Ultrasound Obstet Gynecol 2023; **61**: 431–435
Published online in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.26156

Opinion

Audit of fetal biometry: understanding sources of error to improve our practice

I. PAPASTEFANOU[1,2]*, K. H. NICOLAIDES[1] and L. J. SALOMON[3,4]

[1]*Fetal Medicine Research Institute, King's College Hospital, London, UK;* [2]*Department of Women and Children's Health, Faculty of Life Sciences & Medicine, King's College London, London, UK;* [3]*Department of Obstetrics, Fetal Medicine and Surgery, Necker–Enfants Malades Hospital, AP–HP, Paris, France;* [4]*URP FETUS 7328 and LUMIERE Platform, University of Paris Cité, Institut Imagine, Paris, France*
**Correspondence. (e-mail: jdpap2@yahoo.co.uk; ioannis.papastefanou@kcl.ac.uk)*

Introduction

Fetal growth assessment using ultrasound is an integral part of antenatal care[1–3]. It starts early in pregnancy with the measurement of crown–rump length (CRL) and continues during the course of pregnancy with the measurement of abdominal circumference (AC), femur length (FL), biparietal diameter and head circumference (HC)[2,3]. These measurements have a major impact on pregnancy care for several reasons: first, dating of pregnancy is based on CRL in the first trimester or fetal biometry in the second trimester; second, small or large babies are at higher risk of adverse perinatal outcome; and, third, small or large babies are associated with maternal conditions such as pre-eclampsia and diabetes. The antenatal identification of fetal growth divergence may reduce the rate of maternal and fetal complications[4–6].

Fetal biometry by ultrasound scan is the translation of a biological phenomenon such as fetal growth to measurable quantities. Apart from the obvious physiological variability between individuals, everything that involves a measurement has variability, which consists of two elements: the inherent statistical error and measurement bias attributed to conditions of the measurement, such as operator's performance and equipment. In pregnancy, it becomes even more complex because the measurement should be adjusted for gestational age using reference ranges, which introduces another important source of error. All these types of variability have an additive effect that is reflected in clinical practice.

A way to identify and mitigate, in part, the issue of measurement error is to apply measurement-error theory and quantify the bias and spread of the measurement through a specific audit for each operator[7]. We also believe that simply being aware of the different sources of error and bias can lead to improved practice.

In the present Opinion, we address the issue of measurement error in fetal biometry, discuss the role of audit in safeguarding proper fetal growth assessment and demonstrate an approach to implementation of fetal biometry audit in everyday practice, with the aim of achieving a beneficial impact on antenatal care.

Clinical impact of crown–rump length measurement error

CRL measurement is an early form of assessment of fetal growth and is considered to be the optimal way of dating pregnancy at the time of the first-trimester nuchal translucency (NT) scan[2,8]. Additionally, all first-trimester biomarkers are standardized directly based on CRL or indirectly by correcting for gestational age that has been determined using CRL. Therefore, incorrect and inconsistent CRL measurements affect the basis of antenatal care and have impact on pregnancy dating, standardization of biophysical and biochemical indices, screening for trisomies, assessment of fetal growth and, ultimately, obstetric decisions[9–12].

Despite the extensive use of CRL, there is considerable debate regarding the charts used, and researchers strive to increase awareness of the need for standardization of CRL measurement[8–12]. The impact of CRL measurement error is clear and profound, affecting the incidence of clinical outcomes and, therefore, inducing a different clinical reality. The rates of preterm and post-term birth, small-for-gestational age (SGA) and large-for-gestational age (LGA) are highly dependent on correct CRL measurement.

Repeatability studies revealed that, in 95% of cases, the differences between CRL measurements by two operators were roughly within ±5 mm or 2.5 days of gestation[9,12,13]. The clinical impact of this measurement error has been examined by simulation studies, which demonstrated that the performance of prenatal screening for Down syndrome relies on accurate pregnancy dating using CRL measurement[9,12]. The necessity to account for gestational-age-dependent changes for the markers used introduces the CRL measurement error into screening[9,12]. A relatively small under- or overestimation of CRL results in considerable respective under- or overestimation of patient-specific risk for trisomy 21. Underestimating CRL lowers the expected NT and the expected maternal serum pregnancy-associated plasma protein-A concentration, while the expected free β-human chorionic gonadotropin concentration becomes higher. The NT-based risk increases and the biochemical risk decreases, resulting in an overall reduction of the combined risk[9,12]. Overestimating CRL has the opposite effect, increasing the combined risk. Therefore, a systematic under- or overestimation of CRL diminishes the operator-specific screening performance[9,12].

A recent study revealed that even a clinically reasonable and relatively small measurement error of −2 mm would shift an estimated fetal weight (EFW) on the 10th percentile at the 20-week scan to around the 20th percentile, and a CRL measurement error of +2 mm would shift EFW on the 10th percentile to around the 5th percentile[10].

Another nationwide cross-sectional study revealed that the dating formula used has a significant impact on the evaluation of fetal biometry later in pregnancy[11]. Consequently, there is an urgent need to increase quality awareness, reach a consensus about reference ranges used and launch auditing policies for this crucial measurement[10,11].

Clinical impact of measurement error in fetal biometry

The main purpose of the assessment of fetal growth is the quantification of the true fetal weight at the time of the scan. There are five important sources of variation in EFW distribution. The first is the biological variation, the second is the measurement error for the biometric indices that can be further influenced by the expected values[14], the third is the error introduced by the formula used for the computation of EFW, the fourth is the error due to dating by CRL and the fifth is the one related to the choice of the reference range or standard used to evaluate EFW computed according to the measurements at the given estimated gestational age. Reduction of the second and fourth sources of variation should be the objective of an effective audit policy, reduction of the third may potentially be accomplished by improving the algorithm, while reduction of the biological variation may be impossible to achieve. Finally, the reduction of the fifth source of variation should be through selection of the most methodologically robust reference chart, in view of the multiple biases discussed in this manuscript, and not just through selection of the chart that would yield 10% of small individuals in a given population.

The initial step needed to improve effectiveness of fetal growth assessment is the standardization of biometric measurements[3]. Consistency and adherence to common techniques is of paramount importance to ensure quality control. The second step is to deal with the measurement error. A recent simulation study demonstrated that the measurement error in fetal biometry causes substantial error in EFW, resulting in misclassification of SGA and LGA fetuses[15]. Assuming random Gaussian errors for AC, HC and FL, when the 10th and 90th percentiles of EFW are used to identify true SGA and LGA fetuses, only 78% of SGA and LGA will be classified correctly at the time of the scan[15]. However, measurement error cannot be considered as having a simple random Gaussian distribution given that it has been shown that expected-value bias frequently occurs[14]. This conclusion should increase awareness of the fact that fetal biometry is not equivalent to fetal weight but rather is a good approximation of it if essential technical prerequisites are met.

Another important clinical question is how to translate fetal biometry into clinical practice, considering the inevitable measurement error. Using EFW percentiles as a single, fixed diagnostic criterion for SGA or LGA misclassifies a significant proportion of fetuses at the time of assessment[4,15]. There is also accumulating evidence that EFW is the best predictor of SGA in the context of prediction models applied as early as 20 weeks of gestation[5,6]. In the newly shaped framework of personalized care, EFW is a powerful continuous biomarker, and we should avoid using it as a fixed, arbitrary classification criterion[4–6,15]. Significant progress was made after switching from a single NT threshold (3 or 3.5 mm) to continuous use of this variable, within a multivariate model[16–19]. The same should now be applied to EFW in screening for growth restriction[4–6]. Probabilistic continuous models that also include maternal history and other biomarkers of impaired placentation should improve the prediction of both imminent and later smallness[4–6]. Therefore, the key aims are, first, to reduce measurement error by standardization and adequate training; second, to use EFW as a continuous biomarker in the context of effective models, taking into account the inherent and unavoidable statistical error; and, third, to apply audit policies to ensure application of the abovementioned strategies.

Reference ranges and formulas

Prediction models that aim to personalize care utilize the information from fetal biometry after adjustment for gestational age[4–6]. This adjustment requires the use of appropriate reference ranges for gestational age. Choosing charts for fetal and neonatal weight is a crucial decision. The chart for neonatal weight is the one that affects the incidence of the outcome of interest, while the chart for EFW affects the performance of a model, considering that EFW is the best predictor of growth deviation[4–6].

Salomon *et al.* recognized that including preterm births when constructing reference ranges is misleading because preterm birth is associated with growth restriction[20]. The authors recommended to use EFW charts to avoid underestimation of fetal growth restriction. However, another issue is the higher variation of birth weight compared with EFW. An alternative method to resolve these issues is to use EFW for preterm gestational ages and birth weight for the term ones, assuming a common median and different spread around the mean for birth weight and EFW[21]. This method overcomes the underestimation of preterm growth restriction, and it was the basis for the Fetal Medicine Foundation fetal and neonatal growth charts[21]. Another interesting approach to standardizing measurements and reducing the risk of bias is the one of the INTERGROWTH-21st project. This consortium created homogeneous charts for pregnancy dating based on CRL and biometry, using data from many different countries collected by specifically trained and audited practitioners blinded to the measurements they were taking[22,23]. The authors also reported that, although they included preterm births, they avoided including high-risk pregnancies[24].

When using local descriptive or ethnicity-specific charts, this will artificially set 10% of the population below the 10th percentile, but it is also likely that

being on the 10th percentile of an American or Indian reference chart does not have the same meaning[25,26]. The adoption of a global standard would be a major step forward, allowing meaningful evaluation, comparison and collaboration. This would in no way prevent different practices between different countries for the same percentile, depending on local health and economic characteristics. The current international recommendations are stubbornly attempting to reach a consensus regarding practices for percentiles that have completely different definitions and meanings in different countries, making the objective an unattainable one.

Estimation of fetal weight using ultrasound requires a regression formula that combines fetal biometry measurements. Various formulas have been developed and compared in several studies. It has been recently reported that Hadlock's three-parameter (HC, AC and FL) formula has the least error when assessing the agreement with the true weight of the fetus[27,28]. We would, therefore, recommend the universal use of Hadlock's three-parameter formula[27].

Audit

Audit may have several forms, including reproducibility studies with the Bland–Altman analysis, image quality control by a panel of experts or application of the CUSUM method. An important methodology that has been used widely with great success for auditing NT measurements is the annual assessment of the distributional properties of the operator's measurements[7]. This method quantifies the systematic error (bias) and random error (spread) of measurements. Detection of a significant divergence from the expected distribution increases awareness of potential clinical implications, indicating the need for further training and recertification[7].

Mathematical background and performance indices of measurements

The first step required to move forward to implementation of auditing policies is to elaborate on the mechanics of measurement error. The distribution of a given biometric measurement X is defined as: X = mean for gestational age + biological variation + measurement error.

Biological variation is a distribution with a mean of zero and a SD similar to the SD of the residuals. Residual is the difference between the true measurement and the mean for the gestational age. Biological variation represents how the measurement varies because of natural variation of fetal size, and it usually increases with gestational age. This increase is called heteroscedasticity, and it can be modeled effectively. The measurement error is considered, in the absence of additional expected-value bias, to have a Gaussian distribution with a mean, which is called bias, and a SD, which is called precision. In this definition, the bias represents the average deviation from the mean for gestational age and the precision represents the variation of the measurement error (Figure 1). The total variance is the sum of variances due to biological variation and precision of the measurement.

The audit can be based on the quantification of systematic-error bias (the average difference between the measured and the expected value for gestational age) and the spread of the measurement due to the total variance. The quality control of a given biometric measurement assumes, of course, that its variability is known and, therefore, that a perfectly defined standard is used, with equations allowing to calculate the expected value, SD and *Z*-score of any observed value at any gestational age. Additionally, data should be assessed for any trend for changing bias for different gestational ages. The operators' distributional properties should be compared against the acceptable limits of bias, spread and trend (Figure 2).

A previous simulation study described the impact of bias on the detection of SGA and LGA babies at the time of assessment[11]. The next step should be the quantification of the effect of deviation of biometry on the early prediction of SGA and relevant stratification of pregnancy care.

Personalization of care

The implementation of prenatal screening for Down syndrome by measuring NT thickness caused a major shift from predetermined, arbitrary and ineffective criteria, such as increased maternal age, to a personalized risk assessment[1]. Ongoing audit of the measurement

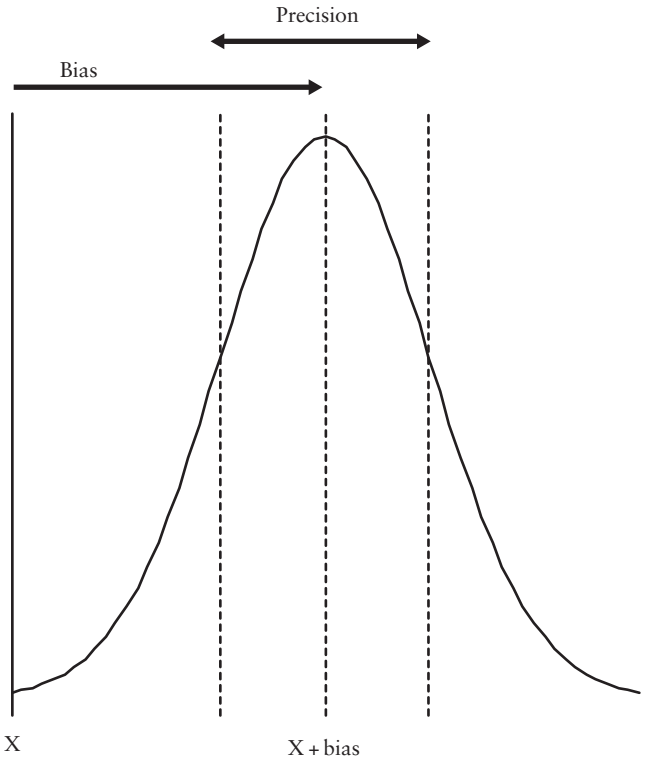

Figure 1 Graphical representation of measurement-error model. Assuming Gaussian errors, bias is mean and precision is spread of such distribution. Presence of error distribution modifies the mean of the actual measurement (X) and also contributes to the total variance.

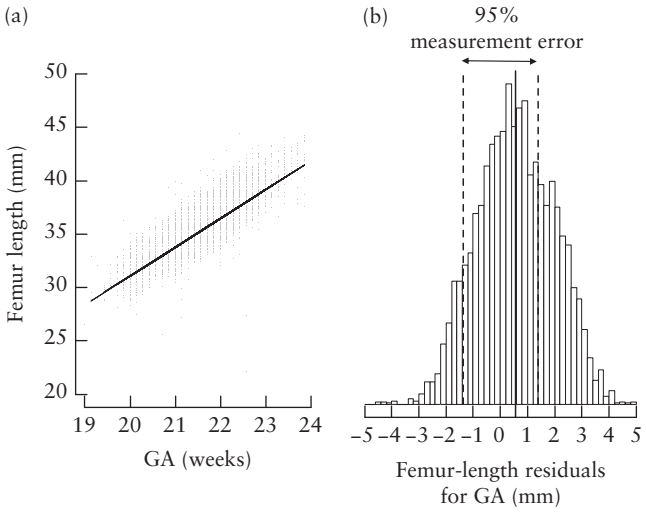

Figure 2 Example of an audit for femur length. (a) Distribution of femur-length measurements according to gestational age at measurement (GA) with the 50th percentile superimposed. (b) Gaussian distribution of femur-length residuals (femur length − predicted mean for GA). Operator-specific bias = 0.56 mm, spread = 1.47 mm and trend = 0.03.

of NT has had a pivotal role in the quality assurance and widespread application of the method[7] and was made possible by the use of a single, universally applicable NT standard that provides a solid foundation for subsequent combined screening. Obstetric care is increasingly based on algorithms that combine maternal factors with biophysical measurements of the mother and fetus[4–6,29,30]. The management of the major obstetric conditions can be based on prediction models applicable at different stages of pregnancy[1]. Specifically, for growth restriction, the contemporary trend is personalized care at different pregnancy stages, in the context of policies that can be tailored to the needs and resources of the healthcare system[4–6,30]. Also, another form of personalization is to use customized growth standards that may improve the detection of adverse perinatal outcome[31].

Personalization of care has been erroneously considered to be the process of having a different approach to each case. On the contrary, personalization of care advocates a universal integrated approach that can be tailored in terms of prediction, prognostication and decision-making. Individualization of antenatal care involves applying universal models on a large scale, leveraging biomarkers measured according to specific criteria. This method may be prone to increased measurement error, as previous studies have demonstrated[7,9–12].

The basis of the new era of precision medicine that emerges is the use of biomarkers in the milieu of continuous survival models applied in a Bayesian framework[4–6,29,30]. Inevitably, the performance of these algorithms will depend on the correct assessment of biomarkers, and the operators' measurement error may have a profound impact on decision-making and consequently on perinatal outcome.

Artificial intelligence and fetal biometry

Artificial intelligence (AI) is based on the hypothesis that a machine can simulate human learning and intelligence[32]. We have entered a new era that fulfils three basic requirements for AI application: big data, computational power and AI algorithms[33]. In medicine, non-symbolic AI simulates learning, perception and pattern recognition. Machine learning with artificial neural networks algorithms or deep learning is the main AI approach used, especially in obstetrics and fetal medicine[33]. The main characteristic of this type of AI is that the internal process that takes place within the algorithm remains to a great extent ambiguous to the user[33].

Research attempts have been focused on combining deep-learning algorithms with ultrasound scan[34]. The aim is to automatically acquire, measure and store standard fetal biometric planes[34]. The application of AI in antenatal ultrasound biometry has some important challenges. For the highly trained operators, AI must demonstrate usefulness in reducing scanning time and improving the clinical workflow. AI should be compared with standard assessment with regard to measurement error. Will AI reduce variability associated with biometric measurement? For operators lacking specialist skills, the crucial question is whether AI will substitute the lack of training and experience. Moreover, it will be interesting to see whether AI will transform ultrasound scan practice and training by shifting our attention from the AI-acquired biometry to other aspects of imaging. Another important application is the use of deep-learning algorithms to assist a basic scan, including fetal biometry, considering the limited capacity of current databases[35].

Alan Turing laid the foundation for AI by introducing the homonymous Turing test in 1950[36]. The Turing test, otherwise known as the imitation game, assesses a machine's ability to exhibit intelligence indistinguishable from that of a human. The machines that pass this test are considered to be AI machines. It is probable that, sooner or later, the Turing test will be inverted because AI will be at the minimum equivalent to human intelligence. In fetal medicine, biometry may be a field in which this inversion of the Turing's theorem is possible.

Conclusion and future direction

The transition to a risk-based obstetric management is rapid, and algorithm-driven individualized pregnancy care is becoming the new clinical reality[1]. The increasing worldwide use of prediction algorithms for fetal trisomies, pre-eclampsia, growth restriction and preterm birth is anticipated to improve fetal, maternal and perinatal outcomes.

The decisions for interventions in fetal growth disorders should be based on a personalized probabilistic framework. We must move forward to an integrated early-risk assessment for smallness/growth restriction and macrosomia, in which EFW is a potent biomarker amongst others[1,4–6]. This should allow effective stratification and

personalization of care[1,4–6]. A prerequisite for this new era of precision medicine is correct ultrasound measurements according to strict criteria and stringent standardization of our practices and standards at each of the steps (pregnancy dating, biometric measurements, EFW formulas, prescriptive standards used). We must go beyond the current debate regarding the type of reference ranges used and the role of population characteristics. Measurement accuracy is becoming the cornerstone of modern obstetrics, and we anticipate that the implementation of audit policies should stimulate a discussion that will probably lead to a consensus for the standardization of measurements. Despite the increasing use of fetal biometry and the fact that the need for quality control in fetal biometry has long been recognized, fetal biometry is used without any systematic quality control. An important need for our clinical practice is the introduction of vigorous audit. This will ensure the required quality for the implementation of algorithms, which may have a profound impact on perinatal outcome.

Improving perinatal outcome is an objective that will remain unfulfilled unless we adopt common universal standards and efficient quality-control policies. The ultimate goal through a process of universal practice and audit is to ensure adequate training with adherence to common techniques. The implementation of systematic auditing should enhance quality of care and awareness amongst health providers.

REFERENCES

1. Nicolaides KH. Turning the pyramid of prenatal care. *Fetal Diagn Ther* 2011; **29**: 183–196.
2. International Society of Ultrasound in Obstetrics and Gynecology: Bilardo CM, Chaoui R, Hyett JA, Kagan KO, Karim JN, Papageorghiou AT, Poon LC, Salomon LJ, Syngelaki A, Nicolaides KH. ISUOG Practice Guidelines (updated): performance of 11–14-week ultrasound scan. *Ultrasound Obstet Gynecol* 2023; **61**: 127–143.
3. Salomon LJ, Alfirevic Z, Da Silva Costa F, Deter RL, Figueras F, Ghi T, Glanc P, Khalil A, Lee W, Napolitano R, Papageorghiou A, Sotiriadis A, Stirnemann J, Toi A, Yeo G. ISUOG Practice Guidelines: ultrasound assessment of fetal biometry and growth. *Ultrasound Obstet Gynecol* 2019; **53**: 715–723.
4. Papastefanou I, Thanopoulou V, Dimopoulou S, Syngelaki A, Akolekar R, Nicolaides KH. Competing-risks model for prediction of small-for-gestational-age neonates at 36 weeks' gestation. *Ultrasound Obstet Gynecol* 2022; **60**: 612–619.
5. Papastefanou I, Nowacka U, Syngelaki A, Dragoi V, Karamanis G, Wright D, Nicolaides KH. Competing-risks model for prediction of small-for-gestational-age neonates from estimated fetal weight at 19–24 weeks' gestation. *Ultrasound Obstet Gynecol* 2021; **57**: 917–924.
6. Papastefanou I, Nowacka U, Syngelaki A, Mansukhani T, Karamanis G, Wright D, Nicolaides KH. Competing risks model for prediction of small-for-gestational-age neonates from biophysical markers at 19 to 24 weeks' gestation. *Am J Obstet Gynecol* 2021; **225**: 530.e1–19.
7. Kagan KO, Wright D, Etchegaray A, Zhou Y, Nicolaides KH. Effect of deviation of nuchal translucency measurements on the performance of screening for trisomy 21. *Ultrasound Obstet Gynecol* 2009; **33**: 657–664.
8. Ioannou C, Sarris I, Hoch L, Salomon LJ, Papageorghiou AT; International Fetal and Newborn Growth Consortium for the 21st Century. Standardisation of crown-rump length measurement. *BJOG* 2013; **120** Suppl 2: 38–41.
9. Kagan KO, Hoopmann M, Baker A, Huebner M, Abele H, Wright D. Impact of bias in crown–rump length measurement at first-trimester screening for trisomy 21. *Ultrasound Obstet Gynecol* 2012; **40**: 135–139.
10. Gadsbøll K, Wright A, Kristensen SE, Verfaille V, Nicolaides KH, Wright D, Petersen OB. Crown–rump length measurement error: impact on assessment of growth. *Ultrasound Obstet Gynecol* 2021; **58**: 354–359.
11. Fries N, Dhombres F, Massoud M, Stirnemann JJ, Bessis R, Haddad G, Salomon LJ. The impact of optimal dating on the assessment of fetal growth. *BMC Pregnancy Childbirth* 2021; **21**: 167.
12. Salomon LJ, Bernard M, Amarsy R, Bernard JP, Ville Y. The impact of crown–rump length measurement error on combined Down syndrome screening: a simulation study. *Ultrasound Obstet Gynecol* 2009; **33**: 506–511.
13. Souka AP, Pilalis A, Papastefanou I, Salamalekis G, Kassanos D. Reproducibility study of crown-rump length and biparietal diameter measurements in the first trimester. *Prenat Diagn* 2012; **32**: 1158–1165.
14. Drukker L, Droste R, Chatelain P, Noble JA, Papageorghiou AT. Expected-value bias in routine third-trimester growth scans. *Ultrasound Obstet Gynecol* 2020; **55**: 375–382.
15. Wright D, Wright A, Smith E, Nicolaides KH. Impact of biometric measurement error on identification of small- and large-for-gestational-age fetuses. *Ultrasound Obstet Gynecol* 2020; **55**: 170–176.
16. Nicolaides KH, Azar G, Byrne D, Mansur C, Marks K. Fetal nuchal translucency: ultrasound screening for chromosomal defects in first trimester of pregnancy. *BMJ* 1992; **304**: 867–869.
17. Pandya PP, Goldberg H, Walton B, Riddle A, Shelley S, Snijders RJ, Nicolaides KH. The implementation of first-trimester scanning at 10-13 weeks' gestation and the measurement of fetal nuchal translucency thickness in two maternity units. *Ultrasound Obstet Gynecol* 1995; **5**: 20–25.
18. Snijders RJ, Noble P, Sebire N, Souka A, Nicolaides KH. UK multicentre project on assessment of risk of trisomy 21 by maternal age and fetal nuchal-translucency thickness at 10-14 weeks of gestation. Fetal Medicine Foundation First Trimester Screening Group. *Lancet* 1998; **352**: 343–346.
19. Wright D, Kagan KO, Molina FS, Gazzoni A, Nicolaides KH. A mixture model of nuchal translucency thickness in screening for chromosomal defects. *Ultrasound Obstet Gynecol* 2008; **31**: 376–383.
20. Salomon LJ, Bernard JP, Ville Y. Estimation of fetal weight: reference range at 20–36 weeks' gestation and comparison with actual birth-weight reference range. *Ultrasound Obstet Gynecol* 2007; **29**: 550–555.
21. Nicolaides KH, Wright D, Syngelaki A, Wright A, Akolekar R. Fetal Medicine Foundation fetal and neonatal population weight charts. *Ultrasound Obstet Gynecol* 2018; **52**: 44–51.
22. Papageorghiou AT, Kennedy SH, Salomon LJ, Ohuma EO, Cheikh Ismail L, Barros FC, Lambert A, Carvalho M, Jaffer YA, Bertino E, Gravett MG, Altman DG, Purwar M, Noble JA, Pang R, Victora CG, Bhutta ZA, Villar J; International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). International standards for early fetal size and pregnancy dating based on ultrasound measurement of crown–rump length in the first trimester of pregnancy. *Ultrasound Obstet Gynecol* 2014; **44**: 641–648.
23. Papageorghiou AT, Ohuma EO, Altman DG, Todros T, Cheikh Ismail L, Lambert A, Jaffer YA, Bertino E, Gravett MG, Purwar M, Noble JA, Pang R, Victora CG, Barros FC, Carvalho M, Salomon LJ, Bhutta ZA, Kennedy SH, Villar J; 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.
24. Stirnemann J, Villar J, Salomon LJ, Ohuma E, Ruyan P, Altman DG, Nosten F, Craik R, Munim S, Cheikh Ismail L, Barros FC, Lambert A, Norris S, Carvalho M, Jaffer YA, Noble JA, Bertino E, Gravett MG, Purwar M, Victora CG, Uauy R, Bhutta Z, Kennedy S, Papageorghiou AT. International estimated fetal weight standards of the INTERGROWTH-21st Project. *Ultrasound Obstet Gynecol* 2017; **49**: 478–486.
25. Tarca AL, Romero R, Gudicha DW, Erez O, Hernandez-Andrade E, Yeo L, Bhatti G, Pacora P, Maymon E, Hassan SS. A new customized fetal growth standard for African American women: the PRB/NICHD Detroit study. *Am J Obstet Gynecol* 2018; **218**: S679–S691.e4.
26. Ioannou C, Talbot K, Ohuma E, Sarris I, Villar J, Conde-Agudelo A, Papageorghiou AT. Systematic review of methodology used in ultrasound studies aimed at creating charts of fetal size. *BJOG* 2012; **119**: 1425–1439
27. Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the use of head, body, and femur measurements—a prospective study. *Am J Obstet Gynecol* 1985; **151**: 333–337.
28. Hammami A, Mazer Zumaeta A, Syngelaki A, Akolekar R, Nicolaides KH. Ultrasonographic estimation of fetal weight: development of new model and assessment of performance of previous models. *Ultrasound Obstet Gynecol* 2018; **52**: 35–43.
29. Wright D, Wright A, Nicolaides KH. The competing risk approach for prediction of preeclampsia. *Am J Obstet Gynecol* 2020; **223**: 12–23.e7.
30. Papastefanou I, Wright D, Syngelaki A, Souretis K, Chrysanthopoulou E, Nicolaides KH. Competing-risks model for prediction of small-for-gestational-age neonates from biophysical and biochemical markers at 11–13 weeks' gestation. *Ultrasound Obstet Gynecol* 2021; **57**: 52–61.
31. Kabiri D, Romero R, Gudicha DW, Hernandez-Andrade E, Pacora P, Benshalom-Tirosh N, Tirosh D, Yeo L, Erez O, Hassan SS, Tarca AL. Prediction of adverse perinatal outcome by fetal biometry: comparison of customized and population-based standards. *Ultrasound Obstet Gynecol* 2020; **55**: 177–188.
32. McCarthy J, Minsky M, Rochester N, Shannon C. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine* 2006; **27**: 12.
33. Dhombres F, Bonnard J, Bailly K, Maurice P, Papageorghiou AT, Jouannic JM. Contributions of Artificial Intelligence Reported in Obstetrics and Gynecology Journals: Systematic Review. *J Med Internet Res* 2022; **24**: e35465.
34. Drukker L, Noble JA, Papageorghiou AT. Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. *Ultrasound Obstet Gynecol* 2020; **56**: 498–505.
35. Arroyo J, Marini TJ, Saavedra AC, Toscano M, Baran TM, Drennan K, Dozier A, Zhao YT, Egoavil M, Tamayo L, Ramos B, Castaneda B. No sonographer, no radiologist: New system for automatic prenatal detection of fetal biometry, fetal presentation, and placental location. *PLoS One* 2022; **17**: e0262107.
36. Turing AM. I–Computing Machinery and Intelligence. *Mind* 1950; **LIX**: 433–460.

