Blinded ultrasonic fetal biometry at 36 weeks and the risk of emergency caesarean delivery: a prospective cohort study of 3,047 low risk nulliparous women

Short title: EFW and caesarean risk

Dr Ulla Sovio, PhD,[1]

Prof Gordon C.S. Smith, MD, PhD, DSc.[1]

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

Correspondence to:

Prof GCS Smith, Department of Obstetrics and Gynaecology, University of Cambridge, Box 223 The Rosie Hospital, Cambridge, CB2 2SW, UK.

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

ABSTRACT

Objectives

We studied the risk of emergency caesarean delivery (CD) using blinded ultrasonographic estimated fetal weight (EFW) at 36 weeks of gestational age (wkGA): (1) to compare the association for customised and non-customised EFW, (2) to determine whether adding ultrasonic EFW improved prediction based on maternal characteristics alone, and (3) to determine whether women at high predicted risk of emergency CD had higher risks of maternal and perinatal morbidity than other women.

Methods

We studied 3,047 low risk women (no pre-existing medical conditions or acquired complications of pregnancy) from the Pregnancy Outcome Prediction study (Cambridge UK) who had ultrasonic EFW at ~36 weeks gestational age, where women and clinicians were blinded to the result.

Results

Blinded EFW was strongly associated with the risk of emergency CD (coefficient for a 1 standard deviation increase in EFW = 0.39 [95% CI 0.30 to 0.48], odds ratio [OR] = 1.48 [95% CI 1.35 to 1.62]). The coefficient for customised EFW was similar (0.42 [95% CI 0.33 to 0.51], OR = 1.53 [95% CI 1.39 to 1.67]), hence, for simplicity, non-customised EFW was subsequently employed. Maternal characteristics (age, height, body mass index, and weight gain between 12 and 36 weeks) when combined in a multivariate logistic regression model were moderately predictive for emergency CD (AUROCC = 0.68). Adding blinded EFW to the model increased the AUROCC to 0.71 and this model was more predictive (P<0.0001). When using this model and defining screen positive as a predicted risk of emergency CD ≥40%, 189

(6.2%) women screened positive and the proportion delivered by caesarean was 48%. Compared with screen negative women, they had elevated risks (relative risk [95% CI]) of severe postpartum hemorrhage (2.49 [1.83 to 3.38]), any adverse neonatal outcome (1.86 [1.22 to 2.82]), and severe adverse neonatal outcome (4.03 [1.35 to 12.03]). The risks of these events were also higher compared to women who had a term CD for breech presentation. The model was similarly predictive of the risk of emergency CD and perinatal morbidity when evaluated using routinely collected data from 55,337 births in Scotland between 2003 and 2008.

Conclusions

Ultrasonic EFW at 36 weeks, combined with maternal characteristics, identifies women who are at increased risk of subsequent emergency CD. These women were at increased risk of maternal and perinatal morbidity compared with women at low risk of emergency CD and with women having CD for breech presentation at term.

KEYWORDS

Adverse pregnancy outcomes, emergency caesarean delivery, estimated fetal weight, prediction model, pregnancy, ultrasound

INTRODUCTION

The WHO Global Survey demonstrated that spontaneous vaginal birth was associated with lower rates of maternal and offspring morbidity and mortality compared with other types of birth (operative vaginal delivery, planned caesarean delivery [CD] and emergency CD).[1] However, the risks of adverse maternal and perinatal outcomes were analysed on the basis of the actual rather than the intended mode of delivery. A woman who attempts a normal birth may ultimately require emergency CD, which is associated with higher rates of adverse maternal and perinatal outcome than a planned CD.[1] Consequently, the risk of complications among women attempting vaginal birth will, inevitably, be higher than the risk of complications among the women who are successful, and the risk will increase with the probability that the mother will require emergency CD.

It follows that being able to identify women at high risk of emergency CD may allow interventions which would improve outcome. It is well recognised that the risk of emergency CD increases with increasing birth weight, and this suggests that ultrasonic estimated fetal weight (EFW) might be a useful predictor. However, multiple studies have demonstrated that suspected macrosomia is a risk factor for emergency CD even when the baby is normally grown. Moreover, it is not clear whether EFW should be "customised" for maternal characteristics, and there are no data on whether EFW adds information to other maternal characteristics which are also associated with the risk of macrosomia, such as obesity.[2]

We used data from 3,047 low risk women from a prospective cohort study of nulliparous women, who all had a blinded ultrasonic EFW at ~36 weeks. The aims of the present study were: (1) to quantify the association between customised and non-customised EFW and the risk of emergency CD, (2) to determine whether adding ultrasonic EFW improved prediction based on maternal characteristics alone, (3) to assess the screening performance of a multivariate model using both EFW and maternal characteristics, and (4) to determine whether women at high predicted risk of emergency CD had higher risks of maternal and perinatal morbidity than other women.

METHODS

Design

The Pregnancy Outcome Prediction (POP) study was a prospective cohort study conducted at the Rosie Hospital, Cambridge (UK). The study has previously been described in detail.[3;4] In brief, nulliparous women attending for their dating ultrasound scan between 14th January 2008 and 31st July 2012 were eligible. Women with a viable singleton pregnancy were approached by a research midwife. Following written consent, women were given follow up appointments at approximately 20, 28 and 36wkGA. Ultrasound scans were performed at all three follow up appointments. At 28 and 36wkGA, fetal biometry was performed, and the results were not revealed to the women or the professionals providing care. However, important incidental findings were revealed, including non-cephalic presentation at 36wkGA. The outcome of the pregnancy was obtained both by individual review of the paper case record by research midwives, and by record linkage to the hospital's databases. The study was approved by the Cambridgeshire 2 Research Ethics Committee, reference number 07/H0308/163. The reporting of this study conforms to the STROBE (The Strengthening the Reporting of Observational Studies in Epidemiology) statement.

Study population

Women who had preterm birth, non-cephalic presentation at the 36wkGA scan, pre-labour CD, antepartum stillbirth, pre-existing diabetes or hypertension were excluded. Women were also excluded if they had developed gestational hypertension, pre-eclampsia or gestational diabetes prior to their 36wkGA ultrasound scan. We also excluded women who

withdrew, were lost to follow-up, did not attend the 20 or 36wkGA scan or had no information on the outcome, or one or more predictor variables.

Selection of predictors

The first aim of this analysis was to determine the diagnostic effectiveness of ultrasonic estimation of fetal weight (EFW), combined with maternal characteristics, in the prediction of emergency CD in otherwise low risk women at ~36wkGA. Other characteristics which were potential predictors of emergency CD were selected based on previously published studies[5-16] and their availability in the POP study at 36wkGA: maternal age, height, body mass index (BMI), weight gain from 12 to 36wkGA, and fetal sex. Details of the prediction models and validation are given in the Statistical methods and the Appendix S1.

Definition of exposures

Biometric measurements were performed as previously described.[4;17-19] Importantly, women and clinicians were blinded to the results of biometry at the 36wkGA scan. Biometric measurements taken at the 20 and 36wkGA scans included head circumference (HC), biparietal diameter (BPD), femur length (FL) and abdominal circumference (AC). AC growth velocity was expressed as the change in GA adjusted z scores between 20 and 36wkGA scans.[4] EFW was calculated from the four biometric measurements obtained at the 36wkGA scan using published formulae.[20] Where the head measurements could not be made, the equation employing AC and FL alone was applied. The EFW was expressed as a z score adjusted for GA at measurement.[4] Additionally, customised EFW was calculated (GROW v6.7.3_13 [UK], Gestation Network [www.gestation.net]) and converted into customised EFW z scores.[21] Maternal age was defined as age at recruitment (~12wkGA).

Height was measured at ~20wkGA. Maternal weight was measured at ~12wkGA and ~36wkGA, and maternal weight gain between these measurements was transformed into a z score, adjusted for the exact GA at the 36wkGA measurement (Appendix S1). The weight measurement used in the BMI calculation was made at ~12wkGA.

Definition of outcomes

Emergency CD was defined as delivery by caesarean method where the date of delivery had not been pre-arranged. We confined analysis of postpartum haemorrhage (PPH) to major blood loss (≥1000 mL). We also included acquired postnatal anaemia, which was defined as hemoglobin (Hb) <8 g/dL within a week after delivery where the prenatal Hb (≥1 day before the day of delivery) was >10 g/dL. Neonatal morbidity was defined as ≥1 of the following: (1) metabolic acidosis, defined as cord blood pH <7.10 and base deficit of >10mmol/L, (2) 5 minute Apgar <7, (3) neonatal unit admission within 48 hours from birth for at least 48 hours. Severe neonatal morbidity was defined as ≥1 of the following: (1) neonatal death, (2) hypoxic ischemic encephalopathy, (3) use of inotropes, (4) mechanical ventilation, (5) severe metabolic acidosis, defined as pH<7.00 and a base deficit of >12mmol/L.

Statistical methods

Numerical data were compared using a two-sample Wilcoxon rank-sum test, and categorical data were compared using a Pearson Chi-square test, Chi-square test for trend, and Fisher's exact test, as appropriate. Continuous predictors were transformed into z scores, with adjustment for GA at measurement where appropriate. Univariable and bivariable associations with interactions were estimated for all candidate predictors. Predictive models were then generated using multivariable logistic regression allowing for nonlinear

associations using fractional polynomials (see details in Appendix S1). Two-way interactions were tested using the likelihood-ratio test. The model selection was performed by backward elimination using a likelihood-ratio test p-value threshold of 0.05. The performance of models was described using the area under the receiver operating characteristic curve (AUROCC, i.e. the C statistic). Difference between the nested models that included/excluded the ultrasonic fetal biometry was tested using a likelihood-ratio test as recommended.[22;23] Optimism was assessed using 100-fold bootstrapping.[24] Further validation was performed by dividing the dataset into model development and validation groups using study epochs (Appendix S1).

External validation was performed using routinely collected data from Scotland, from a previously described cohort.[25] We identified nulliparous women in the dataset who had a singleton pregnancy in a cephalic presentation and who delivered a liveborn infant by a means other than planned caesarean at term. The dataset included the exposures maternal age, height and weight. Multiple imputation using predictive mean matching[26] (m=10 imputations, k=10 nearest neighbours) was performed to estimate values of maternal weight gain and EFW (Appendix S1). Imputation of EFW was aided by the presence of a highly correlated proxy (actual birth weight).

As previously described,[6] women with an estimated ≥40% risk of emergency CD were defined as screen positive, and screening statistics were estimated using this threshold. Additionally, we sub-divided screen negative women into moderate (≥20 to <40%) and low (<20%) risk groups. The predictive ability of models was also analysed using the predicted probability of emergency CD as a continuous variable.

We compared the risk of maternal and neonatal complications in the screen positive women with women who had a CD following diagnosis of breech presentation at the 36wkGA scan

in the POP study. This group included women who declined external cephalic version, or had a failed attempt and went on to have a planned CD, and those who were delivered by emergency CD (e.g. because of the onset of labour prior to a scheduled date for external cephalic version or planned caesarean). In the Scottish data, we were able to compare the risk of neonatal morbidity only, and this was done comparing screen positive women with women who had a CD (planned or emergency) for breech presentation at term.

In the comparison of outcomes by predicted risk, relative risks (95% CI) were used. Risk differences and numbers needed to treat (NNT) were calculated to compare outcomes in screen positive women and women who had a planned CD. All statistical analyses were performed using Stata, version 13.1 (StataCorp LP, College Station, Texas).

RESULTS

Main cohort selection

Among the 4,512 recruited women, 4,011 (89%) attended for the 36wkGA scan (see Sovio et al,[4] for flow diagram). From this group, we excluded 885 (22%) women with one or more of the pre-defined exclusion criteria (Appendix S1). Of the remaining 3,126 women, 79 (2.5%) had a missing value in one or more potential predictor variables. This resulted in a study group of 3,047 low risk women. The prevalence of emergency CD in this group was 18.7%.

Descriptive statistics

The distributions of all candidate predictors, except fetal sex, significantly differed between the women who underwent emergency CD and the women who had a vaginal delivery (Table 1). The women in the emergency CD group were older, shorter and heavier, and they gained more weight during pregnancy. The growth velocity of the fetal AC was greater, and their EFW was higher at 36wkGA. They were also more likely to have an induced labour, and give birth at a later GA.

EFW and the risk of emergency CD

There was a linear relationship between the EFW z score at 36 weeks and the risk of emergency CD (Table 2, Appendix S1). The increase in the log odds of emergency CD for a 1 standard deviation increase in EFW was similarly predictive comparing non-customised EFW with EFW customised for maternal characteristics (coefficients = 0.39 [95% CI 0.30 to 0.48] versus 0.42 [95% CI 0.33 to 0.51], respectively, P=0.65), therefore, for simplicity, non-customised EFW was subsequently used.

Prediction models

We considered six other candidate predictors (maternal age, height, 1/BMI, weight gain, AC growth velocity and fetal sex) (Table 2, Appendix S1). The strongest associations were with maternal height and EFW. A predictive model excluding EFW and employing maternal age, height, 1/BMI and weight gain had an AUROCC of 0.68. Adding EFW to the model increased the AUROCC to 0.71 and this model was more predictive of emergency CD (likelihood-ratio test P<0.0001). All subsequent analysis used a model including EFW, maternal age, height, 1/BMI and weight gain. All variables were entered as linear terms without interactions, as there was no evidence supporting non-linearity or interactions.

Evaluation of model performance

Correction for optimism using bootstrapping had a negligible influence (reduction of 0.003 in the AUROCC) and observed risks were close to predicted risks (Appendix S1). The screening performance of the selected model is presented for the estimated probability cut-off ≥40%. This cut off identified 16% of the 569 women with emergency CD as screen positive and the positive predictive value was 48% (Appendix S1). A risk calculator for emergency CD was developed (Appendix S2). For example, for a 37-year old 165 cm tall woman with BMI 30 kg/m^2 who put on 13 kg weight and had a scan at 36wkGA + 1 day with an EFW of 3400 g is estimated to have a 50% risk of emergency CD.

Internal and external validation of the model

We developed a model from births between 2008 and 2010 (n=1,436), and tested the model in births from 2011 to 2013 (n=1,611). Women delivering 2011-2013 who had a predicted

risk of emergency CD ≥40% had an observed CS risk of 47%. We performed external validation on Scottish data of 55,537 eligible births. The prevalence of emergency CD was 16.6% (n=9,212). The C statistic for the model was 0.71, and the observed risk of emergency CD in the screen positive women was 44%. Screening summary statistics, calibration plots and univariable associations between predictors and emergency CD are presented in the Appendix S1.

Predicted and observed risk by gestational age and induction of labour

The association between predicted and observed risk of emergency CD was further analysed in the POP study in relation to GA and induction of labour. Screen positive women had observed emergency CD between 39% to 61% at all gestational ages (Figure 1). Among women at low or intermediate risk, there was a striking increase in the risk of emergency CD with advancing GA. In all groups of predicted risk, the proportion of emergency CD was higher among women whose labour was induced.

Maternal and neonatal complications by predicted risk

Screen positive women within the POP study had a higher subsequent risk of maternal and neonatal complications than screen negative women (Table 3). They had 2.5-fold risk of postpartum haemorrhage, a 2-fold risk of any neonatal morbidity and a 4-fold risk of severe neonatal morbidity. Screening statistics of maternal and neonatal complications by predicted risk of emergency CD are presented in Table 4. Maternal and neonatal outcomes were then compared between screen positive women and women in the cohort who had a breech presentation diagnosed at the 36wkGA scan and ultimately had a CD (n=128, either planned, generally at 39 wkGA, or emergency, generally if labour started prior to the

scheduled date for planned caesarean) (Table 3). The risk of postpartum haemorrhage and any neonatal morbidity were lower in the breech group. There were no cases of severe neonatal morbidity in the breech group and the sample size was too small to detect a difference. In the Scottish dataset, neonatal morbidity was more frequent among the screen positive women both in comparison to screen negative women and women who had a planned caesarean for breech presentation (Table 5).

Risk of morbidity in screen positive women, stratified by actual mode of delivery

Among women delivered by emergency CD, those who were screen positive had a higher risk (relative risk [95% CI]) of postpartum haemorrhage (2.12 [1.49, 3.00]), and severe neonatal morbidity (3.55 [1.02, 12.32]), but there was no association with any neonatal morbidity (1.43 [0.83, 2.47]). Among women delivered vaginally, those who were screen positive had a similar risk of postpartum haemorrhage (1.21 [0.61, 2.39]) and neonatal morbidity (1.51 [0.76, 3.00]), and there were no cases of severe neonatal morbidity. However, there was a higher risk of metabolic acidosis (6.01 [1.72, 20.95]).

DISCUSSION

We found that a model for predicting emergency CD using a blinded ultrasonic EFW at 36wkGA, maternal age, height, BMI and weight gain, correctly identified women at increased risk of the procedure. The actual proportion of emergency CD in screen positive (≥40% predicted risk) women was 48%. When we corrected for optimism and performed internal and external validation, the model was similarly predictive. These data suggest that this approach robustly identifies women at high risk of emergency CD.

Previous studies have shown that emergency intrapartum CD is associated with higher risks of adverse maternal and perinatal outcome than both vaginal birth and planned CD.[1] In the context of vaginal breech delivery, it had previously been calculated that a decision to deliver all women by planned caesarean may be neutral for maternal morbidity when the risk of emergency CD was between 16-30%.[27] We speculated, therefore, that women with a high predicted risk of emergency CD and a cephalic presentation may be at increased risk of both adverse maternal and perinatal outcome compared with other women. Consistent with this, we found that women with a predicted risk of emergency CD of ≥40% (screen positive) had a 2.5-fold risk of postpartum haemorrhage, an almost 2-fold risk of any neonatal morbidity, and a 4-fold risk of severe neonatal morbidity. This analysis compared all screen positive women, irrespective of the eventual mode of delivery. Interestingly, when we analysed all women who were ultimately delivered by emergency CD, those identified as screen positive had higher rates of complications than other women delivered by emergency CD. Hence, the association between screen positive for emergency CD using our model and maternal and perinatal morbidity was explained by: (i) women delivered by

emergency CD have higher rates of complications, and (ii) the model identified a sub-group of women experiencing emergency CD who were at particularly high risk of complications. We speculate that the mediator for this association is likely to be slow progress in labour and/or obstructed labour, as factors such as high BMI and advanced age have been associated with increased risks of labour dystocia.[28;29] These observations also indicate that although the model had low sensitivity (16%), it did identify a sub-group of women who had an emergency CD who were at particularly high risk of associated complications.

The current analysis suggests that low risk nulliparous women might be screened for their risk of emergency CD at 36wkGA. However, screening is only justified when an intervention exists which might mitigate associated risks. We performed further analyses of the dataset to help inform the question of candidate interventions. The WHO survey found that the risk of perinatal and maternal morbidity was lower among women delivered by planned CD than emergency CD and this, therefore, might be regarded as a possible intervention in a future trial. Hence, we next compared the rate of maternal and perinatal complications in screen positive women with women who had a breech presentation diagnosed at 36wkGA and who were ultimately delivered by caesarean (planned or emergency). The risk of both maternal and perinatal adverse outcome was lower in the women who had a breech presentation diagnosed at 36wkGA than among women who were screen positive at 36wkGA. The data suggest, but do not prove, that planning CD for 39wkGA, and performing an emergency CD if labour starts prior to the scheduled date for CD, may be an intervention which would mitigate the maternal and perinatal risks associated with a high predicted risk of emergency CD at 36wkGA. However, confirming this hypothesis will require a randomized controlled clinical trial.

We also considered induction of labour as a possible intervention, as meta-analyses indicate that routine induction of labour may slightly reduce the risk of emergency CD.[30] Among screen positive women, the actual proportion of emergency CD was even higher among those who ultimately had labour induced. Moreover, the observed proportion of emergency CD was consistently ≥40% for all weeks of GA from 37wkGA onwards. The data suggest, but do not prove, that offering early term induction of labour is unlikely to mitigate the high predicted risk of emergency CD.

The information from the model also has immediate clinical application, in addition to informing a future trial. Some women opt to deliver at home or in units which lack the facilities to perform emergency CD, such as low risk birthing units. Transfer during labour is known to be a high risk situation. The predicted risk of emergency CD may help inform decisions around the place of birth. This is of particular importance for nulliparous women as there is high quality evidence that planned home delivery is associated with a higher rate of complications in nulliparous, but not multiparous women.[31]

One of the strengths of the present study is that we confined the analysis to data which were available at the time of the 36wkGA prenatal assessment. It is known that increasing birth weight is associated with the risk of emergency CD. However, birth weight is clearly only known post-delivery and cannot, therefore, be included in prenatal risk assessment and decision making. Ultrasonic fetal biometry is correlated with birth weight, but the average absolute error is ~7% even when the measurements are made within a week of delivery.[32] Moreover, the relationship between ultrasonic EFW and the risk of CD is complicated by the

fact that knowledge of the EFW is a determinant of the risk of emergency CD independently of the actual birth weight: multiple studies have shown that a high EFW is associated with an increased risk of CD even when the birth weight is normal.[33-35] This association is attributed to bias on the part of the attending staff in labour. We acknowledge that we could not blind the assessment of maternal predictors. However, one of the strengths of our study is that we had blinded ultrasonic EFW. Hence, the current analysis presents the true association between EFW and the risk of CD, rather than the association due to biases based on knowledge of the EFW.

ACKNOWLEDGMENTS

The work was supported by the NIHR Cambridge Comprehensive Biomedical Research Centre and the Stillbirth and Neonatal Death Society. GE donated two Voluson i ultrasound systems for this study. The study was also supported by the NIHR Cambridge Clinical Research Facility, where all visits at ~20, ~28 and ~36 weeks took place. None of the bodies providing direct or indirect support for the project was involved in any aspect of preparation of this work for publication. We are grateful to the Perinatal Institute for providing a bulk calculator for customised percentiles of estimated fetal weight. We are extremely grateful to all the women who participated in the study, and to all the staff in the Rosie Hospital and NIHR Cambridge Clinical Research Facility who provided direct or indirect assistance for the study. We are grateful to the sonographers who performed all the measurements in this study: Alison Dacey, Amanda Rowley, Maxine Snaith, Sam Everett, Sonia Cummings and Jan Nelder and to the hospital's lead sonographer, Trish Chudleigh. Finally, we thank Professor Ian White for statistical advice.

REFERENCES

(1) Lumbiganon P, Laopaiboon M, Gulmezoglu AM, Souza JP, Taneepanichskul S, Ruyan P, Attygalle DE, Shrestha N, Mori R, Nguyen DH, Hoang TB, Rathavy T, Chuyun K, Cheang K, Festin M, Udomprasertgul V, Germar MJ, Yanqiu G, Roy M, Carroli G, Ba-Thike K, Filatova E, Villar J; World Health Organization Global Survey on Maternal and Perinatal Health Research Group. Method of delivery and pregnancy outcomes in Asia: the WHO global survey on maternal and perinatal health 2007-08. *Lancet* 2010; 375(9713):490-499.

(2) Vinturache AE, Chaput KH, Tough SC. Pre-pregnancy body mass index (BMI) and macrosomia in a Canadian birth cohort. *J Matern Fetal Neonatal Med* 2017; 30(1):109-116.

(3) Pasupathy D, Dacey A, Cook E, Charnock-Jones DS, White IR, Smith GC. Study protocol. A prospective cohort study of unselected primiparous women: the pregnancy outcome prediction study. *BMC Pregnancy Childbirth* 2008; 8(1):51.

(4) Sovio U, White IR, Dacey A, Pasupathy D, Smith GCS. 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* 2015; 386:2089-2097.

(5) Seshadri L, Mukherjee B. A predictive model for cesarean section in low risk pregnancies. *Int J Gynaecol Obstet* 2005; 89(2):94-98.

(6) Smith GCS, Dellens M, White IR, Pell JP. Combined logistic and Bayesian modeling of cesarean section risk. *Am J Obstet Gynecol* 2004; 191:2029-2034.

(7) Smith GC, White IR, Pell JP, Dobbie R. Predicting cesarean section and uterine rupture among women attempting vaginal birth after prior cesarean section. *PLoS Med* 2005; 2(9):e252.

(8) Schuit E, Kwee A, Westerhuis ME, Van Dessel HJ, Graziosi GC, Van Lith JM, Nijhuis JG, Oei SG, Oosterbaan HP, Schuitemaker NW, Wouters MG, Visser GH, Mol BW, Moons KG, Groenwold RH. A clinical prediction model to assess the risk of operative delivery. *BJOG* 2012; 119(8):915-923.

(9) Kozinszky Z, Orvos H, Zoboki T, Katona M, Wayda K, Pal A, Kovacs L. Risk factors for cesarean section of primiparous women aged over 35 years. *Acta Obstet Gynecol Scand* 2002; 81(4):313-316.

(10) Rane SM, Guirgis RR, Higgins B, Nicolaides KH. The value of ultrasound in the prediction of successful induction of labor. *Ultrasound Obstet Gynecol* 2004; 24(5):538-549.

(11) Usha Kiran TS, Hemmadi S, Bethel J, Evans J. Outcome of pregnancy in a woman with an increased body mass index. *BJOG* 2005; 112(6):768-772.

(12) Morris JM, Thompson K, Smithey J, Gaffney G, Cooke I, Chamberlain P, Hope P, Altman D, MacKenzie IZ. The usefulness of ultrasound assessment of amniotic fluid in predicting adverse outcome in prolonged pregnancy: a prospective blinded observational study. *BJOG* 2003; 110(11):989-994.

(13) Peregrine E, O'Brien P, Omar R, Jauniaux E. Clinical and ultrasound parameters to predict the risk of cesarean delivery after induction of labor. *Obstet Gynecol* 2006; 107(2 Pt 1):227-233.

(14) Chen G, Uryasev S, Young TK. On prediction of the cesarean delivery risk in a large private practice. *Am J Obstet Gynecol* 2004; 191(2):616-624.

(15) Saunders N, Amis S, Marsh M. The prognostic value of fetal ultrasonography before induction of labour. *Br J Obstet Gynaecol* 1992; 99(11):903-906.

(16) Young TK, Woodmansee B. Factors that are associated with cesarean delivery in a large private practice: the importance of prepregnancy body mass index and weight gain. *Am J Obstet Gynecol* 2002; 187(2):312-318.

(17) Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 2. Head measurements. *Br J Obstet Gynaecol* 1994; 101(1):35-43.

(18) Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 3. Abdominal measurements. *Br J Obstet Gynaecol* 1994; 101(2):125-131.

(19) Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 4. Femur length. *Br J Obstet Gynaecol* 1994; 101(2):132-135.

(20) 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(3):333-337.

(21) Gardosi J, Chang A, Kalyan B, Sahota D, Symonds EM. Customised antenatal growth charts. *Lancet* 1992; 339(8788):283-287.

(22) Vickers AJ, Cronin AM, Begg CB. One statistical test is sufficient for assessing new predictive markers. *BMC Med Res Methodol* 2011; 11:13.

(23) Pepe MS, Kerr KF, Longton G, Wang Z. Testing for improvement in prediction model performance. *Stat Med* 2013; 32(9):1467-1482.

(24) Smith GC, Seaman SR, Wood AM, Royston P, White IR. Correcting for optimistic prediction in small data sets. *Am J Epidemiol* 2014; 180(3):318-324.

(25) Wood AM, Pasupathy D, Pell JP, Fleming M, Smith GC. Trends in socioeconomic inequalities in risk of sudden infant death syndrome, other causes of infant mortality, and stillbirth in Scotland: population based study. *BMJ* 2012; 344:e1552.

(26) Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Med Res Methodol* 2014; 14:75.

(27) Bingham P, Lilford RJ. Management of the selected term breech presentation: assessment of the risks of selected vaginal delivery versus cesarean section for all cases. *Obstet Gynecol* 1987; 69(6):965-978.

(28) Smith GC, Cordeaux Y, White IR, Pasupathy D, Missfelder-Lobos H, Pell JP, Charnock-Jones DS, Fleming M. The Effect of Delaying Childbirth on Primary Cesarean Section Rates. *PLoS Med* 2008; 5(7):e144.

(29) Cedergren MI. Non-elective caesarean delivery due to ineffective uterine contractility or due to obstructed labour in relation to maternal body mass index. *Eur J Obstet Gynecol Reprod Biol* 2009; 145(2):163-166.

(30) Mishanina E, Rogozinska E, Thatthi T, Uddin-Khan R, Khan KS, Meads C. Use of labour induction and risk of cesarean delivery: a systematic review and meta-analysis. *CMAJ* 2014; 186(9):665-673.

(31) Brocklehurst P, Hardy P, Hollowell J, Linsell L, Macfarlane A, McCourt C, Marlow N, Miller A, Newburn M, Petrou S, Puddicombe D, Redshaw M, Rowe R, Sandall J, Silverton L, Stewart M for the Birthplace in England Collaborative Group. Perinatal and maternal outcomes by planned place of birth for healthy women with low risk pregnancies: the Birthplace in England national prospective cohort study. *BMJ* 2011; 343:d7400.

(32) Smith GCS, Smith M-FS, McNay MB, Fleming J-EE. The relation between fetal abdominal circumference and birthweight: Findings in 3512 pregnancies. *Br J Obstet Gynaecol* 1997; 104:186-190.

(33) Little SE, Edlow AG, Thomas AM, Smith NA. Estimated fetal weight by ultrasound: a modifiable risk factor for cesarean delivery? *Am J Obstet Gynecol* 2012; 207(4):309-6.

(34) Blackwell SC, Refuerzo J, Chadha R, Carreno CA. Overestimation of fetal weight by ultrasound: does it influence the likelihood of cesarean delivery for labor arrest? *Am J Obstet Gynecol* 2009; 200(3):340-343.

(35) Parry S, Severs CP, Sehdev HM, Macones GA, White LM, Morgan MA. Ultrasonographic prediction of fetal macrosomia. Association with cesarean delivery. *J Reprod Med* 2000; 45(1):17-22.

SUPPORTING INFORMATION ON THE INTERNET

Appendix S1. Supplementary Information. Details of model selection and validation are given in Supplementary Methods and Supplementary Results.

Appendix S2. Emergency caesarean section risk calculator.

Table 1. Characteristics of the study sample (N=3,047), stratified by mode of delivery.

Characteristic*	**Spontaneous or assisted vaginal, N=2,478**	**Intrapartum emergency caesarean, N=569**	**P†**
Maternal characteristics			
Age, years	30 (26 to 33)	31 (28 to 34)	<0.001
<20	104 (4.2)	9 (1.4)	
20 to 24.9	358 (14)	62 (11)	
25 to 29.9	812 (33)	157 (28)	<0.001
30 to 34.9	904 (36)	236 (41)	
35 to 39.9	279 (11)	91 (16)	
≥40	21 (0.9)	14 (2.5)	
Height, cm	166 (162 to 170)	163 (159 to 167)	<0.001
<160	377 (15)	144 (25)	
160 to 164.9	644 (26)	191 (34)	<0.001
165 to 169.9	752 (30)	147 (26)	
≥170	705 (28)	87 (15)	
BMI, kg/m^2	23 (22 to 26)	25 (23 to 28)	<0.001
<25	1610 (65)	268 (47)	
25 to 29.9	629 (25)	204 (36)	
30 to 34.9	183 (7.4)	81 (14)	<0.001
35 to 39.9	46 (1.9)	11 (1.9)	
≥40	10 (0.4)	5 (0.9)	
Maternal weight gain‡, kg	12 (10 to 14)	12 (10 to 15)	<0.001

Birth characteristics			
Gestational age (wkGA)	40.4 (39.6 to 41.1)	41.0 (40.1 to 41.7)	<0.001
<39	345 (14)	45 (7.9)	
39	503 (20)	72 (13)	<0.001
40	826 (33)	154 (27)	
≥41	804 (32)	298 (52)	
Fetal sex			
Male	1222 (49)	299 (53)	0.16
Induced labour	639 (26)	281 (49)	<0.001
Ultrasound measurements‡			
AC growth, 20-36 wkGA, mm	162 (151 to 175)	169 (158 to 182)	<0.001
EFW at 36 weeks, g	2713 (2514 to 2960)	2838 (2591 to 3115)	<0.001

*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 binary and a score test for trend of odds for categorical ordered variables (summary P value across the range of categories of the given variable).

‡For maternal weight gain and ultrasound measurements, unadjusted measurements are reported in kg, mm or g.

AC, abdominal circumference; BMI, body mass index; EFW, estimated fetal weight, wkGA, weeks of gestational age.

Table 2. Coefficients with 95% CI from univariable and multivariable logistic regression models, n=3,047.

	Model			
Predictor variable	*Univariable models*	*Full model*	*Selected multivariable model*	*Selected model excluding US*
Age	0.28 (0.18 to 0.37)	0.32 (0.22 to 0.42)	0.32 (0.22 to 0.42)	0.34 (0.24 to 0.44)
Height	-0.42 (-0.51 to -0.32)	-0.52 (-0.63 to -0.42)	-0.53 (-0.63 to -0.43)	-0.46 (-0.56 to -0.36)
1/BMI	-0.37 (-0.46 to -0.28)	-0.30 (-0.40 to -0.20)	-0.30 (-0.40 to -0.20)	-0.37 (-0.47 to -0.28)
Weight gain	0.17 (0.09 to 0.26)	0.21 (0.11 to 0.30)	0.21 (0.11 to 0.30)	0.27 (0.18 to 0.36)
Fetal sex: female	-0.13 (-0.31 to 0.05)	-0.10 (-0.29 to 0.10)	N/A	N/A
EFW at 36 week scan	0.39 (0.30 to 0.48)	0.35 (0.24 to 0.47)	0.39 (0.29 to 0.49)	N/A
AC growth velocity	0.29 (0.20 to 0.39)	0.06 (-0.05 to 0.18)	N/A	N/A

All variables except fetal sex are expressed as z scores, adjusted for gestational age at measurement where appropriate. BMI denotes body mass index, AC denotes abdominal circumference, EFW denotes estimated fetal weight, US denotes ultrasonic assessment.

Table 3. Maternal and neonatal complications by predicted risk of intrapartum emergency caesarean delivery (CD) using 40% risk cut-off point and in women who had CD due to breech presentation.

	Screen positive		Screen negative		Breech		Comparison: Screen positive versus screen negative			Comparison: Screen positive versus breech			
Outcome	**n/N**	**%**	**n/N**	**%**	**n/N**	**%**	**RR***	**95% CI**	**P†**	**RD‡**	**95% CI‡**	**P†**	**NNT**
Emergency CD	90/189	48	479/2858	17	-		2.84	2.40, 3.37	<0.0001	-		-	-
Assisted vaginal delivery[§]	41/99	41	759/2379	32	-		1.30	1.02, 1.65	0.062	-		-	-
Postpartum hemorrhage[¶]	39/189	21	237/2858	8	0/128	0	2.49	1.83, 3.38	<0.0001	20.6	14.7, 27.0	<0.0001	4.8
Metabolic acidosis	6/189	3	23/2858	1	0/128	0	3.94	1.63, 9.57	0.0075	3.2	-0.2, 6.8	0.085	31.5
5 minute Apgar <7	4/189	2	21/2858	1	0/128	0	2.88	1.00, 8.31	0.065	2.1	-1.1, 5.3	0.15	47.3
Neonatal unit admission at term	18/189	10	152/2858	5	5/128	4	1.79	1.12, 2.85	0.021	5.6	-0.4, 11.1	0.077	17.8
Any neonatal morbidity[¶]	22/189	12	179/2858	6	5/128	4	1.86	1.22, 2.82	0.0091	7.7	1.5, 13.5	0.023	12.9
Severe neonatal morbidity at term[¶]	4/189	2	15/2858	1	0/128	0	4.03	1.35, 12.03	0.027	2.1	-1.1, 5.3	0.15	47.3

*RR for the presence of each outcome is for screen positive (≥40% predicted risk) vs. screen negative (<40% predicted risk) group.

†P-values are from 2-sided Fisher's exact test.

‡RD is expressed as a percentage. Newcombe Method 10 CIs are used for RD.

§N is restricted to the vaginal deliveries (N=2,478).

¶Postpartum hemorrhage is defined as Hb<8 g/dL or estimated blood loss ≥ 1000 mL. Hb <8 g/dL within a week from delivery but >10 g/dL antepartum (before the day of delivery). Any neonatal morbidity is defined as ≥1 of the following: (1) metabolic acidosis, defined as cord blood pH <7.1 and base deficit of >10mmol/L, (2) 5 minute Apgar <7, (3) neonatal unit admission within 48 hours from birth at term for at least 48 hours. Severe neonatal morbidity is defined as ≥1 of the following: (1) neonatal death, (2) hypoxic ischemic encephalopathy, (3) use of inotropes, (4) mechanical ventilation, (5) severe metabolic acidosis, defined as pH<7.0 and a base deficit of >12mmol/L at term.

CD, caesarean delivery; RR, relative risk; RD, risk difference; CI, confidence interval; NNT, numbers needed to treat.

Table 4. Screening statistics of maternal and neonatal complications by predicted risk of intrapartum emergency caesarean delivery (CD) using 40% risk cut-off point.

Outcome	Se (%)	95% CI	Sp (%)	95% CI	PPV (%)	95% CI	NPV (%)	95% CI	LR+	95% CI	LR-	95% CI
Emergency CD	15.8	12.8, 18.8	96.0	95.2, 96.8	47.6	40.5, 54.7	83.2	81.9, 84.6	4.0	3.0, 5.2	0.88	0.85, 0.91
Assisted vaginal delivery*	5.1	3.6, 6.7	96.5	95.7, 97.4	41.4	31.7, 51.1	68.1	66.2, 70.0	1.5	1.0, 2.2	0.98	0.96, 1.00
Postpartum hemorrhage†	14.1	10.0, 18.2	94.6	93.7, 95.4	20.6	14.9, 26.4	91.7	90.7, 92.7	2.6	1.9, 3.6	0.91	0.86, 0.95
Metabolic acidosis	20.7	5.9, 35.4	93.9	93.1, 94.8	3.2	0.7, 5.7	99.2	98.9, 99.5	3.4	1.7, 7.1	0.84	0.70, 1.02
5 minute Apgar <7	16.0	1.6, 30.4	93.9	93.0, 94.7	2.1	0.1, 4.2	99.3	99.0, 99.6	2.6	1.1, 6.5	0.89	0.75, 1.06
Neonatal unit admission at term	10.6	6.0, 15.2	94.1	93.2, 94.9	9.5	5.3, 13.7	94.7	93.9, 95.5	1.8	1.1, 2.8	0.95	0.90, 1.00
Any neonatal morbidity†	10.9	6.6, 15.3	94.1	93.3, 95.0	11.6	7.1, 16.2	93.7	92.8, 94.6	1.9	1.2, 2.8	0.95	0.90, 0.99
Severe neonatal morbidity at term†	21.1	2.7, 39.4	93.9	93.0, 94.7	2.1	0.1, 4.2	99.5	99.2, 99.7	3.4	1.4, 8.3	0.84	0.67, 1.06

*N is restricted to the vaginal deliveries (N=2,478).

†Postpartum hemorrhage is defined as Hb<8 g/dL or estimated blood loss ≥ 1000 mL. Hb <8 g/dL within a week from delivery but >10 g/dL antepartum (before the day of delivery). Any neonatal morbidity is defined as ≥1 of the following: (1) metabolic acidosis, defined as cord blood pH <7.1 and base deficit of >10mmol/L, (2) 5 minute Apgar <7, (3) neonatal unit admission within 48 hours from birth at term for at least 48 hours. Severe neonatal morbidity is defined as ≥1 of the following: (1) neonatal death, (2) hypoxic ischemic encephalopathy, (3) use of inotropes, (4) mechanical ventilation, (5) severe metabolic acidosis, defined as pH<7.0 and a base deficit of >12mmol/L at term.

CD, caesarean delivery; Se, sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio; CI, confidence interval.

Table 5. Maternal and neonatal complications by predicted risk of intrapartum emergency caesarean delivery (emergency CD) using 40% risk cut-off point and in women who had CD due to breech presentation in 55,512 term livebirths in Scottish validation data, 2003-2008. *

	Comparison: Screen positive versus screen negative			Comparison: Screen positive versus breech		
Outcome	**RR†**	**95% CI†**	**P†**	**RR†**	**95% CI†**	**P†**
Emergency CD	3.01	2.82, 3.21	<0.001	-		-
5 minute Apgar <7	1.32	0.95, 1.82	0.097	5.59	2.00, 15.63	0.001
Neonatal unit admission at term	1.59	1.29, 1.96	<0.001	1.95	1.22, 3.11	0.005
Any neonatal morbidity‡	1.49	1.24, 1.79	<0.001	2.31	1.51, 3.53	<0.001

*Predicted risk estimates are from the selected model where a simple shrinkage has been applied.

†RR (95% CI) and p-values are obtained from a generalized linear model with a log-link.

‡Any neonatal morbidity is defined as 5 minute Apgar <7 or neonatal unit admission at term.

CD, caesarean delivery; RR, relative risk; CI, confidence interval.

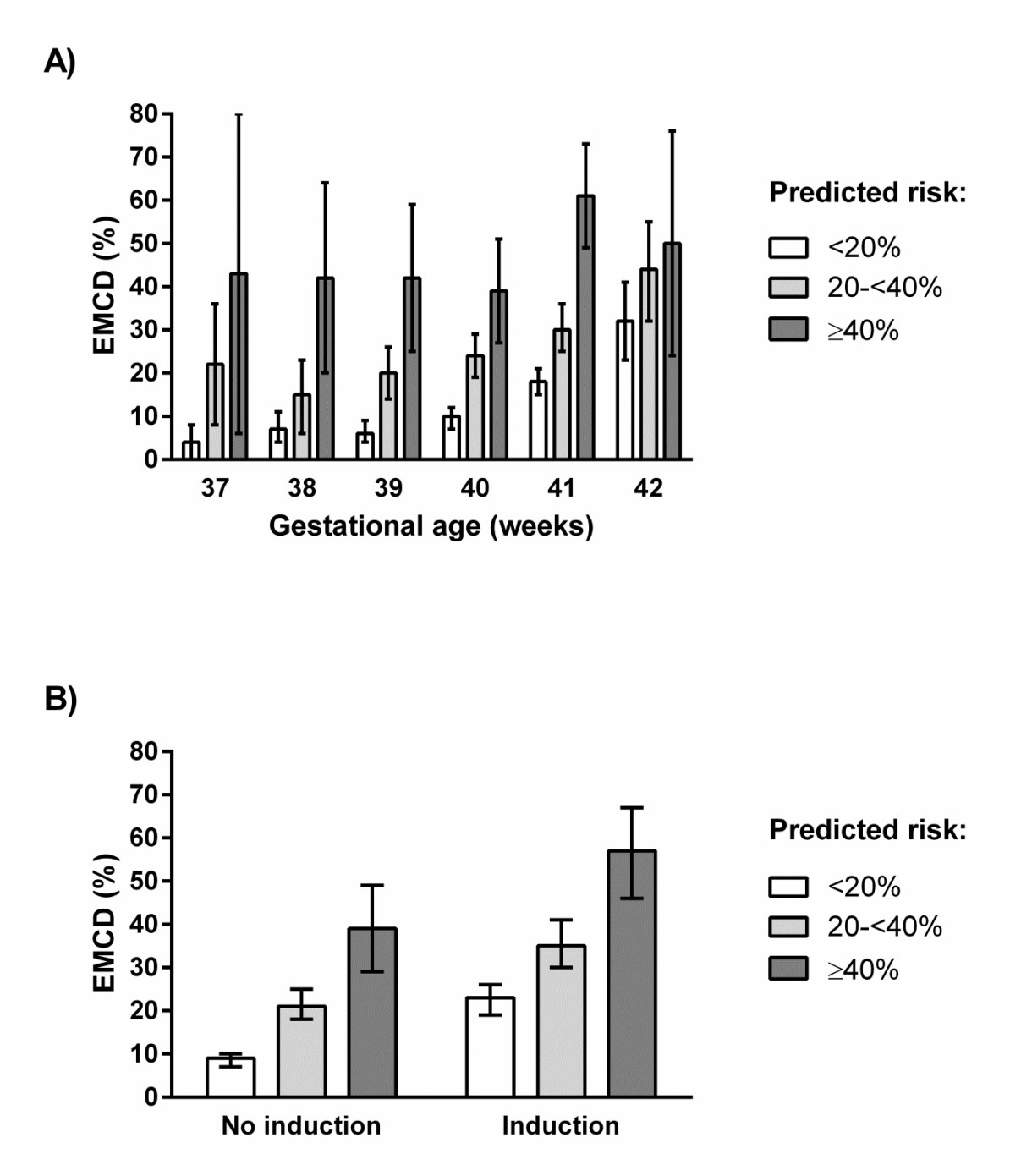


Figure 1. Intrapartum emergency caesarean delivery (%) by predicted risk from the model in relation to A) GA at delivery and B) induction of labour in the POP study. The error bars represent 95% confidence intervals. EMCD denotes emergency caesarean delivery.

Appendix S1. Supplementary Information.

Supplementary Methods

Preliminary analyses of association

Potential continuous predictors were categorised into equal sized groups and patterns of association between each predictor and the outcome were explored graphically to detect nonlinearities of association. The distributions of continuous predictors were examined and a transformation to improve normality was applied where necessary. Correlations between predictors were examined using the Spearman correlation coefficient.

Details of the fractional polynomial (FP) models

Association between intrapartum emergency caesarean delivery (CD) and the predictors was analysed using logistic regression. For continuous predictors, the odds ratios are expressed for one SD difference in the predictor. Linearity of the association between each continuous predictor and the outcome was formally tested using univariable fractional polynomial (FP) models. Fractional powers were chosen from the set (-2, -1, -0.5, 0, 0.5, 1, 2, 3) and models up to degree-2 were considered. Two-way interactions between all pairs of predictors were tested and the linearity of association between any identified interaction and intrapartum emergency CD was tested using FP models. The final model was chosen by fitting multivariable FP models. P-value threshold of 0.05 was applied for variable selection by backward elimination.

Comparison between logistic regression and Poisson regression

The multivariable model was first fitted using logistic regression. For comparison, the selected multivariable model was fitted using Poisson regression with a robust variance estimator. Model fit was assessed using Akaike's information criterion (AIC) and Bayesian information criterion (BIC).

Model validation

We give details of model validation in this section, which are additional to the description of model validation in the method section of the main manuscript. A temporal validation study was performed within the Pregnancy Outcome Prediction (POP) study cohort using two study epochs. A model including ultrasound measurements was first built in the group of women who delivered in 2008-2010 and the coefficients from this model were applied to the group of women who delivered in 2011-2013. Model discrimination and calibration were assessed in both groups.

Calibration of the selected model in the POP study was assessed using tables and calibration plots of observed and predicted values in the study population by the decile of the linear predictor. Additionally, calibration was performed stratified by clinical scans, i.e. by whether

the woman had at least one clinically indicated ultrasound scan in addition to the research scans at ≥28 weeks of gestation and/or her 28 week or 36 week research scan results were revealed, resulting in a referral to a clinical scan. Model calibration was formally assessed using Hosmer-Lemeshow goodness-of-fit tests.

The POP study data and the external Scottish validation data were combined into a single dataset for the imputation model development. Weight gain and estimated fetal weight were imputed using predictive mean matching (m=10 imputations, k=10 donors). The imputation model included all the five predictors, the outcome, and additionally the population-based birth weight z score and fetal sex in order to improve the prediction of EFW and maternal weight gain. The rank correlation between EFW and birth weight z scores was 0.63 in the POP study and the correlation was similar in the imputed validation dataset (0.67). The birth weight z score varied more in the validation data (SD=0.93) than in the POP study (SD=0.84). Also, the correlation between maternal weight gain and birth weight z score was also preserved in the imputation; the rank correlation coefficient was 0.21 in the POP study and 0.20 in the validation data.

Regression coefficients were multiplied by a shrinkage factor calculated in the POP study using a heuristic formula [s = (model LR – df)/model LR] before calculating the predictions in the imputed external validation data. Model discrimination and calibration were then evaluated in the external validation dataset. External validation was supplemented by the comparison of neonatal complications between women who were screen positive (≥40% predicted risk) of intrapartum emergency CD and all other women. Neonatal morbidity was defined as 5 minute Apgar <7 or neonatal unit admission at term for at least 48 hours in the validation data. Additionally, the relative risk of neonatal complications in the women who were screen positive compared to the women who had a planned CD due to breech presentation was estimated in the validation data. The information on planned CD was only available at the time of birth in the validation data. Generalized linear models with a log-link were fitted to calculate and combine relative risks in the imputed external validation data using Rubin's rules.

To further evaluate the external validity of the model when applied to diverse populations, we assessed its predictive ability in the POP study women who were excluded due to pre-existing condition but who otherwise fulfilled the eligibility criteria. We also examined the observed risks of maternal and neonatal complications in categories of predicted risk using cut-offs 10%, 20%, 30% and 40%.

Supplementary Results

The selection of the study group is outlined in **Figure S1**. Among the 4,512 recruited women, 4,011 (89%) attended for the 36wkGA scan. From this group, we excluded 188 (4.7%) with a non-cephalic presentation at the 36wkGA scan, 6 (0.1%) who failed to attend the 20wkGA scan, 133 (3.3%) who were lost to follow-up or did not have data on the mode of delivery,

354 (8.8%) who had a pre-labour caesarean (82 [2.0%] of these were emergency and 272 [6.8%] were elective), 7 (0.2%) who had an antepartum stillbirth, 202 (5.0%) who had pre-existing diabetes or hypertension, and 230 (5.7%) with gestational hypertension, pre-eclampsia or gestational diabetes diagnosed before their 36wkGA scan. A total of 885 (22%) of these women had one or more of the exclusion criteria and additionally 79 (2.5% of the remaining 3,126 women) had a missing value in one or more potential predictor variables. The final study population consisted of 3,047 low risk women (76% of those scanned at 36wkGA) who had a research ultrasound scan at 36wkGA, had complete data and who were delivered by a means other than planned CD at term. Out of the 3,047 women included in the study population, 603 (19.8%) had missing HC and/or BPD measurement, and their EFW was solely based on the measurements of AC and FL. The prevalence of emergency CD in this group was 18.7% (n=569). For comparison, among all recruited women who were not lost to follow-up and had information on the outcome available (n=4,192), the prevalence of emergency CD was 17.3% (n=726).

Among the 569 women who were not excluded from the study and had emergency CD, the indications for CD were fetal distress (n=253, 44%), failure to progress (n=273, 48%), both fetal distress and failure to progress (n=9, 0.3%), other (n=33, 5.8%; for example malpresentation, failed forceps, suboptimal cardiotocography, and maternal pyrexia with fetal tachycardia). In one case (0.2%) the indication for CD was unknown.

Equations for mean and standard deviation (SD) estimated within the POP study by gestational age interval have been previously given for AC and EFW (Sovio et al, Lancet 2015). Maternal weight did not vary by gestational age at 12 week scan but it varied at 36 week scan, performed between 34 and 38 completed weeks: mean = -9.886 + 0.6107*GA, SD = -9.387 + 0.3702*GA. GA-specific z scores were calculated as (observed value – fitted mean) / fitted SD. For all other continuous predictors, a simple z score (observed value – sample mean) / sample SD was calculated. To improve normality of the distribution, an inverse transformation was applied to BMI at 12 week scan before converting it to a z score. For the calculation of simple z scores, the means (SDs) in the POP study women were: age 29.68 (4.973) years, height 165.4 (6.317) cm, 1/BMI 0.04163 (0.006422) m^2/kg.

Correlations between continuous predictors were weak or moderate. The highest correlation coefficient of 0.48 was observed between AC growth velocity and EFW. This level of correlation is unlikely to cause problems of colinearity. The correlation coefficient between maternal height and 1/BMI was only 0.07, i.e. the correlation between height and BMI was -0.07. This is explained by the fact that BMI represents weight adjusted for height (BMI = weight[kg] / height$[m]^2$). Similarly to our study, in most populations height and BMI have a weak negative correlation, which means that taller people are slightly slimmer than average. For comparison, the correlation between maternal weight and height was 0.38, but maternal weight was not included among the candidate predictors.

Univariable associations of all candidate predictors (maternal age, height, 1/BMI, weight gain, AC growth velocity, EFW and fetal sex) were estimated, followed by the estimation of bivariable associations and pairwise interactions. There were approximately 20 events (=569/28) per candidate predictor (7 predictors + 21 pairwise interactions).

Univariable analysis using fractional polynomials suggested a linear association for all predictors. For EFW there was a suggestion towards a quadratic association (degree-1, fractional power 2) but the quadratic model was not statistically significantly better than the linear model (p=0.062). Adjustment for all other predictors in the model improved the linearity of EFW (quadratic vs. linear model p=0.10). No pairwise interactions were observed between any of the linear predictors (p>0.05).

The initial multivariable model included all seven candidate predictors. Maternal age, height, 1/BMI, weight gain and EFW were selected into the multivariable model as linear terms since there was no evidence for nonlinearity of associations (p>0.05). Fetal sex was not predictive of emergency CD in any of the models and AC growth velocity was no longer a statistically significant predictor when it was adjusted for other predictors.

Comparison between logistic regression and Poisson regression

The model fit of the selected multivariable model was better with logistic regression than Poisson regression: logistic regression resulted in AIC of 2680 and BIC of 2716, whereas Poisson regression resulted in AIC of 2854 and BIC of 2890. Hence, results from logistic regression are reported.

Model calibration and validation

The data set was divided into 10 deciles of estimated probabilities from the selected model: Hosmer-Lemeshow goodness-of-fit tests did not show evidence of poor fit of either model in the whole population or stratified by clinical scans (p>0.4 for all tests). A calibration plot in the whole population is presented in **Figure S2**. Calibration plots are also given stratified by clinical scans (**Figure S3**). The observed risks were close to predicted risks in all these analyses. The AUROCC was 0.70 (95% CI: 0.67-0.73) in women who did not have clinical scans at ≥28 weeks and 0.72 (95% CI: 0.68-0.76) in women who had clinical scans at ≥28 weeks and/or a revealed research scan result.

At first, the models were validated internally: correction for optimism had a negligible effect on the AUROCC decreasing it from 0.7057 (**Figure S4**) to 0.7027. Secondly, temporal validation was performed. The same variables were included in the model selection as previously for the final model. Calibration plots for the model which was developed in the group of women who delivered in 2008-2010 and was validated in the group of women who delivered in 2011-2013 showed reasonably good fit in both groups (**Figure S5**).

The selected multivariable model (**Table 1**) was fitted in the POP study and the estimated odds of intrapartum emergency CD was exp(-1.665 + 0.3174*age - 0.5299*height - 0.2999*(1/BMI) + 0.2077*weightgain + 0.3881*EFW). The output of logistic regression models was expressed in the main manuscript as coefficients rather than odds ratios, as odds ratios cannot be easily interpreted for common outcomes, such as caesarean delivery. However, we have provided the odds ratios in **Table S1**.

Regression coefficients were multiplied by a shrinkage factor s=0.9812 in the Scottish validation data. The selected model discriminated equally well in the validation data as in the POP study: C statistic was 0.71 (**Figure S6, Table S2**). The observed risks of intrapartum emergency CD in the screen positive women were 44% in the imputed validation data and 48% in the POP study (**Table S2**). Model calibration was very good in low risk women in the validation data, but there was a gradual downward deviation of observed probabilities of intrapartum emergency CD from the predicted probabilities towards higher predicted risks (**Figure S7**).

The actual proportions of emergency CD were plotted against predictor variables in the POP study and in the Scottish validation data, divided in quintiles calculated in the POP study (**Figure S8**). The trends were similar in both cohorts for the three predictors observed in both data sets. The imputed EFW had a similar trend and the imputed weight gain had a slightly weaker trend with emergency CD in the Scottish data than in the POP study cohort. In addition, the association between EFW and the observed probability of emergency CD was described using Kernel-weighted local polynomial smoothing and Lowess smoothing in the POP study (**Figure S9**). The plot illustrates the statistically non-significant suggestion towards a quadratic association which was identified also in the univariable fractional polynomial analysis.

Predictive ability of the model in excluded women

The aim of the study was to determine whether we could predict the risk of emergency CD in low risk women, hence, we had excluded a total of 308 women who had pre-existing diabetes or hypertension, or gestational hypertension, pre-eclampsia or gestational diabetes diagnosed before their 36 week research visit, but who otherwise fulfilled the eligibility criteria. We subsequently assessed its predictive ability in these high risk women who were excluded. Coefficients from the POP study model were multiplied by the shrinkage factor of 0.9812. A total of 42 (14%) these women had a ≥40% risk of emergency CD and their actual rate of emergency CD was 62%. The rate of emergency CD in the 266 women with a <40% predicted risk was 22%.

Maternal and neonatal complications in categories of predicted risk

The risk of assisted vaginal delivery was highest in women who had at least 30% predicted risk of emergency CD (**Table S3**). Also, the risk of postpartum hemorrhage was already elevated (14%) when the risk of emergency CD was 30 to <40%, and it increased from to

21% in women who had ≥40% risk of emergency CD. The risk of any neonatal morbidity increased 1%-unit by category (from 5% to 8%) when the predicted risk of emergency CD was <40% and it increased to 12% in the group with ≥40% predicted risk. A clear elevation in the risk of metabolic acidosis and severe neonatal morbidity at term was also observed at the point when the risk of emergency CD increased to ≥40%.

Table S1. Odds Ratios (OR) with 95% CI from univariable and multivariable models, n=3,047.

		Model		
Predictor variable	*Univariable models*	*Full model*	*Selected multivariable model*	*Selected model excluding US*
Age	1.32 (1.20 to 1.45)	1.37 (1.24 to 1.52)	1.37 (1.24 to 1.52)	1.40 (1.27 to 1.55)
Height	0.66 (0.60 to 0.72)	0.59 (0.53 to 0.66)	0.59 (0.53 to 0.65)	0.63 (0.57 to 0.70)
1/BMI	0.69 (0.63 to 0.76)	0.74 (0.67 to 0.82)	0.74 (0.67 to 0.82)	0.69 (0.62 to 0.76)
Weight gain	1.19 (1.09 to 1.30)	1.23 (1.12 to 1.35)	1.23 (1.12 to 1.35)	1.31 (1.20 to 1.44)
Fetal sex: female	0.88 (0.73 to 1.05)	0.91 (0.75 to 1.10)	N/A	N/A
EFW at 36 week scan	1.48 (1.35 to 1.62)	1.42 (1.27 to 1.60)	1.47 (1.34 to 1.63)	N/A
AC growth velocity	1.34 (1.22 to 1.47)	1.07 (0.95 to 1.20)	N/A	N/A

All variables except fetal sex are expressed as z scores, adjusted for gestational age at measurement where appropriate. BMI denotes body mass index, AC denotes abdominal circumference, EFW denotes estimated fetal weight, US denotes ultrasonic assessment.

Table S2. Diagnostic effectiveness of the selected model using predicted risk cut-off 40% in the POP study and the Scottish data.

Diagnostic effectiveness measure	*POP study*	*Scottish data*
Sensitivity (%)	16	18
Specificity (%)	96	95
Positive predictive value (%)	48	44
Negative predictive value (%)	83	85
False positive rate (%)	4.0	4.5
False negative rate (%)	84	82
Positive likelihood ratio	4.0	4.0
Negative likelihood ratio	0.88	0.86
Area under the ROC curve*	0.71	0.71

*Calculated using the continuous linear predictor for intrapartum emergency caesarean delivery. The model includes maternal age, height, 1/BMI, maternal weight gain and estimated fetal weight. The model was developed in the POP study and validated in the Scottish data. Maternal weight gain and estimated fetal weight were imputed in the validation data. Coefficients from the model developed in the POP study have been multiplied by the shrinkage factor of 0.9812 in the calculation of predicted risks in the Scottish data.

Table S3. Maternal and neonatal complications by category of predicted risk of intrapartum emergency caesarean delivery (CD) in the POP study (n=3,047)**.**

Predicted risk of emergency CD	**<10%**	**10 to <20%**	**20 to <30%**	**30 to <40%**	**≥40%**
Outcome	**n/N**	**n/N**	**n/N**	**n/N**	**n/N**
Emergency CD	52/782 (7%)	184/1151 (16%)	136/630 (22%)	107/295 (36%)	90/189 (48%)
Assisted vaginal delivery*	195/730 (27%)	301/967 (31%)	167/494 (34%)	96/188 (51%)	41/99 (41%)
Postpartum hemorrhage	50/782 (6%)	88/1151 (8%)	59/630 (9%)	40/295 (14%)	39/189 (21%)
Metabolic acidosis	5/782 (0.6%)	10/1151 (0.9%)	4/630 (0.6%)	4/295 (1%)	6/189 (3%)
5 minute Apgar <7	5/782 (0.6%)	6/1151 (0.5%)	6/630 (1%)	4/295 (1%)	4/189 (2%)
Neonatal unit admission at term	37/782 (5%)	55/1151 (5%)	42/630 (7%)	18/295 (6%)	18/189 (10%)
Any neonatal morbidity	42/782 (5%)	65/1151 (6%)	47/630 (7%)	25/295 (8%)	22/189 (12%)
Severe neonatal morbidity at term	5/782 (0.6%)	7/1151 (0.6%)	3/630 (0.5%)	0/295 (0.0%)	4/189 (2%)

*N is restricted to the vaginal deliveries (N=2,478).

CD denotes caesarean delivery. Postpartum hemorrhage is defined as Hb<8 g/dL or estimated blood loss ≥ 1000 mL. Hb <8 g/dL within a week from delivery but >10 g/dL antepartum (before the day of delivery). Any neonatal morbidity is defined as ≥1 of the following: (1) metabolic acidosis, defined as cord blood pH <7.1 and base deficit of >10mmol/L, (2) 5 minute Apgar <7, (3) neonatal unit admission within 48 hours from birth at term for at least 48 hours. Severe neonatal morbidity is defined as ≥1 of the following: (1) neonatal death, (2) hypoxic ischemic encephalopathy, (3) use of inotropes, (4) mechanical ventilation, (5) severe metabolic acidosis, defined as pH<7.0 and a base deficit of >12mmol/L at term.

Supplementary Figure Legends

Figure S1. Flow chart of the study cohort.

Figure S2. Calibration plot for the final model (n=3,047) in the POP study.

Figure S3. Calibration plots for the final model stratified by clinically indicated scans in the POP study. A. No clinically indicated scans at ≥28 weeks and blinded research scans at 28 and 36 weeks (n=2,041). B. At least one clinically indicated scan at ≥28weeks or research scan result revealed at 28 or 36 weeks (n=1,006).

Figure S4. Receiver operating characteristic curve for the selected model in the POP study. Area under the ROC curve (95% confidence interval) is 0.706 (0.682-0.729). Predicted risks of 10%, 20%, 30% and 40% are displayed on the ROC curve.

Figure S5. Temporal validation of the final model within the POP study. A. Calibration plot in the model development dataset from 2008-2010 (n=1,436). B. Calibration plot in the temporal model validation dataset from 2011-2013 (n=1,611).

Figure S6. Receiver operating characteristic curve for the selected model with a simple shrinkage in the imputed external validation dataset. Area under the ROC curve (95% confidence interval) is 0.709 (0.697-0.721). Predicted risks of 10%, 20%, 30% and 40% are displayed on the ROC curve.

Figure S7. Calibration plot for the selected model with a simple shrinkage in the external validation dataset.

Figure S8. Univariable associations between predictors and emergency CD in both the POP study and Scottish data. Data are plotted for the 5 quintiles of the given measurement, with the median value within the quintile for the variable on the X axis and the proportion of women delivered by emergency caesarean on the Y axis. The overall proportion of emergency CD was 18.7% in the POP study and 16.6% in the Scottish data. Weight gain and estimated fetal weight were imputed using predictive mean matching (m=10 imputations, k=10 donors). EMCD denotes emergency caesarean delivery, POPs denotes Pregnancy Outcome Prediction study, SMR 02 denotes Scottish Morbidity Record 02, BMI denotes body mass index, EFW denotes estimated fetal weight, CI denotes confidence interval.

Figure S9. Univariable association between EFW and emergency CD in the POP study using Kernel-weighted local polynomial smoothing (dashed line) and Lowess smoothing (solid line). EMCD denotes emergency caesarean delivery and EFW denotes estimated fetal weight.

Figure S1.

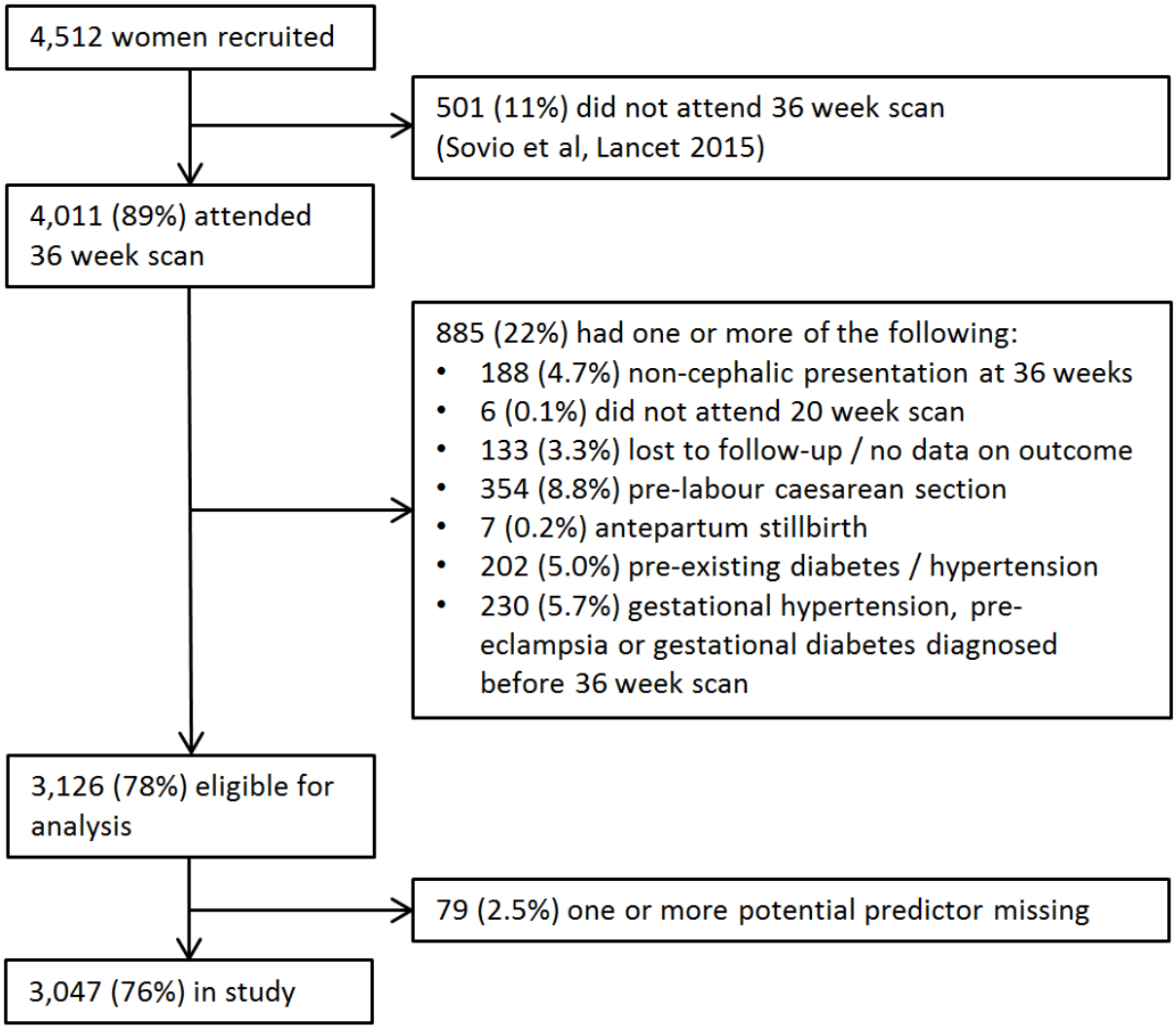

4,512 women recruited
501 (11%) did not attend 36 week scan
(Sovio et al, Lancet 2015)
4,011 (89%) attended
36 week scan
885 (22%) had one or more of the following:
• 188 (4.7%) non-cephalic presentation at 36 weeks
• 6 (0.1%) did not attend 20 week scan
• 133 (3.3%) lost to follow-up / no data on outcome
• 354 (8.8%) pre-labour caesarean section
• 7 (0.2%) antepartum stillbirth
• 202 (5.0%) pre-existing diabetes / hypertension
• 230 (5.7%) gestational hypertension, pre-eclampsia or gestational diabetes diagnosed before 36 week scan
3,126 (78%) eligible for analysis
79 (2.5%) one or more potential predictor missing
3,047 (76%) in study

Figure S2.

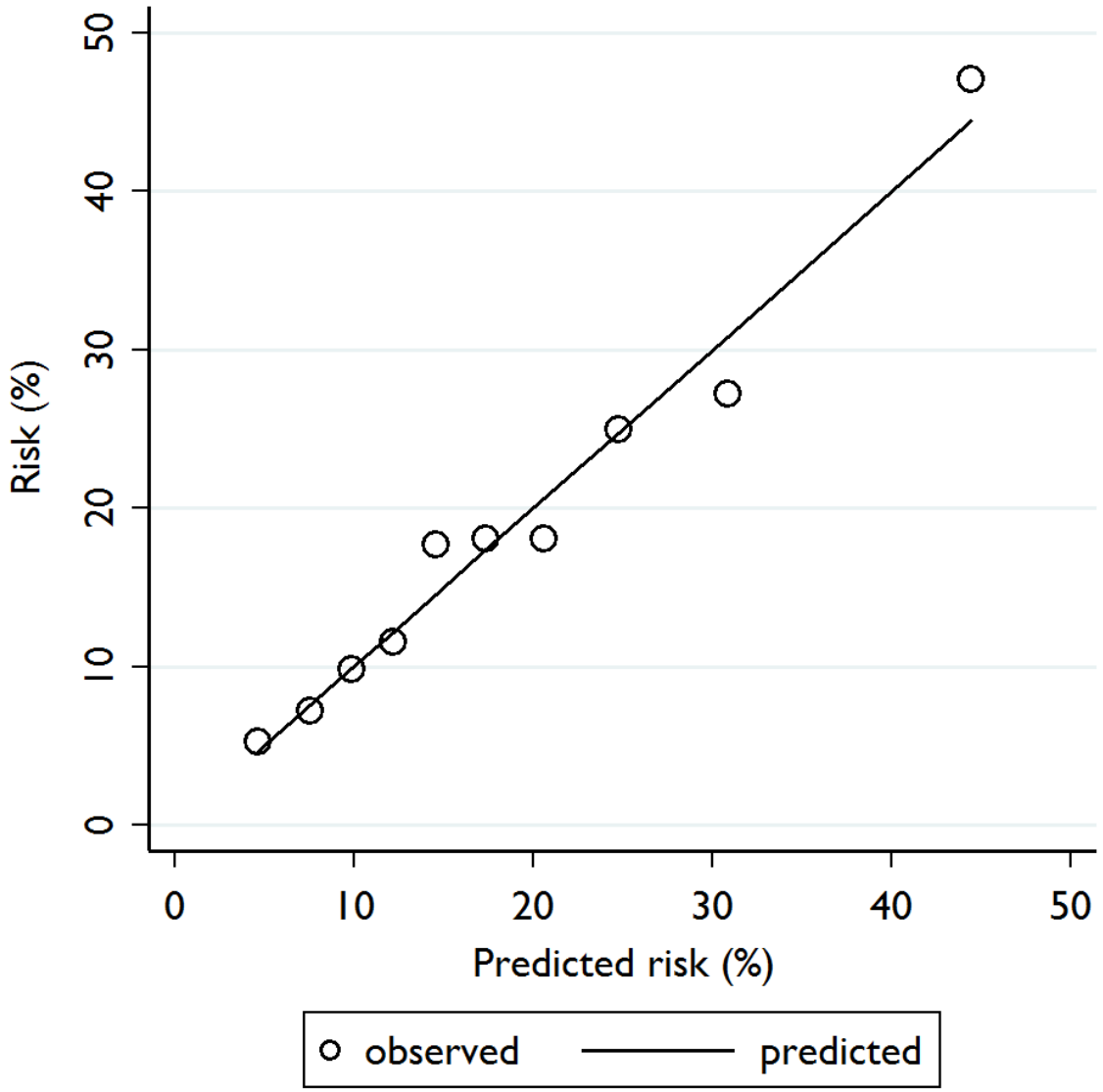

Risk (%)
0
10
20
30
40
50
0
10
20
30
40
50
Predicted risk (%)
observed
predicted

Figure S3 A.

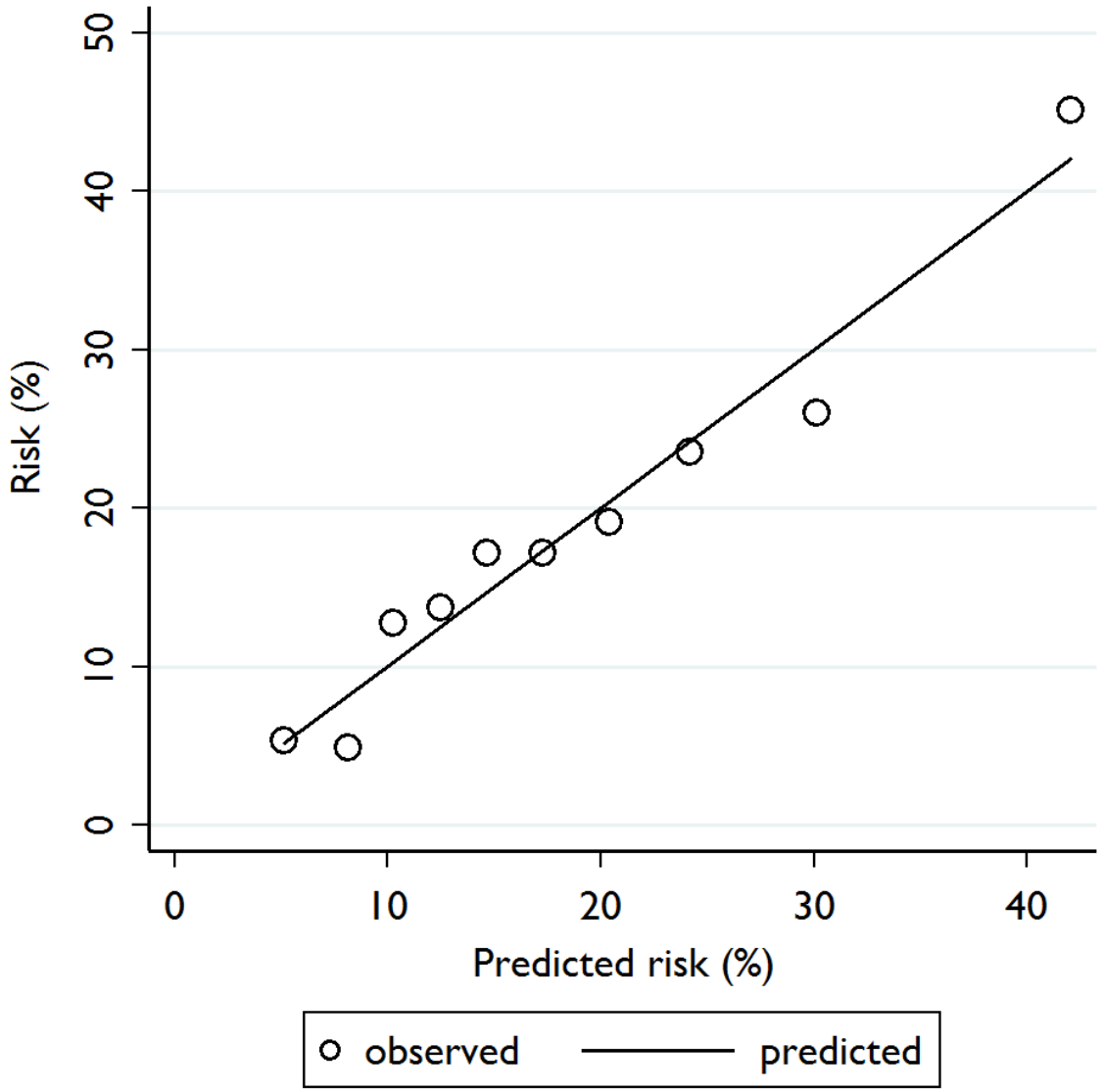

Risk (%)
0
10
20
30
40
50
0
10
20
30
40
Predicted risk (%)
observed
predicted

Figure S3 B.

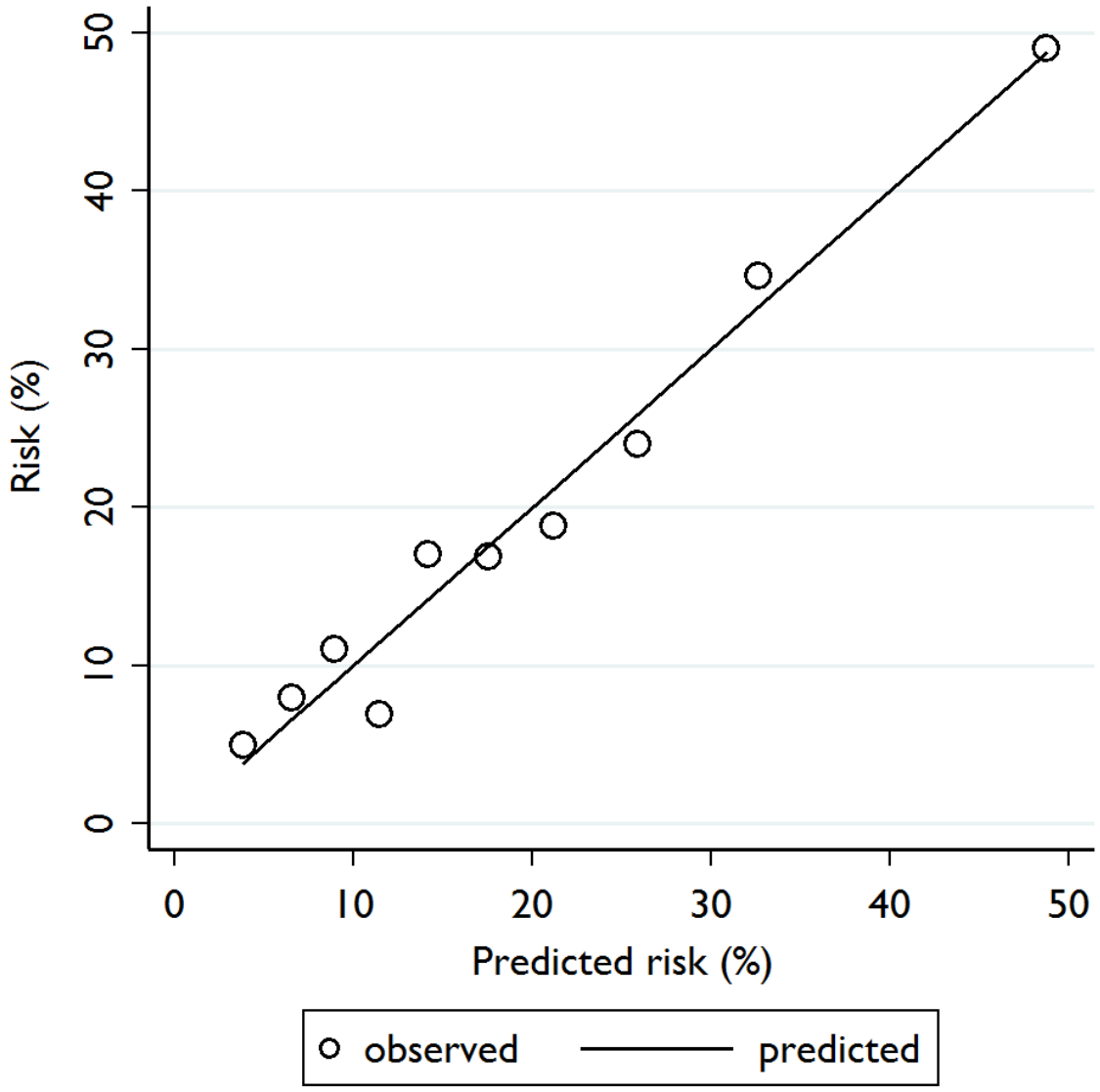

Risk (%)
0
10
20
30
40
50
0
10
20
30
40
50
Predicted risk (%)
observed
predicted

Figure S4.

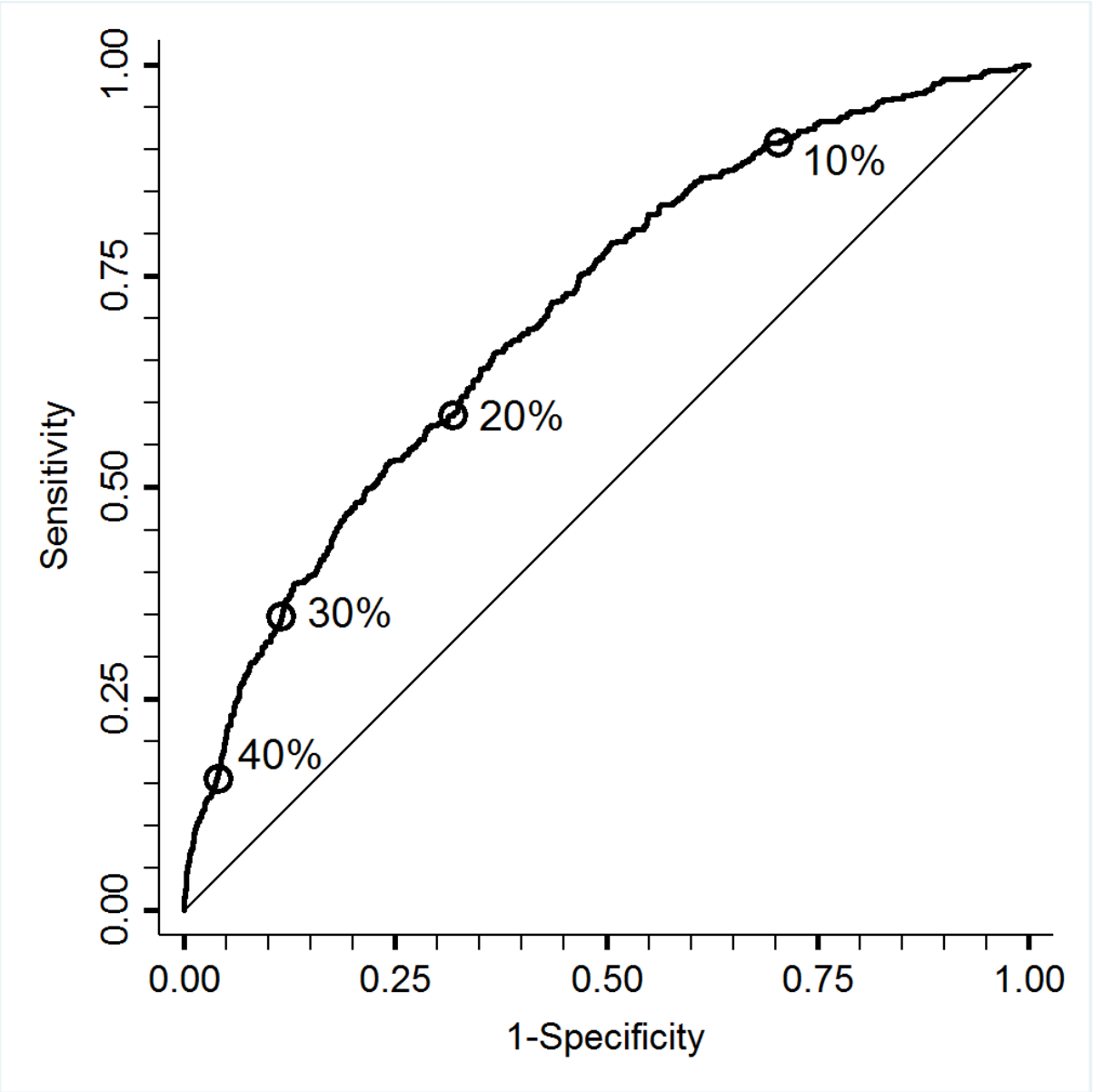


Figure S5 A.

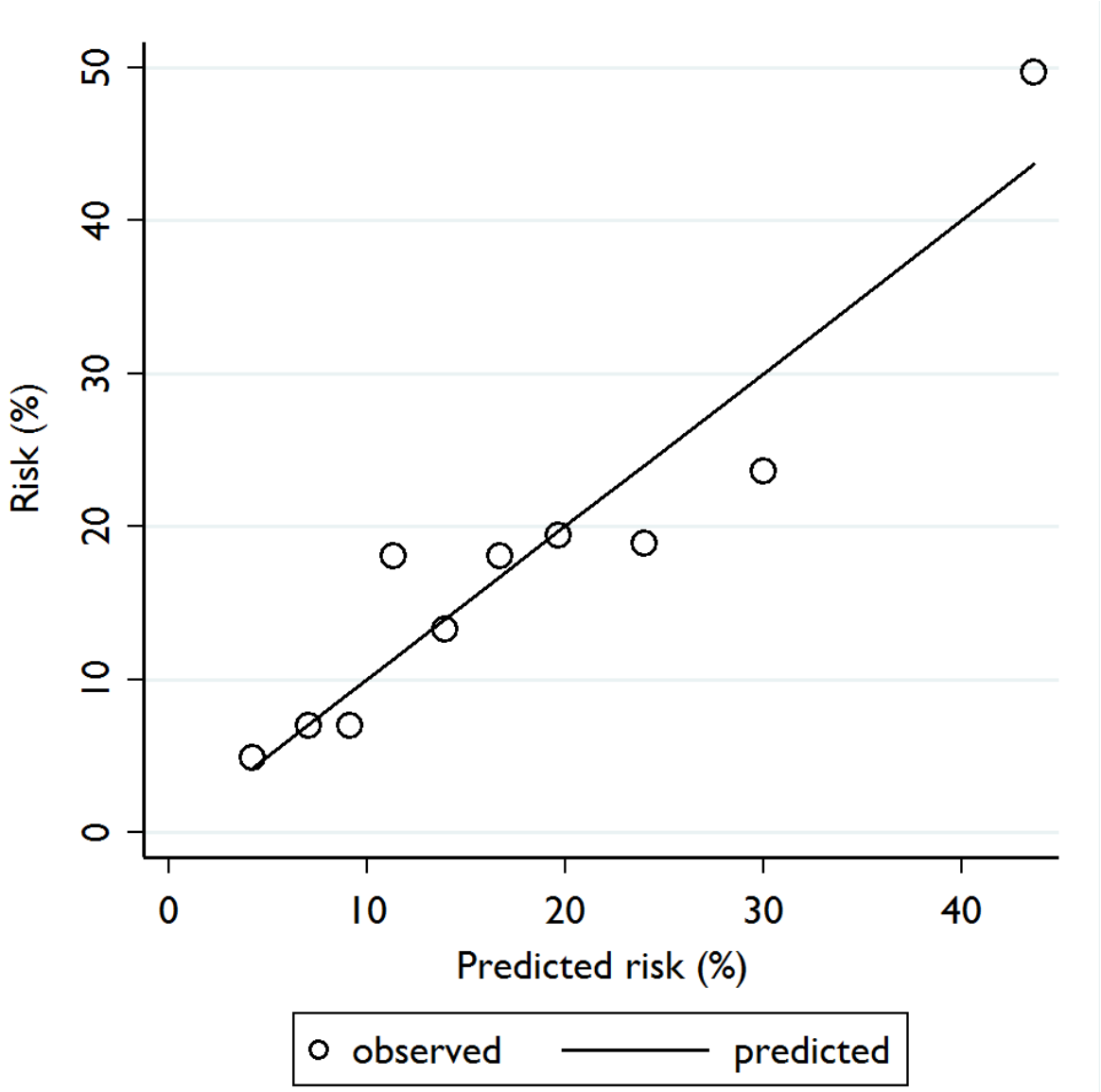

Risk (%)
0
10
20
30
40
50
0
10
20
30
40
Predicted risk (%)
observed
predicted

Figure S5 B.

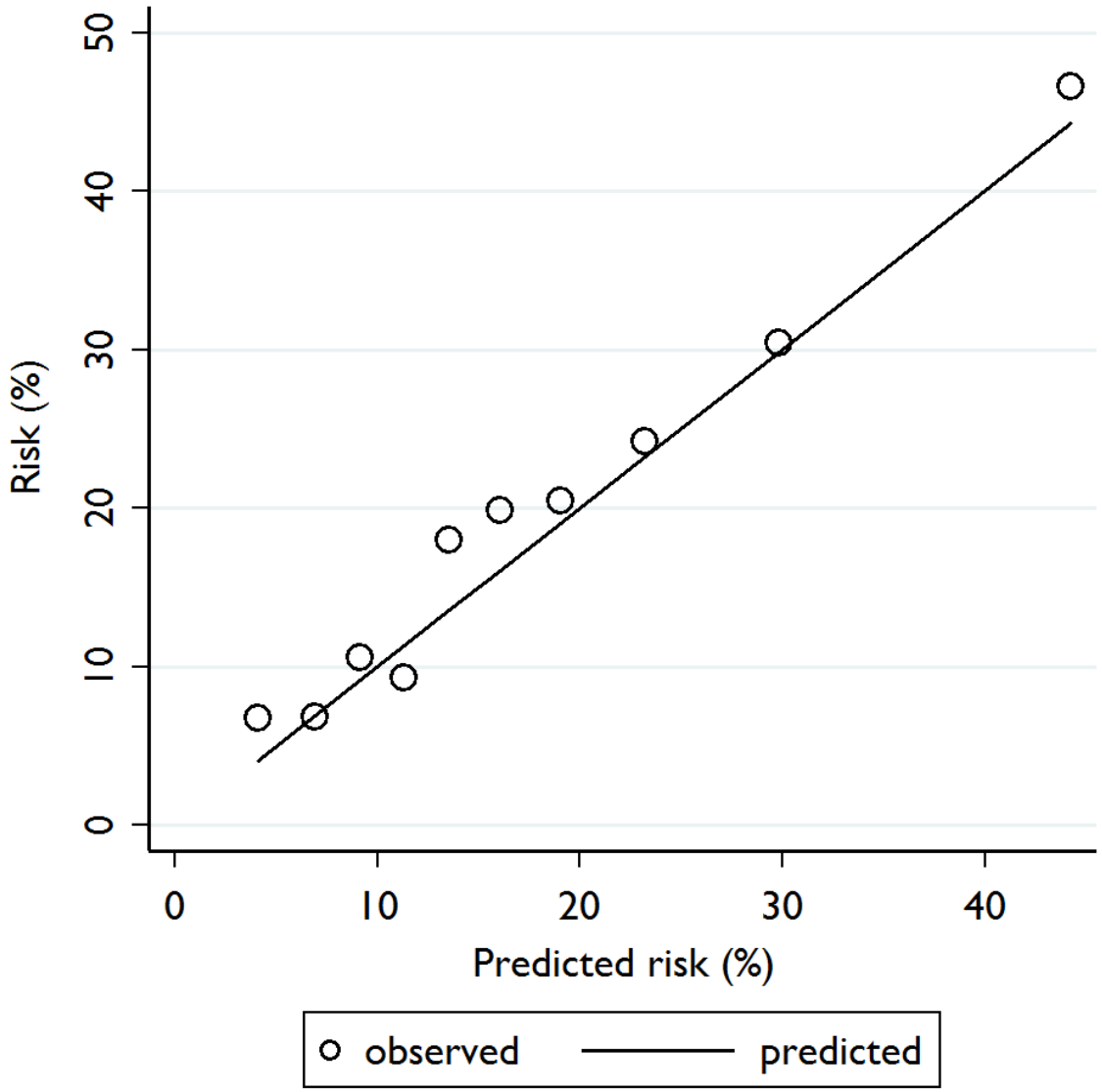

Risk (%)
0
10
20
30
40
50
0
10
20
30
40
Predicted risk (%)
observed
predicted

Figure S6.

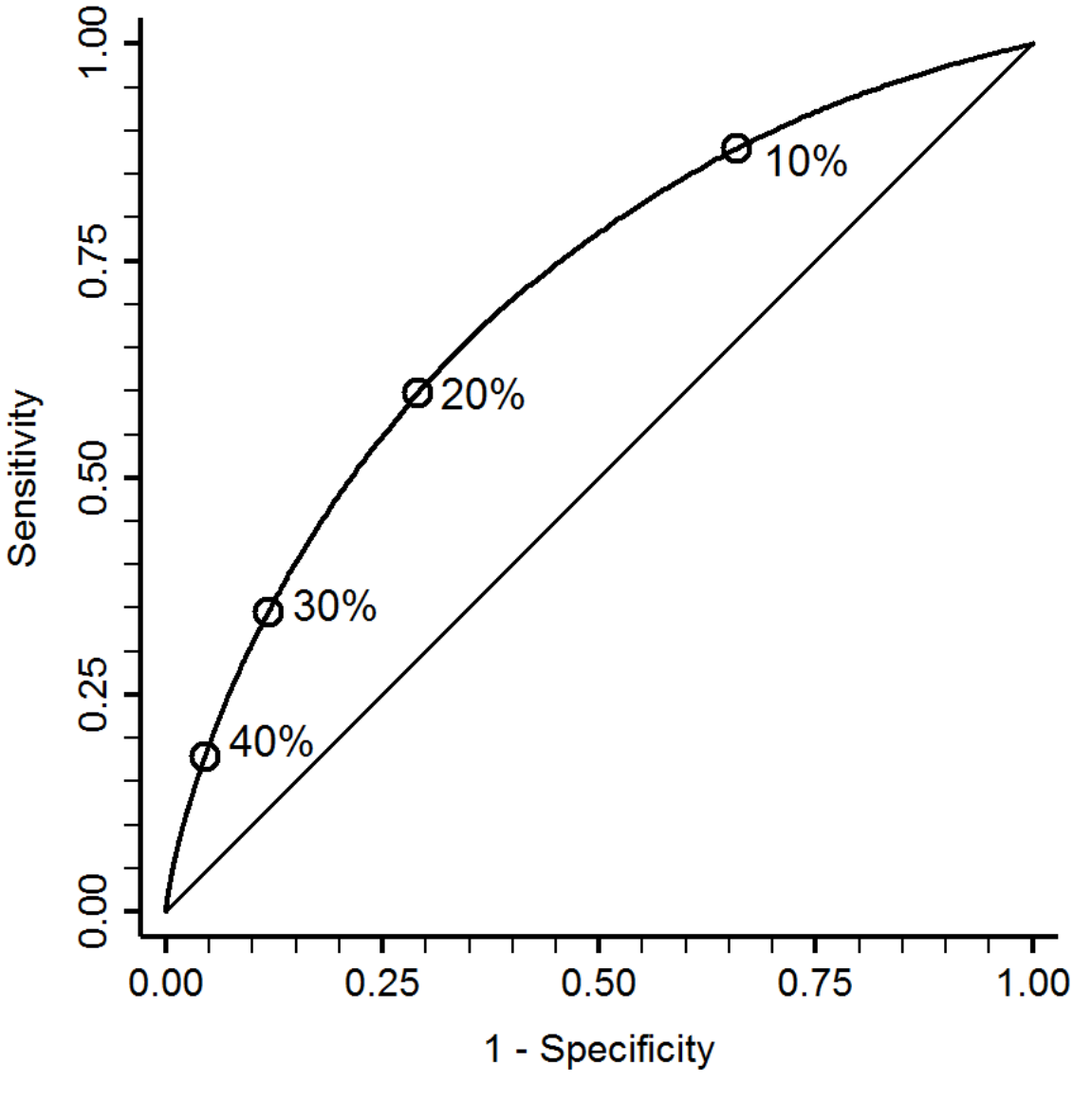

Sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 - Specificity
10%
20%
30%
40%

Figure S7.

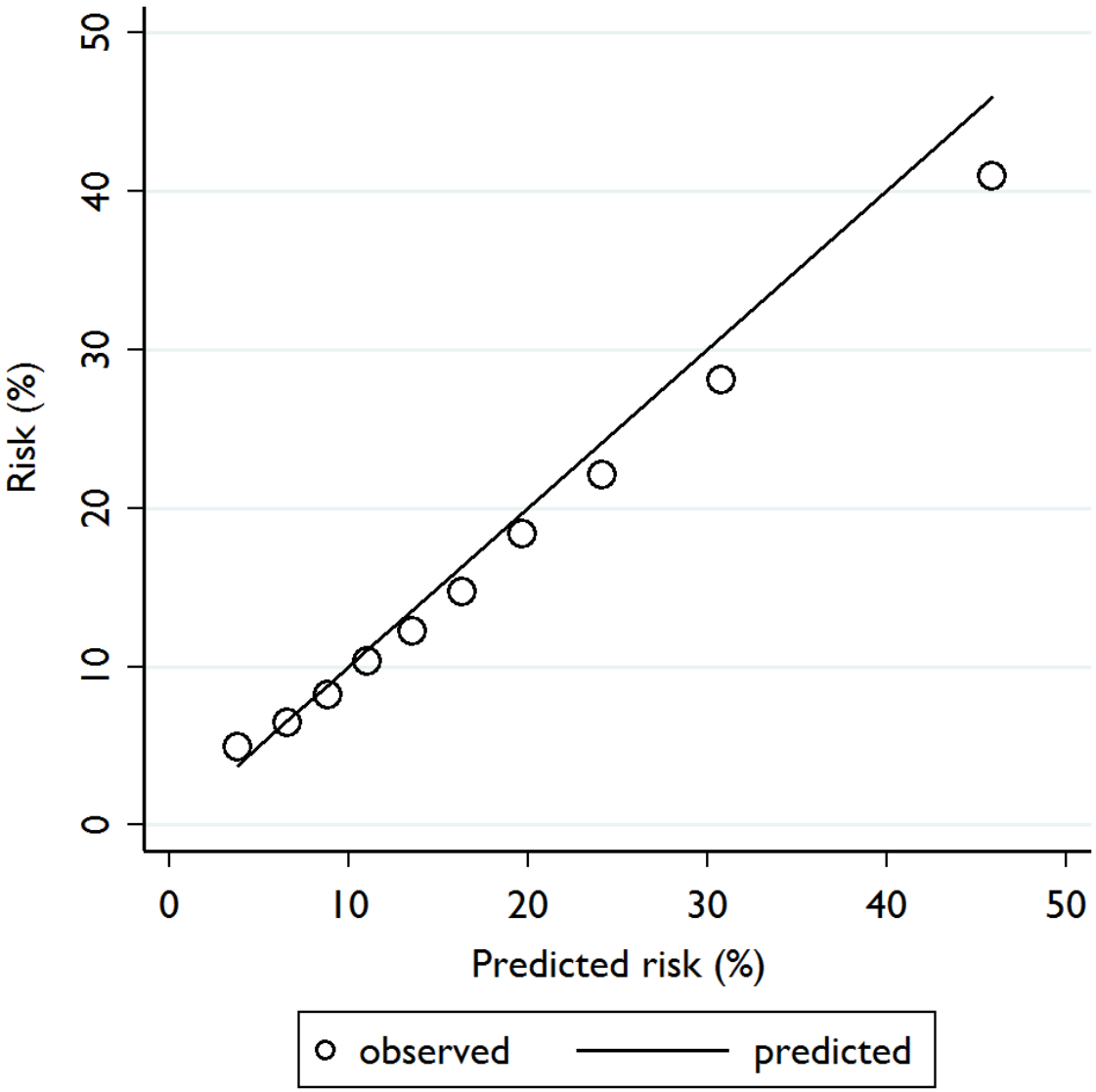

Risk (%)
0
10
20
30
40
50
Predicted risk (%)
0
10
20
30
40
50
observed
predicted

Figure S8.

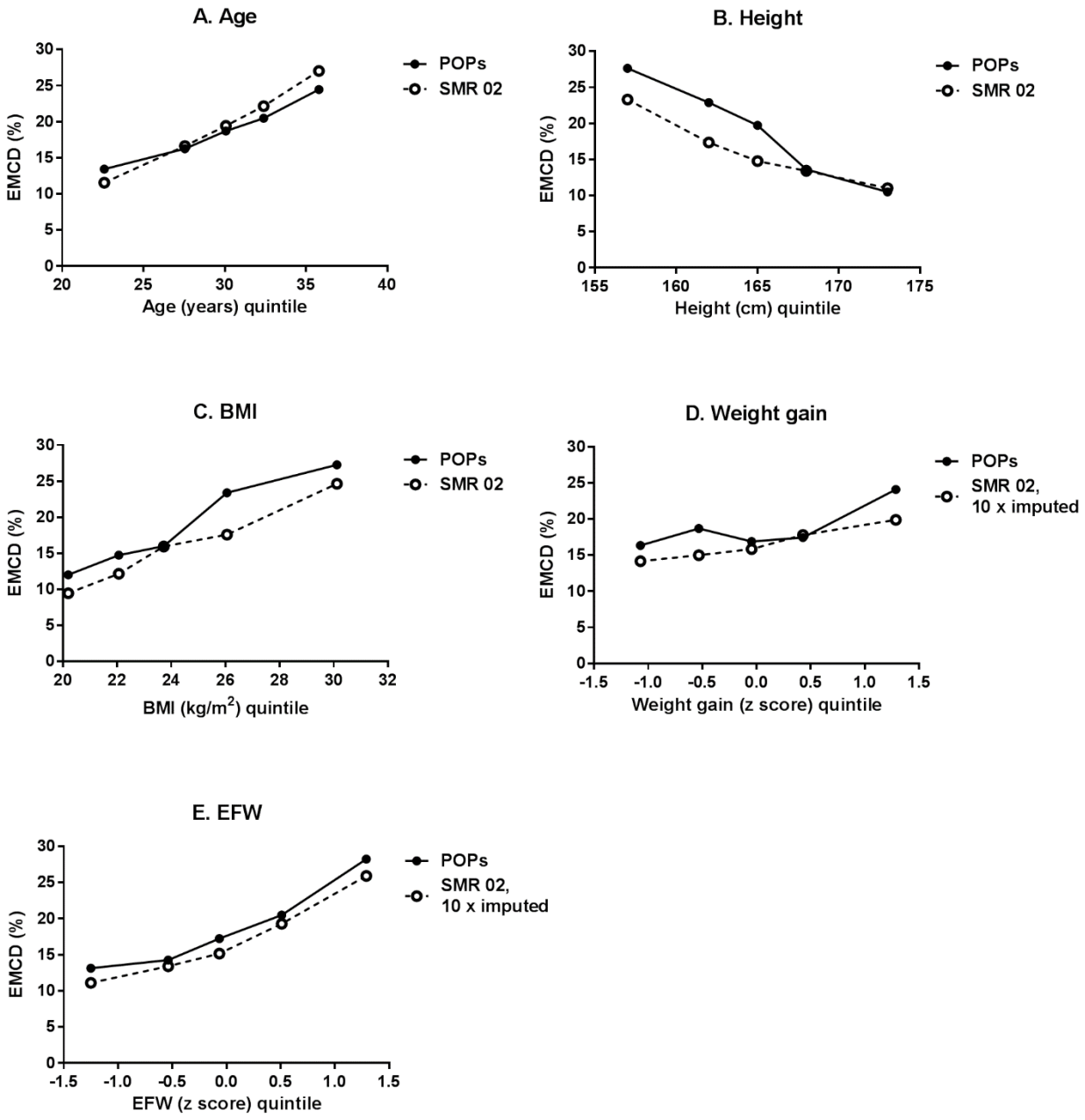
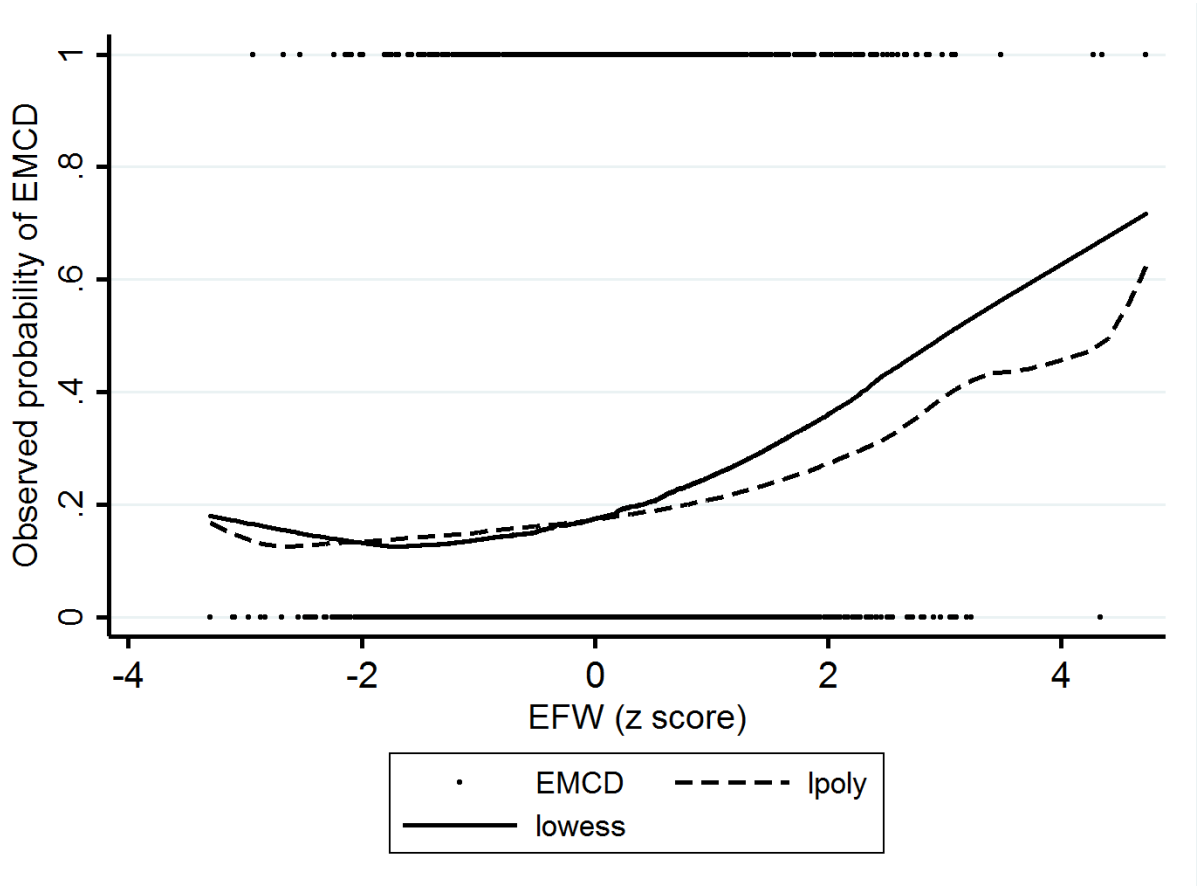

A. Age
EMCD (%)
Age (years) quintile
POPs
SMR 02
B. Height
Height (cm) quintile
C. BMI
BMI (kg/m2) quintile
D. Weight gain
Weight gain (z score) quintile
SMR 02,
10 x imputed
E. EFW
EFW (z score) quintile

Figure S9.


Observed probability of EMCD
0
.2
.4
.6
.8
1
-4
-2
0
2
4
EFW (z score)
EMCD
lpoly
lowess

Appendix S2. Emergency caesarean section risk calculator.

Please fill in the green boxes to get a risk estimate from the calculator.

POP study (n=3047):

Distributions	**age (y)**	**height (cm)**	**1/BMI (m^2/kg)**
Mean	29.70	165.5	0.04166
SD	4.956	6.310	0.006394

Model: logit	**constant**	**age**	**height**	**1/BMI**	**weightgain**	**EFW**	**Shrinkage factor:**
Beta	-1.66485	0.3173724	-0.5298984	-0.2999465	0.2077034	0.388135	0.9812

Inputs: maternal age, height, BMI at dating scan, weight gain from dating scan to 36 week scan, EFW, EDD and date of 36 week scan.

variable	**value**	**z-score**
age (y)	37	1.473
height (cm)	165	-0.07214
BMI (kg/m^2)	30	
1/BMI (m^2/kg)	0.0333	-1.303
weightgain (kg)	13	0.2037
EFW (g)	3400	1.9545

EDD (dd/mm/yyyy): 30/10/2015

Date (dd/mm/yyyy): 03/10/2015

Calculated GA (weeks): 36.14

mean: 12.19 **SD:** 3.993

mean: 2741 **SD:** 337.1

The above equations for mean and SD were estimated within the POP study.

Odds(EMCS) 1.00

Risk(EMCS) 50%

Abbreviations

BMI: body mass index

EDD: estimated date of delivery

EFW: estimated fetal weight

EMCS: emergency caesarean section

GA: gestational age

POP: Pregnancy Outcome Prediction

SD: standard deviation