The management of error in ultrasound fetal growth monitoring

Nicholas J Dudley (iD)

Abstract
It is important to understand the uncertainty in fetal measurements when using them in the management of pregnancy. The aim of this essay is to provide background on errors and uncertainty, describing error sources and their potential impact, with guidance on improving accuracy. Errors can be systematic or random, arising from equipment, image plane selection, measurement method and caliper placement and influenced by image quality, training and experience. The uncertainty in measurements is larger than clinically significant differences in fetal size and growth. Errors can be reduced by implementing equipment acceptance testing, written procedures, training and audit.

Introduction

The errors in individual fetal measurements and estimated fetal weight (EFW) have been well documented. For example, 95% confidence intervals (95% CI) for abdominal circumference (AC), head circumference (HC) and femur length (FL) of $\pm9\%$, $\pm5\%$ and $\pm11\%$, respectively, and for EFW of up to $\pm50\%$ have been reported.[1,2] Analysis of three studies using the EFW formula of Hadlock et al.[3] (AC, HC, FL) on a total of 1028 patients gave a combined 95% CI of $\pm21\%$.[4–6]

The aim of this essay is to provide some basic background on error and uncertainty, to describe possible sources of error and their potential impact on the monitoring of fetal growth and to provide guidance on improving accuracy.

Error and uncertainty

The error in a measurement is the difference between the measurement and the true value and is unknown unless we know the true value. It should be noted that in fetal biometry we rarely know the true value of a measurement, the single exception being EFW performed immediately before delivery. Uncertainty describes the range of possible errors in some way. If we know the size of errors for a cohort of patients, e.g. by comparing EFW with birthweight, we may then make an estimate of the uncertainty in future measurements. If the size of errors is unknown, we may estimate uncertainty from multiple measurements made by one or more observers.

The simplest way to express uncertainty is to give the range of errors or differences, i.e. the largest negative and positive errors or differences; this describes the worst case (and most unlikely) errors. An alternative is to calculate the mean absolute error (the average error, ignoring the $\pm$ signs); this has been widely used in papers on the accuracy of EFW but, depending on the error distribution, approximately half of all errors may be larger than this.

In scientific papers it is common to express uncertainty as ±1 standard deviation (SD). SD, usually symbolised as σ, is a statistical measure of the spread of errors about a true or mean value; the definition is

Radiation Protection & Radiology Physics, Lincoln County Hospital, Lincoln, UK

Corresponding author:
Nicholas J Dudley, Lincoln County Hospital, Greetwell Road, Lincoln LN2 5QY, UK.
Email: nick.dudley@ulh.nhs.uk

widely available in texts and on web sites. ±1 SD contains 68% of errors, so that approximately one in three errors falls outside this range. The 95% CI (1.96σ) is also often used to express uncertainty and, by definition, contains 95% of errors, so that only 1 in 20 errors falls outside this range.

In manufacturing or science where a very low failure rate or a high degree of accuracy is required, the concept of 'six sigma' is applied, giving products or measurements a 99.99966% probability of accuracy.

Types of error

Systematic errors are consistent in direction and size. They may arise from poor equipment calibration, calculations programmed incorrectly into an ultrasound machine or consistent over or under measurement by the observer. Due to their consistency, systematic errors can be difficult to detect. In some circumstances their consistency means that they have no consequences; if systematic errors are present and consistent in normal ranges and clinical practice, there is no impact on the interpretation of measurements. They may be reduced by implementing controls, such as equipment quality assurance (QA) including caliper accuracy checks and checks of calculations made by the ultrasound machine and reporting systems, training and written procedures to ensure that measurements are made correctly and consistently, and audit.

Random errors are accidental; multiple measurements will give values above and below the true value. In ultrasound these are most often due to observer inconsistency in measurement plane selection or placing measurement calipers. Since measurements may be above or below the true value, the error can be reduced by averaging several measurements. It is worth noting that averaging may not always be the best approach; for example if three measurements are made and two are on sub-optimal image planes, it may be better to accept a single measurement on the optimal image.

Errors may be consistent within a group of staff. For example, if staff are trained to make a measurement in a particular way that leads to systematic error, they will all generate that error. Some errors may be personal to an individual where their practice differs from others.

Figure 1 shows an example of over measurement of HC. If the operator always places the ellipse in this manner it generates a systematic error. If the operator is careless and sometimes under measures, it may be a random error. Measurements often have both systematic and random error.

Compound errors occur when measurements are combined, e.g. in EFW. There are several possible approaches to estimating compound errors; for more
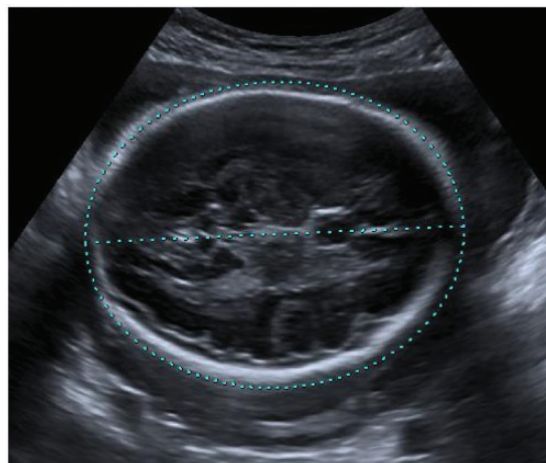

Figure 1. Over measurement of a fetal HC. The ellipse extends outside the perimeter of the fetal skull both proximally and distally.

complex combinations of measurements, the square root of the sum of the squares of individual errors may give a reasonable estimate. Compounding the potential errors in individual measurements reported in the literature (AC: ±9%; HC: ±5%; FL: ±11%)[1] the estimated 95% CI for EFW is ±15%, but in practice errors may be much larger; a systematic review of the accuracy of EFW reported 95% CI for EFW of up to ±50%.[2]

Sources of error

Errors due to poor equipment calibration have been reported.[7] If ultrasound machines are subject to acceptance testing and regular QA then they should not be a source of significant systematic errors.[8]

The most likely sources of error are the choice of image plane, measurement method and caliper placement which may all be influenced by image quality, training and experience and the level of standardisation within and between departments.

EFW formulae also contribute to measurement error since they are derived from populations where fetal proportions vary, so that there may be systematic errors when applied to individuals. However, the focus of this essay is on error sources within the control of the ultrasound practitioner.

Image plane

In a multicentre study auditing against image plane criteria, Dudley and Chapman[9] found that 87% (range 78–95%) of HC images met all quality criteria and only 60% (range 45–75%) of AC images met all quality criteria. They also compared measurements where an optimal and sub-optimal image were

available in the series of three measurements made as standard in their centre, finding that the 95% CI of the differences was −15 to 8 mm for AC. Sub-optimal image planes can therefore easily lead to errors of 5% or more in an individual measurement.

Measurement method

With the exception of FL, which is a linear measurement, a range of measurement methods is available. The widely used circumference charts developed by Chitty et al.[10,11] are based on the measurement of two diameters: bi-parietal diameter (BPD) and occipito-frontal diameter (OFD) for HC and antero-posterior and transverse diameters for AC. Hadlock et al.[3] used a mixture of traced circumferences and the two diameter method. INTERGROWTH-21st recommend ellipse fitting.[12] The two diameter and ellipse methods are not interchangeable in practice, with 95% CI of the differences of ±6% for AC and HC and ±12% for EFW in the late third trimester.[13] Where the charts of Chitty et al. are widely used, as they have been in the UK, it is important to use the two diameter method for circumferences.

Caliper placement

Caliper placement is the critical final stage in making a measurement. No matter how much effort is invested in ensuring accurate equipment and obtaining an optimum image plane, if the caliper placement is wrong the measurement is wrong. It is important to match the criteria used in developing the charts employed and to ensure correct end points are identified, e.g. outer bone surface for BPD and OFD; skin surface for AC diameters.

Impact of errors on growth monitoring

In order to provide context for errors in fetal measurement and EFW, it is important to understand the differences or changes in fetal size that may lead to intervention. The 10th, 5th or 3rd centiles of fetal size curves are used as thresholds for detection of the small-for-gestational-age fetus, which may then undergo further monitoring to assess the risk for fetal growth restriction (FGR) and stillbirth. For example, 'Saving Babies' Lives Care Bundle Version 2' published by NHS England suggests that FGR is defined as EFW or AC below the third centile, or EFW or AC below the 10th centile together with Doppler evidence of placental dysfunction and that sub-optimal growth is less than 280 g in 14 days after 34 weeks.[14] Table 1 shows a comparison between measurement uncertainty and key differences in size and growth at 36 weeks' gestation. The difference between the 3rd and 10th centiles is

Table 1. Comparison between measurement uncertainty and key differences in size and growth at 36 weeks' gestation using uncertainty from literature[1,3–6] and the growth charts of Hadlock et al.[3] and Chitty et al.[11]

Measurement	1 SD	95% CI	10th–3rd centiles	280 g
AC	4.5%	9%	4%	–
EFW	10.5%	21%	8%	10%

AC: abdominal circumference; CI: confidence interval; EFW: estimated fetal weight.

8% of EFW and 4% of AC, and 280 g is 10% of EFW. If 1 in 20 errors in EFW is greater than 21% (9% for AC), and one in three errors is greater than 10.5% (4.5% for AC), these differences in size or growth are difficult to reliably detect.

Figure 2 shows the potential impact of a 10.5% error on EFW size and growth trajectory. The 'true' growth follows the 15th centile. At the 32-week scan a −10.5% error would result in a measurement on the third centile, potentially resulting in a diagnosis of FGR. If the 32-week scan had no error, or a positive error, then a −10.5% error at the 35-week scan would indicate poor growth and may result in intervention and early delivery. Errors will also have an impact for the larger fetus, potentially leading to inappropriate decisions regarding intervention.

Improving measurement accuracy

The first step in minimising errors is to ensure that equipment is correctly calibrated and that calculations performed by the ultrasound machine and any external reporting systems are correctly programmed. The former can be easily checked using a test object with nylon filament targets, making measurements using each relevant method, i.e. linear and circumference; results should be within 1% of expected.[8] Calculations can be checked on the ultrasound machine by making realistic measurements, and on reporting systems by entering data, and comparing the results with manual or spreadsheet calculations; EFW results should be identical but a 1% difference is acceptable.

The other sources of error described here, namely image plane, measurement method and caliper placement, are entirely within the control of the ultrasound practitioner. A key element in reducing errors is the implementation of written standards and procedures, accompanied by training. Periodic audit is important in ensuring ongoing adherence to procedures and maintenance of standards.
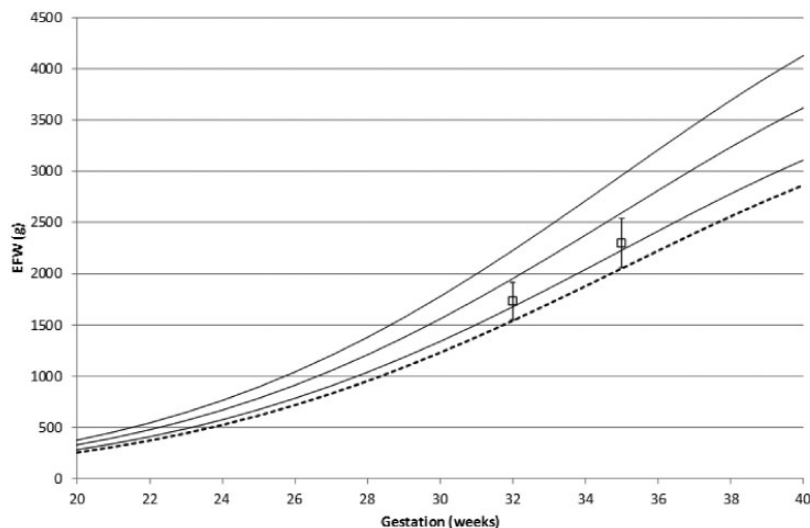
Figure 2. Modelled fetal growth on the 15th centile. Solid lines show the 90th, 50th and 10th centiles, respectively; dashed line shows the third centile. Squares represent the 'true' measurement value. Error bars are ±10.5% of EFW. EFW: estimated fetal weight.

Procedures

Features of the correct image planes for measurement are widely agreed and should be described in written procedures, or reference made to national standards, e.g. the NHS Fetal Anomaly Screening Programme Handbook.[15]

A factor often neglected in written procedures is ultrasound machine settings. The optimum settings for particular applications, e.g. first trimester scanning, fetal anomaly screening, growth scans, should be saved for recall at the start of each examination. There should be agreed departmental settings for each model of ultrasound machine to ensure consistency between operators. Procedures should then guide adjustment of controls to optimise each image, e.g. scale/depth settings and appropriate focal depth, for accurate measurement. The most highly trained and experienced operators, e.g. those with formal postgraduate ultrasound qualifications, should usually make these adjustments without any prompting, but those with less training or experience may benefit from some written guidance.

Measurements

Measurement methods and caliper placement should be agreed and documented to ensure a consistent approach. In particular, circumference measurements should be made according to the charts in use, either the two diameter method or ellipse fitting. Ellipse fitting is more subject to operator interpretation of fit than the two diameter method, where end points are clearly defined, since fetal circumferences are rarely truly elliptical.[13]

An important and potentially controversial decision is the number of repeat measurements and whether to record an average. Where random error is likely to dominate, the average of three measurements should be the most accurate; this may be the case for FL. Where systematic error is likely to dominate, the measurement of the image most closely meeting standard criteria and where measurement points are most clearly visualised should be the most accurate; this may be the case for AC and HC.

Audit

Audit is essential in maintaining standards. One approach is to compare scan measurements with measurements of the neonate. For AC, HC and FL, reproducing the ultrasound measurement methods on a baby with a tape measure is not possible. EFW close to birth may be compared with birthweight but, owing to the known large uncertainty in EFW, a large number of results (hundreds) are required in order to obtain a statistically significant result; with a small number of results any differences may be due to chance.

A proven approach is assessment of the quality of each individual measurement.[9,16] In this method a series of images is assessed against quality criteria: magnification; caliper placement; antero-posterior alignment (AC, HC); lateral alignment (AC, HC); presence of landmarks; angle to ultrasound beam. Operators are encouraged to indicate if they believe an image to be unsuitable for measurement. Images are judged satisfactory where all relevant criteria are met; this is important as failure against any single criterion may result in an inaccurate measurement. Quality criteria will never be
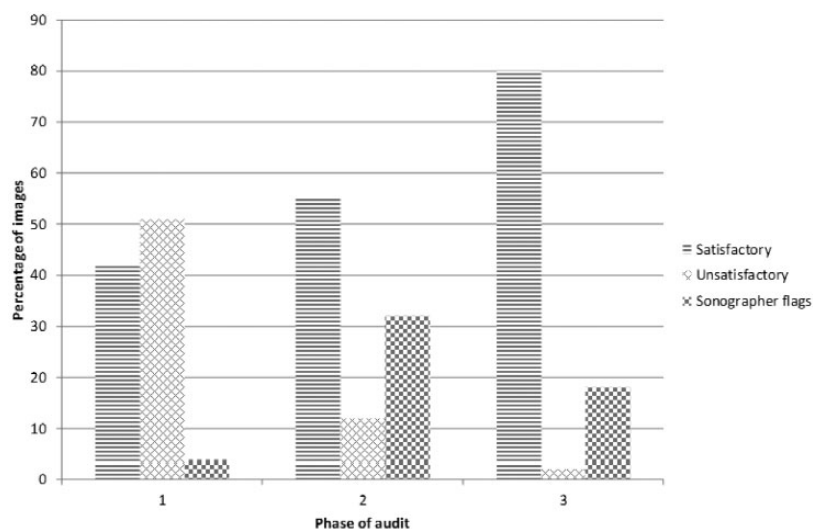
Figure 3. Three phases of measurement quality audit. Satisfactory images met all quality criteria. Unsatisfactory images failed to meet one or more criteria. Sonographers were asked to flag images they judged to be unsatisfactory for measurement.

met for 100% of measurements; the most important function of such an audit is to facilitate improvement. Figure 3 shows the results of three phases of audit. Following feedback from phase 1 sonographers improved their recognition of quality criteria, with a small increase in the number of satisfactory images. Following phase 2 sonographers developed their technical skills and significantly improved measurement quality. This improvement may require coaching, but many sonographers are able to develop their skills solely on the basis of feedback from the audit.

Summary

It is important to understand the uncertainty associated with individual measurements and with EFW when using them in the management of pregnancy. The uncertainty in measurements is larger than clinically significant differences in fetal size and growth. Errors may arise from the ultrasound machine, the operator and reporting systems. Errors associated with ultrasound machines and reporting systems can be prevented by careful acceptance testing. Operator errors can be reduced by controls including procedures, training and audit.

Contributors
NJD conceived the study, performed modelling and data analysis, drafted and revised the manuscript.

Guarantor

NJD.

ORCID iD

Nicholas J Dudley ⓘD https://orcid.org/0000-0002-0057-3760

References

1. Sarris I, Ioannou C, Chamberlain P, et al. Intra- and interobserver variability in fetal ultrasound measurements. *Ultrasound Obstet Gynecol* 2012; 39: 266–273.
2. Dudley NJ. A systematic review of the ultrasound estimation of fetal weight. *Ultrasound Obstet Gynecol* 2005; 25: 80–89.
3. Hadlock FP, Harrist RB, Sharman RS, et al. Estimation of fetal weight with the use of head, body and femur measurements: a prospective study. *Am J Obstet Gynecol* 1985; 151: 333–337.
4. Dudley NJ. Selection of appropriate ultrasound methods for the estimation of fetal weight. *Br J Radiol* 1995; 68: 385–388.

5. Sabbagha RE, Minogue J, Tamura RK, et al. Estimation of birth weight by use of ultrasonographic formulas targeted to large-, appropriate-, small-for-gestational-age fetuses. *Am J Obstet Gynecol* 1989; 160: 854–860.

6. Simon NV, Levisky JS, Shearer DM, et al. Influence of fetal growth patterns on sonographic estimation of fetal weight. *J Clin Ultrasound* 1987; 15: 376–383.

7. Dudley NJ and Griffith K. The importance of rigorous testing of circumference measuring calipers. *Ultrasound Med Biol* 1996; 22:1117–1119.

8. Institute of Physics and Engineering in Medicine (IPEM). *Quality assurance of ultrasound imaging systems*. York: IPEM, 2010.

9. Dudley NJ and Chapman E. The importance of quality management in fetal measurement. *Ultrasound Obstet Gynecol* 2002; 19: 190–196.

10. Chitty LS, Altman DG, Henderson A, et al. Charts of fetal size: 2. Head measurements. *Br J Obstet Gynaecol* 1994; 101: 35–43.

11. Chitty LS, Altman DG, Henderson A, et al. Charts of fetal size: 3. Abdominal measurements. *Br J Obstet Gynaecol* 1994; 101: 125–131.

12. Papageorghiou AT, Ohuma EO, Altman DG, et al. International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. *Lancet* 2014; 384: 869–879.

13. Dudley NJ. Are ultrasound foetal circumference measurement methods interchangeable? *Ultrasound* 2019; 27: 176–182.

14. NHS England. *Saving babies' lives care bundle version 2*. Leeds: NHS England, 2019.

15. Public Health England. *NHS fetal anomaly screening programme handbook*. London: Public Health England, 2018.

16. Dudley NJ and Potter R. Quality assurance in obstetric ultrasound. *Br J Radiol* 1993; 66: 865–870.