cosgrove C, Fathima S, Salomon L, Papageorgiou A; International Fetal and Newborn Growth Consortium for the 21<sup>st</sup> Century (INTERGROWTH-21<sup>st</sup>). Standardisation and quality control of ultrasound measurements taken in the Intergrowth-21<sup>st</sup> Project. *BJOG* 2013; 120 (Suppl 2): 33–37.
- Westergaard HB, Langhoff-Roos J, Lingman G, Marsál K, Kreiner S. A critical appraisal of the use of umbilical artery Doppler ultrasound in high-risk pregnancies: use of meta-analyses in evidence-based obstetrics. *Ultrasound Obstet Gynecol* 2001; 17: 466–476.
- Khalil A, Morales-Roselló J, Townsend R, Morlando M, Papageorgiou A, Bhide A, Thilaganathan B. Value of third-trimester cerebroplacental ratio and uterine artery Doppler indices as predictors of stillbirth and perinatal loss. *Ultrasound Obstet Gynecol* 2016; 47: 74–80.
- Ruiz-Martinez S, Volpe G, Vannuccini S, Cavallaro A, Impey L, Ioannou C. An objective scoring method to evaluate image quality of middle cerebral artery Doppler. *J Matern Fetal Neonatal Med* 2020; 33: 421–426.
- Figueras F, Gratacos E, Rial M, Gull I, Krofta L, Lubusky M, Rogelio CM, Mónica CL, Miguel MR, Socias P, Aleuanlli C, Cordero MCP. Revealed versus concealed criteria for placental insufficiency in an unselected obstetric population in late pregnancy (RATIO37): Randomised controlled trial study protocol. *BMJ Open* 2017; 7: e014835.
- Bhide A, Acharya G, Bilardo CM, Brezinka C, Cafici D, Hernandez-Andrade E, Kalache K, Kingdom J, Kiserud T, Lee W, Lees C, Leung KY, Malinger G, Mari G, Prefumo F, Sepulveda W, Trudinger B. ISUOG practice guidelines: use of Doppler ultrasonography in obstetrics. *Ultrasound Obstet Gynecol* 2013; 41: 233–239.
- Marasini D, Quatto P, Ripamonti E. Assessing the inter-rater agreement for ordinal data through weighted indexes. *Stat Methods Med Res* 2016; 25: 2611–2633.
- Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990; 43: 551–558.
- Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. *Psychol Bull* 1969; 72: 323–327.
- Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Int J Nurs Stud* 2011; 48: 661–671.
- Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol* 2003; 22: 85–93.
- Roland M, Torgerson DJ. Understanding controlled trials: What are pragmatic trials? *BMJ* 2011; 316: 285–285.
- Coelho Neto MA, Roncato P, Nastro CO, Martins WP. True Reproducibility of UltraSound Techniques (TRUST): Systematic review of reliability studies in obstetrics and gynecology. *Ultrasound Obstet Gynecol* 2015; 46: 14–20.
- Salvi S, Badade A, Khatal K, Bhide A. Reliability of Doppler assessment of the middle cerebral artery in the near and far fields in healthy and anemic fetuses. *J Ultrasound Med* 2015; 34: 2037–2042.

## SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article:



**Figure S1** Bland–Altman graphs of difference between raters in quality score for Doppler images of fetal middle cerebral artery and umbilical artery. Solid lines represent mean difference and dotted lines represent 95% limits of agreement.