Universal versus selective ultrasonography to screen for large for gestational age infants and associated morbidity.

Ulla Sovio, Ph.D.[1]

Alexandros A. Moraitis, M.Sc.[1]

Hilary S. Wong, Ph.D.[2]

Gordon C. S. Smith, D.Sc.[1]

[1]Department of Obstetrics and Gynaecology, University of Cambridge; NIHR Cambridge Comprehensive Biomedical Research Centre, Cambridge, UK.

[2]Department of Paediatrics, University of Cambridge, Cambridge, UK.

Corresponding author:

Professor Gordon C. S. Smith, Department of Obstetrics and Gynaecology, University of Cambridge, Box 223 The Rosie Hospital, Cambridge, CB2 0SW, UK.

Tel: +44 (0)1223 336871, Fax: +44 (0)1223 215327, E-mail: gcss2@cam.ac.uk

Short title: Universal USS screening for LGA infants

Keywords Large for gestational age, macrosomia, screening, ultrasound, neonatal morbidity

Word count (text): 3,116 words.

Abstract

Objective

To compare the diagnostic effectiveness of selective versus universal ultrasonography as a screening test for large for gestational age (LGA) infants, and to determine whether previously described ultrasonic markers of excessive fetal growth could identify which suspected LGA fetuses were at increased risk of neonatal morbidity.

Methods

We analysed data from a prospective cohort study of nulliparous women, the Pregnancy Outcome Prediction study. All women had clinically indicated scans as per routine care. Additionally, all women had blinded ultrasonic estimated fetal weight (EFW) performed at around 36 weeks of gestational age (wkGA). Screen positive for LGA was defined as an EFW >90th percentile ≥34wkGA.

Results

The current analysis included 3,866 eligible women. Of these, 177 (5%) infants had a birth weight >90th percentile. 1,354 (35%) women had a clinically indicated ultrasonography ≥34wkGA. The sensitivity of selective ultrasonography was 27% and the sensitivity of universal ultrasonography was 38%. The specificity of both approaches was high (99% and 97%, respectively). Using universal ultrasonography, neonatal outcome differed (P for interaction) by abdominal circumference growth velocity (ACGV) for both any neonatal morbidity (P=0.08) and severe adverse neonatal outcome (P=0.03). LGA fetuses with increased ACGV had a relative risk (95% CI, P) of any neonatal morbidity of 2.0 (1.1-3.6, P=0.04) and severe adverse neonatal outcome of 6.5 (2.0-21.1, P=0.01), whereas LGA fetuses with normal ACGV were not at increased risk.

Conclusion

Screening using universal ultrasonographic fetal biometry increases the detection of LGA infants and combined with ACGV identifies infants at increased risk of adverse neonatal outcome.

Introduction

A large for gestational age (LGA) infant is defined as one with birthweight higher than the 90th percentile for the given week of pregnancy. LGA infants are at higher risk of morbidity, including shoulder dystocia and brachial plexus injury,(1) as well as mortality including both antepartum stillbirth and delivery related perinatal death.(2) Ultrasonic fetal biometry can be used as a means to identify suspected LGA fetuses. The two obvious candidate interventions following this diagnosis are planned caesarean delivery, which may prevent the risk of birth injury, and early induction of labor, which may reduce birth weight by abbreviating the duration of pregnancy. A cost-benefit analysis indicated that caesarean delivery for non-diabetic women with suspected macrosomia is not justified.(3) Until recently, there has been no direct evidence for a beneficial effect of induction of labor.(4) However, an RCT published in 2015 suggested that early induction of labor (between 37+0 to 38+6 weeks' gestation) for ultrasonically suspected LGA reduced a composite of shoulder dystocia and perinatal morbidity by about 70% without increasing the risk of caesarean section.(5)

Currently, clinical guidelines in the UK(6) and the US(7) recommend that women should not be routinely screened using ultrasound in the last third of pregnancy, as there is no clear evidence of benefit from a meta-analysis of randomized controlled trials (RCTs),(8) and false positive ultrasonic diagnoses have the potential to cause harm through unnecessary intervention. However, the UK Guideline recommended further research on the diagnostic effectiveness of universal ultrasound. We undertook a prospective cohort study between 2008 and 2013, with a design to generate Level 1 evidence of the diagnostic effectiveness of universal serial ultrasound, i.e. where the results were blinded to the women and their carers. We have previously reported our results on screening for fetal growth restriction.(9) The aims of the present study were: 1. to compare the diagnostic effectiveness of selective

versus universal ultrasound as a screening test for LGA. 2. to determine which, if any, of a series of previously described ultrasonic markers of excessive fetal growth could identify LGA infants which were at increased risk of adverse neonatal outcome.

Methods

Study design

The Pregnancy Outcome Prediction study was a prospective cohort study conducted at the Rosie Hospital, Cambridge (UK) and has previously been described in detail.(9, 10) In brief, nulliparous women attending for their dating ultrasound scan between 14/01/2008 and 31/07/2012 with a viable singleton pregnancy were eligible. Women who agreed to participate signed a consent form and were given follow up appointments at approximately 20, 28 and 36 weeks gestational age (wkGA) in the NIHR Cambridge Clinical Research Facility. Women were selected for clinically indicated ultrasound scans in the third trimester as per routine clinical care using local and national guidelines, and the results of these scans were reported (selective ultrasonography). In contrast, women and clinicians were blinded to the results of the research ultrasound scans (universal ultrasonography). The study was designed to generate level 1 evidence of diagnostic effectiveness, as defined by the UK's National Institute for Health and Care Excellence (NICE).(11) The reporting of this study conforms to the STARD (Standards for Reporting Diagnostic accuracy studies) guidelines.(12) Ethical approval for the study was given by the Cambridgeshire 2 Research Ethics Committee (reference number 07/H0308/163). The inclusion criteria for the present analysis were that women attended their 36 week research scan and had a live birth at the Rosie Hospital. Women who delivered prior to their 36 week scan appointment were excluded.

Selective and universal ultrasonography

The results of clinically indicated scans was ascertained by linkage of the research data to the hospital's electronic ultrasonography database (Astraia, Munich, Germany). In both selective (clinically indicated) and universal (research) ultrasonography, fetal biometry included measurement of fetal biparietal diameter, head circumference (HC), abdominal

circumference (AC) and femur length (FL) using standard techniques. An estimated fetal weight (EFW) percentile was calculated using the Hadlock equations and reference standard.(13, 14) Where all four measurements were available, the formula employing all measurements was used: $EFW = 10^{(1.3596 - 0.00386*AC*FL + 0.0064*HC + 0.00061*BPD*AC + 0.0424*AC + 0.174*FL)}$. Where the head measurements were missing, the formula based on AC and FL was used: $EFW = 10^{(1.304 + 0.05281*AC + 0.1938*FL - 0.004*AC*FL)}$. Following delivery, the results of the research scans were un-blinded and their associations with outcome were assessed.

Screening status in relation to EFW was classified on the basis of the last scan prior to birth (for universal ultrasonography this was the 36 week scan). Screen positive was defined as an EFW>90th percentile using an externally derived reference range(13, 14) (both selective and universal). Screen negative was defined as an EFW≤90th percentile (both selective and universal), or when no clinically indicated scan had been performed ≥34 weeks gestational age (selective only). Customised percentiles of EFW were also calculated using published methods,(15) but employing co-efficients from the most recent model (GROW v6.7.3_13 [UK], Gestation Network [www.gestation.net]). The associations between population-based and customised EFW>90th percentile and neonatal morbidity were compared.

Analysis of ultrasonic indicators of overgrowth was performed by comparing the association between an EFW>90th percentile and neonatal morbidity in the presence or absence of the given factor. HC:AC and AC:FL ratios were classified by the last measurement performed prior to birth. All measurements were quantified as gestational age adjusted z scores, to take into account variation in the exact timing of ultrasound scans (Supplementary Tables 1 & 2 in Sovio et al(9)). AC growth velocity (ACGV) was quantified as the difference in AC z score comparing the 36 week scan and the 20 week scan. For these three indices, deciles were generated using the distribution within the study cohort. The lowest decile of HC:AC and the highest deciles of AC:FL and AC growth velocity were defined as abnormal. No other growth

indices were studied to reduce the possibility of chance findings due to repeated hypothesis tests.

Outcome data

The outcome of the pregnancy was ascertained by individual review of all paper case records by research midwives, and by linkage of the research data to the hospital's electronic databases of delivery (Protos, iSoft, Banbury, UK), biochemical tests (Meditech, Westwood MA, USA) and neonatal intensive care (Badgernet, Clevermed Ltd, Edinburgh, UK). The gold standard for LGA was birth weight >90th percentile for sex and gestational age, calculated using a UK reference.(16) Macrosomia was defined as birth weight >4000g and severe macrosomia was defined as birth weight >4500g. Neonatal morbidity was defined as ≥1 of the following: a 5 minute Apgar score less than 7, delivery with metabolic acidosis (defined as a cord blood pH <7.1 and a base deficit of >10mmol/L) or admission to the neonatal unit at term (defined as admission <48 hours after birth at ≥37 weeks gestational age and discharge ≥48 hours after admission). Severe adverse neonatal outcome was defined as term live birth associated with neonatal death, hypoxic ischemic encephalopathy, use of inotropes, mechanical ventilation, or severe metabolic acidosis (defined as a cord blood pH <7.0 and a base deficit of >12mmol/L). Shoulder dystocia and neonatal hypoglycaemia were documented in the electronic delivery record. Additional cases of diagnosed hypoglycaemia were obtained from the neonatal intensive care database, which was also used to identify cases of neonatal jaundice. These conditions were sub-classified on the basis of whether they were associated with neonatal morbidity or severe adverse neonatal outcome, as defined above.

Statistics

Continuous variables were compared using a two-sample Wilcoxon rank-sum test and categorical variables were compared using the Pearson Chi-square test (with a trend test where appropriate) or Fisher's exact test where numbers were small. Sensitivity and

specificity were compared using McNemar's test, positive and negative predictive values were compared using weighted generalized score tests,(17) and likelihood ratios were compared using regression model based tests.(18) Analyses were repeated adjusting for pre-existing diabetes and gestational diabetes using exact logistic regression. Interactions between EFW and ultrasonic markers of overgrowth in their associations with neonatal morbidity were tested using the Mantel-Haenszel test or exact logistic regression, as appropriate. Conditional probabilities test was used to calculate p-values from the exact logistic regression(19) since the exact probabilities are analogous to the exact p-values obtained from a Fisher's exact test.(20) Statistical significance was assumed at $P<0.05$ (two sided). Analyses were performed using Stata version 14.1. and R version 3.0.2.

Results

In total, 4,512 (56%) women were recruited to the study and provided written informed consent.(9) We excluded women who withdrew from the study or defaulted from their 36 week research scan (n=326), delivered before the 36 week scan (n=176) or had missing biometric measurements (n=12). We excluded further 127 women who were lost to follow-up or did not deliver in the Rosie Hospital and 5 women who had a stillbirth after their 36 week scan (Supplementary Figure 1). The study group for the present analysis consisted of 3,866 women (86% of all recruited). A total of 1,354 of these women (35%) had a clinically indicated scan including biometry ≥34 wkGA (Table 1). Women having clinically indicated scans were more likely to be at extremes of maternal age, to have discontinued education earlier in life, to have a body mass index >30, to have had previous miscarriages, and to have pre-existing diabetes or to develop gestational diabetes than the women who did not have clinically indicated scans. The average birth weight of their infants was lower, and they had a greater proportion of LGA infants, births < 40 wkGA, induced labors and pre-labor cesarean deliveries.

A total of 177 (4.6%) infants had a birth weight >90th percentile. The last clinically indicated scan (selective) before birth recorded an EFW>90th percentile in 47 of these cases yielding a sensitivity of 27%. The research 36 week ultrasound scan (universal) recorded an EFW of >90th percentile in 67 of these cases yielding a sensitivity of 38% (67/177). The specificity was high for both approaches, but was slightly higher for selective compared with universal ultrasonography (99% versus 97%, respectively). Screening summary statistics for universal and selective ultrasonography are presented (Table 2). The area under the receiver operating characteristic curve for LGA detected by selective ultrasonography was 0.72 and for universal ultrasonography was 0.87 (Figure 1).

There was no evidence for association between an EFW >90th percentile from universal ultrasound and the risk of neonatal morbidity using either population based or customised reference percentiles (Table 3, raw data (n/N) are shown in Supplementary Table 1). The association between an EFW >90th percentile and the risk of neonatal morbidity was then assessed in relation to three previously described indices of overgrowth (Figure 2). The only measurement where there was evidence for an interaction was with increased (i.e. top decile) of AC growth velocity. An interaction was observed for both any morbidity (P=0.08) and severe adverse neonatal outcome (P=0.03). There was no clear indication of an increased overall risk of adverse neonatal outcome where the EFW was >90th and the ACGV was not in the top decile, unless the baby was LGA at birth (Table 3). However, in the cases where universal ultrasonography demonstrated an LGA fetus with increased ACGV, there was a doubling in the risk of any neonatal morbidity (relative risk 2.0, 95% CI 1.1 to 3.6, P=0.04) and greater than 6-fold risk of severe adverse neonatal outcome (relative risk 6.5, 95% CI 2.0 to 21.1, P=0.01). When the outcome was confined to cases of neonatal morbidity where the baby was actually confirmed to be LGA, ultrasonic LGA was associated with a 10-fold risk and the combination of LGA and top decile of ACGV was associated with a greater than 20-fold risk. The associations remained very similar after adjustments for pre-existing diabetes and gestational diabetes (Table 3).

All analyses of ACGV were repeated using AC growth charts generated by the Fetal Growth Longitudinal Study component of the INTERGROWTH-21st Project,(21) an international consortium which constructed fetal growth standards using methods recommended by the WHO and the associations were virtually unchanged (Supplementary Table 2). None of the indices of overgrowth were associated with adverse outcome when the EFW was ≤90th percentile (Supplementary Table 3). In addition, screening summary statistics for universal ultrasonography for detecting macrosomia and severe macrosomia are presented (Supplementary Table 4). The area under the receiver operating characteristic curve for macrosomia was 0.83 (95%CI 0.81-0.85) and for severe macrosomia was 0.87 (95%CI 0.82-

225 0.91). Among infants who had EFW >90th percentile in the universal ultrasound, 41% were
226 delivered through intrapartum emergency caesarean section, whereas the proportion was
227 17% when the EFW was ≤90th percentile (risk ratio 2.50 [95%CI 2.08 to 3.00]). Finally, there
228 were no significant associations between ultrasonic suspicion of LGA, with or without
229 increased ACGV, and the risk of shoulder dystocia (Table 4). The risk of neonatal
230 hypoglycaemia was elevated when there was a combination of ultrasonic suspicion of LGA
231 and increased ACGV (Supplementary Table 5) but the risk of jaundice was not elevated in
232 any of the groups (Supplementary Table 6).

Discussion

The main findings of the current analysis were (i) that universal ultrasonography increased the detection of LGA infants from 27% to 38%, and (ii) that the only ultrasonic marker of fetal overgrowth that discriminated between LGA infants at increased risk of neonatal complications was the ACGV. LGA fetuses with a normal ACGV were not at increased risk of adverse outcome. However, LGA fetuses with accelerated ACGV were at increased risk of adverse neonatal outcome, including severe outcome.

The present study has immediate implications for obstetric care. Many women have late pregnancy ultrasound with indications including prior risk factors and acquired pregnancy complications. LGA will be diagnosed in a proportion of these women. The current study indicates that, where this diagnosis is made, assessment of the ACGV helps assess the risk of associated complications. Diagnosis of LGA with normal ACGV did not appear to be associated with an increased risk of adverse neonatal outcome, whereas diagnosis of LGA with accelerated ACGV was significantly associated with an increased risk of any neonatal morbidity. This diagnostic combination was also significantly associated with severe adverse neonatal outcome but as the latter occurred in only 0.6% of all infants, the association has relatively wide confidence intervals (Figure 2b). Importantly, we used the AC measurement at the routine 20 week anomaly scan as the baseline measurement. This means that an assessment of growth velocity can be made even when a woman has had only a single scan in late pregnancy.

The aim of this study was to assess the diagnostic effectiveness of late pregnancy scanning, hence we have excluded women that delivered before reaching late pregnancy. An interesting finding, which we also noted in our previous study on universal screening for small-for-gestational age infants,(9) was that the positive likelihood ratio was significantly

higher in the selective screening group. We believe that the result from selective screening probably reflects both the indication for doing the scan and the scan result itself, whereas the likelihood ratio from the universal screening reflects simply the scan result.

Interestingly, the association between ultrasonic diagnosis of fetal overgrowth and neonatal morbidity was not mediated through associations with shoulder dystocia. No combination of LGA or ACGV was significantly associated with shoulder dystocia, considering either any documentation of the condition at the time of delivery, or shoulder dystocia associated with neonatal morbidity. Therefore, the association between fetal overgrowth and adverse neonatal outcome was mediated by other causes. This is consistent with the view that macrosomia associated with pathological fetal overgrowth has multiple adverse effects on the fetus, in addition to predisposing to birth injuries.(22) Serious shoulder dystocia only affected 1.6 per 1,000 pregnancies in the current study and this analysis was underpowered to address this outcome. A recent randomised controlled trial demonstrated improved outcome following induction of labor at 37-38 wkGA for suspected macrosomia. That study employed women who had ultrasound scans for clinically suspected macrosomia and used an EFW threshold of >95th percentile, and these features may explain the high rates of shoulder dystocia and severe morbidity. However, the current study demonstrates that these findings should be applied cautiously to women who are suspected to have a LGA fetus in the absence of clinically suspected macrosomia, and that macrosomia may be associated with adverse neonatal outcome through mechanisms other than shoulder dystocia. Our findings are also consistent with a preliminary report from another prospective cohort study using blinded ultrasonic EFW in nulliparous women, which showed no association with shoulder dystocia.(23) Finally, this study was underpowered to address the association between LGA and specific neonatal outcomes such as metabolic acidosis or low Apgar score as these were present in <1% of the cohort.

We found no evidence to suggest that use of a customised EFW resulted in a stronger association between LGA and adverse neonatal outcome although the present study was underpowered to address serious shoulder dystocia as an outcome. The aim of ultrasonic assessment of growth is to differentiate pathological LGA from healthy LGA. Customisation of EFW attempts to achieve this by correcting the estimated fetal size for the maternal characteristics. Assessment of the ACGV uses the fetal AC in earlier pregnancy as the reference for later measurements, rather than using a reference modified for maternal characteristics. We used the highest decile to describe abnormal growth velocity since it is easy to use and interpret. The disadvantages are that it is specific to our cohort in its use of data-driven cut-off points, and, similarly to any categorisation of a continuous trait, assumes an unrealistic step-function of risk and within-group homogeneity.(24) In both the current analysis and our previous analysis of SGA and fetal growth restriction, we found that the ACGV was better than customisation in identifying fetuses in the extremes of the distribution of EFW which were at increased risk of neonatal complications. However, we did find that the estimated association between customised EFW and shoulder dystocia was stronger, although statistically non-significant. In that case, the outcome is determined by the interaction between the size of the fetus and the size of the mother, and it is plausible that customisation might perform better in that situation. We also used the INTERGROWTH-21st Project reference centiles which performed similarly to population and customised centiles.

In conclusion, the present study found that universal ultrasonographic fetal biometry increases the detection of LGA infants and combined with ACGV stratifies infants to those who are at increased risk of adverse neonatal outcome. The immediate clinical implication of the study is that once the fetus is diagnosed LGA, assessment of the ACGV gives further information on the risk of associated complications which helps in planning obstetric care in late pregnancy. The current study is also the first to provide level 1 evidence of the diagnostic effectiveness of universal ultrasound to detect LGA. However, a randomized

controlled trial would be required prior to clinical implementation of screening and intervention.

Acknowledgement: We thank the Perinatal Institute for providing a bulk calculator for customised percentiles of estimated fetal weight. We thank all the women who participated in the study, and all the staff in the Rosie Hospital (Cambridge, UK) and NIHR Cambridge Clinical Research Facility who provided direct or indirect assistance for the study.

Disclosure of interest: GCSS reports grants from the National Institute for Health Research (NIHR), Stillbirth and Neonatal Death Society, and MRC, and reports other from GE and NIHR Cambridge Clinical Research Facility, during the study; reports personal fees and other from GlaxoSmithKline and Roche, and reports other from Chiesi, outside of the study. US and AAM declare no competing interests.

Contributors: GCSS created the study concept and design. US, AAM, HW and GCSS did the data analysis and interpretation. US, AAM, and GCSS drafted the manuscript. All authors contributed to the critical revision of the manuscript for important intellectual content and approved the final version to be published. US and GCSS had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Details of ethics approval: Ethical approval for the study was given by the Cambridgeshire 2 Research Ethics Committee (reference number 07/H0308/163).

Funding: The work was supported by the National Institute for Health Research (NIHR) Cambridge Comprehensive Biomedical Research Centre (Women's Health theme), and a project grant from the Stillbirth and neonatal death society (Sands). The study was also supported GE Healthcare (donation of two Voluson i ultrasound systems for this study), and by the NIHR Cambridge Clinical Research Facility, where all research visits took place.

Legends for figures.

Figure 1. Receiver operating characteristic (ROC) curves for screening for an LGA infant using ultrasonic estimated fetal weight (EFW), comparing selective ultrasonography (dashed line), and universal ultrasonography (solid line). When the results of selective ultrasonography were analysed, 65% (2512/3866) of women did not have a clinically indicated scan at or after 34 weeks gestation. In this group, EFW centile was imputed using a sex-specific population median (46.30 in males, 38.93 in females). Areas under the ROC curves (95% confidence interval) are 0.72 (0.68-0.76) for selective scan and 0.87 (0.85-0.90) for universal scan. P<0.0001 for the comparison of the two approaches.

Figure 2. Stratified analyses of perinatal outcome associated with diagnosis of large for gestational age (LGA) using universal ultrasonography in relation to ultrasonographic indicators of fetal overgrowth. **A.** Any neonatal morbidity. **B.** Severe adverse neonatal outcome. The three previously described indices of fetal overgrowth were classified as the extreme decile associated with fetal overgrowth (highest or lowest, as appropriate) compared with the other nine deciles in the cohort. Z score cut-off point is 1.4285 for the highest decile of ACGV, 1.2789 for the highest decile of AC:FL ratio and -1.2484 for the lowest decile of HC:AC ratio. Points are relative risks of any neonatal morbidity (A) or odds ratios (B) associated with an ultrasonic diagnosis of a large for gestational age (LGA) infant at the 36 week scan. P values are from Mantel-Haenszel test of interaction (A) or from conditional probabilities test for interaction which is analogous to the Fisher's exact test (B). The upper confidence limit of the odds ratio is infinity for the highest decile of ACGV and the lowest decile of HC:AC ratio. The odds ratio axis has been truncated to 1000. AC, abdominal circumference; GV, growth velocity; FL, femur length; HC, head circumference.

References

1. Esakoff TF, Cheng YW, Sparks TN, Caughey AB. The association between birthweight 4000 g or greater and perinatal outcomes in patients with and without gestational diabetes mellitus. American journal of obstetrics and gynecology. 2009;200(6):672.e1-4.

2. Moraitis AA, Wood AM, Fleming M, Smith GC. Birth weight percentile and the risk of term perinatal death. Obstetrics and gynecology. 2014;124(2 Pt 1):274-83.

3. Rouse DJ, Owen J, Goldenberg RL, Cliver SP. The effectiveness and costs of elective cesarean delivery for fetal macrosomia diagnosed by ultrasound. Jama. 1996;276(18):1480-6.

4. Irion O, Boulvain M. Induction of labour for suspected fetal macrosomia. The Cochrane database of systematic reviews. 2000(2):Cd000938.

5. Boulvain M, Senat MV, Perrotin F, Winer N, Beucher G, Subtil D, Bretelle F, Azria E, Hejaiej D, Vendittelli F, Capelle M, Langer B, Matis R, Connan L, Gillard P, Kirkpatrick C, Ceysens G, Faron G, Irion O, Rozenberg P. Induction of labour versus expectant management for large-for-date fetuses: a randomised controlled trial. Lancet (London, England). 2015;385(9987):2600-5.

6. NICE. Clinical guideline 62: antenatal care. London: National Institute of Health and Care Excellence; 2008.

7. ACOG Practice Bulletin No. 101: Ultrasonography in pregnancy. Obstetrics and gynecology. 2009;113(2 Pt 1):451-61.

8. Bricker L, Neilson JP, Dowswell T. Routine ultrasound in late pregnancy (after 24 weeks' gestation). The Cochrane database of systematic reviews. 2008(4):Cd001451.

9. Sovio U, White IR, Dacey A, Pasupathy D, Smith GC. Screening for fetal growth restriction with universal third trimester ultrasonography in nulliparous women in the Pregnancy Outcome Prediction (POP) study: a prospective cohort study. Lancet (London, England). 2015;386(10008):2089-97.

386 10. Pasupathy D, Dacey A, Cook E, Charnock-Jones DS, White IR, Smith GC. Study protocol. A
387 prospective cohort study of unselected primiparous women: the pregnancy outcome prediction
388 study. BMC pregnancy and childbirth. 2008;8:51.
389 11. NICE. Guideline development methods: Information for National Collaborating Centres and
390 Guideline Developers. London: National Institute of health and Care Excellence.; 2004.
391 12. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D,
392 Rennie D, de Vet HC, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA, Cohen JF.
393 STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ
394 (Clinical research ed). 2015;351:h5527.
395 13. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic
396 weight standard. Radiology. 1991;181(1):129-33.
397 14. Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the
398 use of head, body, and femur measurements--a prospective study. American journal of obstetrics
399 and gynecology. 1985;151(3):333-7.
400 15. Gardosi J, Mongelli M, Wilcox M, Chang A. An adjustable fetal weight standard. Ultrasound
401 in obstetrics & gynecology : the official journal of the International Society of Ultrasound in
402 Obstetrics and Gynecology. 1995;6(3):168-74.
403 16. Freeman JV, Cole TJ, Chinn S, Jones PR, White EM, Preece MA. Cross sectional stature and
404 weight reference curves for the UK, 1990. Archives of disease in childhood. 1995;73(1):17-24.
405 17. Kosinski AS. A weighted generalized score statistic for comparison of predictive values of
406 diagnostic tests. Statistics in medicine. 2013;32(6):964-77.
407 18. Gu W, Pepe MS. Estimating the capacity for improvement in risk prediction with a marker.
408 Biostatistics (Oxford, England). 2009;10(1):172-86.
409 19. Mehta CR, Patel NR. Exact logistic regression: theory and examples. Statistics in medicine.
410 1995;14(19):2143-60.

411 20. Stokes ME DC, Koch GG Categorical Data Analysis Using SAS. Third edition ed: SAS Institute;
412 2012.
413 21. Papageorghiou AT, Ohuma EO, Altman DG, Todros T, Cheikh Ismail L, Lambert A, Jaffer YA,
414 Bertino E, Gravett MG, Purwar M, Noble JA, Pang R, Victora CG, Barros FC, Carvalho M, Salomon LJ,
415 Bhutta ZA, Kennedy SH, Villar J. International standards for fetal growth based on serial ultrasound
416 measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. Lancet
417 (London, England). 2014;384(9946):869-79.
418 22. Boney CM, Verma A, Tucker R, Vohr BR. Metabolic syndrome in childhood: association with
419 birth weight, maternal obesity, and gestational diabetes mellitus. Pediatrics. 2005;115(3):e290-6.
420 23. Galvin DM, Burke N, Burke G, Breathnach F, McAuliffe F, Morrison J, Turner M, Dornan S,
421 Higgins J, Cotter A, Geary M, Cody F, Mulcahy C, Daly S, Dicker P, Tully E, Malone F. 94: Accuracy of
422 prenatal detection of macrosomia >4,000g and outcomes in the absence of intervention: results of
423 the prospective multicenter genesis study. American journal of obstetrics and gynecology.
424 2017;216(1, Supplement):S68.
425 24. Bennette C, Vickers A. Against quantiles: categorization of continuous variables in
426 epidemiologic research, and its discontents. BMC medical research methodology. 2012;12:21.

427

Table 1. Characteristics of the study cohort (N=3,866).

Characteristic	No clinically indicated scan ≥34 weeks (N=2512)	≥1 clinically indicated scan ≥34 weeks (N=1354)	P Value	Overall baseline characteristics (N=3866)
Maternal characteristics				
Age, years				
<20	71 (3%)	65 (5%)		136 (4%)
20 to 24.9	350 (14%)	161 (12%)		511 (13%)
25 to 29.9	821 (33%)	371 (27%)	<0.0001	1192 (31%)
30 to 34.9	947 (38%)	488 (36%)		1435 (37%)
35 to 39.9	299 (12%)	222 (16%)		521 (13%)
≥40	24 (1%)	47 (3%)		71 (2%)
Age stopped FTE, years				
<19	800 (32%)	480 (35%)		1280 (33%)
19 to 22	889 (35%)	454 (34%)	0.03	1343 (35%)
≥23	756 (30%)	377 (28%)		1133 (29%)
Missing	67 (3%)	43 (3%)		110 (3%)
Deprivation quartile				
1 (lowest)	611 (24%)	332 (24%)		943 (24%)
2	593 (24%)	324 (24%)	0.92	917 (24%)
3	602 (24%)	329 (24%)		931 (24%)
4 (highest)	592 (24%)	325 (24%)		917 (24%)
Missing	114 (5%)	44 (3%)		158 (4%)
Postcode area				
Central Cambridge city	775 (31%)	413 (30%)		1188 (31%)
Peripheral Cambridge city	558 (22%)	322 (24%)	0.08	880 (23%)
Cambridgeshire, outside city	605 (24%)	363 (27%)		968 (25%)
Outside Cambridgeshire	502 (20%)	234 (17%)		736 (19%)
Missing	72 (3%)	22 (2%)		94 (2%)
White ethnicity	2336 (93%)	1261 (93%)	0.76	3597 (93%)
Missing	45 (2%)	19 (1%)		64 (2%)
Married	1713 (68%)	933 (69%)	0.65	2646 (68%)
Smoker	115 (5%)	66 (5%)	0.67	181 (5%)
Any alcohol consumption	123 (5%)	57 (4%)	0.33	180 (5%)
Missing	1 (<1%)	0 (0%)		1 (<1%)
BMI, kg/m^2				
<25	1535 (61%)	737 (54%)		2272 (59%)
25 to 29.9	713 (28%)	361 (27%)		1073 (28%)
30 to 34.9	238 (9%)	133 (10%)	<0.0001	371 (10%)
35 to 39.9	25 (1%)	79 (6%)		104 (3%)
≥40	2 (<1%)	43 (3%)		45 (1%)
Missing	0 (0%)	1 (<1%)		1 (<1%)
≥1 previous miscarriage	223 (9%)	166 (12%)	0.001	389 (10%)

Diabetes				
Type 1 or type 2 DM	0 (0%)	12 (1%)		12 (<1%)
Gestational DM	2 (<1%)	153 (11%)	<0.0001	155 (4%)
Birth outcomes				
Birth weight, g	3485 (3190 to 3780)	3350 (3040 to 3680)	<0.0001	3440 (3130 to 3750)
LGA (>90th)	93 (4%)	84 (6%)	0.0004	177 (5%)
Severe LGA (>97th)	12 (<1%)	25 (2%)	<0.0001	37 (1%)
Macrosomia (>4000g)	303 (12%)	125 (9%)	0.007	428 (11%)
Severe macrosomia (>4500g)	36 (1%)	22 (2%)	0.64	58 (2%)
Gestational age, weeks	40.6 (39.7 to 41.3)	39.9 (38.9 to 40.9)	<0.0001	40.4 (39.3 to 41.1)
<37	18 (1%)	27 (2%)		45 (1%)
37	77 (3%)	104 (8%)		181 (5%)
38	225 (9%)	217 (16%)		442 (11%)
39	460 (18%)	355 (26%)	<0.0001	815 (21%)
40	806 (32%)	322 (24%)		1128 (29%)
41	766 (30%)	271 (20%)		1037 (27%)
≥ 42	160 (6%)	58 (4%)		218 (6%)
Induction of labor	715 (29%)	532 (39%)	<0.0001	1247 (32%)
Mode of delivery				
Spontaneous vaginal	1325 (53%)	557 (41%)		1882 (49%)
Assisted vaginal	649 (26%)	284 (21%)	<0.0001	933 (24%)
Intrapartum cesarean	458 (18%)	230 (17%)		688 (18%)
Pre-labor cesarean	76 (3%)	277 (20%)		353 (9%)
Missing	4 (<1%)	6 (<1%)		10 (<1%)

Data are expressed as median (IQR) or n (%) as appropriate. P-values are for difference between groups calculated using the two-sample Wilcoxon rank-sum (Mann-Whitney) test for continuous variables and the Pearson Chi-square test for categorical variables, with trend test as appropriate. The missing category was not included in statistical tests. For fields where there is no category labelled "missing", data were 100% complete.

Maternal age was defined as age at recruitment. All other maternal characteristics were defined by self-report at the 20 weeks questionnaire, from examination of the clinical case record, or linkage to the hospital's electronic databases. Socio-economic status was quantified using the Index of Multiple Deprivation (IMD) 2007, which is based on census data from the area of the mother's postcode.[Noble 2008, The English Indices of Deprivation 2007] The median (IQR) gestational age for the clinically indicated scan was 36.4 (36.0 to 37.9) weeks.

Abbreviations: FTE denotes full time education, BMI denotes body mass index, DM denotes diabetes mellitus, LGA denotes large for gestational age.

Table 2. Comparison of diagnostic effectiveness of selective versus universal ultrasonography for detection of LGA infants.

	Selective	**Universal**	**P***
True positive/ False positive	47/ 49	67/ 127	N/A
False negative/ True negative	130/ 3640	110/ 3562	N/A
Sensitivity (%)	27 (20-33)	38 (31-45)	0.005
Specificity (%)	99 (98-99)	97 (96-97)	<0.0001
Positive likelihood ratio	20 (14-29)	11 (9-14)	0.002
Negative likelihood ratio	0.74 (0.68-0.81)	0.64 (0.57-0.72)	0.01
Positive predictive value (%)	49 (39-60)	35 (28-41)	0.002
Negative predictive value (%)	97 (96-97)	97 (96-98)	0.01

*Statistical comparison by DeLong, McNemar, or weighted generalised score tests, as appropriate. LGA denotes large for gestational age. LGA is defined as birth weight >90th percentile. Estimated fetal weight (EFW) measurement was taken from the last scan prior to birth. "Selective" reports the results of clinically indicated scans. If a woman did not have a clinically indicated scan at ≥34 weeks, she was defined as screen negative by selective ultrasonography. "Universal" reports the results of the 36 week research scan. All values were calculated with EFW >90th percentile as screen positive. 95% confidence intervals are given in brackets.

Table 3. The relationship between estimated fetal weight (EFW) >90th percentile, abdominal circumference growth velocity (ACGV) and perinatal outcome using universal ultrasonography, total n=3,866.

	Perinatal outcome											
	Any neonatal morbidity (n=267)		*Metabolic acidosis (n=37)*		*5 Minute Apgar <7 (n=31)*		*Neonatal unit admission (n=229)*		*Severe adverse neonatal outcome (n=26)*		*LGA at birth + any neonatal morbidity (n=11)*	
Research scan result	*RR (95% CI)*	*P*	*RR (95% CI)*	*P*	*RR (95% CI)*	*P*	*RR (95% CI)*	*P*	*RR (95% CI)*	*P*	*RR (95% CI)*	*P*
UNIVARIABLE ANALYSIS												
EFW>90th												
Population	1.2 (0.7-2.0)	0.47	1.1 (0.3-4.5)	0.71	2.0 (0.6-6.6)	0.20	1.3 (0.8-2.2)	0.27	2.5 (0.7-8.2)	0.14	10.8 (3.2-36.6)	0.002
Customised	1.3 (0.9-1.8)	0.19	1.2 (0.4-3.3)	0.77	2.3 (1.0-5.6)	0.06	1.3 (0.9-1.9)	0.20	1.8 (0.6-5.1)	0.30	5.5 (1.6-18.8)	0.01
EFW>90th + Normal ACGV	0.7 (0.3-1.6)	0.58	0.9 (0.1-6.3)	>0.99	0.0* --	>0.99	0.7 (0.3-1.7)	0.55	0.0* --	>0.99	4.4 (0.5-35.3)	0.23
EFW>90th + Highest decile ACGV	2.0 (1.1-3.6)	0.04	1.4 (0.2-10.2)	0.51	5.3 (1.7-17.1)	0.02	2.3 (1.3-4.2)	0.01	6.5 (2.0-21.1)	0.01	21.3 (5.6-80.6)	0.0008
MULTIVARIABLE ANALYSIS												
	Adj OR (95% CI)	*P*	*Adj OR (95% CI)*	*P*	*Adj OR (95% CI)*	*P*	*Adj OR (95% CI)*	*P*	*Adj OR (95% CI)*	*P*	*Adj OR (95% CI)*	*P*
EFW>90th + Highest decile ACGV adjusted for DM&GDM†	2.1 (0.9-4.1)	0.04	1.4 (0.0-8.3)	0.53	5.4 (1.0-18.1)	0.02	2.4 (1.1-4.9)	0.02	6.5 (1.2-22.2)	0.02	21.0 (3.4-95.4)	0.001

*Number of exposed cases = 0, therefore 95% confidence interval (CI) for relative risk (RR) is not defined.
†Adjusted for pre-existing diabetes mellitus (DM) and gestational diabetes mellitus (GDM): odds ratio (OR) and 95% CI from exact logistic regression are given instead of RR.

All estimated fetal weights (EFWs) are based on population-based percentiles, unless stated otherwise. All RRs and ORs are referent to infants with an EFW of ≤90th percentile by population-based standards, except for the RRs for customised EFW >90th percentile, which are referent to infants with an EFW of the ≤90th percentile by customised standards. Large for gestational age (LGA) is defined as birthweight of >90th percentile by population standards. Abdominal circumference growth velocity (ACGV) is based on the change in the gestational age adjusted *Z* score, comparing the result at the 20 week scan with the 36 week scan. Z score cut-off point of the highest decile of ACGV is 1.4285. Any neonatal morbidity is a composite outcome—ie, one or more of these three outcomes: metabolic acidosis (defined as pH <7.1 and base deficit >10 mmol/L), 5 min Apgar score less than 7, and neonatal unit admission (defined as admission to the neonatal intensive care unit, the high dependency unit, or the special care baby unit). Severe adverse neonatal outcome is a composite outcome—ie, one or more of the following outcomes specified: neonatal death at term (not due to congenital anomaly), hypoxic ischemic encephalopathy at term, use of inotropes at term, mechanical ventilation at term, severe metabolic acidosis at term (defined as pH <7.0 and base deficit >12 mmol/L). Customized percentiles of EFW were calculated with the Gestation-Related Optimal Weight Customised Weight Centile Calculator (version 6.7 [UK]). P values for RRs are from Fisher's exact test and p-values for ORs are from conditional probabilities test. All p-values are two-sided.

Table 4. The relationship between estimated fetal weight (EFW), abdominal circumference growth velocity (ACGV) and shoulder dystocia from universal ultrasonography, total n=3,866.

	Outcome							
	Shoulder dystocia (n=62)				*Shoulder dystocia + any neonatal morbidity (n=6)*			
Research scan result	*TP/FP*	*TN/FN*	*LR+ (95% CI)*	*P*	*TP/FP*	*TN/FN*	*LR+ (95% CI)*	*P*
EFW>90th Population	2/192	3612/60	0.6 (0.2-2.5)	0.77	0/194	3666/6	0.0* --	>0.99
EFW>90th Customised	10/352	3452/52	1.7 (1.0-3.1)	0.08	1/361	3499/5	1.8 (0.3-10.7)	0.45
EFW>90th + Highest decile ACGV	0/74	3715/62	0.0* --	0.63	0/74	3771/6	0.0* --	>0.99
Highest decile ACGV	6/380	3409/56	1.0 (0.4-2.1)	>0.99	1/385	3460/5	1.7 (0.3-10.0)	0.47
EFW>80th + ACGV>1SD	4/217	3572/58	1.1 (0.4-2.9)	0.78	1/220	3625/5	2.9 (0.5-17.5)	0.30

*Number of exposed cases = 0, therefore 95% confidence interval (CI) for positive likelihood ratio (LR+) is not defined. All estimated fetal weights (EFWs) are based on population-based percentiles, unless stated otherwise. LR+s are referent to all other infants. Abdominal circumference growth velocity (ACGV) is based on the change in the gestational age adjusted *Z* score, comparing the result at the 20 week scan with the 36 week scan. Z score cut-off point of the highest decile of ACGV is 1.4285. Any neonatal morbidity is a composite outcome—ie, one or more of these three outcomes: metabolic acidosis (defined as pH <7.1 and base deficit >10 mmol/L), 5 min Apgar score less than 7, and neonatal unit admission (defined as admission to the neonatal intensive care unit, the high dependency unit, or the special care baby unit). Customised percentiles of EFW were calculated with the Gestation-Related Optimal Weight Customised Weight Centile Calculator (version 6.7 [UK]). P values are from Fisher's exact test. All p-values are two-sided.
TP, true positive; FP, false positive; TN, true negative; FN, false negative.