Original article

Standardisation and quality control of ultrasound measurements taken in the INTERGROWTH-21st Project

I Sarris,[a] C Ioannou,[a] EO Ohuma,[a,b] DG Altman,[b] L Hoch,[a] C Cosgrove,[a] S Fathima,[a,c] LJ Salomon,[d] AT Papageorghiou,[a] for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st)

[a] Nuffield Department of Obstetrics & Gynaecology and Oxford Maternal & Perinatal Health Institute, Green Templeton College, University of Oxford, Oxford, UK [b] Centre for Statistics in Medicine, University of Oxford, Oxford, UK [c] Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, UK [d] Université Paris V—René Descartes, Maternité Hôpital Necker-Enfants Malades, AP-HP, Paris, France
Correspondence: Dr AT Papageorghiou, Nuffield Department of Obstetrics & Gynaecology, University of Oxford, Women's Centre, Level 3, John Radcliffe Hospital, Oxford, OX3 9DU, UK. Email aris.papageorghiou@obs-gyn.ox.ac.uk

Accepted 30 April 2013.

Meticulous standardisation and ongoing monitoring of adherence to measurement protocols during data collection are essential to ensure consistency and to minimise systematic error in multicentre studies. Strict ultrasound fetal biometric measurement protocols are used in the INTERGROWTH-21st Project so that data of the highest quality from different centres can be compared and potentially pooled. A central Ultrasound Quality Unit (USQU) has been set up to oversee this process. After initial training and standardisation, the USQU monitors the performance of all ultrasonographers involved in the project by continuously assessing the quality of the images and the consistency of the measurements produced. Ultrasonographers are identified when they exceed preset maximum allowable differences. Corrective action is then taken in the form of retraining or simply advice regarding changes in practice. This paper describes the procedures used, which can form a model for research settings involving ultrasound measurements.

Keywords Fetal biometry, INTERGROWTH-21st, quality control, standardisation, ultrasound measurements.

Introduction

The International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st) is a large-scale, population-based, multicentre project of fetal and newborn growth currently underway in eight sites across the world.[1] One component of the project, the Fetal Growth Longitudinal Study (FGLS), involves carrying out serial fetal growth scans every 5 ± 1 weeks from recruitment at 9^{+0}–13^{+6} weeks of gestation until, but not beyond, 42^{+0} weeks of gestation. The primary aim of the study is to develop new 'prescriptive' standards describing optimal fetal growth. Therefore, the measurement values used to construct these standards should reflect the actual distribution of fetal growth and must be minimally influenced by other sources of variation.

The health institutions participating in the INTER-GROWTH-21st Project are diverse and employ different pathways and protocols for scanning pregnant women in their routine clinical practice. For the data collected to be comparable within and between the study sites, all ultrasound measurements had to be performed in a standardised manner. This enables the data collected at each site to be compared and potentially combined into a single data set for the purposes of generating the growth references. Standardisation and quality control of measurements are, therefore, necessary to ensure that ultrasonographers measure all fetal biometric parameters in an identical fashion.

Standardisation of practice, quality assurance and control of this component of the INTERGROWTH-21st Project is the remit of the Ultrasound Quality Unit (USQU). This paper describes the methodology and work performed by the USQU to achieve these objectives.

Overview of the USQU, standardisation and quality control

All participating sites and ultrasonographers are provided with an ultrasound manual describing measurement techniques, as well as protocols and procedures for data collection (see Supporting Information, Appendix S1).[2]

The USQU is coordinated by a lead ultrasound specialist (ATP) assisted by an external expert (LS) and two clinical research fellows (IS, CI). Day-to-day data management is performed by data managers (LH, CC). Ultrasound images and volumes are managed in a specially designed database by a biomedical engineer with expertise in image analysis (SF). Statistical advice is provided by the study statistician (EO) under the supervision of the INTERGROWTH-21st Project senior statistician (DA). The USQU is responsible for:

1 Development of standard operating procedures.
2 Initial standardisation (involving training, assessment and certification) of ultrasonographers.
3 Site standardisation exercises.
4 Quality control (QC) monitoring of routine replicate measurements by re-measurement and quality assessment of a random 10% sample of images.
5 Analysis and reporting of ultrasound data quality.
6 Identification of retraining needs.

Standardisation

Local ultrasonographers taking part in FGLS already have a high standard of training. The goals of standardisation are, therefore, to ensure that all ultrasonographers fully understand the study protocol and take measurements in an identical fashion, and that they are familiar with the equipment used. The protocol is described in detail elsewhere.[2] Briefly, ultrasonographers are required to measure fetal biparietal diameter (BPD), occipito-frontal diameter (OFD), head circumference (HC), transverse abdominal diameter (TAD), anterior-posterior abdominal diameter (APAD), abdominal circumference (AC) and femur length (FL); assess and measure amniotic fluid volume; and document placental localisation and fetal presentation.

Standardisation operates through a period of training, assessment and certification. It involves two steps with similar processes: an initial exercise at the Project Coordinating Unit in Oxford for the lead ultrasonographer from each study site, followed by sessions at each site.

Initial standardisation exercise

Before initiating FGLS, each site sent a lead ultrasonographer to the Project Coordinating Unit in Oxford to take part in a 3 day standardisation training exercise. The effect of this exercise on each lead ultrasonographer's measurements and the resultant improvement over time when compared with the reference standard (i.e. measurements made by the lead trainer, ATP), have been reported elsewhere.[3]

Site-specific standardisation

All local ultrasonographers are recruited on the basis of being motivated, reliable and well trained already in ultrasound, as well as their ability to speak the local language(s) and work positively within a team structure. Before participating in the study, they must all first read and familiarise themselves with the ultrasound manual (Appendix S1). Thereafter, individual and group theoretical training takes place to understand the study protocol and QC system.

The local standardisation exercise was held over 1–2 days. In each case, the trainer was the lead ultrasonographer who had participated in the initial standardisation exercise at the Project Coordinating Unit. The only difference was that the reference standard was taken to be the overall mean of all observations for each biometric parameter, rather than the measurements made by the trainer.

The training involved scanning women at various gestational ages to enable the ultrasonographers to become familiar with the equipment and scanning protocol, as well as with data storage and extraction. Each ultrasonographer obtains three measurements of each biometric variable from three fetuses. These measurements are repeated by the lead ultrasonographer at each site. All measurements are taken in a blinded fashion and all measurements are concealed. The measurement data and image quality are then assessed by the USQU. First, all the measurements are used to assess the intraobserver and inter-observer variability. Second, all images are scored for quality purposes using a system proposed by Salomon et al.[4] (Table 1). Briefly, each plane has a set of specific criteria that score one point each: the maximum score is six points for both the cephalic and abdominal planes, and four points for the femoral plane.

For an ultrasonographer to be certified as having passed the standardisation process the following must be fulfilled: all three test scan measurements should be within one standard deviation (SD) of the trainer's measurements, and an average image score of >67% of the maximum score (i.e. four for cephalic and abdominal planes, and three for femoral planes) must be achieved. If these criteria are not met, certification is withheld.

Standardisation of new ultrasonographers

If ultrasonographers subsequently join the study, they are either standardised locally during an USQU site visit, or visit the USQU at the Coordinating Unit for training and standardisation.

USQU site visits

During the study, the USQU team arranges site visits specifically to assess ultrasound issues (Appendix S2). The purpose

Table 1. Image scoring criteria used for standardisation and quality control, based on Salomon et al.[4]

Cephalic plane (max. 6 points)	Abdominal plane (max. 6 points)	Femoral plane (max. 4 points)
1 Symmetrical plane	1 Symmetrical plane	1 Both ends of the bone clearly visible
2 Thalami visible	2 Stomach bubble visible	2 Angle <45°
3 Cavum septi pellucidi visible	3 Portal sinus visible	3 Femur occupying at least 30% of image
4 Cerebellum not visible	4 Kidneys not visible	4 Callipers placed correctly
5 Head occupying at least 30% of image	5 Abdomen occupying at least 30% of image	
6 Callipers/ellipse placed correctly	6 Callipers/ellipse placed correctly	

is to support local teams; calibrate and ensure correct operation of the ultrasound equipment; verify adherence to protocols, and perform re-standardisation exercises. These visits also provide an opportunity for feedback and discussion between the USQU team and the lead ultrasonographers about data-quality statistics and, if retraining is needed, how it should be organised.

Quality control

The USQU is responsible for ensuring adherence to the protocol and ongoing QC assessment. This includes a pilot reproducibility study, quality assessment of images, assessment of collected data, and evaluation and repetition of ultrasound measurements.

Pilot reproducibility study

Quality control of ultrasound measurements is primarily based on the comparison of repeat measurements by the same or different observers. Routine data-quality assessment is guided by a system of maximum allowable differences between replicates. However, little is known about what the maximum allowable differences for ultrasound measurements should be for each fetal biometric variable at different stages of pregnancy; nor is it known how exactly fetal size and gestational age influence measurement error

in ultrasound. In addition, it is not clear how much measurement error in ultrasound is influenced by different parameters such as fetal position, fetal activity, maternal body mass index, order of measurement and level of experience. Therefore, a pilot study to determine the variability of fetal ultrasound measurements from 14^{+0} to 42^{+0} weeks of gestation, using the same equipment and protocol as for the study, was performed by the USQU team at Oxford at the start of the study.[5] The results provide a reference standard against which the performance of the study's ultrasonographers is monitored.

Qualitative image quality control

Images taken by ultrasonographers are qualitatively controlled and scored according to the scheme described by Salomon et al.[4] (Table 1). The monitoring process consists of ultrasonographers self-scoring all their images (only the best image will be scored and this score will be uploaded onto the study database) and independent scoring by the USQU of a random sample of 10% of all images.

For self-scoring of images
1 All scores of ≤3 (for abdominal circumference, AC and head circumference, HC); or ≤2 (for femur length, FL) are identified and the images are retrieved; these are then reviewed by the USQU for independent scoring.
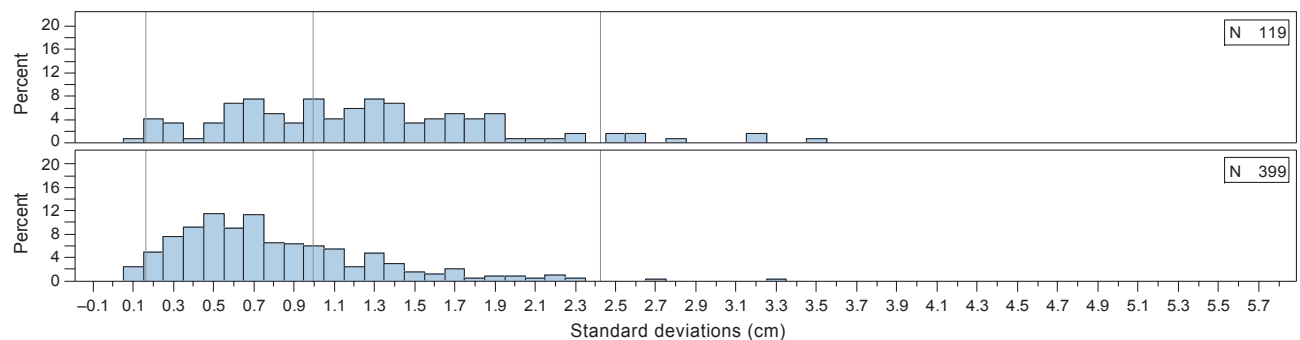

Figure 1. Distribution of standard deviations (expressed as a percentage of the mean*) of triplicate measurements of the head circumference from two ultrasonographers. The lines indicate the 97.5th centile value (2.42), mean and 2.5th centile (0.16) taken from the combined data of the reference standard ultrasonographers, allowing visual comparison between ultrasonographers.
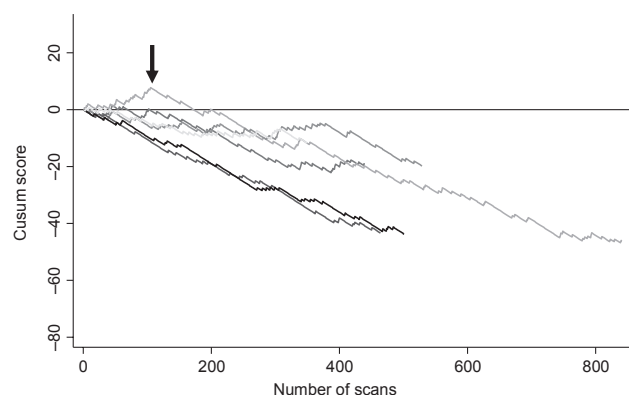
Figure 2. Cumulative sum (CUSUM) chart of consistency of measurement. Each line represents a different ultrasonographer, and the number of scans each has performed is on the *x*-axis. When an ultrasonographer had a z-score value >1.28 SD, this was considered a 'failure' resulting in a positive CUSUM score; whereas a z-score within 1.28 gave a negative CUSUM score. Note how one operator (arrow) had a higher than expected failure rate, which improved after corrective action was taken.

2 Those scores confirmed as ≤3 for AC and HC, or ≤2 for FL, are apportioned to specific ultrasonographers by the Project Coordinating Unit. The proportion of low scoring images is calculated and any ultrasonographer with >10% rejected images in any given 4 week period or overall, is identified as not fulfilling the QC criteria. This may lead to the ultrasonographer requiring re-training and, if persistently poor, his/her certification is withdrawn.

For independent scoring of images
Ten percent of all scans per site are randomly chosen for QC by the USQU. Assessors are blinded to both the site and the identity of the ultrasonographer. Images are scored according to the system described above. Any ultrasonographer with >10% of low scoring images is identified as requiring re-training and, if persistently poor, his/her certification is withdrawn.

Quantitative QC: intraobserver variability
Intraobserver reliability is continuously and prospectively assessed based on the three concealed measurements taken for every fetal biometric variable during each scan. For each ultrasonographer, the range of the values for each triplicate measurement is calculated. For a given fetal biometric variable at a given gestational age, no more than 10% of these ranges should be above two SD of the expected triplicate range based on the equivalent data derived from the pilot variability study. The aim of the QC process is, therefore, to identify ultrasonographers that consistently have intraobserver differences above the allowable thresholds. The monitoring process is by assessment of distributions of SDs of the triplicate measurements (Figure 1) and by using cumulative sum (CUSUM) graphs (Figure 2). Scan visits containing extreme outliers (defined as a triplicate measurement range >4 SD of the expected) are also reviewed by the USQU. The monitoring process may identify the ultrasonographer as requiring retraining and, if persistently poor, his/her certification is withdrawn.
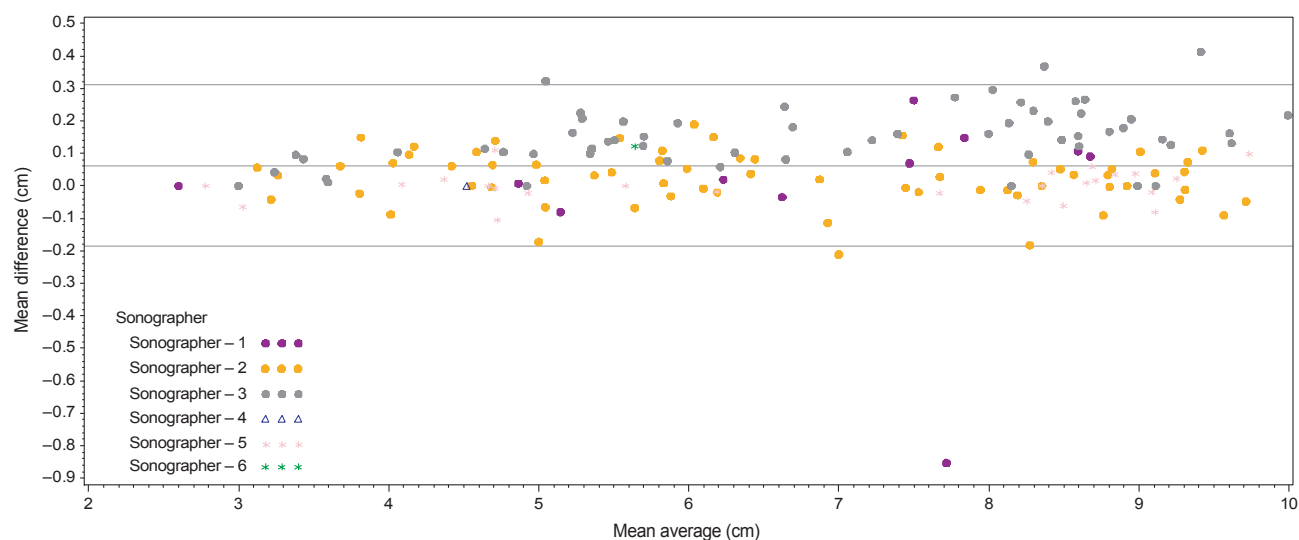

Figure 3. Bland–Altman plot showing differences between the original ultrasonographer measurement of biparietal diameter (BPD) versus re-measurement by the quality control team (calliper placement repeatability). Each ultrasonographer is represented in a different colour allowing individual assessment of systematic (bias) and random error.

Quantitative QC: random re-measurement of images

To assess correct calliper placement, a random sample of 10% of all scans is selected. The measurements are then repeated offline. The assessor is blinded both to the original measurement and to his own measurement. This allows production and evaluation of inter-observer variability and bias of calliper placement (Figure 3), which is done for each ultrasonographer. No more than 10% of repeated measurements by an USQU assessor should vary by more than two SD of the expected inter-observer calliper placement for that fetal biometric variable at a given gestational age. Furthermore, the systematic bias should not exceed 0.5 SD.

Reporting of QC data

The USQU team meets every month and produces data quality statistics based on standardisation sessions, evaluation of routine replicates and the QC re-measurements. This includes the production of statistics and trend plots of intraobserver reliability of individual ultrasonographers; CUSUM charts of intraobserver reliability; inter-observer reliability of calliper placement of individual ultrasonographers against the USQU; comparison among sites, including site specific bias against the rest of the study. Areas of potential concern are highlighted and, if appropriate, site visits and retraining are arranged.

Conclusion

In this paper, we have described the procedures used by the INTERGROWTH-21st Project to achieve consistency and high-quality data. They can form a model for research settings involving fetal ultrasound measurements.

Disclosure of interests
None.

Contribution to authorship
All authors are members of the Ultrasound Quality Unit of the INTERGROWTH-21st Project. All authors approved the final version.

Details of ethics approval
The INTERGROWTH-21st protocol was approved by the Oxfordshire Research Ethics Committee 'C' (reference: 08/H0606/139) and the research ethics committees of the individual participating institutions and corresponding health authorities where the Project was implemented.

Funding

Acknowledgements

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. INTERGROWTH-21st Project. The Ultrasound Operations Manual

Appendix S2. INTERGROWTH-21st Project: The Ultrasound Quality Unit: Site Visit Manual ∎

References

1 Villar J, Altman DG, Purwar M, Noble JA, Knight HE, Ruyan P, et al. for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). The objectives, design and implementation of the INTERGROWTH-21st Project. *BJOG* 2013; DOI: 10.1111/1471-0528.12047.

2 Papageorghiou AT, Sarris I, Ioannou C, Todros T, Carvalho M, Pilu G, et al. for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). Ultrasound methodology used to construct the fetal growth standards in the INTERGROWTH-21st Project. *BJOG* 2013; DOI: 10.1111/1471-0528.12313.

3 Sarris I, Ioannou C, Dighe M, Mitidieri A, Oberto M, Qingqing W, et al. for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). Standardisation of fetal ultrasound biometry measurements: improving the quality and consistency of measurements. *Ultrasound Obstet Gynecol* 2011;38:681–7.

4 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.

5 Sarris I, Ioannou C, Chamberlain P, Ohuma E, Roseman F, Hoch L, et al. for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). Intra- and interobserver variability in fetal ultrasound measurements. *Ultrasound Obstet Gynecol* 2012;39:266–73.