Ultrasound Obstet Gynecol 2015; **46**: 14–20
Published online 4 June 2015 in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.14654

True Reproducibility of UltraSound Techniques (TRUST): systematic review of reliability studies in obstetrics and gynecology

M. A. COELHO NETO*, P. RONCATO*†, C. O. NASTRI* and W. P. MARTINS*

**Department of Obstetrics and Gynecology, Ribeirao Preto Medical School, University of Sao Paulo (DGO-FMRP-USP), Ribeirao Preto, Brazil; †School of Health Technology – Ultrasonography School of Ribeirao Preto (FATESA-EURP), Ribeirao Preto, Brazil*

KEYWORDS: gynecology; obstetrics; reproducibility of results; ultrasound

ABSTRACT

Objectives *To examine the quality of methods used and the accuracy of the interpretation of agreement in existing studies that examine the reliability of ultrasound measurements and judgments in obstetrics and gynecology.*

Methods *A systematic search of MEDLINE was performed on 25 March 2014, looking for studies that examined the reliability of ultrasound measurements and judgments in obstetrics and gynecology with evaluation of concordance (CCC) or intraclass (ICC) correlation coefficients or kappa as a main objective.*

Results *Seven hundred and thirty-three records were examined on the basis of their title and abstract, of which 141 full-text articles were examined completely for eligibility. We excluded 29 studies because they did not report CCC/ICC/kappa, leaving 112 studies that were included in our analysis. Two studies reported both ICC and kappa and were counted twice, therefore, the number used as the denominator in the analyses was 114. Only 16/114 (14.0%) studies were considered to be well designed (independent acquisition and blinded analysis) and to have interpreted the results properly. Most errors occurring in the studies are likely to overestimate the reliability of the method examined.*

Conclusions *The vast majority of published studies examined had important flaws in design, interpretation and/or reporting. Such limitations are important to identify as they might create false confidence in the existing measurements and judgments, jeopardizing clinical practice and future research. Specific guidelines aimed at improving the quality of reproducibility studies that examine ultrasound methods should be encouraged.*

INTRODUCTION

Clinical measurements and ratings are crucial for a correct diagnosis, however, it must be borne in mind that such measurements and ratings are always subject to error, and simple repetition of the examination is likely to provide a different result[1]. Examining the relevance of such errors, or the reproducibility of a method, is important clinically because using non-reproducible methods poses a risk for patients and results in a waste of financial/human resources, leading, for example, to unnecessary research[2].

The reproducibility of a method is frequently examined by two different concepts: reliability and agreement. Reliability is the ability of a measurement/rating to differentiate among subjects, while agreement quantifies how similar the measurements or ratings are[1,3]. Although both concepts are important, we aimed to assess only reliability in this review, which is more comparable between methods, since agreement is usually assessed on the same scale as are the measurements themselves[1]. The most common means of assessing reliability uses kappa statistics for nominal/ordinal data and intraclass correlation coefficients (ICC) or concordance correlation coefficients (CCC) for continuous data[3–6]. Kappa is intended to give a quantitative measure of magnitude of agreement between observers, and its calculation is based on the difference between how much agreement is actually present (observed agreement) and how much agreement would be expected to be present by chance alone (expected agreement)[7]. ICC and CCC estimate the amount of total variance that might be attributed to the 'true' variance between subjects, while the difference between 1.0 and ICC/CCC is the proportion of the total variance that can be attributed to errors: for example, a value of 0.90 means that genuine differences between subjects was responsible for 90% of the total variance

Correspondence to: Prof. W. P. Martins, Av. Bandeirantes, 3900 – 8 Andar - HCRP - Campus Universitario, Ribeirao Preto, Sao Paulo, Brazil 14048–900 (e-mail: wpmartins@gmail.com)

Accepted: 22 August 2014

observed, while the other 10% was caused by errors during the measurement process[1,4–6]. CCC seems to be better for examining reliability; there are several types of ICC and choosing the appropriate method can be difficult[4]. Additionally, ICC is only valid when ANOVA assumptions are present (normality of distributions and homogeneity of variances), which does not occur frequently[5]. However, as can be seen in this review, ICC is used more often than CCC for this purpose.

Studies that evaluate the reproducibility of tests may help us to discriminate between the available instruments, leading to an informed decision when choosing one. However, these studies need to be well designed, well performed and properly interpreted, otherwise their results might be misleading[8]. One important issue regarding the quality of such studies is a tendency to overrate the reliability of the methods analyzed in order to facilitate publication. Some flaws that are frequently observed in reproducibility studies examining ultrasound techniques are:

- Not examining the variability caused by acquisition: if the same images, videoclip or three-dimensional dataset are examined more than once by the same or different observers this will overestimate reproducibility of the method since several inherent sources of variability (e.g. movements or pressure applied to the probe) are not taken into account[9,10].
- Not blinding the observers during acquisition/analysis: the knowledge of how the image was acquired and how it was analyzed tend to overestimate the reproducibility of the method, as the observer is more likely to repeat exactly what was just performed[2].
- Interpretation of the results: frequently researchers use low cut-off values, particularly for ICC/CCC, as this will facilitate an interpretation of good reproducibility[4].

The aim of this systematic review was to evaluate the proportion of studies that assess reproducibility of ultrasound techniques in obstetrics and gynecology that have errors during their execution (i.e. using the same acquisition or not blinding) or in the interpretation of reproducibility.

METHODS

Eligibility criteria and search strategy

Only studies examining the reliability of ultrasound measurements and judgments in obstetrics and gynecology that evaluated CCC, ICC or kappa as a main objective were considered eligible. We performed a systematic search of MEDLINE on 25 March 2014 using the following terms: (ICC OR CCC OR (intraclass correlation) OR (concordance correlation) OR kappa OR reliability OR reproducibility OR agreement) AND (ultrasound OR ultrasonography) AND (fetus* OR fetal OR embryo* OR uterus OR uterine OR ovary OR ovari*). We did not impose any limitations on publication date or language for the searches.

Study selection

Titles and abstracts of the search results were reviewed independently by two review authors (M.A.C.N. and W.P.M.), who checked for duplicates and used pre-established criteria for inclusion. Full-text manuscripts were evaluated for eligibility by two authors (M.A.C.N. and P.R.) and disagreements were resolved by consulting a third author (W.P.M.).

Data collection

Data were extracted from included studies using a data-extraction form, designed and pilot-tested by the authors, in a standardized manner by two authors (M.A.C.N. and P.R.). Disagreements were resolved by consulting a third author (W.P.M.). We extracted data regarding the main characteristics of the study based on the following questions: Was the ultrasound measurement fetal or non-fetal (if fetal, how dependent were the data on gestational age)? How was data acquisition performed (unclear, single acquisition for multiple interpretations, acquisition by different observers, different acquisitions by the same observer)? Was there blinding in the analysis? Which measurement was used to evaluate reliability (ICC/CCC/kappa)? What was the lowest value observed? How did the study authors interpret such values? Which kind of ICC was used (only for studies reporting ICC)? We considered the type of ICC to be reported properly when details on one-way random effects *vs* two-way random effects *vs* two-way mixed effects, single measurements *vs* average measurements and consistency *vs* absolute agreement were included[11].

Synthesis of results

We assessed the proportion of studies in which: (1) observers performed independent acquisition; (2) observers were blinded to the results; (3) authors interpreted the results properly; and (4) authors reported the type of ICC that was used. The precision of the estimates was assessed by their 95% CIs. When assessing whether authors had interpreted their results properly, we compared the interpretation reported by the authors with that of studies that used published cut-off values. For interpretation of ICC and CCC values, we used recently published cut-off values: ICC/CCC < 0.70, very poor; 0.70–0.90, poor; 0.90–0.95, moderate; 0.95–0.99, good; and > 0.99, very good[4]. These values were chosen because they are conservative and in agreement with a recently published guideline for reporting reproducibility studies[3]. Although the use of different cut-off values is suggested when examining fetal measurements that are highly dependent on gestational age, owing to the large variability between individuals[4], we used the same cut-off points for these studies but planned to judge the interpretation as being proper even when these studies underrated the observed reliability. For the interpretation of kappa, we used the following cut-off values: < 0.20, very poor;

0.21–0.40, poor; 0.41–0.60, moderate; 0.61–0.80, good; and ≥ 0.81, very good[3].

Studies that interpreted their observed reproducibility statistics in agreement with the terminology and cut-offs that we have listed were considered to have satisfactory interpretation. The number of levels of difference between the authors' interpretation and the interpretation following from use of the abovementioned cut-off points was evaluated in the analysis (i.e. for levels defined as: Level 1, very poor; Level 2, poor; Level 3, moderate; Level 4, good; Level 5, very good).

RESULTS

Study selection

The last electronic search was performed of MEDLINE on 25 March 2014. We examined 733 records on the basis of abstract and title and excluded 592 records because they clearly did not meet eligibility criteria. One hundred and forty-one full-text articles were completely examined for eligibility; 29 were excluded because they did not report CCC/ICC/kappa[12–40]. We included 112 full-text articles in our analysis[41–152]; two reported both ICC and kappa and were counted twice[55,56]. Therefore, the number of included studies used as the denominator in the analyses was 114.

Main results

The included studies were heterogeneous with regard to the main outcome measure, its design and how the interpretation of its results had been performed (Table 1). Regarding the main outcome of the study, 73 (64.0%) studies examined fetal measurements, of which 28 (38.4%) were considered as very dependent on gestational age and the remaining 45 (61.6%) as not very dependent, and 41 studies (36.0%) examined non-fetal measurements. Concerning acquisition of measurements, the majority of studies ($n = 62$; 54.4%) used the same acquisition to examine reliability, nine (7.9%) used different acquisitions performed by the same observer, 40 (35.1%) used completely independent acquisitions, and acquisition was unclear in three (2.6%). The majority ($n = 71$; 62.3%) of the included studies performed blinded analysis, while analysis was not blinded in five (4.4%) and the use of blinding was unclear in the remaining 38 (33.3%). Regarding interpretation of reliability, a large proportion of the studies overrated the observed reliability ($n = 43$; 37.7%), 59 (51.8%) performed a proper interpretation, 12 (10.5%) did not interpret their results and no study underrated the observed reliability. We observed only 16 (14.0%) studies that had independent acquisitions and blinded analysis, and interpreted their results properly.

Additional analyses

Further analysis of the studies identified only one study that reported CCC; it did not interpret its results. Concerning the 32 studies that examined kappa, 25 (78.1%) interpreted their results properly, five (15.6%) overrated the observed reliability and two (6.3%) did not interpret their results (Table 2). Regarding the 81 studies that reported ICC, 34 (42.0%) interpreted their

Table 1 Characteristics of 114 studies examining reproducibility of ultrasound techniques in obstetrics and gynecology

Characteristic	n *(% (95% CI))*
Study outcome	
Fetal measurement highly dependent on gestational age	28 (24.6 (17.6–33.2))
Other fetal measurement	45 (39.5 (31.0–48.6))
Non-fetal measurement	41 (36.0 (27.7–45.1))
Acquisition of measurement	
Single acquisition for multiple interpretations	62 (54.4 (45.2–63.2))
Different acquisitions by same observer	9 (7.9 (4.2–14.3))
Completely independent acquisitions	40 (35.1 (26.9–44.2))
Unclear	3 (2.6 (0.9–7.5))
Analysis	
Blinded	71 (62.3 (53.1–70.6))
Not blinded	5 (4.4 (1.9–9.9))
Unclear	38 (33.3 (25.3–42.3))
Interpretation of results	
Overrated observed reliability	43 (37.7 (29.4–46.9))
Proper interpretation	59 (51.8 (42.7–60.7))
No interpretation	12 (10.5 (6.1–17.5))
Independent acquisitions, blinded analyses, proper interpretation	16 (14.0 (8.8–21.6))

Table 2 Interpretation of reliability in studies examining reproducibility of ultrasound techniques that reported kappa and intraclass correlation coefficient (ICC)

Interpretation of reliability	n/N *(% (95% CI))*
Interpretation of kappa	
Satisfactory	25/32 (78.1 (61.2–89.0))
Overrated	5/32 (15.6 (6.9–31.8))*
No interpretation	2/32 (6.3 (1.7–20.1))
Interpretation of ICC	
Satisfactory	34/81 (42.0 (31.8–52.8))
Overrated	38/81 (46.9 (36.4–57.7))†
No interpretation	9/81 (11.1 (6.0–19.8))
Type of ICC	
Not reported	58/81 (71.6 (61.0–80.3))
Reported partially	21/81 (25.9 (17.6–36.4))
Reported completely	2/81 (2.5 (0.7–8.6))

Only one study examined concordance correlation coefficient; it did not provide any interpretation. *One study overrated reproducibility by three levels (examined measurements highly dependent on gestational age) and four studies overrated reproducibility by one level (non-fetal measurements ($n = 1$), other fetal measurements ($n = 3$)). †Eight studies overrated reproducibility by three levels (measurements highly dependent on gestational age ($n = 1$), other fetal measurements ($n = 3$), non-fetal measurements ($n = 4$)); 12 studies overrated reproducibility by two levels (measurements highly dependent on gestational age ($n = 2$), other fetal measurements ($n = 5$), non-fetal measurements ($n = 5$)); 18 studies overrated reproducibility by one level (measurements highly dependent on gestational age ($n = 4$), other fetal measurements ($n = 8$), non-fetal measurements ($n = 6$)).

results properly, 38 (46.9%) overrated the observed reliability and nine (11.1%) did not interpret their results. Regarding the type of ICC used in the analysis, only two (2.5%) studies reported fully the type of ICC used, while 58 (71.6%) did not report the type of ICC used and 21 (25.9%) partially reported the type of ICC used.

DISCUSSION

Summary of evidence

Only 16 (14.0%) of the 114 included studies were considered to be well-designed (independent acquisition and blinded analysis) and to have interpreted the results properly. The majority of studies that used ICC to examine reliability did not provide a complete description of the type of ICC used in the analysis.

Limitations

This review has some limitations. One limitation includes the restriction of the literature search to only one database (MEDLINE), but we believe that, although the search was not comprehensive, we were able to obtain an unbiased and representative sample of the eligible published studies. Additionally, we used arbitrary cut-off values for the interpretation of kappa, ICC and CCC, which are not widely accepted.

Interpretation

Most of the recently published studies examining the reliability of ultrasound techniques in obstetrics and gynecology contain some flaws in design or interpretation that are likely to overrate the 'true' reproducibility of these methods. Another important point that should be mentioned is that most studies did not report properly the type of ICC used during their analysis; choosing the wrong ICC (e.g. consistency instead of agreement or average measures instead of single measures) will result in higher ICC values. Additionally, most of the reproducibility studies were performed by experts, in academic centers, using the same machine settings; all these factors are likely to lead to better results than those that would be observed in regular clinical practice. As a consequence, we are prone to feeling overconfident when applying these methods in practice.

Another issue that raises concern is that a large number of reliability studies in obstetrics and gynecology evaluated fetal measurements that were highly dependent on gestational age. In these studies, the authors usually used the absolute observed measurement to assess ICC/CCC and, as the real variability among fetuses can be high, interpretation of these results is easily overrated[4]. A better approach to compensate for the high natural variability is to assess the reliability of the percentiles of the fetal measurements, for a given gestational age. The calculation of fetal measurement ICC/CCC based on percentile enables them to be comparable with ICC/CCC from other types of measurement, and the same cut-off values could be applied. Additionally, percentiles of fetal measurements are used widely in clinical practice, and are easier to use in comparisons. However, despite these advantages, none of the included studies examined the reproducibility of the percentile; we believe the reason for not using the percentile is the need for transforming the absolute measurements into percentiles, which requires proper reference curves and some data manipulation.

One important point to be considered concerns the potential causes of faults in the study design and the tendency to overestimate the reproducibility results. Regarding the study design, performing multiple analyses using only the same acquisition is much simpler and would avoid intensive or time-consuming examinations for both patient and sonographer; additionally such studies can be performed retrospectively, using stored images or datasets. The problem lies with acquisition as a potential source of variability, and the estimates obtained when examining the same acquisition will not represent the total variability that would be observed by repeating the examination; this is the most important factor to consider when making decisions. Regarding errors in interpretation, we have two main hypotheses for the cause: first, ICC values of > 0.60 or > 0.80 are used frequently as the minimum standard for reliability coefficients, and researchers often consider a value over such a limit as indicating good reliability. However, this threshold is only valid when considering the method for research purposes; ICC values should be > 0.90 or > 0.95 to consider employing the method in clinical practice[3]. Second, researchers might fear that they will encounter difficulties in publishing their study if they report poor reproducibility of the method, as it is less likely to be used in clinical practice.

Conclusions

The great majority of the published studies have important flaws in study design, interpretation and/or reporting. Such problems are important, as they might create false confidence in existing measurements and judgments, jeopardizing clinical practice and future research. Physicians and patients might use misleading information when choosing the best diagnostic test; moreover, they may assume a diagnosis or make clinically important decisions, such as surgery, based on the results of unreliable methods. Researchers can spend a lot of resources in studies examining unreliable methods for either the diagnosis or prediction of a disease and may postpone or have less interest in new studies that examine the impact of technical refinements aimed at improving the reliability of a method. Specific guidelines to improve the quality of reproducibility studies examining ultrasound methods should be encouraged.

REFERENCES

1. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol* 2008; **31**: 466–475.
2. Khan KS, Chien PF. Evaluation of a clinical test. I: assessment of reliability. *BJOG* 2001; **108**: 562–567.
3. 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.
4. Martins WP, Nastri CO. Interpreting reproducibility results for ultrasound measurements. *Ultrasound Obstet Gynecol* 2014; **43**: 479–480.
5. Chen CC, Barnhart HX. Comparison of ICC and CCC for assessing agreement for data without and with replications. *Comput Stat Data Anal* 2008; **53**: 554–564.
6. McBride GB. A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. *NIWA Client Report* 2005; HAM 2005–062: National Institute of Water & Atmospheric Research Ltd: Hamilton, New Zealand, May 2005.
7. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005; **37**: 360–363.
8. Martins WP, Raine-Fenning NJ. Analysis and acquisition reproducibility of 3D power Doppler. *Ultrasound Obstet Gynecol* 2010; **36**: 392–393; author reply 393–394.
9. Martins WP. Comparing reliability between different ultrasound techniques. *Ultrasound Obstet Gynecol* 2012; **39**: 482–485; author reply 485–486.
10. Martins WP, Nastri CO. Reproducibility of 3D power Doppler placental vascular indices. *Arch Gynecol Obstet* 2011; **283**: 403–404.
11. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996; **1**: 30–46.
12. Akmal S, Tsoi E, Nicolaides KH. Intrapartum sonography to determine fetal occipital position: interobserver agreement. *Ultrasound Obstet Gynecol* 2004; **24**: 421–424.
13. Amer A, Hammadeh ME, Kolkailah M, Ghandour AA. Three-dimensional versus two-dimensional ultrasound measurement of follicular volume: are they comparable? *Arch Gynecol Obstet* 2003; **268**: 155–157.
14. Amso NN, Watermeyer SR, Pugh N, O'Brien S, D'Angelo A. Quantification of power Doppler energy and its future potential. *Fertil Steril* 2001; **76**: 583–587.
15. Barbera AF, Pombar X, Perugino G, Lezotte DC, Hobbins JC. A new method to assess fetal head descent in labor with transperineal ultrasound. *Ultrasound Obstet Gynecol* 2009; **33**: 313–319.
16. Blaas HG, Taipale P, Torp H, Eik-Nes SH. Three-dimensional ultrasound volume calculations of human embryos and young fetuses: a study on the volumetry of compound structures and its reproducibility. *Ultrasound Obstet Gynecol* 2006; **27**: 640–646.
17. Boito SM, Laudy JA, Struijk PC, Stijnen T, Wladimiroff JW. Three-dimensional US assessment of hepatic volume, head circumference, and abdominal circumference in healthy and growth-restricted fetuses. *Radiology* 2002; **223**: 661–665.
18. Bojikian KD, de Moura CR, Tavares IM, Leite MT, Moron AF. Fetal ocular measurements by three-dimensional ultrasound. *J AAPOS* 2013; **17**: 276–281.
19. Brett S, Bee N, Wallace WH, Rajkhowa M, Kelsey TW. Individual ovarian volumes obtained from 2-dimensional and 3-dimensional ultrasound lack precision. *Reprod Biomed Online* 2009; **18**: 348–351.
20. Bujold E, Effendi M, Girard M, Gouin K, Forest JC, Couturier B, Giguere Y. Reproducibility of first trimester three-dimensional placental measurements in the evaluation of early placental insufficiency. *J Obstet Gynaecol Can* 2009; **31**: 1144–1148.
21. Chan LW, Fung TY, Leung TY, Sahota DS, Lau TK. Volumetric (3D) imaging reduces inter- and intraobserver variation of fetal biometry measurements. *Ultrasound Obstet Gynecol* 2009; **33**: 447–452.
22. Chang FM, Hsu KF, Ko HC, Yao BL, Chang CH, Yu CH, Liang RI, Chen HY. Fetal heart volume assessment by three-dimensional ultrasound. *Ultrasound Obstet Gynecol* 1997; **9**: 42–48.
23. Chen M, Lee CP, Leung KY, Hui PW, Tang MH. Pilot study on the midsecond trimester examination of fetal nasal bone in the Chinese population. *Prenat Diagn* 2004; **24**: 87–91.
24. Chen M, Wang HF, Leung TY, Fung TY, Chan LW, Sahota DS, Lao TH, Lau TK. First trimester measurements of nasal bone length using three-dimensional ultrasound. *Prenat Diagn* 2009; **29**: 766–770.
25. De Jong-Pleij EA, Ribbert LS, Tromp E, Bilardo CM. Three-dimensional multiplanar ultrasound is a valuable tool in the study of the fetal profile in the second trimester of pregnancy. *Ultrasound Obstet Gynecol* 2010; **35**: 195–200.
26. Deb S, Campbell BK, Clewes JS, Raine-Fenning NJ. Quantitative analysis of antral follicle number and size: a comparison of two-dimensional and automated three-dimensional ultrasound techniques. *Ultrasound Obstet Gynecol* 2010; **35**: 354–360.
27. Deb S, Kannamannadiar J, Campbell BK, Clewes JS, Raine-Fenning NJ. The interovarian variation in three-dimensional ultrasound markers of ovarian reserve in women undergoing baseline investigation for subfertility. *Fertil Steril* 2011; **95**: 667–672.
28. Herman A, Maymon R, Dreazen E, Caspi E, Bukovsky I, Weinraub Z. Image magnification does not contribute to the repeatability of caliper placement in measuring nuchal translucency thickness. *Ultrasound Obstet Gynecol* 1998; **11**: 266–270.
29. Jurkovic D. Re: Intra- and interobserver agreement when describing adnexal masses using the International Ovarian Tumor Analysis terms and definitions: a study on three-dimensional ultrasound volumes. P. Sladkevicius and L. Valentin. Ultrasound Obstet Gynecol 2013; 41: 318–327. *Ultrasound Obstet Gynecol* 2013; **41**: 245–246.
30. Lazanakis M, Marsh M, Brockbank E, Economides D. Assessment of the cervix in the third trimester of pregnancy using transvaginal ultrasound scanning. *Eur J Obstet Gynecol Reprod Biol* 2002; **105**: 31–35.
31. Liang RI, Prapas N, Detti L, Cosmi E, Copel JA, Mari G. Assessment of blood flow velocity waveforms of the pulmonary circulation by multigate spectral Doppler scanning and traditional pulsed Doppler ultrasonography. *J Ultrasound Med* 2002; **21**: 31–37.
32. Mari G, Abuhamad AZ, Cosmi E, Segata M, Altaye M, Akiyama M. Middle cerebral artery peak systolic velocity: technique and variability. *J Ultrasound Med* 2005; **24**: 425–430.
33. Paul C, Krampl E, Skentou C, Jurkovic D, Nicolaides KH. Measurement of fetal nuchal translucency thickness by three-dimensional ultrasound. *Ultrasound Obstet Gynecol* 2001; **18**: 481–484.
34. Peixoto-Filho FM, Sa RA, Lopes LM, Velarde LG, Marchiori E, Ville Y. Three-dimensional ultrasound fetal urinary bladder volume measurement: reliability of rotational (VOCAL) technique using different steps of rotation. *Arch Gynecol Obstet* 2007; **276**: 345–349.
35. Prapas N, Mari G, Liang RI, Copel JA. Assessment of Doppler velocimetry of the fetal umbilical artery by multigate spectral Doppler scanning and traditional pulsed Doppler ultrasonography plus color flow mapping. *J Ultrasound Med* 1999; **18**: 831–835.
36. Reid S, Lu C, Casikar I, Mein B, Magotti R, Ludlow J, Benzie R, Condous G. The prediction of pouch of Douglas obliteration using offline analysis of the transvaginal ultrasound 'sliding sign' technique: inter- and intra-observer reproducibility. *Hum Reprod* 2013; **28**: 1237–1246.
37. Sabogal JC, Becker E, Bega G, Komwilaisak R, Berghella V, Weiner S, Tolosa J. Reproducibility of fetal lung volume measurements with 3-dimensional ultrasonography. *J Ultrasound Med* 2004; **23**: 347–352.
38. Sarris I, Ioannou C, Chamberlain P, Ohuma E, Roseman F, Hoch L, Altman DG, Papageorghiou AT. Intra- and interobserver variability in fetal ultrasound measurements. *Ultrasound Obstet Gynecol* 2012; **39**: 266–273.
39. Scheffer GJ, Broekmans FJ, Dorland M, Habbema JD, Looman CW, te Velde ER. Antral follicle counts by transvaginal ultrasonography are related to age in women with proven natural fertility. *Fertil Steril* 1999; **72**: 845–851.
40. Shrimali V, Anand RS, Kumar V. Improved segmentation of ultrasound images for fetal biometry, using morphological operators. *Conf Proc IEEE Eng Med Biol Soc* 2009; **2009**: 459–462.
41. Abele H, Hoopmann M, Wright D, Hoffmann-Poell B, Huettelmaier M, Pintoffl K, Wallwiener D, Kagan KO. Intra- and interoperator reliability of manual and semi-automated measurement of fetal nuchal translucency by sonographers with different levels of experience. *Ultrasound Obstet Gynecol* 2010; **36**: 417–422.
42. Alcazar JL, Laparte C. In vivo validation of the time domain velocity measurement technique of blood flow in human fetuses. *Ultrasound Med Biol* 1998; **24**: 9–13.
43. Alcázar JL, Mercé LT, Manero MG, Bau S, López-García G. Endometrial volume and vascularity measurements by transvaginal 3-dimensional ultrasonography and power Doppler angiography in stimulated and tumoral endometria: an interobserver reproducibility study. *J Ultrasound Med* 2005; **24**: 1091–1098.
44. Amer SA, Li TC, Bygrave C, Sprigg A, Saravelos H, Cooke ID. An evaluation of the inter-observer and intra-observer variability of the ultrasound diagnosis of polycystic ovaries. *Hum Reprod* 2002; **17**: 1616–1622.
45. Araujo Junior E, Nardozza LM, Rodrigues Pires C, Filho HA, Moron AF. Comparison of two- and three-dimensional ultrasonography in lung volume measurement of normal fetuses. *J Perinat Med* 2007; **35**: 415–421.
46. Araujo Júnior E, Martinez LH, Simioni C, Martins WP, Nardozza LM, Moron AF. Delineation of vertebral area on the coronal plane using three-dimensional ultrasonography advanced volume contrast imaging (VCI) Omni view: intrarater reliability and agreement using standard mouse, high definition mouse, and pen-tablet. *J Matern Fetal Neonatal Med* 2012; **25**: 1818–1821.
47. Ata B, Seyhan A, Reinblatt SL, Shalom-Paz E, Krishnamurthy S, Tan SL. Comparison of automated and manual follicle monitoring in an unrestricted population of 100 women undergoing controlled ovarian stimulation for IVF. *Hum Reprod* 2011; **26**: 127–133.
48. Bagheban AA, Zayeri F, Anaraki FB, Elahipanah Z. The reliability and distinguishability of ultrasound diagnosis of ovarian masses. *Indian J Med Sci* 2008; **62**: 217–221.
49. Barra DA, Lima JC, Mauad Filho F, Araujo E Jr, Martins WP. Measuring fetal volume during late first trimester by three-dimensional ultrasonography using virtual organ computer-aided analysis. *Ultrasound Med Biol* 2013; **39**: 1552–1559.
50. Benavides-Serralde A, Hernández-Andrade E, Fernández-Delgado J, Plasencia W, Scheier M, Crispi F, Figueras F, Nicolaides KH, Gratacós E. Three-dimensional sonographic calculation of the volume of intracranial structures in growth-restricted and appropriate-for-gestational age fetuses. *Ultrasound Obstet Gynecol* 2009; **33**: 530–537.
51. Beninni JR, Faro C, Marussi EF, Barini R, Peralta CF. Fetal thigh volumetry by three-dimensional ultrasound: comparison between multiplanar and VOCAL techniques. *Ultrasound Obstet Gynecol* 2010; **35**: 417–425.
52. Bhaduri M, Fong K, Toi A, Tomlinson G, Okun N. Fetal anatomic survey using three-dimensional ultrasound in conjunction with first-trimester nuchal translucency screening. *Prenat Diagn* 2010; **30**: 267–273.
53. Bordes A, Bory AM, Benchaïb M, Rudigoz RC, Salle B. Reproducibility of transvaginal three-dimensional endometrial volume measurements with virtual organ computer-aided analysis (VOCAL) during ovarian stimulation. *Ultrasound Obstet Gynecol* 2002; **19**: 76–80.
54. Borrell A, Perez M, Figueras F, Meler E, Gonce A, Gratacos E. Reliability analysis on ductus venosus assessment at 11–14 weeks' gestation in a high-risk population. *Prenat Diagn* 2007; **27**: 442–446.
55. Borrell A, Costa D, Delgado RD, Martinez JM, Borrell C, Fortuny A. Interobserver variability of midtrimester fetal nuchal thickness. *Eur J Obstet Gynecol Reprod Biol* 1997; **72**: 27–29.

56. Boutin A, Jastrow N, Girard M, Roberge S, Chaillet N, Brassard N, Bujold E. Reliability of two-dimensional transvaginal sonographic measurement of lower uterine segment thickness using video sequences. *Am J Perinatol* 2012; **29**: 527–532.
57. Brodszki J, Gardiner HM, Eriksson A, Stale H, Marsal K. Reproducibility of ultrasonic fetal volume blood flow measurements. *Clin Physiol* 1998; **18**: 479–485.
58. Chang FM, Hsu KF, Ko HC, Yao BL, Chang CH, Yu CH, Chen HY. Three-dimensional ultrasound assessment of fetal liver volume in normal pregnancy: a comparison of reproducibility with two-dimensional ultrasound and a search for a volume constant. *Ultrasound Med Biol* 1997; **23**: 381–389.
59. Chang CH, Chang FM, Yu CH, Ko HC, Chen HY. Three-dimensional ultrasound in the assessment of fetal cerebellar transverse and antero-posterior diameters. *Ultrasound Med Biol* 2000; **26**: 175–182.
60. Chen CY, Su HW, Pai SH, Hsieh CW, Jong TL, Hsu CS, Chou SY. Evaluation of placental maturity by the sonographic textures. *Arch Gynecol Obstet* 2011; **284**: 13–18.
61. Cheong KB, Leung KY, Chan HY, Lee YP, Yang F, Tang MH. Comparison of inter- and intraobserver agreement between three types of fetal volume measurement technique (XI VOCAL, VOCAL and multiplanar). *Ultrasound Obstet Gynecol* 2009; **33**: 287–294.
62. Cheong KB, Leung KY, Li TK, Chan HY, Lee YP, Tang MH. Comparison of inter- and intraobserver agreement and reliability between three different types of placental volume measurement technique (XI VOCAL, VOCAL and multiplanar) and validity in the in-vitro setting. *Ultrasound Obstet Gynecol* 2010; **36**: 210–217.
63. Collins SL, Birks JS, Stevenson GN, Papageorghiou AT, Noble JA, Impey L. Measurement of spiral artery jets: general principles and differences observed in small-for-gestational-age pregnancies. *Ultrasound Obstet Gynecol* 2012; **40**: 171–178.
64. D'Arcy TJ, Jayaram V, Lynch M, Soutter WP, Cosgrove DO, Harvey CJ, Patel N. Ovarian cancer detected non-invasively by contrast-enhanced power Doppler ultrasound. *BJOG* 2004; **111**: 619–622.
65. DeKoninck P, Steenhaut P, Van Mieghem T, Mhallem M, Richter J, Bernard P, De Catte L, Deprest J. Comparison of Doppler-based and three-dimensional methods for fetal cardiac output measurement. *Fetal Diagn Ther* 2012; **32**: 72–78.
66. Delisle MF, Villeneuve M, Boulvain M. Measurement of endometrial thickness with transvaginal ultrasonography: is it reproducible? *J Ultrasound Med* 1998; **17**: 481–484; quiz 485–486.
67. Deurloo K, Spreeuwenberg M, Rekoert-Hollander M, van Vugt J. Reproducibility of 3-dimensional sonographic measurements of fetal and placental volume at gestational ages of 11–18 weeks. *J Clin Ultrasound* 2007; **35**: 125–132.
68. Dietz HP, Lanzarone V. Measuring engagement of the fetal head: validity and reproducibility of a new ultrasound technique. *Ultrasound Obstet Gynecol* 2005; **25**: 165–168.
69. Duin LK, Willekes C, Vossen M, Beckers M, Offermans J, Nijhuis JG. Reproducibility of fetal renal pelvis volume measurement using three-dimensional ultrasound. *Ultrasound Obstet Gynecol* 2008; **31**: 657–661.
70. Duin LK, Willekes C, Vossen M, Offermans J, Nijhuis JG. Reproducibility of fetal renal pelvis volume assessed by three-dimensional ultrasonography with two different measurement techniques. *J Clin Ultrasound* 2013; **41**: 230–234.
71. Emanuel MH, Ankum WM, Verdel MJ, Hart AA. The reproducibility of the results of transvaginal sonography of the uterus in patients with abnormal uterine bleeding. *Ultrasound Obstet Gynecol* 1996; **8**: 346–349.
72. Fouron JC, Proulx F, Miro J, Gosselin J. Doppler and M-mode ultrasonography to time fetal atrial and ventricular contractions. *Obstet Gynecol* 2000; **96**: 732–736.
73. García-López MA, Bermúdez-Rojas Mde L, Díaz-Aguiar E. [Modified Tei index, interobserver agreement related to sonographer experience level]. *Ginecol Obstet Mex* 2011; **79**: 107–115.
74. Ghi T, Contro E, Farina A, Nobile M, Pilu G. Three-dimensional ultrasound in monitoring progression of labor: a reproducibility study. *Ultrasound Obstet Gynecol* 2010; **36**: 500–506.
75. Guerriero S, Alcazar JL, Pascual MA, Ajossa S, Gerada M, Bargellini R, Virgilio B, Melis GB. Diagnosis of the most frequent benign ovarian cysts: is ultrasonography accurate and reproducible? *J Womens Health (Larchmt)* 2009; **18**: 519–527.
76. Guerriero S, Pilloni M, Alcazar JL, Sedda F, Ajossa S, Mais V, Melis GB, Saba L. Tissue characterization using mean gray value analysis in deep infiltrating endometriosis. *Ultrasound Obstet Gynecol* 2013; **41**: 459–464.
77. Hernandez-Andrade E, Thuring-Jonsson A, Jansson T, Lingman G, Marsal K. Fractional moving blood volume estimation in the fetal lung using power Doppler ultrasound: a reproducibility study. *Ultrasound Obstet Gynecol* 2004; **23**: 369–373.
78. Hernandez-Andrade E, Lopez-Tenorio J, Figueroa-Diesel H, Sanin-Blair J, Carreras E, Cabero L, Gratacos E. A modified myocardial performance (Tei) index based on the use of valve clicks improves reproducibility of fetal left cardiac function assessment. *Ultrasound Obstet Gynecol* 2005; **26**: 227–232.
79. Ioannou C, Sarris I, Javaid MK, Papageorghiou AT. Sphenoidal fontanelle area measurement on rendered three-dimensional ultrasound. *Prenat Diagn* 2012; **32**: 592–595.
80. Jarvela IY, Sladkevicius P, Tekay AH, Campbell S, Nargund G. Intraobserver and interobserver variability of ovarian volume, gray-scale and color flow indices obtained using transvaginal three-dimensional power Doppler ultrasonography. *Ultrasound Obstet Gynecol* 2003; **21**: 277–282.
81. Jastrow N, Antonelli E, Robyr R, Irion O, Boulvain M. Inter- and intraobserver variability in sonographic measurement of the lower uterine segment after a previous Cesarean section. *Ultrasound Obstet Gynecol* 2006; **27**: 420–424.
82. Jayaprakasan K, Walker KF, Clewes JS, Johnson IR, Raine-Fenning NJ. The interobserver reliability of off-line antral follicle counts made from stored three-dimensional ultrasound data: a comparative study of different measurement techniques. *Ultrasound Obstet Gynecol* 2007; **29**: 335–341.
83. Jayaprakasan K, Campbell BK, Clewes JS, Johnson IR, Raine-Fenning NJ. Three-dimensional ultrasound improves the interobserver reliability of antral follicle counts and facilitates increased clinical work flow. *Ultrasound Obstet Gynecol* 2008; **31**: 439–444.
84. Jouannic JM, Rosenblatt J, Demaria F, Jacobs R, Aubry MC, Benifla JL. Contribution of three-dimensional volume contrast imaging to the sonographic assessment of the fetal uterus. *Ultrasound Obstet Gynecol* 2005; **26**: 567–570.
85. Kalache KD, Espinoza J, Chaiworapongsa T, Londono J, Schoen ML, Treadwell MC, Lee W, Romero R. Three-dimensional ultrasound fetal lung volume measurement: a systematic study comparing the multiplanar method with the rotational (VOCAL) technique. *Ultrasound Obstet Gynecol* 2003; **21**: 111–118.
86. Kist WJ, Slaghekke F, Papanna R, Johnson A, Vandenbussche FP, Wolterbeek R, Oepkes D. Sonography-based Automated Volume Count to estimate fetal urine production in twin-to-twin transfusion syndrome: comparison with Virtual Organ Computer-aided AnaLysis. *Am J Obstet Gynecol* 2011; **205**: 574.e1–5.
87. Kusanovic JP, Nien JK, Goncalves LF, Espinoza J, Lee W, Balasubramaniam M, Soto E, Erez O, Romero R. The use of inversion mode and 3D manual segmentation in volume measurement of fetal fluid-filled structures: comparison with Virtual Organ Computer-aided AnaLysis (VOCAL). *Ultrasound Obstet Gynecol* 2008; **31**: 177–186.
88. Lee C, Salim R, Ofili-Yebovi D, Yazbek J, Davies A, Jurkovic D. Reproducibility of the measurement of submucous fibroid protrusion into the uterine cavity using three-dimensional saline contrast sonohysterography. *Ultrasound Obstet Gynecol* 2006; **28**: 837–841.
89. Lee W, DeVore GR, Comstock CH, Kalache KD, McNie B, Chaiworapongsa T, Conoscenti G, Treadwell MC, Johnson A, Huang R, Romero R. Nasal bone evaluation in fetuses with Down syndrome during the second and third trimesters of pregnancy. *J Ultrasound Med* 2003; **22**: 55–60.
90. Lima JC, Miyague AH, Filho FM, Nastri CO, Martins WP. Biometry and fetal weight estimation by two-dimensional and three-dimensional ultrasonography: an intraobserver and interobserver reliability and agreement study. *Ultrasound Obstet Gynecol* 2012; **40**: 186–193.
91. Lujan ME, Chizen DR, Peppin AK, Kriegler S, Leswick DA, Bloski TG, Pierson RA. Improving inter-observer variability in the evaluation of ultrasonographic features of polycystic ovaries. *Reprod Biol Endocrinol* 2008; **6**: 30.
92. Marret H, Sauget S, Giraudeau B, Body G, Tranquart F. Power Doppler vascularity index for predicting malignancy of adnexal masses. *Ultrasound Obstet Gynecol* 2005; **25**: 508–513.
93. Martins WP, Barra DA, Gallarreta FM, Nastri CO, Filho FM. Lower uterine segment thickness measurement in pregnant women with previous Cesarean section: reliability analysis using two- and three-dimensional transabdominal and transvaginal ultrasound. *Ultrasound Obstet Gynecol* 2009; **33**: 301–306.
94. Martins WP, Raine-Fenning NJ, Leite SP, Ferriani RA, Nastri CO. A standardized measurement technique may improve the reliability of measurements of endometrial thickness and volume. *Ultrasound Obstet Gynecol* 2011; **38**: 107–115.
95. Martins WP, Lima JC, Welsh AW, Araujo Junior E, Miyague AH, Filho FM, Raine-Fenning NJ. Three-dimensional Doppler evaluation of single spherical samples from the placenta: intra- and interobserver reliability. *Ultrasound Obstet Gynecol* 2012; **40**: 200–206.
96. Mattsson N, Rosendahl H, Luukkaala T. Good accuracy of ultrasound estimations of fetal weight performed by midwives. *Acta Obstet Gynecol Scand* 2007; **86**: 688–692.
97. Mavrides E, Holden D, Bland JM, Tekay A, Thilaganathan B. Intraobserver and interobserver variability of transabdominal Doppler velocimetry measurements of the fetal ductus venosus between 10 and 14 weeks of gestation. *Ultrasound Obstet Gynecol* 2001; **17**: 306–310.
98. Merce LT, Gomez B, Engels V, Bau S, Bajo JM. Intraobserver and interobserver reproducibility of ovarian volume, antral follicle count, and vascularity indices obtained with transvaginal 3-dimensional ultrasonography, power Doppler angiography, and the virtual organ computer-aided analysis imaging program. *J Ultrasound Med* 2005; **24**: 1279–1287.
99. Mercé LT, Alcázar JL, Engels V, Troyano J, Bau S, Bajo JM. Endometrial volume and vascularity measurements by transvaginal three-dimensional ultrasonography and power Doppler angiography in stimulated and tumoral endometria: intraobserver reproducibility. *Gynecol Oncol* 2006; **100**: 544–550.
100. Merce LT, Barco MJ, Bau S. Three-dimensional volume sonographic study of fetal anatomy: intraobserver reproducibility and effect of examiner experience. *J Ultrasound Med* 2008; **27**: 1053–1063.
101. Modugno F, Weissfeld JL, Hill LM. Reproducibility in the assessment of postmenopausal ovaries with transvaginal ultrasound. *Gynecol Oncol* 2000; **77**: 289–292.
102. Molina FS, Terra R, Carrillo MP, Puertas A, Nicolaides KH. What is the most reliable ultrasound parameter for assessment of fetal head descent? *Ultrasound Obstet Gynecol* 2010; **36**: 493–499.
103. Moratalla J, Pintoffl K, Minekawa R, Lachmann R, Wright D, Nicolaides KH. Semi-automated system for measurement of nuchal translucency thickness. *Ultrasound Obstet Gynecol* 2010; **36**: 412–416.
104. Naftalin J, Hoo W, Nunes N, Mavrelos D, Nicks H, Jurkovic D. Inter- and intraobserver variability in three-dimensional ultrasound assessment of the endometrial–myometrial junction and factors affecting its visualization. *Ultrasound Obstet Gynecol* 2012; **39**: 587–591.
105. Pasquini L, Tondi F, Rizzello F, Pontello V, Paoletti E, Fontanarosa M. Impact of tissue harmonic imaging on measurement of nuchal translucency thickness. *Ultrasound Obstet Gynecol* 2010; **36**: 423–426.
106. Perni SC, Chervenak FA, Kalish RB, Magherini-Rothe S, Predanic M, Streltzoff J, Skupski DW. Intraobserver and interobserver reproducibility of fetal biometry. *Ultrasound Obstet Gynecol* 2004; **24**: 654–658.
107. Pexsters A, Luts J, Van Schoubroeck D, Bottomley C, Van Calster B, Van Huffel S, Abdallah Y, D'Hooghe T, Lees C, Timmerman D, Bourne T. Clinical implications of

intra- and interobserver reproducibility of transvaginal sonographic measurement of gestational sac and crown–rump length at 6–9 weeks' gestation. *Ultrasound Obstet Gynecol* 2011; **38**: 510–515.
108. Pistorius LR, Stoutenbeek P, Groenendaal F, de Vries L, Manten G, Mulder E, Visser G. Grade and symmetry of normal fetal cortical development: a longitudinal two- and three-dimensional ultrasound study. *Ultrasound Obstet Gynecol* 2010; **36**: 700–708.
109. Quarello E, Stirnemann J, Ville Y, Guibaud L. Assessment of fetal Sylvian fissure operculization between 22 and 32 weeks: a subjective approach. *Ultrasound Obstet Gynecol* 2008; **32**: 44–49.
110. Raine-Fenning NJ, Campbell BK, Clewes JS, Kendall NR, Johnson IR. The interobserver reliability of three-dimensional power Doppler data acquisition within the female pelvis. *Ultrasound Obstet Gynecol* 2004; **23**: 501–508.
111. Raine-Fenning NJ, Campbell BK, Clewes JS, Johnson IR. The interobserver reliability of ovarian volume measurement is improved with three-dimensional ultrasound, but dependent upon technique. *Ultrasound Med Biol* 2003; **29**: 1685–1690.
112. Rijken MJ, Lee SJ, Boel ME, Papageorghiou AT, Visser GH, Dwell SL, Kennedy SH, Singhasivanon P, White NJ, Nosten F, McGready R. Obstetric ultrasound scanning by local health workers in a refugee camp on the Thai–Burmese border. *Ultrasound Obstet Gynecol* 2009; **34**: 395–403.
113. Rizzo G, Capponi A, Cavicchioni O, Vendola M, Arduini D. Fetal cardiac stroke volume determination by four-dimensional ultrasound with spatio-temporal image correlation compared with two-dimensional and Doppler ultrasonography. *Prenat Diagn* 2007; **27**: 1147–1150.
114. Rolo LC, Nardozza LM, Araujo Júnior E, Simioni C, Zamith MM, Moron AF. [Assessment of the fetal mitral and tricuspid valves areas development by three-dimensional ultrasonography]. *Rev Bras Ginecol Obstet* 2010; **32**: 426–432.
115. Rousian M, Groenenberg IA, Hop WC, Koning AH, van der Spek PJ, Exalto N, Steegers EA. Human embryonic growth and development of the cerebellum using 3-dimensional ultrasound and virtual reality. *Reprod Sci* 2013; **20**: 899–908.
116. Rutten MJ, Pistorius LR, Mulder EJ, Stoutenbeek P, de Vries LS, Visser GH. Fetal cerebellar volume and symmetry on 3-d ultrasound: volume measurement with multiplanar and vocal techniques. *Ultrasound Med Biol* 2009; **35**: 1284–1289.
117. Salama S, Arbo E, Lamazou F, Levaillant JM, Frydman R, Fanchin R. Reproducibility and reliability of automated volumetric measurement of single preovulatory follicles using SonoAVC. *Fertil Steril* 2010; **93**: 2069–2073.
118. Salim R, Woelfer B, Backos M, Regan L, Jurkovic D. Reproducibility of three-dimensional ultrasound diagnosis of congenital uterine anomalies. *Ultrasound Obstet Gynecol* 2003; **21**: 578–582.
119. Salman MM, Twining P, Mousa H, James D, Momtaz M, Aboulghar M, El-Sheikhah A, Bugg GJ. Evaluation of offline analysis of archived three-dimensional volume datasets in the diagnosis of fetal brain abnormalities. *Ultrasound Obstet Gynecol* 2011; **38**: 165–169.
120. 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**: 822–827.
121. Scheffer GJ, Broekmans FJ, Bancsi LF, Habbema JD, Looman CW, Te Velde ER. Quantitative transvaginal two- and three-dimensional sonography of the ovaries: reproducibility of antral follicle counts. *Ultrasound Obstet Gynecol* 2002; **20**: 270–275.
122. Schoonderwaldt EM, Groenenberg IA, Hop WC, Wladimiroff JW, Steegers EA. Reproducibility of echocardiographic measurements of human fetal left ventricular volumes and ejection fractions using four-dimensional ultrasound with the spatio-temporal image correlation modality. *Eur J Obstet Gynecol Reprod Biol* 2012; **160**: 22–29.
123. Senat MV, Bernard JP, Boulvain M, Ville Y. Intra- and interoperator variability in fetal nasal bone assessment at 11–14 weeks of gestation. *Ultrasound Obstet Gynecol* 2003; **22**: 138–141.
124. Senat MV, Quarello E, Levaillant JM, Buonumano A, Boulvain M, Frydman R. Determining chorionicity in twin gestations: three-dimensional (3D) multiplanar sonographic measurement of intra-amniotic membrane thickness. *Ultrasound Obstet Gynecol* 2006; **28**: 665–669.
125. Sladkevicius P, Valentin L. Intra- and interobserver agreement when describing adnexal masses using the International Ovarian Tumor Analysis terms and definitions: a study on three-dimensional ultrasound volumes. *Ultrasound Obstet Gynecol* 2013; **41**: 318–327.
126. Smeets NA, van de Ven J, Oei SG. Inter- and intra-observer variation of fetal volume measurements with three-dimensional ultrasound in the first trimester of pregnancy. *J Perinat Med* 2011; **39**: 539–543.
127. Spandorfer SD, Arrendondo-Soberon F, Loret de Mola JR, Feinberg RF. Reliability of intraobserver and interobserver sonographic endometrial stripe thickness measurements. *Fertil Steril* 1998; **70**: 152–154.
128. Sur SD, Clewes JS, Campbell BK, Raine-Fenning NJ. Embryo volume measurement: an intraobserver, intermethod comparative study of semiautomated and manual three-dimensional ultrasound techniques. *Ultrasound Obstet Gynecol* 2011; **38**: 516–523.
129. Sur SD, Jayaprakasan K, Jones NW, Clewes J, Winter B, Cash N, Campbell B, Raine-Fenning NJ. A novel technique for the semi-automated measurement of embryo volume: an intraobserver reliability study. *Ultrasound Med Biol* 2010; **36**: 719–725.
130. Suwanrath C, Pruksanusak N, Kor-Anantakul O, Suntharasaj T, Hanprasertpong T, Pranpanus S. Reliability of fetal nasal bone length measurement at 11–14 weeks of gestation. *BMC Pregnancy Childbirth* 2013; **13**: 7.
131. Tekay A, Jouppila P. Intraobserver reproducibility of transvaginal Doppler measurements in uterine and intraovarian arteries in regularly menstruating women. *Ultrasound Obstet Gynecol* 1996; **7**: 129–134.
132. Tekay A, Jouppila P. Intraobserver variation in transvaginal Doppler blood flow measurements in benign ovarian tumors. *Ultrasound Obstet Gynecol* 1997; **9**: 120–124.
133. Thomas JT, Muller P, Baghurst P, Wilkinson C. Interobserver variability in the measurement of fetal middle cerebral artery peak systolic velocity in a tertiary fetal medicine unit. *Ultrasound Obstet Gynecol* 2008; **32**: 77–81.
134. Torkildsen EA, Salvesen KA, Eggebo TM. Agreement between two- and three-dimensional transperineal ultrasound methods in assessing fetal head descent in the first stage of labor. *Ultrasound Obstet Gynecol* 2012; **39**: 310–315.
135. Turan OM, Turan S, Buhimschi IA, Funai EF, Campbell KH, Bahtiyar OM, Harman CR, Copel JA, Baschat AA, Buhimschi CS. Comparative analysis of 2-D versus 3-D ultrasound estimation of the fetal adrenal gland volume and prediction of preterm birth. *Am J Perinatol* 2012; **29**: 673–680.
136. Uittenbogaard LB, Haak MC, Peters RJ, van Couwelaar GM, Van Vugt JM. Validation of volume measurements for fetal echocardiography using four-dimensional ultrasound imaging and spatiotemporal image correlation. *Ultrasound Obstet Gynecol* 2010; **35**: 324–331.
137. Valentin L, Sladkevicius P, Bland M. Intraobserver reproducibility of Doppler measurements of uterine artery blood flow velocity in premenopausal women. *Ultrasound Obstet Gynecol* 2001; **17**: 431–433.
138. Van den Bosch T, Valentin L, Van Schoubroeck D, Luts J, Bignardi T, Condous G, Epstein E, Leone FP, Testa AC, Van Huffel S, Bourne T, Timmerman D. Detection of intracavitary uterine pathology using offline analysis of three-dimensional ultrasound volumes: interobserver agreement and diagnostic accuracy. *Ultrasound Obstet Gynecol* 2012; **40**: 459–463.
139. Ventura W, De Paco C, Delgado JL, Blanco JE, Penalver C, Parrilla JJ. Reliability of examining the external iliac artery with Doppler ultrasound in the first trimester and its relationship with maternal blood pressure and uterine artery blood flow. *Eur J Obstet Gynecol Reprod Biol* 2012; **165**: 42–46.
140. Verburg BO, Mulder PG, Hofman A, Jaddoe VW, Witteman JC, Steegers EA. Intra- and interobserver reproducibility study of early fetal growth parameters. *Prenat Diagn* 2008; **28**: 323–331.
141. Verwoerd-Dikkeboom CM, Koning AH, Hop WC, Rousian M, Van Der Spek PJ, Exalto N, Steegers EA. Reliability of three-dimensional sonographic measurements in early pregnancy using virtual reality. *Ultrasound Obstet Gynecol* 2008; **32**: 910–916.
142. von Kaisenberg CS, Kuhling-von Kaisenberg H, Fritzer E, Schemm S, Meinhold-Heerlein I, Jonat W. Fetal transabdominal anatomy scanning using standard views at 11 to 14 weeks' gestation. *Am J Obstet Gynecol* 2005; **192**: 535–542.
143. Wang PH, Chen GD, Lin LY. Imaging comparison of basic cardiac views between two- and three-dimensional ultrasound in normal fetuses in anterior spine positions. *Int J Cardiovasc Imaging* 2002; **18**: 17–23.
144. Wax JR, Royer D, Mather J, Chen C, Aponte-Garcia A, Steinfeld JD, Ingardia CJ. A preliminary study of sonographic grading of fetal intracardiac echogenic foci: feasibility, reliability and association with aneuploidy. *Ultrasound Obstet Gynecol* 2000; **16**: 123–127.
145. Wax JR, Cartin A, Pinette MG, Blackstone J, Michaud J, Byers S, Boutin N. Sonographic grading of fetal intracardiac echogenic foci in a population at low risk of aneuploidy. *J Clin Ultrasound* 2003; **31**: 31–38.
146. Yaman C, Sommergruber M, Ebner T, Polz W, Moser M, Tews G. Reproducibility of transvaginal three-dimensional endometrial volume measurements during ovarian stimulation. *Hum Reprod* 1999; **14**: 2604–2608.
147. Yang F, Leung KY, Lee YP, Chan HY, Tang MH. Fetal biometry by an inexperienced operator using two- and three-dimensional ultrasound. *Ultrasound Obstet Gynecol* 2010; **35**: 566–571.
148. Yazbek J, Ameye L, Testa AC, Valentin L, Timmerman D, Holland TK, Van Holsbeke C, Jurkovic D. Confidence of expert ultrasound operators in making a diagnosis of adnexal tumor: effect on diagnostic accuracy and interobserver agreement. *Ultrasound Obstet Gynecol* 2010; **35**: 89–93.
149. Yazbek J, Ameye L, Timmerman D, Testa AC, Valentin L, Holland TK, Van Holsbeke C, Jurkovic D. Use of ultrasound pattern recognition by expert operators to identify borderline ovarian tumors: a study of diagnostic performance and interobserver agreement. *Ultrasound Obstet Gynecol* 2010; **35**: 84–88.
150. Youssef A, Salsi G, Bellussi F, Arcangeli T, Farina A, Contro E, Maroni E, Pilu G, Rizzo N, Ghi T. Three-dimensional ultrasound is an accurate and reproducible technique for fetal crown–rump length measurement. *Prenat Diagn* 2012; **32**: 220–227.
151. Youssef A, Maroni E, Ragusa A, De Musso F, Salsi G, Iammarino MT, Paccapelo A, Rizzo N, Pilu G, Ghi T. Fetal head–symphysis distance: a simple and reliable ultrasound index of fetal head station in labor. *Ultrasound Obstet Gynecol* 2013; **41**: 419–424.
152. Yuan Y, Leung KY, Ouyang YS, Yang F, Tang MH, Chau AK, Dai Q. Simultaneous real-time imaging of four-chamber and left ventricular outflow tract views using xPlane imaging capability of a matrix array probe. *Ultrasound Obstet Gynecol* 2011; **37**: 302–309.

