Ultrasound Obstet Gynecol 2012; **39**: 563–568
Published online in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.10081

Advantages of the population-based approach to pregnancy dating: results from 23 020 ultrasound examinations

I. ØKLAND*†, J. NAKLING‡, H. K. GJESSING§¶, P. GRØTTUM**and S. H. EIK-NES†#

**Department of Obstetrics and Gynecology, Stavanger University Hospital, Stavanger, Norway; †National Center for Fetal Medicine, Department of Obstetrics and Gynecology, St Olavs University Hospital, Trondheim, Norway; ‡Department of Obstetrics and Gynecology, Innlandet Hospital Trust, Lillehammer, Norway; §Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway; ¶Department of Public Health and Primary Health Care, University of Bergen, Bergen, Norway; **Section of Medical Informatics, University of Oslo, Oslo, Norway; #Department of Laboratory Medicine, Children's and Women's Health, Norwegian University of Science and Technology, Trondheim, Norway*

KEYWORDS: date of delivery; fetal measurements; gestational age; post-term pregnancy; second trimester; term prediction

ABSTRACT

Objective *To confirm the results from two previous evaluations of term prediction models, including two sample-based models and one population-based model, in a third population.*

Methods *In a study population of 23 020 second-trimester ultrasound examinations, data were prospectively collected and registered over the period 1988–2009. Three different models for ultrasonically estimated date of delivery were applied to the measurements of fetal biparietal diameter (BPD) and two models were applied to the femur length (FL) measurements; the resulting term estimations were compared with the actual time of delivery. The difference between the actual and the predicted dates of delivery (the median bias) was calculated for each of the models, for three BPD/FL-measurement subgroups and for the study population as a whole.*

Results *For the population-based model, the median bias was + 0.4 days for the BPD-based predictions and −0.4 days for the FL-based predictions, and the biases were stable over the inclusion ranges. The biases of the two traditional models varied with the size of the fetus at examination; median biases were −0.87 and + 2.2 days, respectively, with extremes −4.2 and + 4.8 days for the BPD-based predictions, and the median bias was + 1.72 days with range −0.8 to + 4.5 days for FL-based predictions. The disagreement between the two sample-based models was never less than 2 days for the BPD-based predictions.*

Conclusion *This study confirms the results from previous studies; median biases were negligible with term predictions from the population-based model, while those from the traditional models varied substantially. The biases, which have clinical implications, seem inevitable with the sample-based models, which, even if overall biases were removed, will perform unsatisfactorily. Copyright © 2012 ISUOG. Published by John Wiley & Sons, Ltd.*

INTRODUCTION

In modern pregnancy care, it is recommended to date pregnancies by ultrasound in the first trimester or the second trimester[1,2]. Thus, it is imperative that the dating models are reliable. To assess prediction quality and reveal potential systematic biases, the models must be evaluated in large populations. In earlier studies, we evaluated two traditional, sample-based models for term prediction and demonstrated significant and nearly identical biases when the models were applied to different populations[3,4]. We also validated a population-based model that avoided the biases[4], which generally appear to be model-dependent.

The terms 'assessment of gestational age (GA)' and 'estimation of date of delivery' are considered almost synonymous. However, the calculations in fact concern totally different times in a pregnancy. Traditional models primarily estimate a last menstrual period (LMP) from second-trimester fetal measurements; thus, the estimated date of delivery (EDD) is actually an indirect and secondary issue[3,5,6]. Conversely, the new population-based models are constructed from observations of the verified date of delivery, predicting the remaining time of pregnancy and EDD directly from first- or second-trimester fetal measurements[5,7].

First-trimester screening tests[8,9] and the management of extremely preterm deliveries close to the limits of viability[10,11] rely on an accurately calculated GA, while the scheduling of invasive procedures and interventions in later pregnancy and determining when to induce in

Correspondence to: Dr I. Økland, Department of Obstetrics and Gynecology, Stavanger University Hospital, PO Box 8100, N-4068 Stavanger, Norway (e-mail: iok@lyse.net)

Accepted: 18 August 2011

post-term pregnancies[12,13], depend on knowledge of a reliable EDD. Therefore, it is still relevant to demonstrate prediction biases in the range of 2–5 days.

Because the EDD is model-dependent, recommending post-term induction practices without a uniform system for pregnancy dating is futile[12,14,15]. Moreover, different term prediction routines and induction practices make it impossible to compare important perinatal quality indicators. This has major consequences for comparison of data between countries, regions and even hospitals[12].

The significantly biased term predictions that resulted from applying the sample-based models[3,4] were avoided with a direct, LMP-independent model[4]. In this study, we extend previous findings to a population examined using slightly different routines and different sonographers. We also explore the mechanisms that make systematic biases almost inevitable with the sample-based models.

SUBJECTS AND METHODS

Subjects

The data were collected over a period of 22 years, from 1988 to 2009. They comprise routine fetal ultrasound examinations performed in Oppland County, Norway, on an unselected population of pregnant women. There are two maternity wards in the county, at the Gjøvik and Lillehammer Hospitals, each handling between 800 and 1100 births every year. Most of the women in the study were examined prenatally and subsequently gave birth at these hospitals. In addition, data were collected from two smaller midwifery units at other locations in the same county.

Pregnancies with a fetal biparietal diameter (BPD) in the range of 38–60 mm or a femur length (FL) in the range of 21–42 mm at the routine ultrasound examination were included. Multiple pregnancies, pregnancies complicated by stillbirth, diagnosed anomalies, induction of labor for reasons other than post-term pregnancy or elective Cesarean sections were not included. In accordance with the post-term managing scheme of the maternity wards during the study period, inductions at 11 or more days past the EDD were defined as post-term inductions.

In total, fetal measurements from 23 020 second-trimester routine ultrasound examinations were included.

Prediction models

Three different models for estimating the date of delivery were evaluated in this study. The obstetric wheel 'Snurra' (referred to here as 'Trondheim–1984')[16] predicts GA from second-trimester BPD measurements between 38 and 60 mm only, with EDD derived using a pregnancy duration of 282 days. The model was developed from ultrasound examinations of 90 carefully selected women with reliable menstrual data and anticipated normal pregnancies, included in a prospective, longitudinal study. A fourth-order polynomial regression analysis was used to establish the curves. This was the only dating method in use in Norway from 1984 until 2005.

The second sample-based method, 'Terminhjulet' ('Bergen–2004')[17,18], used a similar statistical model based on fractional polynomial regression analysis[19], with a newer data sample. It predicts GA from measurements of BPD (14–60 mm), FL (2–44 mm) or head circumference (not considered in this study) (50–134 mm), with EDD derived using a pregnancy duration of 282 days. The model was constructed from a cross-sectional study of 650 highly selected, healthy women with reliable menstrual data, assumed uncomplicated, singleton pregnancies, and with fewer than 14 days' disagreement between the LMP-based and the ultrasound-based EDD.

The third prediction model 'eSnurra' ('Trondheim–2007')[5] employs the new population-based approach with direct prediction of date of delivery. It is based on second-trimester fetal measurements from an unselected population of approximately 37 000 singleton pregnancies. The median remaining time of pregnancy was estimated using a local linear quantile regression model. Trondheim–2007 predicts the date of delivery from BPD (25–60 mm) or from FL (11–42 mm) measurements.

The application of each model's prediction table is described elsewhere[3,4]. For the Trondheim–1984 and the Bergen–2004 models, the date of delivery is estimated by adding 282 days to the estimated LMP date. For Trondheim–2007, predicted remaining time of pregnancy is found from the published tables[5].

Ultrasound examinations

The ultrasound examinations were mostly performed by specially trained midwives at the hospitals or at the midwifery units, and the remaining examinations (10–15% per year) were performed by doctors at the hospitals or in private practice. There were 23 different examiners altogether. Four of the midwives each performed between 3200 and 7000 of the included ultrasound examinations. The data were prospectively registered in a database. A large proportion of the included data were measurements from pregnancy weeks 17–19. In general, all clinical problems were managed according to the EDD predicted by the Trondheim–1984 model.

The management of post-term pregnancies in the departments was modified during the study period. In the first years, induction of labor was scheduled around 14 days past the EDD (≥ 296 days), while in later years the post-term inductions were gradually scheduled earlier, from 7 to 11 days past term.

The BPD and the FL were measured according to the standard method for fetal ultrasound biometry in Norway[3,4]: BPD was measured from the outer to the outer contour of the parietal bones, and the mean of three BPD measurements was used. The FL was measured as the length of the ossified part of the femoral diaphysis in a longitudinal section, using the longest of three measurements[20]. The Bergen–2004 model used the mean of three FL measurements: this issue has been addressed previously[3].

Statistical methods

The three models for ultrasound-based prediction of date of delivery were applied to data collected from the ultrasound examinations and the subsequent deliveries. The 22 815 measurements with a BPD in the range of 38 to 60 mm were used for EDD calculations with all three models. The 22 553 FL measurements between 21 and 42 mm were used with the Bergen–2004 and Trondheim–2007 models.

To correct for the narrowed beam width in newer ultrasound scanners, which is demonstrated to shorten measurements in the lateral direction[21], a correction for the time period[5] that applies to the FL measurements was included for the two newer models. The collection of the data in this study started in 1988, and newer prediction models should not be unrestrictedly applied to older data.

The resulting term predictions were compared with the actual time of delivery, and the disagreement was assessed in terms of the median bias for each model. The median bias reflects the systematic error of the term predictions and indicates the calibration level of the model as related to the study population. Predicting term too early results in a positive bias and an increased rate of apparently post-term pregnancies[3]. The median biases were calculated for subgroups with different fetal ages, as well as for the study population as a whole, because a varying bias may be missed if only the overall median bias is computed[22].

To assess the differences between the LMP-estimated GA at the actual time of the deliveries and the EDD as predicted from the BPD measurements with each model, data from women with available LMP information were used. From 1999 onwards, the registry included information on whether the LMP information was certain or not; in this period only women with certain LMP data were included in the subanalysis. As a result, 19 131 measurements were available. LMP was defined as reliable when the woman was certain about the exact LMP date.

P-values for testing a non-zero median bias were computed using permutation tests with 2000 permutations. All analyses and graphics were produced in the R statistical programming environment[23].

RESULTS

Table 1 shows the percentage of pregnancies ongoing at 4, 7, 11 and 14 days after the EDD predicted from BPD and FL measurements by each model. Depending on the prediction model, there is a considerable difference in the percentage of pregnancies classified as post-term. This shows that the choice of dating model has a strong impact on post-term induction rates, regardless of which day past the EDD is recommended for post-term induction.

The study population of 23 020 pregnancies, with large numbers even in the subgroups, resulted in only one median bias in one subgroup having a non-significant *P*-value, namely the bias of 0.13 days in the FL subgroup 21–26 mm for the Bergen–2004 model ($P = 0.18$). All other *P*-values were zero or < 0.01, indicating biases different from zero (results not shown). Hence, considering the large sample size, *P*-values were not very useful in deciding whether a bias was large enough to be clinically meaningful or not.

BPD-based predictions

Table 2 shows the median biases with 95% CI for the study group as a whole and for three different subgroups with BPD ranges corresponding to a GA of less than 18 weeks (38–43 mm), around 18 weeks (44–46 mm) and more than 18 weeks (47–60 mm). Figure 1 shows the median biases of the three models for each BPD value in

Table 2 Term prediction by the three different models using biparietal diameter (BPD) measurements. Median bias is presented for the three BPD-range groups and for the study group as a whole

BPD (mm)	n	*Model*	*Median bias (days (95% CI))*
38–43	6074	Trondheim–1984	−2.75 (−3.01 to −2.48)
		Bergen–2004	0.22 (0.17 to 1.14)
		Trondheim–2007	0.38 (0.16 to 0.56)
44–46	8682	Trondheim–1984	−0.75 (−0.95 to −0.54)
		Bergen–2004	2.29 (1.52 to 2.29)
		Trondheim–2007	0.57 (0.26 to 0.73)
47–60	8059	Trondheim–1984	0.35 (0.11 to 0.58)
		Bergen–2004	3.68 (3.31 to 3.86)
		Trondheim–2007	0.40 (0.30 to 0.72)
38–60	22 815	Trondheim–1984	−0.87 (−1.01 to −0.74)
		Bergen–2004	2.22 (2.14 to 2.29)
		Trondheim–2007	0.40 (0.30 to 0.57)

Table 1 Percentage of pregnancies still ongoing at 4, 7, 11 and 14 days after the estimated date of delivery (EDD) that was predicted, by each model, from biparietal diameter (BPD) or femur length (FL) measurements

	BPD-based predictions			*FL-based predictions*	
Days past EDD	*Trondheim–1984*	*Bergen–2004*	*Trondheim–2007*	*Bergen–2004*	*Trondheim–2007*
4	30.7	43.6	36.6	41.5	32.9
7	19.9	30.6	24.4	28.4	21.4
11	9.2	16.6	11.9	15.2	9.9
14	4.0	9.2	5.8	7.8	4.4

the span of the study. The biases of the two sample-based models varied substantially, both within and between the models; the bias within a model was related to different BPD values. The bias for the population-based model was stable, essentially within ± 1 day.

Figure 2 shows the GA at the time of the delivery as estimated from the LMP of the women with reliable LMP data, compared with the EDDs predicted from BPD measurements with the three ultrasound models, in the same pregnancies. There was a consistently lower discrepancy between the EDD predictions from the population-based ultrasound model and the LMP-based GAs than there was between the traditional ultrasound models' EDD predictions and the LMP-based GAs.

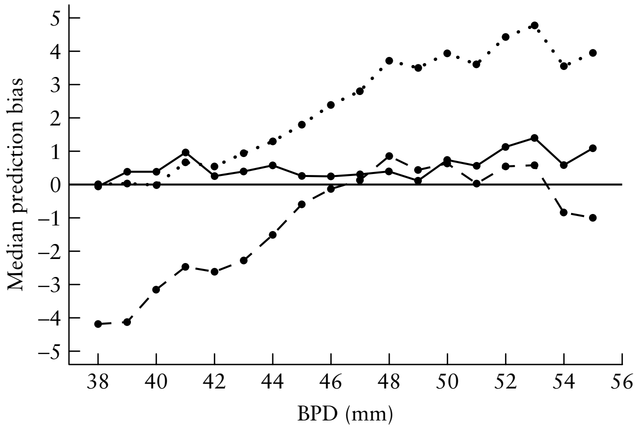

Figure 1 Median bias values for the three models (Trondheim–1984 (– –), Bergen–2004 (. . . .) and Trondheim–2007 (——)) related to different biparietal diameter (BPD) measurements.

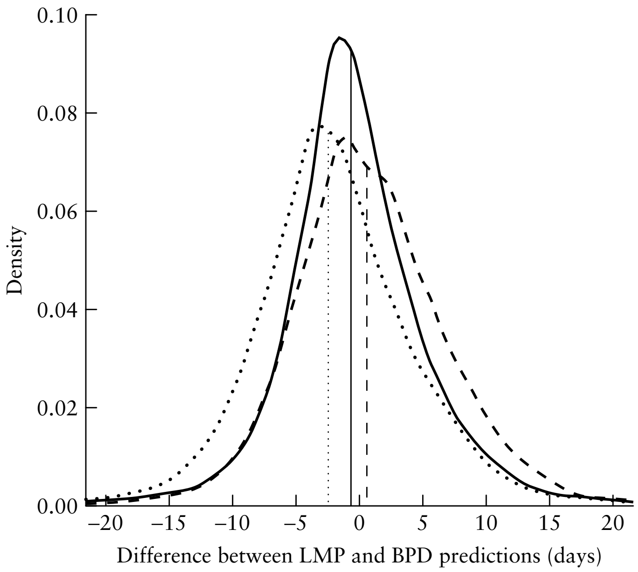

Figure 2 Differences between gestational age at delivery as estimated from last menstrual period (LMP) and date of delivery as predicted from biparietal diameter (BPD) measurements using the three ultrasound models (Trondheim–1984 (– –), Bergen–2004 (. . . .) and Trondheim–2007 (——)). The median difference is marked with vertical lines.

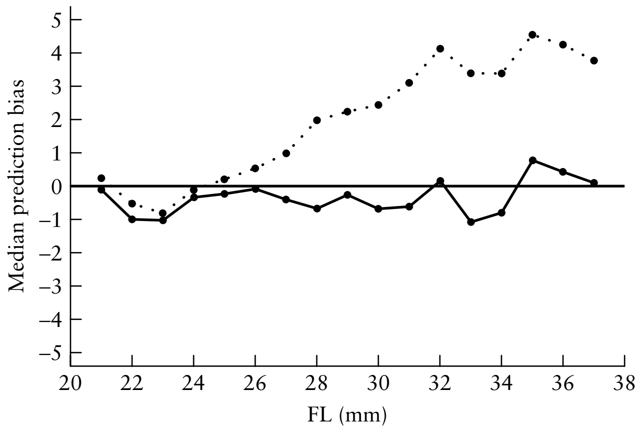

Figure 3 Median biases for the two models (Bergen–2004 (. . . .) and Trondheim–2007 (——)) related to different femur length (FL) measurements.

Table 3 Term prediction by two different models using femur length (FL) measurements. Median bias is presented for the three FL-range groups and for the study group as a whole

FL (mm)	n	*Model*	*Median bias (days (95% CI))*
21–26	5753	Bergen–2004	0.13 (−0.10 to 0.29)
		Trondheim–2007	−0.27 (−0.74 to −0.08)
27–29	9315	Bergen–2004	1.72 (1.28 to 1.72)
		Trondheim–2007	−0.26 (−0.48 to −0.16)
30–42	7485	Bergen–2004	3.39 (2.98 to 3.61)
		Trondheim–2007	−0.46 (−0.63 to −0.41)
21–42	22 553	Bergen–2004	1.72 (1.49 to 1.90)
		Trondheim–2007	−0.40 (−0.48 to −0.26)

FL-based predictions

The two models (Bergen–2004 and Trondheim–2007) were applied to the FL measurements. Figure 3 shows the median biases of the two models for each FL value in the inclusion range. The bias of the sample-based model varied substantially with the size of the fetus at the time of the ultrasound examination, while it was essentially stable for the population-based model. The extent of the biases was similar to that observed with the BPD-based predictions for the same models, and also to the biases found for the same models when evaluated on another population[4].

The median biases with 95% CI for the study group as a whole and for three different subgroups with FL ranges corresponding to a GA of less than 18 weeks (21–26 mm), around 18 weeks (27–29 mm) and more than 18 weeks (30–42 mm) are shown in Table 3.

DISCUSSION

The present evaluation of a population-based model for prediction of date of delivery and the comparison of its predictions with those from two traditional models emphasizes the importance of continuous quality assessment. In this sample of 23 020 second-trimester examinations, the EDD predictions from the BPD and FL measurements of the population-based model were reliable; the median bias was generally within ± 1 day, confirming earlier findings[4]. The biases of the sample-based models were considerable in the study population as a whole and in the subgroups, both for BPD-based and FL-based predictions. The median biases varied substantially both with the fetal size at the time of the examination and between the two models, one generally predicting too early and the other too late (Figures 1 and 3). This also agrees with previous findings[3,4]. Both models performed adequately for a restricted span of fetal measurements. The EDD discrepancy between them was consistently large, and never < 2 days for the BPD-based predictions. For the late FL-based predictions from Bergen–2004 the bias amounted to > 4 days.

In 2006, data collected from the same study population during 1989–1999 were used to evaluate the two sample-based models' BPD-based predictions[24]. That study included only women with reliable LMP dates, and all inductions of labor were excluded, negatively affecting the mean bias[25]. Therefore, an updated study was needed to evaluate the population-based model and the FL measurements, using the more stable median bias as the outcome.

To remove the overall median biases of the sample-based models, one could add or subtract a constant value to all predictions[22]; this would correspond to shifting the curves in Figures 1 and 3 up or down along the y-axis until the median bias was zero. However, the slope of the curves would remain. Particularly for Bergen–2004, as both curves slope upward to the right, a correct overall calibration would result in EDD predictions that are too late for the small fetuses and too early for the large ones. Thus, a simple calibration improving the overall bias would have unfortunate consequences. An optimal calibration should remove the bias over the whole range of measurement values; the population-based model achieves precisely that[4].

The population-based model was constructed from observations of the actual date of delivery to predict the remaining time of pregnancy and EDD from first-trimester or second-trimester fetal measurements[5,7]. However, modern pregnancy care requires a reliable EDD in the late stages of pregnancy and knowledge of GA in the early stages. The traditional sample-based models were devised to estimate a hypothetical LMP (i.e. GA) from second-trimester fetal measurements, and derive the EDD prediction from this[3,5,6]. The population-based model estimates the GA as 283 minus the predicted remaining time of pregnancy. In the reference population[5,26], 283 days is the median time from the LMP to birth. As the traditional models are based on estimating the LMP, one might assume that these methods would provide EDDs and GAs that correspond more closely to those computed from the LMP. Interestingly, this is not the case (Figure 2). The EDD predictions from the population-based method correspond more closely to the GA at delivery computed from the LMPs of women with reliable LMP data, as seen from the narrower distribution curve of the discrepancy between the LMP- and ultrasound-based estimates. An overall calibration to remove these median differences (the shifts of the curves away from zero) would not alter the shape of the curves. Thus, the population-based EDD predictions would still agree better with the LMP-based predictions.

The better corresponding ultrasound- and LMP-based estimates have immediate clinical consequences. First, it is beneficial both for scheduling examinations and reducing concern. Second, it reduces the risk of erroneous dating for fetuses that are small or large at the routine examination[27]. The new population-based method is thus better adapted to the actual target population than are the sample-based methods.

The present analysis, together with two previous studies[3,4], comprises a total of 73 400 examinations in three different populations. These studies demonstrate that both sample-based models give systematically biased EDD predictions and GA calculations. The essential problem is that the traditional models were developed on samples with distributions different from the populations to which they are applied. The population distribution has a strong concentration of examinations at around 17–19 weeks. The considerable numbers of small-for-gestational-age (SGA) and large-for-gestational-age (LGA) fetuses in this central group have correspondingly lower and higher BPD and FL values than expected, and therefore they 'spill over' to measurement intervals with fewer observations and pull their median values toward 17–18 weeks. For the entire population these medians constitute the optimal predictions of remaining time for a given biometric measurement, taking into account both the appropriate-for-GA (AGA) fetuses and the 'spillover' of SGA and LGA fetuses. The sample-based models were developed using data with a flat GA distribution, where the 'spillover' of neighboring SGA and LGA fetuses is balanced and will cancel out. Thus, such models in effect only pay attention to the AGA fetus. Using these models on an unselected population, the large numbers of SGA and LGA fetuses in the central part of the population that are incorrectly interpreted as AGA fetuses of lower or higher age, will produce the reported prediction biases. Conversely, because it is aimed at the actual population, the new method corrects for this effect.

The population-based model will better predict the date of delivery for fetuses with intrauterine growth restriction (IUGR). However, identification of early IUGR fetuses cannot be performed from one single ultrasound examination, irrespective of the prediction model[27]. Any significant difference between reliable LMP-based and ultrasound-based EDD dates indicates a need for further evaluations[27–29].

In uncomplicated pregnancies with spontaneous deliveries close to the EDD, inaccurate dating is of less clinical importance, yet of interest in assessing perinatal outcome or evaluating management protocols. While preterm deliveries are mainly unavoidable, even if occasionally scheduled, iatrogenic postmaturity may follow a biased EDD, leading to unnecessary induction of labor shortly past term[4]. The resulting increase in wrongly identified post-term pregnancies is substantial (Table 1), yet often ignored. A prerequisite for comparison of induction routines is unbiased and uniform EDD predictions with comparable post-term rates[12].

In conclusion, to obtain reliable EDD predictions, the distribution of the population used to develop a model must correspond to the population to which the model is applied. The model must also answer dating questions both in early and late pregnancy. Including this sample of 23 020 examinations, we have now confirmed our findings in a population totaling 73 400 examinations[3,4]. The population-based model is the method of choice for assessing GA and EDD.

ACKNOWLEDGMENT

We would like to thank Nancy Lea Eik-Nes for revising the manuscript.

REFERENCES

1. Whitworth M, Bricker L, Neilson JP, Dowswell T. Ultrasound for fetal assessment in early pregnancy. *Cochrane Database Syst Rev* 2010; Issue **4**: CD007058. DOI: 10.1002/14651858.
2. National Collaborating Centre for Women's and Children's Health. Antenatal care. Routine care for the healthy pregnant woman. National Institute for Health and Clinical Excellence: London, 2010; http://www.nice.org.uk/nicemedia/live/11947/40115/40115.pdf [Accessed 10 March 2011].
3. Økland I, Gjessing HK, Grøttum P, Eik-Nes SH. Biases of traditional term prediction models: results from different sample-based models evaluated on 41 343 ultrasound examinations. *Ultrasound Obstet Gynecol* 2010; **36**: 728–734.
4. Økland I, Gjessing HK, Grøttum P, Eggebø TM, Eik-Nes SH. A new population-based term prediction model vs. two traditional sample-based models: validation on 9046 ultrasound examinations. *Ultrasound Obstet Gynecol* 2011; **37**: 207–213.
5. Gjessing HK, Grøttum P, Eik-Nes SH. A direct method for ultrasound prediction of day of delivery: a new, population-based approach. *Ultrasound Obstet Gynecol* 2007; **30**: 19–27.
6. Hutchon DJ, Ahmed F. Naegele's rule: a reappraisal. *BJOG* 2001; **108**: 775.
7. Salomon LJ, Pizzi C, Gasparrini A, Bernard JP, Ville Y. Prediction of the date of delivery based on first trimester ultrasound measurements: an independent method from estimated date of conception. *J Matern Fetal Neonatal Med* 2010; **23**: 1–9.
8. Hyett JA, Perdu M, Sharland GK, Snijders RS, Nicolaides KH. Increased nuchal translucency at 10–14 weeks of gestation as a marker for major cardiac defects. *Ultrasound Obstet Gynecol* 1997; **10**: 242–246.
9. Snijders RJ, Johnson S, Sebire NJ, Noble PL, Nicolaides KH. First-trimester ultrasound screening for chromosomal defects. *Ultrasound Obstet Gynecol* 1996; **7**: 216–226.
10. Miljeteig I, Markestad T, Norheim OF. Physicians' use of guidelines and attitudes to withholding and withdrawing treatment for extremely premature neonates in Norway. *Acta Paediatr* 2007; **96**: 825–829.
11. Pignotti MS. Extremely preterm births: recommendations for treatment in European countries. *Arch Dis Child Fetal Neonatal Ed* 2008; **93**: 403–406.
12. Zeitlin J, Blondel B, Alexander S, Breart G. Variation in rates of postterm birth in Europe: reality or artefact? *BJOG* 2007; **114**: 1097–1103.
13. Gulmezoglu AM, Crowther CA, Middleton P. Induction of labour for improving birth outcomes for women at or beyond term. *Cochrane Database Syst Rev* 2006; Issue **4**: CD004945. DOI: 10.1002/14651858.
14. Lynch CD, Zhang J. The research implications of the selection of a gestational age estimation method. *Paediatr Perinat Epidemiol* 2007; **21 Suppl 2**: 86–96.
15. Blondel B, Morin I, Platt RW, Kramer MS, Usher R, Breart G. Algorithms for combining menstrual and ultrasound estimates of gestational age: consequences for rates of preterm and postterm birth. *BJOG* 2002; **109**: 718–720.
16. Eik-Nes SH, Grøttum P. *Graviditetskalenderen Snurra*. Scan-Med A/S: Drammen, Norway; 1983.
17. Johnsen SL, Rasmussen S, Sollien R, Kiserud T. Fetal age assessment based on ultrasound head biometry and the effect of maternal and fetal factors. *Acta Obstet Gynecol Scand* 2004; **83**: 716–723.
18. Johnsen SL, Rasmussen S, Sollien R, Kiserud T. Fetal age assessment based on femur length at 10–25 weeks of gestation, and reference ranges for femur length to head circumference ratios. *Acta Obstet Gynecol Scand* 2005; **84**: 725–733.
19. Altman DG, Chitty LS. New charts for ultrasound dating of pregnancy. *Ultrasound Obstet Gynecol* 1997; **10**: 174–191.
20. Hadlock FP, Harrist RB, Deter RL, Park SK. Fetal femur length as a predictor of menstrual age: sonographically measured. *AJR Am J Roentgenol* 1982; **138**: 875–878.
21. Økland I, Bjåstad TG, Johansen TF, Gjessing HK, Grøttum P, Eik-Nes SH. Narrowed beam width in newer ultrasound machines shortens measurements in the lateral direction: fetal measurement charts may be obsolete. *Ultrasound Obstet Gynecol* 2011; **38**: 82–87.
22. Gjessing HK, Grøttum P. Accuracy of second trimester fetal head circumference and biparietal diameter for predicting the time of spontaneous birth. *J Perinat Med* 2007; **35**: 350–351; author reply 351–352.
23. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria, 2010; http://www.R-project.org [Accessed 15 January 2011].
24. Backe B, Nakling J. Term prediction with ultrasound: evaluation of a new dating curve for biparietal diameter measurements. *Acta Obstet Gynecol Scand* 2006; **85**: 156–159.
25. Eik-Nes SH, Grøttum P, Gjessing HK. Regarding "Term prediction with ultrasound: evaluation of a new dating curve for biparietal diameter". *Acta Obstet Gynecol Scand* 2006; **85**: 1276–1278; author reply 1278–1279.
26. Tunon K, Eik-Nes SH, Grøttum P. A comparison between ultrasound and a reliable last menstrual period as predictors of the day of delivery in 15 000 examinations. *Ultrasound Obstet Gynecol* 1996; **8**: 178–185.
27. Verburg BO, Steegers EA, De Ridder M, Snijders RJ, Smith E, Hofman A, Moll HA, Jaddoe VW, Witteman JC. New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population-based cohort study. *Ultrasound Obstet Gynecol* 2008; **31**: 388–396.
28. Tunón K, Eik-Nes SH, Grøttum P. Fetal outcome when the ultrasound estimate of the day of delivery is more than 14 days later than the last menstrual period estimate. *Ultrasound Obstet Gynecol* 1999; **14**: 17–22.
29. Nguyen TH, Larsen T, Engholm G, Møller H. A discrepancy between gestational age estimated by last menstrual period and biparietal diameter may indicate an increased risk of fetal death and adverse pregnancy outcome. *BJOG* 2000; **107**: 1122–1129.

