Universal ultrasound screening in the third trimester

Alexandros A. Moraitis

Gonville and Caius College, University of Cambridge

This thesis is submitted for the degree of Doctor of Philosophy

(September 2021)

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the text.

I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the text.

It does not exceed the prescribed word limit of the Degree Committee of Clinical Medicine and Clinical Veterinary Medicine (60,000 words)

Abstract

Universal Ultrasound Screening in the Third Trimester

Alexandros A. Moraitis

Background

Pregnant women are currently offered two ultrasound scans, one at booking (around 12 weeks' gestation) and one at around 20 weeks' gestation. No further scans are offered unless there are clinical indications. Ultrasound has an important role in the management of high-risk pregnancies. However, there is no clear evidence that it is effective in screening low risk and unselected women. The majority of complications, such as stillbirth and shoulder dystocia occur in low-risk pregnancies, first because most pregnancies are classified as low-risk and second, possibly due to inadequate screening. An effective ultrasound screening programme in late pregnancy combined with an intervention, like induction of labour, for the screen positives could potentially improve pregnancy outcomes. However, the diagnostic accuracy of many ultrasonic features is unknown in low-risk populations and there is a possibility of iatrogenic harm by intervening when it is not necessary.

Objectives

1. To assess the diagnostic effectiveness of late pregnancy ultrasound in nulliparous women based on the existing research literature.
2. To analyse the prospective cohort study, Pregnancy Outcome Prediction Study, for the above ultrasound findings and combine the results with the meta-analyses.
3. Finally, use the results to provide inputs for health economic analyses of the cost-effectiveness of universal ultrasound screening and assess the need, potential design, and acceptability of a future randomised controlled trial.

Methods

The following key ultrasound measurements were identified which might be used in late pregnancy screening: (i) suspected small for gestational age (SGA), (ii) suspected large for gestational age (LGA), (iii) high resistance pattern of umbilical artery Doppler flow velocimetry, (iv) low cerebro-placental ratio (CPR), (v) severe oligohydramnios, (vi) borderline oligohydramnios. I found that there was an on-going Cochrane Diagnostic Test Accuracy review for SGA, hence I focused on the other five measures. The protocol was registered with the PROSPERO register of systematic reviews (CRD42017064093).

Medline, EMBASE, Clinical Trials.gov and the Cochrane library were searched from inception. Studies that performed an ultrasound scan ≥24 weeks of gestational age in unselected, low or mixed risk populations were included, excluding studies which only included high risk pregnancies. The risk of bias in each included study was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS 2) tool. Meta-analysis was performed using the hierarchal summary receiver operating characteristic curve (HSROC) analysis and bivariate logit-normal models. I also performed new analyses on previously unpublished data of the Pregnancy Outcome Prediction (POP) study which was one of the few studies that blinded ultrasound scan results to the clinicians.

Results

41 studies of LGA met our inclusion criteria involving 112,034 patients in total. Ultrasonic suspicion of fetal macrosomia was strongly predictive of the risk of delivering a large baby with the positive LRs (LR+) ranging from 7 to 12. However, it was only weakly predictive of the risk of shoulder dystocia with LR+ around 2. 13 studies of umbilical artery (UA) Doppler that met our inclusion criteria including 67,764 patients in total. UA Doppler had weak/moderate predictive accuracy for detecting SGA and severely SGA (<3rd percentile) infants (LR+ between 2.5 and 3.0). However, it did not predict neonatal morbidity at term. The results were very similar in both the POP study and the meta-analysis (which included the POP study) with the only notable difference being that the association with severe SGA in the POP study was slightly stronger. 16 studies of CPR met the inclusion criteria involving 121,607 patients in total. CPR may be slightly more predictive than UA Doppler in identifying pregnancies at an increased risk of adverse outcome. In the case of SGA, the positive LRs were in the region of 3.5 to 4.0. Moreover, unlike UA Doppler, a low level of CPR was associated with an increased risk of neonatal morbidity. However, the association with morbidity was weaker with positive LRs of <2.0. Furthermore, in both analyses, there was very significant heterogeneity in relation to both SGA and neonatal morbidity. 14 studies of severe oligohydramnios that met our inclusion criteria involving 109,679 patients in total. Diagnosis of severe oligohydramnios was associated with a positive LR for SGA of between 2.5 and 3.0. It was also associated with positive LRs for admission to NICU and emergency caesarean section for fetal distress of between 1.5 and 2.5. However, the study quality was variable and only two studies containing <5% of the patients included in the meta-analysis blinded the results of the scan. 11 studies of borderline oligohydramnios (including the POP study) met our inclusion criteria involving 37,848 patients in total. Borderline oligohydramnios was weakly/moderately predictive of SGA (positive LRs 2.5 to 3.0). This was observed in the meta-analysis of multiple studies of variable quality. There was also a comparable association between borderline

oligohydramnios and severe SGA in the only study where the scan result was blinded, the POP study. Finally, by analysing of the POP cohort We identified the 4.6% of women who had a breech presentation, and for more than half of these, it had not previously been clinically suspected. Most of these women were delivered by planned Caesarean section. No woman in the cohort had a vaginal breech delivery or experienced an intrapartum Caesarean for undiagnosed breech. An introduction of a policy of third trimester ultrasound for fetal presentation would prevent about 5000 emergency Caesarean sections and 8 perinatal deaths annually in the UK. The policy would be cost-effective at a cost of £19.80 per scan.

Conclusion

There is a strong clinical and health economic case for implementing late pregnancy ultrasound screening to assess fetal presentation. Universal ultrasound screening for macrosomia would increase the detection of LGA infants at birth but is unlikely to increase the detection of shoulder dystocia or associated neonatal morbidity in a clinically significant way. Umbilical artery Doppler, CPR, severe oligohydramnios, and borderline oligohydramnios were all weakly predictive of the risk of delivering an SGA infant but either non-predictive or weakly predictive of the risk of neonatal morbidity. They should not be used alone to screen for neonatal morbidity, however a positive result would justify further fetal monitoring due to the association of all above markers with SGA.

Acknowledgements

I would like to thank my supervisor Professor Gordon Smith for all his help, guidance, inspiration, and support in writing this thesis.

I was lucky to collaborate with multiple Principal Investigators throughout the UK including Prof Peter Brocklehurst, Prof Alexander Heazell, Prof Jim Thornton, Prof Aris Papageorghiou, and Prof Stephen Robson. I was also very lucky to have the help of Dr Ulla Sovio with statistics.

I would like to thank the health economists David Wastlund (DW) and Ed Wilson for the productive collaboration.

I would also like to thank my colleagues that I collaborated for the systematic reviews Dr Norman Shreeve (NS), Dr Tom Bainton (TB), Dr Ilianna Armata (IA) and Dexter Hayes (DH).

I would like to thank Dr John Latimer for his help and support with student matters in the college.

Finally, I would like to thank my family for their love and support. My wife Derina who has been extremely supportive throughout this entire process and has made countless sacrifices to help me get to this point, my children Leo and Zoe for understanding why daddy was not always available to play, and my parents Apostolos and Dimitra for their love and support throughout my life.

Table of Contents

List of Tables

List of Figures

List of abbreviations

AFI: Amniotic fluid index

AGA: Appropriate for gestational age

CI: Confidence interval

CP: Cerebral palsy

CPR: Cerebro-placental ratio

CS: Caesarean section

DOR: Diagnostic odds ratio

ECV: External cephalic version

EFW: Estimated fetal weight

EVPI: Expected value of perfect information

EVPPI: Expected value of partial perfect information

FGR: Fetal growth restriction

HTA: Health Technology assessment

HSROC: Hierarchal summary receiver operating characteristic

ICER: Incremental cost-effectiveness ratio

IOL: Induction of labour

LGA: Large for gestational age

LR: Likelihood ratio

OR: Odds ratio

PI: Pulsatility index

PPI: Patient and public involvement

PPV: Positive predictive value

QALY: Quality-adjusted life year

QUADAS: Quality Assessment of Diagnostic Accuracy Studies

RCT: Randomised controlled trial

RR: Relative risk

SEN: Special educational needs

SGA: Small for gestational age

UA: Umbilical artery

US: Ultrasound

VoI: Value of information

wkGA: Weeks of gestational age

WTP: Willingness-to-

Chapter 1. Background.

The current approach to antenatal screening in the UK

Screening is the process which helps identify individuals that are at high risk for certain complications or diseases. Pregnant women in the UK are currently screened for several conditions at different timings during pregnancy.[1]

In the first trimester they are offered screening for trisomies based on maternal history, ultrasound findings such as the nuchal translucency and biochemical markers including the pregnancy-associated plasma protein A (PAPP-A) and free beta subunit of human chorionic gonadotrophin (fβ-hCG) [(https://www.gov.uk/topic/population-screening-programmes/fetal-anomaly]. They are also screened for the risk of deep vein thrombosis (DVT) based on maternal past medical history (e.g. previous DVT) and maternal characteristics (e.g. obesity). Finally, they have an initial screening for pregnancy related conditions such as pre-eclampsia (PET), gestational diabetes, fetal growth restriction (FGR), and preterm birth based on maternal risk factors, past obstetric history, and biochemical markers (PAPP-A for the risk of FGR). When women are classified as high risk for PET or FGR, they are prescribed aspirin and are offered additional ultrasound scans in the second and third trimester.

In the second trimester, women are routinely screened for fetal anomalies and placenta localisation with ultrasound between ≥18 weeks of gestational age (wkGA) and <21wkGA, (https://www.gov.uk/topic/population-screening-programmes/fetal-anomaly). A positive result from this scan might inform decisions around termination of pregnancy (e.g. many women would choose to terminate a pregnancy where the fetus had a severe anomaly) or it might inform the need for targeted follow up and changes to the perinatal care of the infant. For example, identification of a congenital diaphragmatic hernia could lead to invasive testing for aneuploidy, prenatal discussions with the paediatric surgery team and modification to neonatal resuscitation (such as early intubation to avoid expansion of the stomach with air).

Pregnant women are screened for gestational diabetes between 24 and 28 weeks' gestation. They are also offered midwife visits to screen for pre-eclampsia by checking the blood pressure and proteinuria, for breech presentation by abdominal palpation, and for fetal growth using the symphysis-fundal height.

Ultrasound in the third trimester is currently offered in a targeted manner where there is a clinical indication. Such indications could include relevant medical history (e.g. anti-phospholipid antibody syndrome), relevant past obstetric history (e.g. previous fetal growth restriction [FGR]), presentation with symptoms (e.g. antepartum haemorrhage), suspected malpresentation, or abnormal symphysis-fundal height measurements.

The rationale for universal third trimester ultrasound screening.

Ultrasound screening in the third trimester can identify abnormalities in fetal growth and placental function, malpresentation, and previously undiagnosed fetal anomalies or placenta praevia. Abnormalities in fetal growth (including both fetal growth restriction and macrosomia) may account for up to 1 in 3 antepartum stillbirths and 1 in 6 delivery related perinatal deaths (including intrapartum stillbirths and early neonatal deaths) at term.[2] More specifically, the risk of antepartum stillbirth is 10-times higher infants born below the 3rd birthweight centile, 4-times higher for infants with birthweight between the 4th and 10th centile and 2-times higher for infants with birthweight above the 97th centile. Moreover, infants above the 97th centile have 3-times higher risk of delivery related perinatal death which is related to intrapartum anoxia. These deaths could potentially be prevented with a policy of induction of labour at early term gestation (between 37 and 39 weeks' gestation). It has been shown that induction of labour suspected fetal growth restriction at term can be done safely as it does not increase adverse pregnancy outcomes or the rates of instrumental delivery or caesarean section.[3] A meta-analysis of routine induction of labour at term in an unselected population of women compared to standard practice demonstrated a greater than 50% reduction in stillbirth together with a >85% reduction in perinatal death.[4] Moreover, early term induction of labour for suspected macrosomia could reduce shoulder dystocia and associated morbidity by up to 70%.[5]

Fetal size is calculated by measuring the length and circumference of fetal parts such as the head, the abdomen, and the femur. A variety of methods exist for converting these measurements, which are commonly termed fetal biometry, to an estimated fetal weight (EFW)[6] and a number of reference ranges exist for EFW in relation to the exact gestational age.[7,8] The interpretation of EFW can be done in two ways: (i) by positioning the value on the distribution for the given gestational age, and (ii) by assessing the change in the value over serial measurements. Taking the first of these, babies in the smallest 10% of measurements for gestational age are referred to as small for gestational age (SGA) and babies in the largest 10% are referred to as large for gestational age (LGA). The second property

examines the growth velocity across the pregnancy. For example, if a fetus is on the 9th percentile at 36wkGA and it had also been on the 9th percentile at 20wkGA, it would be regarded as SGA but with normal fetal growth velocity. SGA infants with normal growth velocity are often constitutionally small. However, a fetus with a significant drop in the EFW percentile between two scans would be regarded as indicating fetal growth restriction FGR and would be managed differently compared to a case of SGA with the normal growth velocity [9]

One of the ways to assess the placental function is by using Doppler flow velocimetry (referred to as "Doppler", see Hoffman and Galan for review).[10] This is done by assessing both qualitatively and quantitatively the flow velocity waveform of the targeted blood vessel. One of the key blood vessels for study is the umbilical artery. Flow is assessed qualitatively by the direction of flow in end diastole (i.e. immediately prior to the rise in flow that occurs with ventricular contraction – systole). The normal state is forward flow, but there can be absent flow or even reversed flow. However, the flow is most commonly analysed by using a number of indices such as the pulsatility index (PI) and resistance index (RI). The derivation, calculation and detailed interpretation of these indices is described in detail elsewhere.[10] However, raised values in those indices would be considered to show increased resistance to flow in the fetal vascular tree of the placenta. Correlative studies of umbilical artery Doppler and placental microscopy support this interpretation in cases of FGR occurring before 36 weeks' gestation.[11]

Umbilical artery (UA) Doppler has been used to monitor high risk pregnancies, including those with suspected fetal growth restriction (FGR). A systematic review of diagnostic test accuracy showed that UA Doppler can be useful at predicting perinatal mortality and risk of compromise.[12] Moreover, a Cochrane review of randomised controlled trials (RCTs) showed that UA Doppler ultrasound in high-risk pregnancies appears to reduce the number of perinatal deaths and the number of obstetric interventions.[13] However, a similar Cochrane review of RCTs in low risk pregnancies showed no differences in pregnancy outcomes comparing those randomised to UA Doppler with controls.[14] This review included 5 studies designed as routine Doppler vs no Doppler, and there was no consistent management plan for the abnormal results. It included in total 14,185 women but this study design would require >100,000 women to assess rare outcomes such as perinatal death.[15] The authors concluded that there is no adequate evidence that the use of routine UA Doppler ultrasound benefits either the mother or the baby and they recommended future studies that should be designed to address small changes in perinatal outcome.

In more recent years the cerebroplacental ratio (CPR) has been proposed as a more accurate predictor of adverse pregnancy outcome and clinically indicated medical intervention such as caesarean section for fetal distress. The CPR is a measure of cerebral centralization of fetal blood flow and is calculated by dividing the Doppler index (pulsatility index (PI), resistance index (RI), or systolic/diastolic ratio (S/D)) of the middle cerebral artery (MCA) by that of the UA. Physiologically, CPR represents the interaction of alterations in blood flow to the brain, as manifest by increased diastolic flow as a result of cerebrovascular dilatation due to hypoxia and increased placental resistance, leading to decreased diastolic flow in the UA. The CPR appears to be useful in predicting perinatal death in pregnancies with suspected FGR[16] but there is no evidence about its effectiveness in predicting adverse pregnancy outcomes in low or mixed risk populations.

Another important feature which is examined in late pregnancy is the amniotic fluid volume. Reduced amniotic fluid is called “oligohydramnios” and increased amniotic fluid is called “polyhydramnios”. Oligohydramnios could be due to ruptured membranes (which if happens preterm, would increase the risk of early delivery or fetal infection) or fetal anomalies (such as renal agenesis), but it could also indicate reduced urine production by the fetus which, in turn, could be due placental insufficiency. Polyhydramnios causes include gestational diabetes, fetal anomalies with disturbed fetal swallowing of amniotic fluid, fetal infections and other, rare causes. The prognosis of polyhydramnios depends on its cause and severity. Typical symptoms of polyhydramnios include maternal dyspnoea, spontaneous preterm labour, premature rupture of membranes (PPROM), abnormal fetal presentation, cord prolapse and postpartum haemorrhage.[17] Amniotic fluid volume is quantitatively assessed using measurement of the biggest single pool (DVP = deepest vertical pool), or by the sum of the four deepest pools in each of four quadrants of the uterus (AFI = amniotic fluid index).

Finally, undiagnosed breech presentation in labour increases the risk of perinatal morbidity and mortality and represents a challenge for obstetric management. The incidence of breech presentation at term is around 3-4%,[18-20] and fewer than 10% of fetuses who are breech at term revert spontaneously to a vertex presentation.[21] Although breech presentation is easy to detect through ultrasound screening, many women go into labour with an undetected breech presentation.[22] The majority of these women will deliver through emergency Caesarean section, which has high costs and increases risk of morbidity and mortality for both mother and child. The sensitivity of abdominal palpation varies between studies (range: 57-70%), and depends on the skill and experience of the practitioner.[23] In contrast, ultrasound examination provides a quick and safe method of accurately

identifying fetal presentation. Effective interventions exist for the care of women who have breech presentation diagnosed near term. The Royal College of Obstetricians and Gynaecologists recommends "that all women with an uncomplicated breech presentation at term should be offered External Cephalic Version (ECV)".[24]

The risks of third trimester screening.

A screening test would lead to intervention for those that are screened positive. An intervention might benefit those that are correctly classified as high risk (true positives) but might cause harm to those that are classified as high risk despite being normal (false positives). The timing of the screening is crucial because if it is done early such as between 28 and 34 weeks an expedited delivery at these gestations can cause significant harm related to prematurity. An optimal timing for the screening would be around 36 weeks' gestation since an intervention such as induction of labour at 37 weeks could prevent complications at term without causing significant morbidity to the false positives. However, early term deliveries (between 37+0 and 38+6 weeks' gestation) are at higher risk of respiratory distress,[25] prolonged hospitalisation and even higher risk of educational needs later in life.[26] The iatrogenic consequences of antenatal screening were identified in a national study done in France where routine ultrasound screening for fetal growth restriction is routinely offered to all women between 30 and 35 weeks of gestation. The study analysed more that 14,000 pregnancies and showed that the false positives (incorrectly identified as small) had significantly higher rates of provider initiated delivery <37 weeks and worse perinatal outcomes (including 5-minute Apgar score and admission to the neonatal unit) not only compared to the true negatives (correctly identified as normal) but also compared to the false negatives (those that were truly small but were incorrectly classified as normal). This suggests that the harm of preterm screening can be significant and that for any screening policy to be applied there needs to be clear evidence that the benefits outweigh the risks.

Evidence for screening using universal late pregnancy ultrasound

As explained above, the use of ultrasound has showed benefits in high-risk pregnancies. For example, the use of umbilical artery Doppler reduces perinatal mortality by about 30% in high-risk pregnancies.[13] This study showed that its use was also associated with lower rates of induction of labour and Caesarean delivery which could attribute the reduced the risk of perinatal death overall to

the reduction of unnecessary intervention. However, there was also a strong trend to a reduced risk of stillbirth, indicating that Doppler may also have been useful in targeting intervention to the highest risk cases.

However, ultrasound in low-risk populations was studied by a large influential meta-analysis of 13 RCTs including ~35,000 women and did not demonstrate any evidence that routine ultrasound improved outcome.[27] However this systematic review has several flaws that undermine its main conclusions.[28] First, the 13 studies included in the meta-analysis all used different definitions of screen positive. Moreover, the ultrasonic tests used were not consistent. For example, while multiple studies analysed some variant of an estimation of fetal size, one large study assessed placental calcification without any assessment of any other features of the scan. The diagnostic effectiveness of these tests is not comparable as shown in a subsequent systematic review.[29] Second, there was no consistency with the intervention after revealing the scan to the clinicians. This could lead to different management protocol, hence different outcomes. Moreover, the timing of the scans varied across the studies. Given that the primary intervention available to the attending clinicians would have been delivery of the baby, the potential for this resulting in benefit or harm would vary according to the gestational age where the scan was performed this could lead to iatrogenic harm by deciding to deliver preterm. Finally, although the meta-analysis included 35,000 women, it was still underpowered for the key outcome of interest, perinatal death. The risk ratio for perinatal death from the meta-analysis was 1.01 with 95% confidence interval (CI) of 0.67 to 1.54. While these CI might seem quite narrow, the capacity for reducing the rate of an outcome with a screening trial is different from interventional trials in women with established disease. If we identified a screening test for perinatal death with a positive likelihood ratio of 10 with a 5% screen positive rate and if we applied an intervention which reduced the risk by 50%, the estimated relative risk would be 0.76, which is within the 95% CI of the systematic review. Hence, the Cochrane review is underpowered to detect the effect of a highly effective screening test coupled with a highly effective intervention. If we use the 5.8 per 1000 perinatal mortality rate in the control group of the Cochrane review, a power calculation indicates that a sample size of 110,000 women would be required to detect this effect with 90% power.[28]

Parity and the risk of adverse outcome

One of the most important factors that determine whether a pregnancy would be considered as high risk or low risk is past obstetric history, i.e. the outcome of previous pregnancies. Many conditions of

pregnancy have quite high risks of recurrence in subsequent pregnancies, such as preeclampsia,[30] preterm birth[31] stillbirth[32] and FGR[33]. Women that had any of those complications in their first pregnancy would be classified as high risk in subsequent pregnancies and will receive enhanced level of antenatal care. Conversely, previous uncomplicated pregnancies is strongly predictive of a normal outcome would the subsequent pregnancy would be classified as low risk.[34] Past obstetric history is, necessarily, not available for women who have not had prior births. Although maternal characteristics, as described above, are associated with the risk of pregnancy complications, the associations are generally rather weak and perform poorly as a screening test in isolation.[35] Moreover, some conditions such as preeclampsia have higher incidence in first pregnancies and the overall rates of complications are higher compared to second pregnancies.

Summary

The current evidence on the effectiveness of universal ultrasound screening is flawed. An effective screening strategy combined with an intervention for the screen positives could potentially prevent adverse pregnancy outcomes at term. However, there is also risk of causing iatrogenic harm mainly by premature delivery of the baby for a false positive result. This risk could be minimised by screening at about 36 weeks' gestation with a plan to intervene at about 37 weeks' gestation for the screen positives. The need for screening is greatest in the nulliparous population because they have higher background risks of adverse outcome and they lack one of the key discriminating characteristics in risk assessment, namely, knowledge of the outcome of prior births.

Aims

1) To review the published evidence on third trimester ultrasound screening to predict adverse pregnancy outcomes. Conduct meta-analyses of diagnostic test accuracy for the most important ultrasound findings (as selected by healthcare professionals) for tests which lack a recently published systematic review. (Chapters 3-7)

2) To analyse the prospective cohort study, Pregnancy Outcome Prediction Study, for the above ultrasound findings and combine the results with the meta-analyses. (Chapters 4, 7, and 8)

3) Finally, to use the results to provide inputs for a health economist analysis of the cost-effectiveness of universal ultrasound screening and to assess the need, potential design, and acceptability of a future randomised controlled trial. (Chapter 8)

Chapter 2. Protocol design and description of methods

Identifying the research questions

I designed a survey addressed to members of a number of professional organisations with the aim of identifying the ultrasonic features and the pregnancy outcomes considered most important by the relevant health care professionals. The survey got approved by the Ethics Committee of the School of Humanities and Social Sciences at the University of Cambridge and was distributed using SurveyMonkey™. The organisations that agreed to participate were the Royal College of Obstetricians and Gynaecologists, the British Maternal Fetal Medicine Society and the British Association for Perinatal Medicine. It was also distributed locally at the Rosie Hospital in Cambridge. The survey was conducted in May-June 2017.

The survey was completed by 54 respondents including 20 Consultant Obstetricians, 8 Obstetricians in training, 18 Midwives, 5 Sonographers and 3 Consultant Neonatologists. All the replies were anonymous.

The first question was about identifying the most important ultrasonic findings for universal screening in late pregnancy. The most important ultrasonic findings (ranked in order of frequency of response) were:

1. abnormal fetal biometry or growth velocity (83%),
2. malpresentation (63%),
3. abnormal amniotic fluid volume (63%),
4. high resistance pattern of umbilical artery Doppler flow velocimetry (32%),
5. abnormal cerebro-placental ratio or middle cerebral artery doppler (22%).

The second question was about identifying the most important adverse pregnancy outcomes (apart from perinatal death). The most important outcomes (ranked by frequency of response) were:

1. hypoxic ischaemic encephalopathy (69%),
2. fetal asphyxia (low umbilical cord blood pH plus a base deficit consistent with metabolic acidosis; 64%),
3. SGA or severe SGA (51%),
4. severe shoulder dystocia (46%),

5. breech presentation diagnosed in labour (41%),
6. admission to neonatal intensive care unit (28%),
7. low 5-minute Apgar score (21%).

I performed an initial database search (Medline, Embase, ClincalTrials.gov, and Cochrane) to identify any other systematic reviews which might overlap with our aims. I identified a protocol for a Cochrane DTA review of ultrasonic diagnosis of SGA. This Cochrane review was led by one of the co-applicants on the grant that funded the present study (A. Heazell) and was subsequently published in 2019.[29] Hence, I did not include this in my reviews. I also identified a previously published systematic review of DTA on severe oligohydramnios which was published in 2014 and included publications up to 2011. Some of the studies included in this review were performed in low and mixed risk populations. I included these studies in my analysis and then I performed a literature search for eligible studies that have been published subsequent to the search date the 2014 paper. Finally, I did not identify any studies on ultrasound screening for malpresentation. Thus, we decided to use the previously unpublished data from the POP study for the analysis.

Based on the priorities identified by the survey and the concurrent Cochrane DTA review, and on what we believed was feasible in the time scale, we identified the following ultrasonic markers as the priority subjects for systematic review of DTA:

1. Suspected fetal macrosomia
2. High resistance pattern of umbilical artery Doppler flow velocimetry
3. Low cerebro-placental ratio (CPR)
4. Severe oligohydramnios
5. Borderline oligohydramnios

Database search strategy

I was the co-ordinator in all systematic reviews. I designed the search strategy and identified the titles as described below. The literature search, study selection, and analysis ware performed independently by myself and another co-author using Review Manager 5.3. Any differences were resolved in discussion with the senior author (GS).

We searched the literature systematically using the Cochrane database of Systematic Reviews (CDSR), Cochrane Central Register of Controlled Trials (CENTRAL), Medline, EMBASE, and ClinicalTrials.gov. Moreover, we searched the references of all included studies for papers not identified through the literature search. For ongoing clinical trials I searched ClinicalTrials.gov, and for ongoing systematic reviews I searched the Cochrane and PROSPERO databases. Additionally, I applied no restrictions on the language of the report or the location of the study. The dates and the terms used for each systematic review are described in each respective chapter. The exact search strategies are presented in the Appendices.

I had training in the medical library of the University of Cambridge on how to design electronic searches for the systematic reviews. I used several methods ensure that I did not miss any relevant studies. These included:

- Both text words and MEDLINE subject headings (MeSH)
- Synonyms (newborn / neonate) and related terms (e.g. macrosomia / large for gestational age)
- Both full terms and acronyms (e.g. Amniotic fluid index / AFI)
- Variant spellings (e.g. Caesarean / Cesarean)
- Trancation (e.g. ultraso*)

I combined the terms using Boolean operators (AND/OR/NOT). Between terms of similar concept, such as the index text, I used the Boolean 'OR' and to combine terms of different concepts, such as the index text with the target condition, I used the Boolean 'AND'.

Study selection

I included prospective and retrospective cohort studies, randomized clinical trials, and cross-sectional studies. I excluded case-control studies as they tend to overestimate the effect size. I included studies with singleton, non-anomalous, pregnancies which had an ultrasound done after 24 weeks' gestation.

I included all studies in which the ultrasound was done as part of universal ultrasound screening (the ultrasound was offered to all women regardless of indication), studies that were done in low-risk populations (those that excluded pregnancies with any maternal or fetal complication) and studies

with mixed risk population (the ultrasound was offered selectively based on current clinical indications). I excluded studies that were focused only on high risk populations such as pregnancies with fetal growth restriction or pregnancies with gestational diabetes.

Studies were not selected on the basis of the definition of the index test, i.e. the formula and the threshold used. Finally, I included both studies that blinded the results to the clinicians and studies that did not.

Risk of bias and applicability assessment

The risk of bias in each study was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS 2) tool as explained in the Cochrane Handbook of Diagnostic Test Accuracy Studies and revised in 2011. [36]. This tool focuses on four domains that could be sources of bias, three of which could be sources of applicability concerns as well. These include patient selection, index test, reference standard, and flow and timing.

The first domain aims to rule out selection bias by assessing whether a consecutive or random sample of patients was selected, whether case-control design was avoided and whether inappropriate exclusions were avoided. Case-control studies and studies with inappropriate exclusions could overestimate the diagnostic effect. It also assesses concerns about applicability for example by assuring that the population is appropriate for the study question. In our case we excluded studies that included high risk patients only, such as diabetic patients for the assessment of macrosomia.

The second domain about index test which assesses whether the index test (ultrasound finding) was interpreted without knowledge of the results of the reference standard (outcome) and whether the threshold was pre-specified. It also assesses applicability concerns, such as in cases where the formula or the threshold used for an ultrasound finding is no longer used.

The third domain on reference standard assesses whether the reference standard correctly classifies the target condition. It also assesses whether the reference standard was interpreted without the knowledge of the index test, i.e. whether the ultrasound findings were blinded from the clinicians. This is of critical importance as prior knowledge of an abnormal ultrasound result (e.g. suspected LGA) could lead to either earlier intervention (e.g. emergency Caesarean section) or to ascertainment bias,

for example by quickly labelling a case as shoulder dystocia and performing manoeuvres to assist the delivery of the fetus. Regarding applicability concerns this domain assesses whether the reference standard of the study is the same as the reference standard of the review (e.g. different thresholds of cord arterial pH to define neonatal hypoxia).

The fourth domain on flow and timing assesses whether the interval between index test and reference standard was appropriate. In our reviews we used the perspective of late third trimester ultrasound scan (at around 36 weeks' gestation). Hence, we classified studies that performed the scan up to 28 weeks' gestation of after 41 weeks' gestation as high risk. This domain finally assesses whether all the patients were included in the analysis.

For each review we used a pre-designed data extraction form (for a sample form see Appendix) to extract information on study characteristics (year of publication, country, setting, study design, blinding), patient characteristics (inclusion and exclusion criteria, sample size), the index test (gestation at scan, indices, and cut-off values used), reference standard (pregnancy outcome, gestation at delivery, and interval from scan to delivery).

I planned subgroup analyses where possible for: 1) blinding of ultrasound results or not, 2) Parity (nulliparous vs multiparous) 3) Timing from scan to delivery (e.g. <2 weeks or >2weeks). I finally included additional information such as rate of induction of labour (IOL), presence of comorbidities etc, as explained in each review separately.

Data extraction and synthesis

From each study we extracted the 2 x 2 tables for all combinations of index tests and outcomes and we calculated the sensitivity, specificity, positive and negative likelihood ratios (LRs)[37] respectively. For the data synthesis we used the hierarchal summary receiver – operating characteristics (HSROC) model of Rutter and Gatsonis[38] which allows for variation of cut-off points between studies. If a study reported on more than one threshold, the most used cut-off point was selected. Whenever four or more studies were available, estimates of mean sensitivity and specificity and respective variances at a specific threshold were additionally generated using the bivariate logit-normal (Reitsma) model.[39] We also used meta-analysis to obtain a summary of the diagnostic odds ratios (DORs).[40] Publication bias was assessed using the Deeks' funnel plot asymmetry test where data was available from a

sufficient number of studies. Significant asymmetry was assumed at P<0.05.[41] Statistical analyses was performed using STATA version 14 (StataCorp LP, College Station, TX), specifically, its METANDI, METAN and MIDAS packages. Analysis and reporting was performed using the PRISMA guidelines.[42]

The protocol was submitted to the PROSPERO database in June 2017 with the title "Late pregnancy ultrasound screening" (CRD42017064093) and is available at the link below:
https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42017064093

Analysis of the Pregnancy Outcome Prediction (POP) Study

I analysed data from the Pregnancy Outcome Prediction study (POPS), a prospective cohort study that took place at the Rosie Hospital, Cambridge between 2008 and 2012 which has been described before in detail. [9]The study included nulliparous women only, and all women who agreed to participate had two research ultrasound scans at 28 and 36 weeks' gestational age (wkGA) which were blinded to the women and the clinicians. Some of those women had clinically indicated ultrasound scans in the third trimester based on the local and national guidelines.

The ultrasound screening included fetal biometry, amniotic fluid volume, umbilical artery Doppler, presentation, and placental location. The fetal biometry included measurement of biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC) and femur length (FL) using standard techniques. Estimated fetal weight (EFW) percentile was calculated using the Hadlock equations and reference standard. The amniotic fluid volume was assessed using the Amniotic Fluid Index (AFI) and the umbilical artery Doppler was assessed using the Pulsatility Index.

The outcome of the pregnancy was retrieved by 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). 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 perinatal 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). Small for gestational age was

defined a birthweight below the 10th centile and severe SGA as below the 3rd centile for sex and gestational age using a UK reference.[43] Large for gestational age was defined as birthweight above the 90th centile and severe LGA as above the 97th centile respectively.

Chapter 3. Universal ultrasonic screening using fetal macrosomia in the prediction of adverse perinatal outcome.

Introduction

Macrosomia is usually defined as birthweight > 4000g. Large for gestational age (LGA) is commonly defined as birthweight above the 90th centile for sex and gestational age. The terms are sometimes used interchangeably at around full term (40 weeks' gestation). Macrosomic birth weight is associated with the risk of adverse outcomes, including perinatal death[2] and injuries related to traumatic delivery.[44] Ultrasonic estimation of fetal weight (EFW) was first described in 1975[45]. The equation for EFW which is in most widespread use was published by Hadlock et al in 1985[6] and the distribution of EFW in relation to week of gestation was published in 1991[7]. Hence, the diagnostic tools to identify small for gestational age (SGA) and large for gestational age (LGA) fetuses have been available for many years. One of the main complications associated with macrosomia is shoulder dystocia and a Cochrane review of four randomised controlled trials (RCTs), including 1190 women, demonstrated that routine induction of labour (IOL) for suspected LGA may prevent this outcome[46]. However, it remains unclear whether screening and intervention for suspected LGA results in better outcomes.

An RCT of IOL in women with an ultrasonically suspected LGA infant is in progress in the UK (The Big Baby trial, ISRCTN18229892). However, the women recruited to this trial will have been scanned because they were high risk for some reason, as NICE have recommended that women should not be routinely scanned in late pregnancy.[1] Although the trial will confirm whether induction of labour is effective in high risk women, it will not determine whether screening women without risk factors and intervening results in net benefit. It is often the case that screening and intervention programmes which work well in high risk groups do not work as well in low risk populations, and one explanation for this can be that the screening test is less informative in low and mixed risk populations due to the lower prior risk of disease. In this study, we sought to quantify the diagnostic effectiveness of screening for fetal macrosomia and associated complications using universal ultrasonic fetal biometry in late pregnancy.

Methods

Publication statement

This chapter was published as an individual paper in Plos Medicine (Appendix 9A).[47] Some passages were quoted verbatim from this paper since I wrote the text for this publication. The figures and tables have been reproduced under open access licence.

Sources for meta-analysis

The protocol for this review was prospectively written and registered with PROSPERO (the International Prospective Register of Systematic Reviews), and the registration number was CRD42017064093. We searched the literature systematically using the Cochrane database of Systematic Reviews (CDSR), Cochrane Central Register of Controlled Trials (CENTRAL), Medline, EMBASE, and ClinicalTrials.gov from inception to August 2019. An update search was done on the 28th of May 2020. We applied no restrictions on the language of the report or the location of the study. The studies were identified using a combination of words related to "ultrasound", "pregnancy", "estimated fetal weight", "EFW", "birthweight", "macrosomia", "large for gestational age", "shoulder dystocia", and "brachial plexus injury". The exact search strategy is presented in Appendix 1.

Study selection

We set out to include cohort studies where an ultrasound scan was performed ≥24 weeks' gestation (wkGA), excluding multiple pregnancies. We included studies of low risk populations, universal screening, and mixed-risk populations (i.e. included both high-risk and low-risk pregnancies). Studies which included only high risk women, such as patients with pre-existing or gestational diabetes, and those where the ultrasound scan was performed during labour were excluded. Studies were not selected on the basis of the definition of the index test, i.e. the formula and the threshold used. Finally, we included both blinded and un-blinded studies.

Index tests and outcomes

For the purposes of the meta-analysis we defined suspected LGA as a fetus with an EFW >4000g or >90th centile, or with an abdominal circumference (AC) >36 cm or >90th centile. However, we have also documented other thresholds used. The outcomes studied included macrosomic birth weight (>4000g or >90th centile) and severe macrosomic birth weight (>4500g or >97th centile); shoulder dystocia; and perinatal morbidity (neonatal unit admission, 5-minute Apgar score of six or less, metabolic acidosis, neonatal hypoglycaemia and neonatal jaundice).

Study quality assessment and statistical methods

The literature search, study selection and analysis were performed independently by two authors (myself and Norman Shreeve) using Review Manager 5.3. Any differences were resolved in discussion with the senior author (GS). The risk of bias assessment and the statistical methods employed are described in Chapter 2.

Results

Study characteristics and quality assessment of the studies included in the meta-analysis

The literature search flowchart is presented in Figure 1. We identified 41 studies [48-88]that met our inclusion criteria involving 112,034 patients in total. The study characteristics are presented in Table 1. Six studies[51, 60, 66, 69, 70, 85] (N=53,935) included unselected pregnancies, nine studies[56, 62, 64-66, 68, 76, 78, 86, 87] (N= 6436) included only low-risk pregnancies and 26 studies[48-50, 52-55, 57-59, 61, 63, 67, 71-75, 77, 79-84, 88] (N= 51,663) included mixed risk pregnancies.

The assessment of study quality was performed using the QUADAS-2 tool and is summarized in **Error! Reference source not found.**. The Galvin 2017 study [62] was published as an abstract, hence we used a different study from the same cohort (GENESIS study) [89] to assess the risk of bias. Only three studies, Sovio 2018 [85] (POP study), Galvin 2017 [62] (GENESIS study), and Peregrine 2007 [80] blinded the results to the clinicians. Hence, the large majority of studies were at risk of bias in relation to the reference standard. The second most common risk of bias was in relation to flow and timing, as six studies [52, 57, 69, 72, 80, 88] performed the ultrasound either prior to induction of labor or less than 72 hours before delivery, resulting in a very short interval between the scan and delivery. Conversely, two studies [51, 60] had a very long interval (ultrasound <33wkGA). Two studies [50, 53] did not present data on the gestational age at delivery. Finally, three studies [56, 81, 87] were confined to pregnancies progressing beyond 41wkGA, and were classified as having “high applicability concerns due to patient selection”.

Figure 1. PRISMA flow diagram for the systematic review of macrosomia.

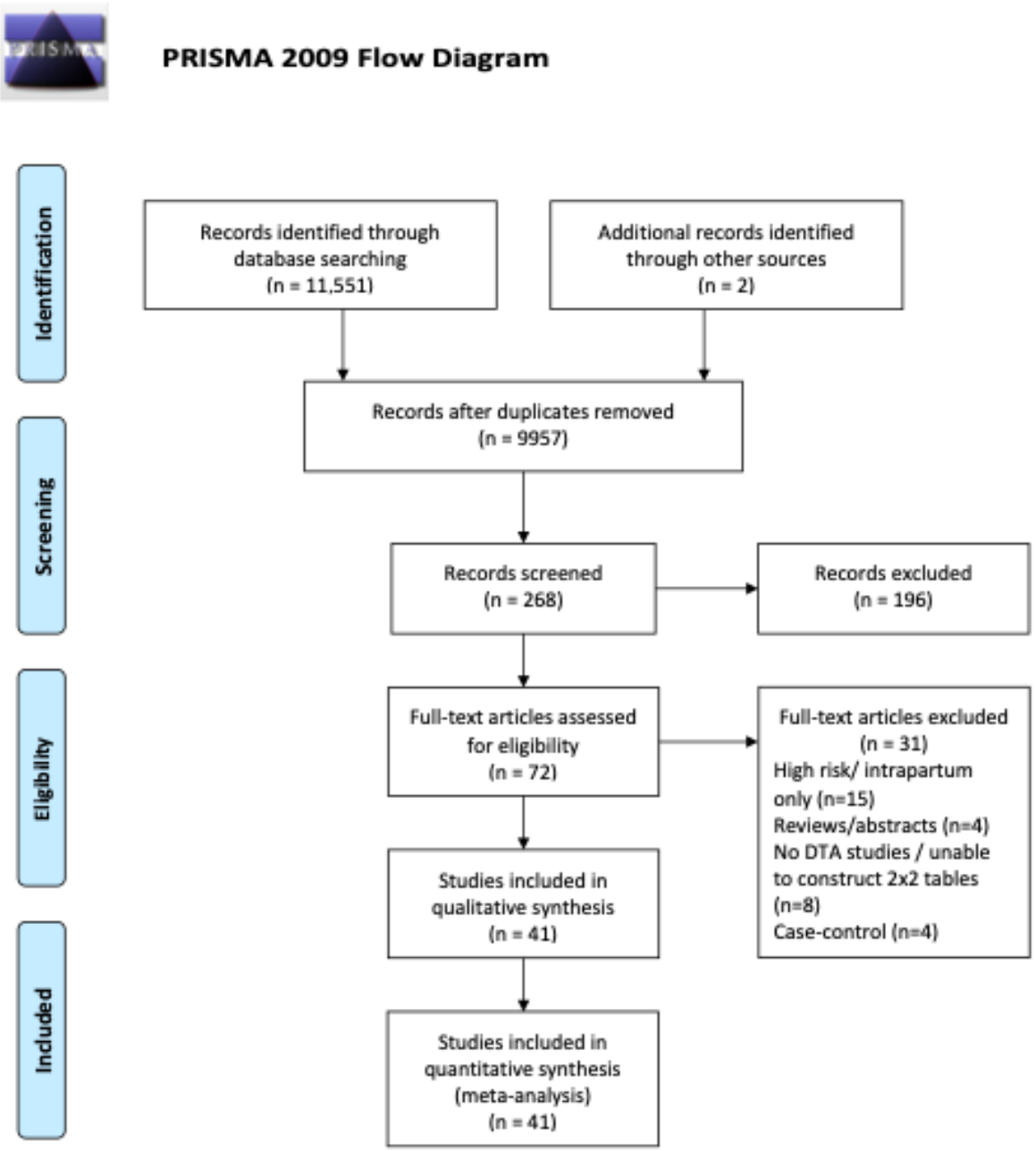


Table 1. Characteristics of the studies included in the meta-analysis of LGA fetuses

First Author (Year)	Type of Study, Setting	Total number of fetuses (number of LGA fetuses at birth), risk, and selection (All singleton, non anomalous unless otherwise stated)	Index test (Blinding)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments (Inclusion of T1DM, T2DM and GDM)
Aviram 2017	Retrospective cohort, Single Hospital, Israel	N= 7996 (1618) Risk: Mixed Selection: Mixed risk, term only. Excluded SGA deliveries, intrapartum and SROM.	EFW (20 formulas) Hadlock (AC/FL/BPD) Hadlock (AC/FL/HC) Hadlock (AC/FL/BPD/HC) Hadlock (AC/FL) Hadlock (AC/BPD) Shepard (AC/BPD) Threshold: >90th centile Blinded: No	Within 1 week from delivery.	BW >90th centile	Mean for LGA group: 39.4 weeks, mean for AGA group: 38.3 weeks	DM/GDM: Included (21% for LGA, 14% for AGA)
Balsyte 2009	Retrospective cohort, Single Hospital, Switzerland	N= 1062 (135) Risk: Mixed Selection: Term only.	EFW Hadlock (AC/FL/HC) Threshold: >4000g Blinded: No	Within 1 week from delivery.	BW >4000g	Mean 39.3 weeks.	DM/GDM: Not reported
Benecerraf 1988	Retrospective cohort, Single hospital, Boston, MA, USA	N= 1301 (324) Risk: Mixed Selection: Included all pregnancies apart from breech and multiples.	EFW (Birnholz) Threshold: Threshold: >4000g, >3800g Blinded: No	Within 1 week from delivery.	BW >4000g	Not specified	DM/GDM: Included
Ben-Haroush 2007	Prospective cohort, Single Hospital, Israel	N= 259 (23) Risk: Universal Selection: Routine scan. Included SGA. Excluded hypertensives and diabetics.	EFW Hadlock (AC/FL/BPD) Threshold: >90th centile Blinded: No	Mean 32 weeks	BW >4000g	Mean 39 weeks.	DM/GDM: Excluded
Ben-Haroush 2008	Retrospective cohort, Single Hospital, Israel	N= 1925 (140) Risk: Mixed Selection: Term only.	EFW Hadlock (AC/FL) EFW + AFI Threshold: EFW >4000g, AFI >95mm (60th centile) Blinded: No	Interval from USS to delivery 2.5 days	BW >4000g	Mean for LGA 40 weeks, Mean for normal BW 39.4 weeks	DM/GDM: Excluded
Benson 1991	Retrospective cohort, Boston, MA, USA	N= 412 (32) Risk: Mixed Selection: Not specified. Excluded diabetics.	EFW Hadlock (AC/FL/BPD) Threshold: >90th centile Blinded: No	Within 1 week from delivery	BW> 90th centile	Not specified	DM/GDM: Excluded

Burkhardt 2014	Retrospective cohort Single Hospital, Zurich, Switzerland	N= 12,794 Risk: Mixed Selection: All term, with vertex presentation with scan with 7days	EFW, AC Hadlock (AC/FL/BPD) Threshold: >4000g, >4500g >35cm, >39cm Blinded: No	Within 1 week from delivery	Shoulder dystocia	281 days fro SD 278 days for no SD	DM/GDM: 7.5% for those with SD 2.7% for those without SD.
Chauhan 2006	Retrospective cohort Single Hospital, Houston, TX, USA	N= 1954 (119) Risk: Mixed Selection: Pregnancies undergoing fetal surveillance. Included SGA, hypertensives (22%) and SROM (5%).	EFW Hadlock (AC/FL/BPD) Threshold: >90th centile Blinded: No	Within 4 weeks from delivery. 64% within 7 days from delivery.	BW >90th centile	34% preterm	DM/GDM: Included (13%)
Chervenak 1989	Prospective cohort Single Hospital, New Jersey, USA	N= 317 (81) Risk: Low Selection: Uncomplicated pregnancies after 41 weeks' gestation.	EFW Hadlock AC/BPD or AC/FL if BPD not available Threshold: >4000g Blinded: Not clear	>41 weeks	BW >4000g	Mean 42 +/- 0.6 weeks	DM/GDM: Excluded
Cohen 2010	Retrospective cohort Single Hospital, Montreal, Canada	N= 1099 (105) Risk: Mixed Selection: Only included pregnancies with USS on the same or next day as delivery	EFW Hadlock (AC/FL/BPD/HC) Threshold: >90th centile Blinded: No	On the same or next day of delivery.	BW >4000g	Mean 275.2 days.	DM/GDM: Included (11.6%)
Crimmins 2018	Retrospective cohort Single hospital, Baltimore, Maryland, USA	N= 945 (40) Risk: Mixed Selection: All pregnancies >34 weeks gestation with normal oGCT.	AFG defined as EFW >90th centile (Hadlock-AC/FL/BPD) or AC >95th centile. Polyhydramnios >25cm Threshold: As above. Blinded: No	>34 weeks	BW >4000g Shoulder dystocia NICU admission	Not specified.	DM/GDM: Excluded
Cromi 2007	Retrospective cohort, 2 hospitals, Swtzerland	N= 1026 (53) Risk: Mixed Selection: All singletons >34 weeks gestation with USS within 4 weeks of delivery. Excluded SROM.	EFW, AC Hadlock (AC/FL/BPD) Threshold: >95th centile Blinded: No	Within 4 weeks of delivery. Mean 37.3 weeks	BW >4000g BW>4500g	>34 weeks Mean 39.2 weeks	DM/GDM: Included (8.8%)
De Reu 2008	Retrospective cohort, Single Hospital, Netherlands	N= 3449 (285) Risk: Universal Selection: Women with no risk factors or pathology. Did not exclude SGA.	AC Threshold: >75th /90th/95th centile Blinded: No	Between 27 and 33 weeks.	BW >90th centile, BW >95th centile	Mean 278.7 days	DM/GDM: Excluded

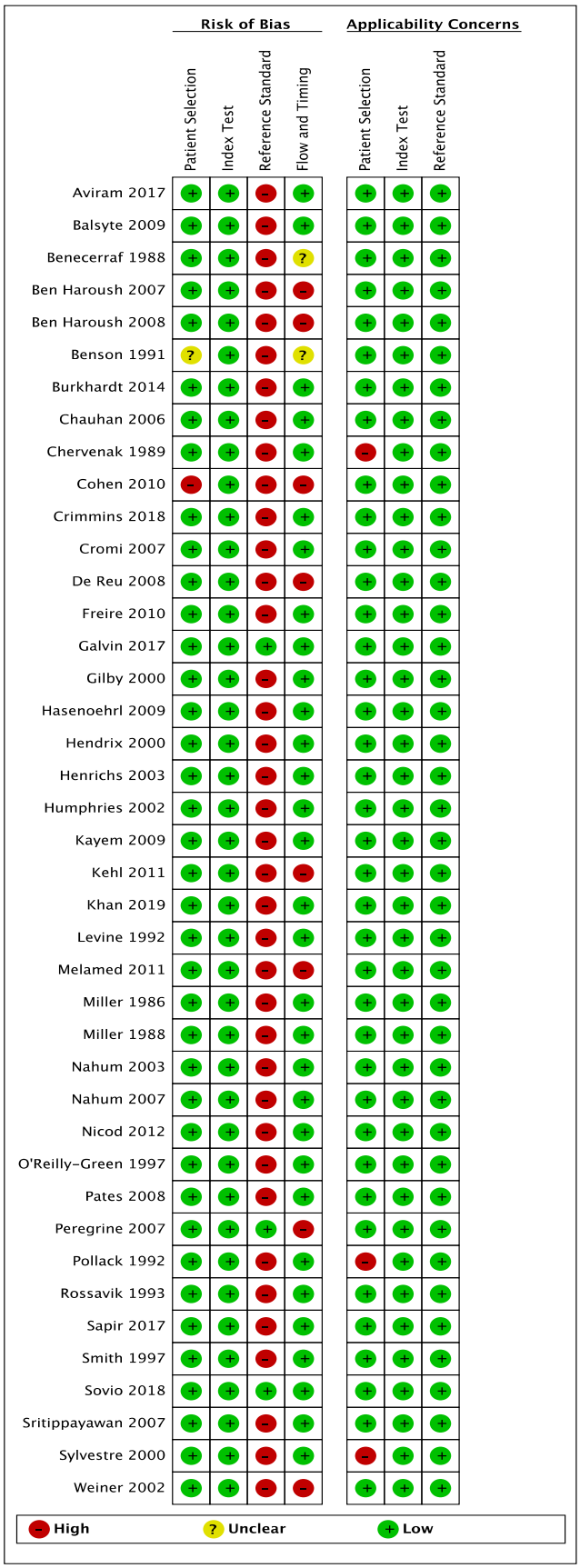
Freire 2010 (Portuguese)	Retrospective cohort, 2 hospitals, Brazil	N= 114 (8) Risk: Mixed Selection: Those with USS within 7 days of delivery	EFW Hadlock (AC/FL/BPD/HC) Threshold: >90th centile Blinded: No	Within 7 days of delivery	BW >90th centile	15.6% preterm, 84.4% at term	DM/GDM: Not reported
Galvin 2017 (GENESIS study) (Abstract)	Prospective cohort Large multi-centre study, Ireland	N= 2336 (not known) Risk: Low Selection: Term, uncomplicated, cephalic only.	EFW (Not specified) Threshold: 4000g Blinded: Yes	Between 39+0 and 40+6 weeks	Shoulder dystocia NICU admission	Not specified.	DM/GDM: excluded
Gilby 2000	Retrospective cohort, Single Hospital, Florida, USA	N= 1996 (318) Risk: Mixed Selection: All singleton >36 weeks with USS within 1 week from delivery.	AC Threshold: >35cm, >38cm Blinded: No	Within 1 week from delivery	BW >4500g	>36 weeks Mean not reported.	DM/GDM: Not reported
Hasenoehrl 2006	Prospective cohort, Single hospital, Austria	N= 200 (33) Risk: Low Selection: Included those with USS within 1 week. Excluded only fetal anomaly.	EFW (Schild) Threshold: >4000g Blinded: No	Mean 39.2 weeks	BW >4000g	Mean interval 2.0 days.	DM/GDM: Not reported
Hendrix 2000	Prospective (RCT) Georgia, USA	N= 367 (39) Risk: Low Selection: Term only.	EFW Hadlock AC/BPD Threshold: >4000g Blinded: No	>37 weeks	BW >4000g	Mean 39.1 weeks	DM/GDM: Not reported
Henricks 2003	Prospective cohort, South Carolina, USA	N= 256 (21) Risk: Universal Selection: Term only.	AC Threshold: >35cm Blinded: No	>37 weeks	BW >4000g	Mean 39.1 weeks	DM/GDM: Not reported
Humphries 2002	Retrospective cohort, South Carolina, USA	N= 238 (29) Risk: Mixed Selection: Term only, with USS within 2 weeks.	EFW Combs (AC/FL/FL) Threshold: >4000g Blinded: No	Within 2 weeks of delivery	BW >4000g	>37 weeks	DM/GDM: Not reported
Kayem 2009	Prospective cohort, Multiple hospitals, France and Belgium	N= 1689 (124) Risk: Low Selection: As part of a prospective cohort for breech. Term only, with USS within 10 days of delivery.	AC Threshold: >36.3cm Blinded: No	Within 10 days of delivery.	BW >4000g	Median 39 weeks	DM/GDM: Not reported
Kehl 2011	Prospective cohort, Single	N= 258 (30) Risk: Universal	AC Threshold: >36cm	Within 3 days of delivery	BW >4000g	40+5 weeks for AC>36cm	DM/GDM: Not reported

	Hospiotal, Germany	Selection: Term only with vertex presentation and USS within 3 days of delivery.	Blinded: No			39+6 weeks for AC <36cm	
Khan 2019	Retrospective cohort, 2 HNS Hospitals, London, UK	N= 45847 (4229) Risk: Universal Selection: Term only.	EFW Hadlock (AC/FL/HC) Threshold: >90th centile Blinded: No	Between 35+0 and 36+6 weeks Mean 36.1 weeks	BW >90th centile BW >97th centile	Mean 39.9 weeks	DM/GDM: T1DM/T2DM Included (0.7% for non-LGA, 2.1% for LGA)
Levine 1992	Retrospective cohort, Single Hospital, New York, USA	N= 406 (68) Risk: Mixed Selection: Term only. Included pregancies with diabetes (22%) and previous CS (20%)	EFW Hadlock (AC/FL/HC) Threshold: >90th centile Blinded: No	5-10 days before delivery	BW >90th centile	Mean 39.4	DM/GDM: Included (22%)
Melamed 2011	Retrospective cohort, Single hospital, Israel	N= 4765 (431) Risk: Mixed Selection: All deliveries with USS within 3 days of delivery. DM/GDM and SROM excluded.	EFW (multiple) and AC Hadlock (AC/FL/BPD) Hadlock (AC/FL/HC) Hadlock (AC/FL/BPD/HC) Hadlock (AC/FL) Shepard (AC/BPD) Threshold: >4000g,>36cm Blinded: No	Within 3 days of delivery	BW >4000g	Mean 38.1	DM/GDM: Excluded
Miller 1986	Retrospective cohort, Single Hospital, Luisiana, USA	N= 150 (28) Risk: Mixed Selection: Term only, included diabetes, PET, prior CS. Excluded SGA	EFW Hadlock (AC/FL) Shepard (AC/BPD) Threshold: >4000g Blinded: No	Within 7 days of delivery	BW >4000g	Term (Mean ga not reported)	DM/GDM: Included
Miller 1988	Retrospective cohort, Single Hospital, Luisiana, USA	N= 382 (58) Risk: Mixed Selection: term only, excluded SROM	EFW and AC Hadlock (AC/FL/BPD) Threshold: EFW >4100g, AC >36.4cm Blinded: No	Within 7 days of delivery. Mean ga 275.8 days	BW >4000g	Mean ga 279.1 days.	DM/GDM: Not reported
Nahum 2003	Retrospective cohort, Single hospital, California, USA	N= 74 (12) Risk: Mixed Selection: Only included Hispanic ethnicity, term only.	EFW (11 formulas) Hadlock (AC/FL/BPD) Hadlock (AC/FL/HC) Hadlock (AC/FL/BPD/HC) Hadlock (AC/BPD) Shepard (AC/BPD) Threshold: >4000g Blinded: No	Within 3 weeks of delivery	BW >4000g	Term (Mean ga not reported)	DM/GDM: Included (23.0%)

Nahum 2007	Retrospective cohort, Single hospital, California, USA	N= 98 (16) Risk: Low risk Selection: Term only, Excluded medical complications (PET, DM)	EFW Hadlock (AC/FL/BPD) Hadlock (AC/BPD) Hadlock (AC/FL) Threshold: >4000g, Blinded: No	Within 3 weeks of delivery	BW >4000g	Term (Mean ga not reported)	DM/GDM: Excluded
Nicod 2012 (French)	Retrospective cohort, Single hospital, Switzerland	N= 708 (141) Risk: Mixed risk Selection: Pregnancies with USS within 7 days of delivery.	EFW Hadlock (AC/FL/BPD/HC) Hadlock (AC/FL) Threshold: >4000g Blinded: No	Within 7 days of delivery	BW >4000g	Not reported	DM/GDM: Not reported
O'Reilly-Green 1997	Retrospective cohort, Single hospital, New York, USA	N= 445 (107) Risk: Low Selection: Prolonged pregnancies defined as ga >40+4.	EFW Hadlock (AC/FL/BPD) Threshold: >4000g, >4500g Blinded: No	Within 3 weeks of delivery	BW >4000g BW >4500g	GA >40+4	DM/GDM: Excluded
Pates 2007	Retrospective cohort, Single hospital, Texas, USA	N= 3115 (239) Risk: Mixed Selection: Those with clinically indicated USS within 7 days of delivery.	EFW and AFI Hadlock (AC/FL/BPD/HC) Threshold: >4000g, AFI >20cm (95th centile) Blinded: No	Within 7 days of delivery	BW >4000g	Not reported	DM/GDM: Included (11%)
Peregrine 2007	Prospective cohort, Single hospital, London, UK	N= 262 (48) Risk: Mixed Selection: Pregnancies with ga >35+6 undergoing IOL, Excluded those withIUD or antepartum haemorrhage.	EFW Hadlock (AC/FL) Shepard (AC/BPD) Threshold: >4000g Blinded: Yes	Exactly before IOL	BW >4000g	Median ga 41 weeks.	DM/GDM: Not reported
Pollack 1992	Retrospective cohort, Single hospital, New York, USA	N= 519 (119) Risk: Mixed Selection: Postdate pregnancies >41 weeks	EFW Hadlock (AC/FL) Threshold: >4000g, >4500g Blinded: No	Within 7 days of delivery	BW >4000g	>41 weeks	DM/GDM: Not reported
Rossavik 1993	Retrospective cohort, Single hospital, Oklahoma, USA	N= 498 (36) Risk: Mixed Selection: Infants with USS within 2 weeks of delivery (if ga >38w) or within 1 week of delivery (if ga <38w)	EFW Hadlock (AC/FL/HC) Threshold: >4000g Blinded: No	Within 2 weeks of delivery (if ga >38w) or within 1 week of delivery (if ga <38w)	BW >4000g	Not reported	DM/GDM: Not reported
Sapir 2017 (Abstract)	Retrospective cohort	N=6214 Risk: Mixed	EFW, AC	Wiothin 1 week of delivery	Shoulder dystocia	Term (not specified)	DM/GDM: Excluded

	Single Hospital, Israel	Selection: term only, no GDM with scan within 7 days of delivery	Threshold: >4000g, >4500g, AC>39cm Blinded: No				
Smith 1997	Retrospective cohort, Single hospital, Glasgow, UK	N= 1213 (16) Risk: Mixed Selection: Non-diabetic pregnancies with USS within 7 days of delivery.	EFW and AC Hadlock (AC/FL) Threshold: >4000g, >4500g, AC >36cm, AC >38cm Blinded: No	Within 7 days of delivery	BW >4500g	Not reported	DM/GDM: Excluded
Sovio 2018	Prospective cohort, Single hospital, Cambridge, UK	N= 3866 (177) Risk: Universal Selection: Unselected n nulliparous women that delivered after 36 weeks.	EFW, ACGV Hadlock (AC/FL/BPD/HC) Threshold: >90th centile (population/customised) Blinded: Yes	Regular research scan at 36 weeks (median 36.4 weeks)	BW >90th centile BW >97th centile BW >4000g, BW >4500g, shoulder dystocia, metabolic acidosis, 5-min Apgar <7, NICU admission, severe neonatal morbidity, neonatal hypoglycaemia, neonatal jauntice	Median 40.4 weeks.	DM/GDM: Included (4.3%)
Sritippayawan 2007	Prospective cohort, Single Hospital, Thailand	N= 328 (3) Risk: Low risk Selection: Pregnancies >34 weeks. Excluded IUFD, any medical complication.	EFW Hadlock (AC/FL/BPD/HC) Threshold: >4000g Blinded: No	>34 weeks Mean interval 16.9 days from delivery	BW >4000g	Mean ga 39.4 weeks.	DM/GDM: Excluded
Sylvestre 2000	Retrospective cohort, Single Hospital, New York, USA	N= 656 (147) Risk: Low risk Selection: Postdate pregnancies only (>41 weeks)	EFW (Hadlock or Shepard/Not specified) Threshold: >4000g Blinded: No	>41 weeks	BW >4000g	41.3 weeks	DM/GDM: Not reported
Weiner 2002	Prospective cohort, Single centre, Israel	N= 315 (134) Risk: Mixed risk Selection: Offered routine clinical screening to all womenat term. Those with suspected EFW >3700g had USS. Only included those with USS with 3 days of delivery.	EFW Shepard (AC/BPD) Threshold: >4000g Blinded: No	USS with 3 days of delivery.	BW >4000g BW >4500g Shoulder dystocia	40.1 weeks for both groups.	DM/GDM: Included (9.2%)

Figure 2. Risk of bias assessment using the QUADAS-2 tool of the studies included in the meta-analysis of LGA fetuses



Meta-analysis results

Full details of the summary diagnostic performance are presented in Table 2. In summary, both definitions of ultrasonically suspected macrosomia (i.e. either EFW >4000g or >90th percentile) had >50% sensitivity for predicting LGA at birth. Many associations were similar regardless of the formula employed but the positive LRs for the Hadlock formulae (ranging between 7.5 and 12) tended to be higher than and for the Shepard formula (around 5). The performance of definitions using just the AC was similar to using an ultrasonic EFW. The sensitivity for predicting severe macrosomia (>4500g) at birth of suspected LGA was around 70%. However, macrosomia (EFW > 4000g or >90th centile) had a lower (22%) sensitivity for predicting shoulder dystocia, although the association was statistically significant and the positive LR was ~2.

Figure 3 has summary ROC curves for shoulder dystocia and macrosomia. For the prediction of macrosomia at birth most of the large studies were close to the point estimate and only a few small studies were outside the prediction intervals. For shoulder dystocia, most studies reported sensitivities below 30% and only one study [88] reported a sensitivity of >50%. However, in this study the total number of shoulder dystocia cases was very small (n=3). Figure 4 and Figure 5 present graphs of the pooling of DORs for macrosomia and shoulder dystocia, respectively. There was significant heterogeneity for the prediction of macrosomia but not for the prediction of shoulder dystocia.

Only three studies, Crimmins 2018, Galvin 2017, and Sovio 2018 reported neonatal unit admission as an outcome and a meta-analysis was not feasible. However, none of the studies reported statistically significant results with positive LRs of 0.73 (95% CI 0.36-1.48), 1.39 (95% CI 0.97-2.00) and 1.33 (95% CI 0.80-2.22) respectively. Only the Sovio 2018 study reported on 5-minute Apgar score of less than 7 and neonatal metabolic acidosis with positive LRs of 1.94 (95% CI 0.66-5.75) and 1.08 (95% CI 0.28-4.18) respectively. Moreover, the Sovio 2018 study was the only one that reported on neonatal hypoglycaemia and neonatal jaundice with positive LRs of 1.9 (95% CI 1.1-3.4) and 1.2 (95% CI 0.6-2.4) respectively. The analysis demonstrated no significant evidence of publication bias (P=0.57), when evaluated using Deeks' funnel plot asymmetry test (Figure 6).

Table 2. Summary diagnostic performance of suspected LGA to predict LGA at birth and shoulder dystocia.

Diagnostic test	Studies	Patients	Summary sensitivity (95% CI)	Summary specificity (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)
Outcome: Birthweight >4000g (or 90th centile)						
EFW (any) >4000g (or 90th centile)	30	80,045	53.2% (47.2-59.1%)	93.9% (91.9-95.5%)	8.74 (6.84-11.17)	0.50 (0.44-0.56)
EFW (Hadlock-AC/FL/HC/BPD)	9	22,073	63.1% (49.1-75.2%)	94.3% (90.9-96.5%)	11.13 (8.24-15.04)	0.39 (0.28-0.55)
EFW (Hadlock- AC/FL/BPD)	10	17,110	55.1% (44.1-65.7%)	92.9% (89.7-95.2%)	7.77 (5.55-10.89)	0.48 (0.38-0.61)
EFW (Hadlock- AC/FL/HC)	7	60,648	55.2% (45.7-64.2)	94.9% (92.4-96.6%)	11.84 (7.46-15.74)	0.47 (0.39-0.58)
EFW (Hadlock- AC/FL)	9	16,736	60.5% (50.7-69.5%)	92.0% (89.4-93.7%)	7.54 (6.13-9.29)	0.43 (0.34-0.54)
EFW (Hadlock- AC/BPD)	6	13,617	62.9% (36.1-83.5%)	93.7% (85.9-97.3%)	9.99 (6.40-15.58)	0.40 (0.21-0.75)
EFW (Shepard)	7	14,060	73.7% (54.4-86.9%)	85.1% (76.5-90.9%)	4.96 (3.29-7.48)	0.31 (0.17-0.56)
AC >36cm (or 90th centile)	5	10,543	57.8% (39.6-74.2%)	92.3% (88.7-94.9%)	7.56 (5.85-9.77)	0.46 (0.30-0.68)
Outcome: Birthweight >4500g (or 97th centile)						
EFW (any) >4000g (or 90th centile)	5	51,686	67.5% (47.8-82.6%)	89.7% (79.1-95.3%)	6.58 (2.78-15.58)	0.36 (0.20-0.65)
Outcome: Shoulder dystocia						
EFW (any) >4000g (or 90th centile)	6	26,264	22.0% (9.9-42.0%)	89.6% (80.8-94.6%)	2.12 (1.34-3.35)	0.87 (0.74-1.02)

EFW: Estimated Fetal Weight; AC: Abdominal circumference; FL: Femur length; HC: Head circumference; BPD: Biparietal diameter; LR: Likelihood ratio; CI: Confidence intervals

Figure 3. Summary ROC curves for the diagnostic performance of EFW > 4000g (or 90th centile) at predicting A. LGA at birth (birthweight above 4000g or above the 90th centile) and B. Shoulder dystocia.

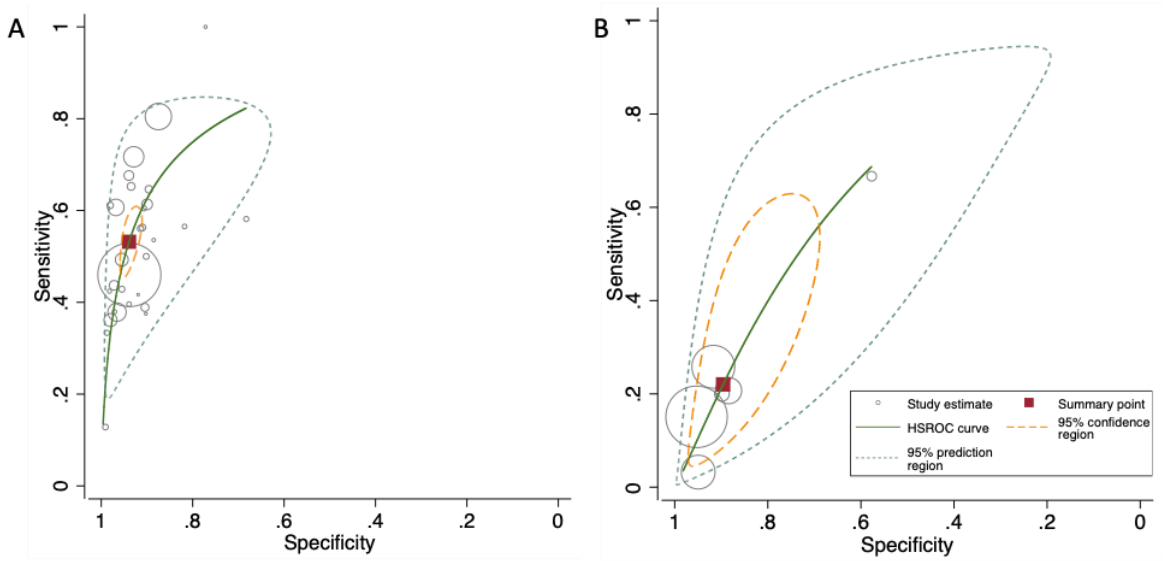


Figure 4. DORs for the diagnostic performance of EFW > 4000g (or 90th centile) at predicting LGA at birth (birthweight above 4000g or above the 90th centile)

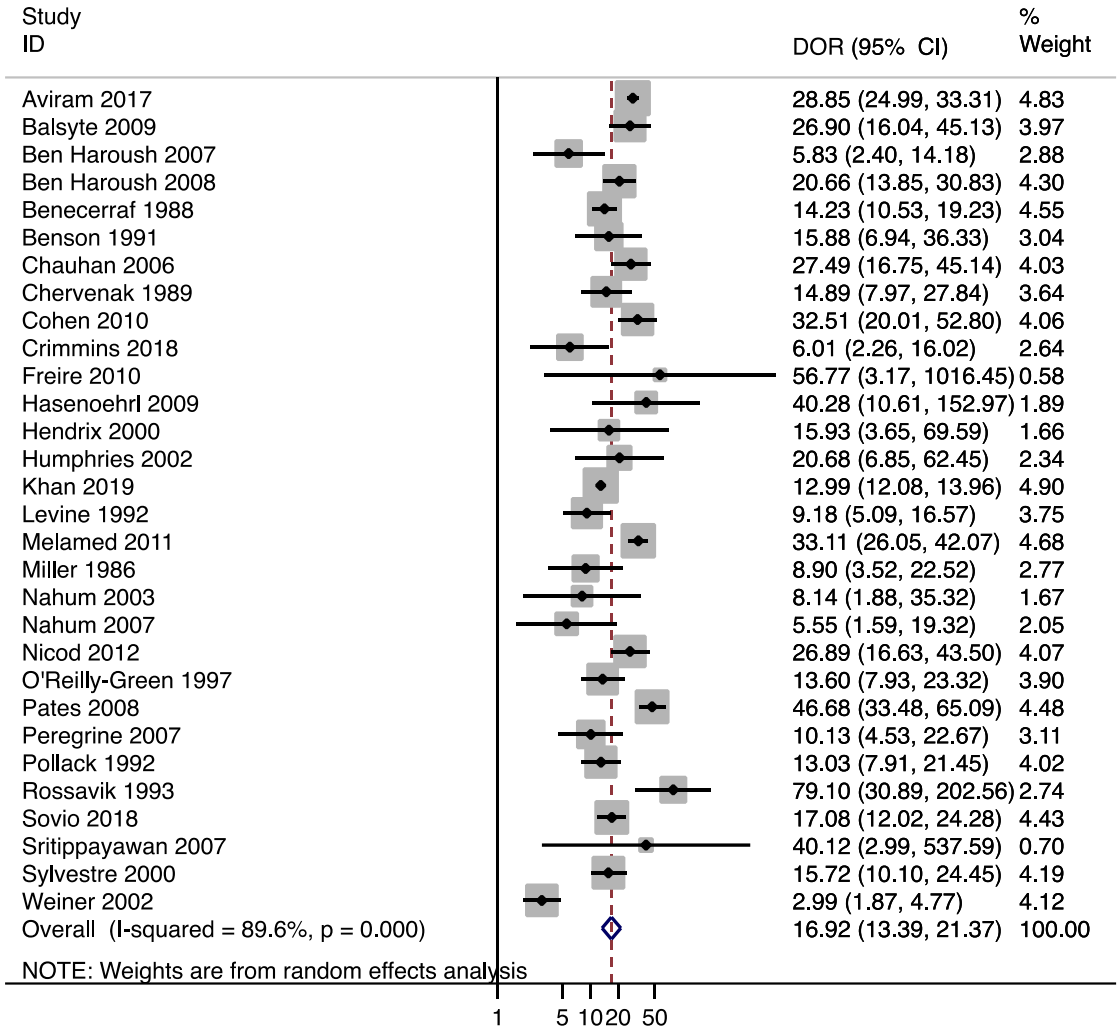


Figure 5. DORs for the diagnostic performance of EFW > 4000g (or 90th centile) at predicting shoulder dystocia

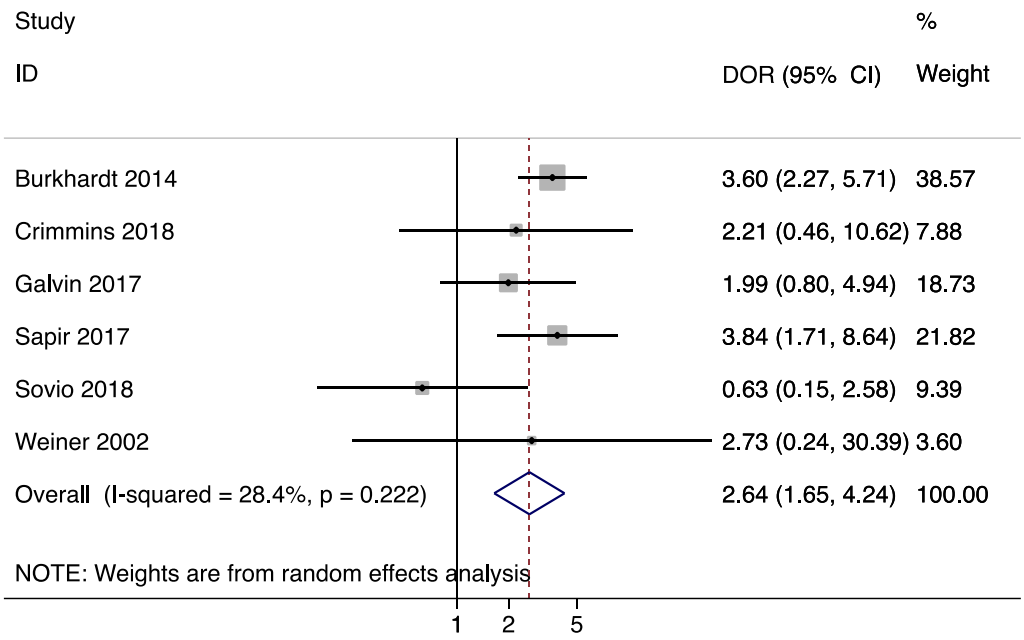
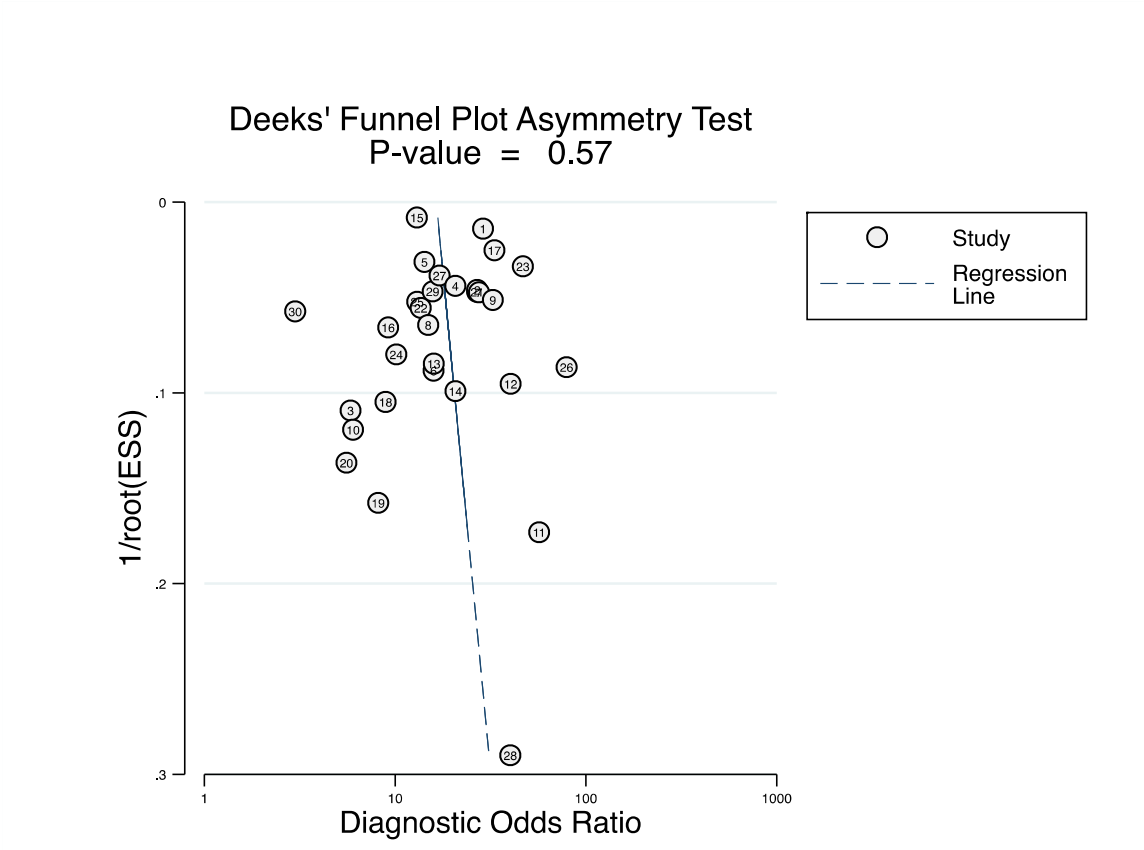


Figure 6. Deeks' funnel plot for publication bias for the prediction of macrosomia (birthweight >4000g or >90th centile).



Discussion

The key findings of the present chapter were that ultrasonic suspicion of fetal macrosomia is strongly predictive of the risk of delivering a large baby but it is only weakly – albeit statistically significantly – predictive of the risk of shoulder dystocia. In the case of delivering an LGA baby using the Hadlock formula, the positive LRs were quite strong, in the region of 7 to 12, whereas in relation to the diagnosis of shoulder dystocia, the positive LR was ~2. The forest plot of DORs indicates that there was significant heterogeneity between the studies in the ability to predict an LGA infant. The source of this heterogeneity is unclear but it could relate to differences in the quality of the performance of the diagnostic test, such as the quality of the imaging equipment, the skill and training of sonographers and the characteristics of the population.

In this meta-analysis we included data from the POP study, as it is particularly applicable to the research question addressed in this report, given that late pregnancy ultrasound was performed in a large number of nulliparous women using contemporary equipment and staff trained using the standards of the English NHS. The POP study analysis of a 36wkGA scan in the diagnosis of macrosomia had previously been published[85] and this was incorporated into the meta-analysis. Interestingly, the DOR (95% CI) from the POP study was 17.1 (12.0 to 24.3) and this was virtually identical to the summary estimate from all of the other studies where it was also 17.1 but with slightly narrower 95% CI (13.3 to 22.0). These data suggest that the results from the POP study are likely to be generalisable.

There has been the lack of blinding in studies of the diagnostic effectiveness of ultrasound in pregnancy screening research. Hence, generally, the POP study has been unique as a contemporary study in late pregnancy in nulliparous women. However, in this analysis there is a second comparable study, the Genesis study. This was a prospective cohort study of 2772 nulliparous pregnant women recruited across seven centres in Ireland between 2012 and 2015. Women had the ultrasound scan ≥39wkGA and <41wkGA, i.e. ~3 to 4 weeks later than the POP study. Although the timing of the scan is slightly later than the research question for the current report, the study design makes it particularly useful.

The analysis of fetal macrosomia from the Genesis study has only been published in abstract form. It did not report the diagnostic effectiveness of EFW as a predictor of LGA birth weight, but it did report shoulder dystocia. Interestingly, the POP study and the Genesis study, which were the only two large studies including more than 1000 women each, did not demonstrate a statistically significant

association between macrosomic EFW and the risk of shoulder dystocia. Overall, the meta-analysis indicated that ultrasound may be weakly predictive. However, as with other analyses in the preceding chapter, these findings could be explained by ascertainment bias. Specifically, if a scan is performed and the fetus is suspected to be macrosomic, the clinical staff attending the birth may be more likely to institute manoeuvres for shoulder dystocia in the event of any delay, or to document a given delay as being due to shoulder dystocia. The potential for such biases may explain why the studies with blinded ultrasound were not significantly associated and why the meta-analysis as a whole was only weakly predictive of shoulder dystocia while it was strongly predictive for macrosomia. A weak association between ultrasonic EFW and the risk of shoulder dystocia is not surprising given that the actual birth weight of the baby is not strongly predictive of shoulder dystocia and that the majority of cases of shoulder dystocia do not involve a macrosomic infant.[90]

Finally, the relationship between a diagnosis of fetal macrosomia and the outcome is an area where there is good evidence around the potential for revealing a scan result to change the experience of complications in women who are false positives. Multiple studies have demonstrated that a false positive diagnosis of fetal macrosomia is an independent risk factor for emergency Caesarean delivery.[91-93] These observations underline the possibilities that screening low risk women has the potential to cause harm and that researching methods of screening using a study design where the results are revealed to the attending clinician has the potential to cause associations which are a consequence of the scan, not a true prediction arising from it.

Chapter 4. Systematic review of the diagnostic effectiveness of universal ultrasonic screening using late pregnancy umbilical artery Doppler flow velocimetry in the prediction of adverse perinatal outcome.

Introduction

High resistance patterns of umbilical artery (UA) Doppler flow velocimetry are thought to reflect placental vascular resistance. A systematic review of diagnostic test accuracy has shown that the UA Doppler can be useful at predicting perinatal mortality and risk of compromise in high-risk pregnancies.[12] A Cochrane review of randomised controlled trials (RCTs) has demonstrated that use of UA Doppler ultrasound in high-risk pregnancies appears to reduce the number of perinatal deaths and the number of obstetric interventions (risk ratio 0.71, 95% confidence interval 0.52 to 0.98).[13] However, a Cochrane review of RCTs in low risk pregnancies failed to demonstrate any difference in outcome comparing pregnancies screened using UA Doppler compared with controls (risk ratio 0.80, 95% confidence interval 0.35 to 1.83).[14] This review included five studies that compared routine Doppler versus no Doppler but there was no consistent management plan for the women with abnormal results. Moreover, although it included 14,185 women it was underpowered to detect an effect on perinatal death using clinically plausible estimates of screening performance and the clinical effectiveness of intervention.[15] The authors concluded that there is no adequate evidence that the routine use of UA Doppler ultrasound benefits either the mother or the baby and they recommended future studies that should be designed to detect smaller changes in adverse perinatal outcome.

In order for a large screening programme to be implemented it needs to meet two prerequisites; first, an index test that can accurately predict adverse outcome and, second, a safe and effective intervention. In this context we have suggested that ultrasound screening at 36 weeks' gestation with a policy of induction of labor (IOL) for those screened positive at 37 weeks could have the potential to be safe and effective. [94] We have previously shown that universal ultrasound screening using biometry and fetal growth velocity can triple the detection of SGA infants and identify infants at increased risk of neonatal morbidity.[9] However, there is no clear evidence of the effectiveness of UA Doppler as part of universal ultrasound screening.

The aim of this chapter was to assess the diagnostic accuracy of third trimester UA Doppler to predict adverse pregnancy outcome at term. We conducted a systematic review and meta-analysis of all studies focusing in low and mixed risk populations. In the above analysis we also included unpublished

data from a prospective cohort study of nulliparous women, the Pregnancy Outcome Prediction (POP) study.[9]

Methods

Publication statement

This chapter was published as an individual paper in the Placenta journal (Appendix 9B).[95] Some passages were quoted verbatim from this paper since I wrote the text for this publication. The figures and tables have been reproduced after gaining permission by the journal.

Analysis of data from the Pregnancy Outcome Prediction study

In the systematic review we included unpublished data from a prospective cohort study, the Pregnancy Outcome Prediction (POP) study, which was conducted at the Rosie Hospital, Cambridge (UK) between 2008 and 2012 and previously described in detail.[96] In brief, the study included nulliparous women only, and all women who agreed to participate had two research ultrasound scans at 28wkGA and 36wkGA which were blinded to the women and the clinicians. About 40% of the women had clinically indicated ultrasound scans in the third trimester based on local and national guidelines. In the present analysis we included women that attended their 36wkGA research scan and had a live birth at the Rosie Hospital. Women who delivered prior to their 36wkGA scan appointment were excluded.

Screen positive was defined as an umbilical artery pulsatility index (PI) >90th percentile. A full description of the conduct of the study, including definition of outcome data, was described in a paper in the Lancet.[9] In brief, 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 arterial 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 perinatal outcome was defined as term perinatal death, hypoxic ischemic encephalopathy, use of inotropes, mechanical ventilation, or severe metabolic acidosis (defined as a cord arterial pH <7.0 and a base deficit of >12mmol/L). Small for gestational age (SGA) and severe SGA were defined as a birthweight <10th percentile and <3rd percentile respectively for sex and gestational age using a UK reference.[43]

Sources for meta-analysis

The protocol for the review was designed a priori and registered with the PROSPERO International Prospective Register of Systematic Reviews (Registration number: CRD42017064093). We searched Medline, EMBASE and the Cochrane library from inception to October 2020. The studies were identified using a combination of words related to “ultrasound”, “Doppler”, “umbilical artery”,

"pregnancy" and "prenatal diagnosis" (see Appendix 2). No restrictions for language or geographic location were applied.

Study selection

Selection criteria included cohort or cross-sectional studies with singleton pregnancies which had an ultrasound performed ≥24wkGA. Case-control studies were excluded as these tend to overestimate the effect size. We included all studies in which the ultrasound was performed as part of universal ultrasound screening (the ultrasound was offered to all women regardless of indication), studies that were done in low-risk populations (those that excluded pregnancies with any maternal or fetal complication) and studies with mixed risk population (the ultrasound was offered selectively based on current clinical indications). We excluded studies that were focused only on high risk populations such as pregnancies with FGR. We included all reported indices of umbilical artery Doppler such as the Pulsatility Index (PI), Resistance Index (RI) or the systolic to diastolic ratio (S/D ratio), as well as all reported cut-off values. Finally, we included studies regardless of blinding of the ultrasound to the clinicians but this was reported in the study characteristics.

Study quality assessment and statistical methods

The literature search, study selection, and analysis ware performed independently by two authors (myself and Tom Bainton) using Review Manager 5.3. Any differences were resolved in discussion with the senior author (GS). The methods are described in chapter 2.

Results

The POP study

The analysis included 3615 women that met the inclusion criteria (Figure 7).[95] All women had a blinded UA ultrasound at 36wkGA and 346 (9.6%) had an UA PI >90th percentile. The maternal age, socio-economic status, ethnicity, BMI, and rates of alcohol consumption and smoking were similar between the two groups (Appendix 2, Table 15). Moreover, the groups had similar rates of pre-existing hypertension, pre-eclampsia, type 1 and 2 diabetes, and gestational diabetes. The gestational age at delivery and rate of induction of labour were similar in both groups which can be attributed to the blinding of the ultrasound. The screening performance of UA PI >90th centile is presented in Table 3. A high resistance pattern of UA Doppler was associated with an increased risk of delivering an SGA infant or a severely SGA infant and the association was stronger for the latter outcome. However, the finding was not strongly predictive with positive LRs between 2.5 and 3.5. A high resistance pattern of UA Doppler was not associated with an increased risk of a range of indicators of neonatal morbidity in the POP study.

Figure 7. POP study inclusion flowchart for UA Doppler

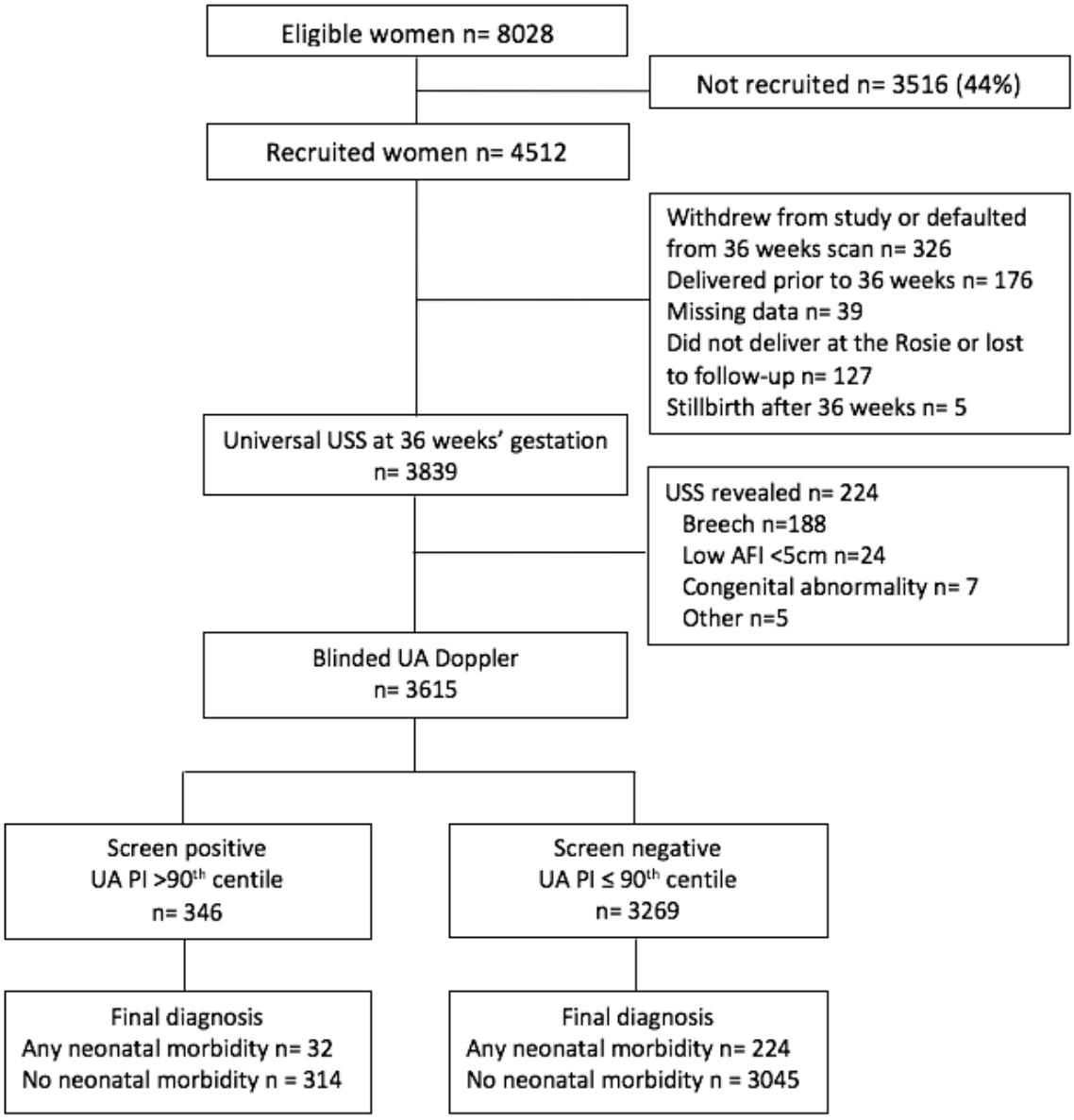


Table 3. Diagnostic performance of UA PI >90th centile at predicting adverse pregnancy outcome in the POP study (N=3615).

Outcome	True Positive / False Positive	True Negative / False Negative	Sensitivity (95% CI)	Specificity (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)
SGA <10th centile	72/274	3016/253	22.2% (17.6-26.7%)	91.7% (90.7-92.6%)	2.66 (2.11-3.36)	0.85 (0.80-0.90)
SGA <3rd centile	23/323	3215/54	29.9% (19.6-40.1%)	90.9% (89.9-91.8%)	3.27 (2.29-4.68)	0.77 (0.67-0.89)
Any neonatal morbidity*	32/314	3045/224	12.5% (8.4-16.6%)	90.7% (89.7-91.6%)	1.34 (0.95-1.88)	0.97 (0.95-1.01)
NICU admission	27/319	3076/193	12.3% (7.9-16.6%)	90.6% (89.6-91.6%)	1.31 (0.90-1.89)	0.97 (0.92-1.02)
5-min Apgar score <7	4/342	3243/26	13.3% (1.2-25.5%)	90.5% (89.5-91.4%)	1.40 (0.56-3.50)	0.96 (0.83-1.10)
Metabolic acidosis	4/342	3237/32	11.1% (0.8-21.4%)	90.4% (89.5-91.4%)	1.16 (0.46-2.95)	0.98 (0.88-1.10)
Severe neonatal morbidity*	3/343	3246/23	11.5% (0.7-23.8%)	90.4% (89.5-91.4%)	1.21 (0.41-3.52)	0.98 (0.85-1.12)

** See Chapter 4, Methods for definitions*

Study characteristics and quality assessment of the studies included in the meta-analysis

The literature search PRISMA flowchart is presented in Figure 8. We identified 13 studies[97-108] that met our inclusion criteria including 67,764 patients in total. The study characteristics are presented in Table 4. Five studies[97, 103, 106, 107] (N=63,436) included unselected pregnancies as part of universal screening, four studies[98, 101, 102, 108] (N=2634) included only low-risk pregnancies and four studies [99, 100, 104, 105](N=1694) included mixed risk pregnancies. Three of the studies[97, 106, 107] that were done in the same hospitals might have had short periods of overlap. Nine studies [98, 99, 101-105, 108] (N=8097) were prospective and four[97, 100, 106, 107] (N=59,687) retrospective. Studies varied in relation to the gestational age at scan (ranging from 28wkGA to 41wkGA), as well as the indices and the cut-off points used. The majority of patients in the included studies delivered at term. The assessment of study quality is presented in Figure 9. Overall, the quality was variable. The main risk of bias was that only six studies[98, 99, 101, 103, 105] (N= 5777) blinded clinicians to the UA Doppler result. However, five of these six studies revealed other features of the scan result, such as fetal biometry. Only the POP study blinded both the utero-placental Doppler and fetal biometry.

Figure 8. Literature search PRISMA flow diagram for the UA systematic review

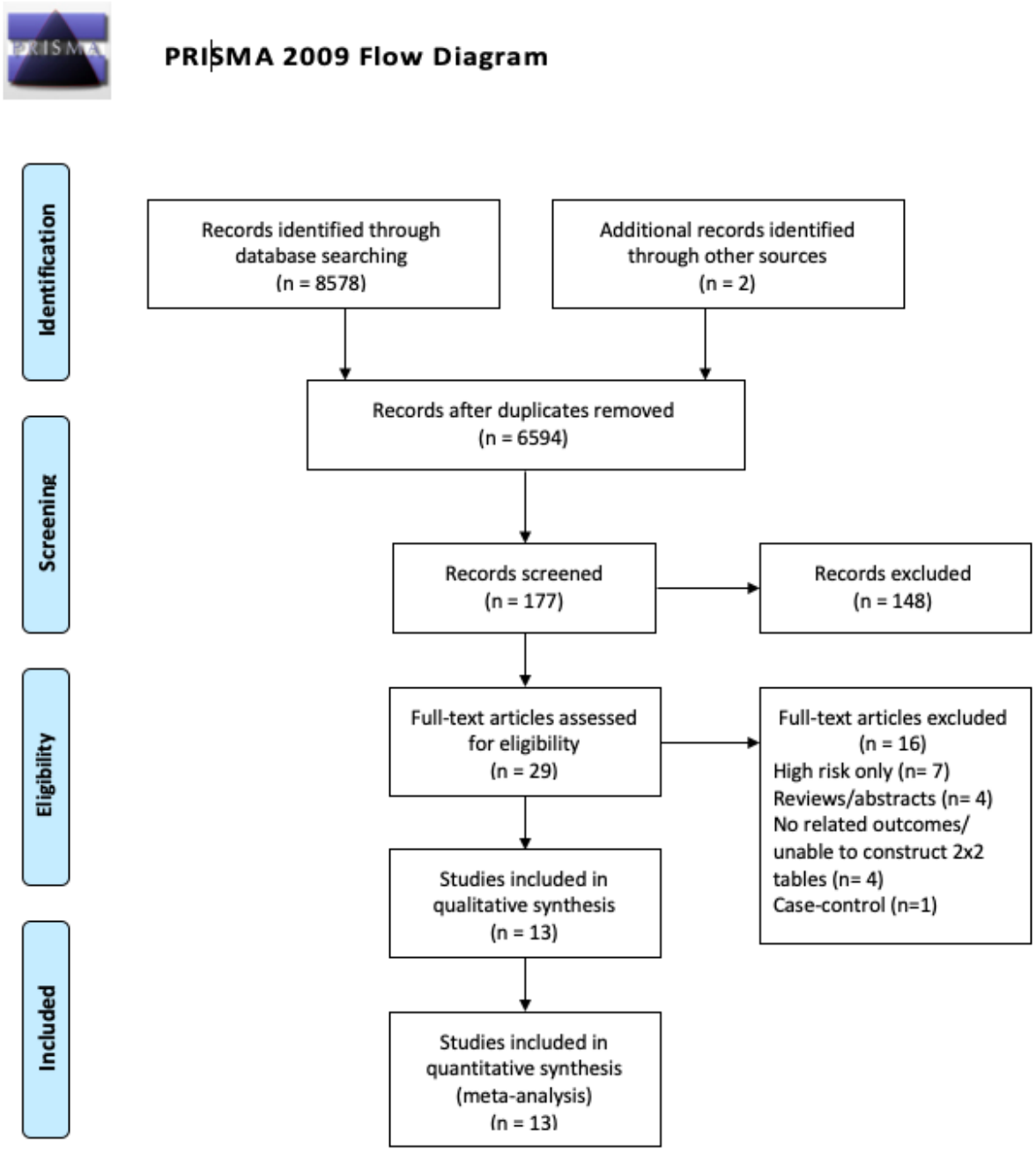


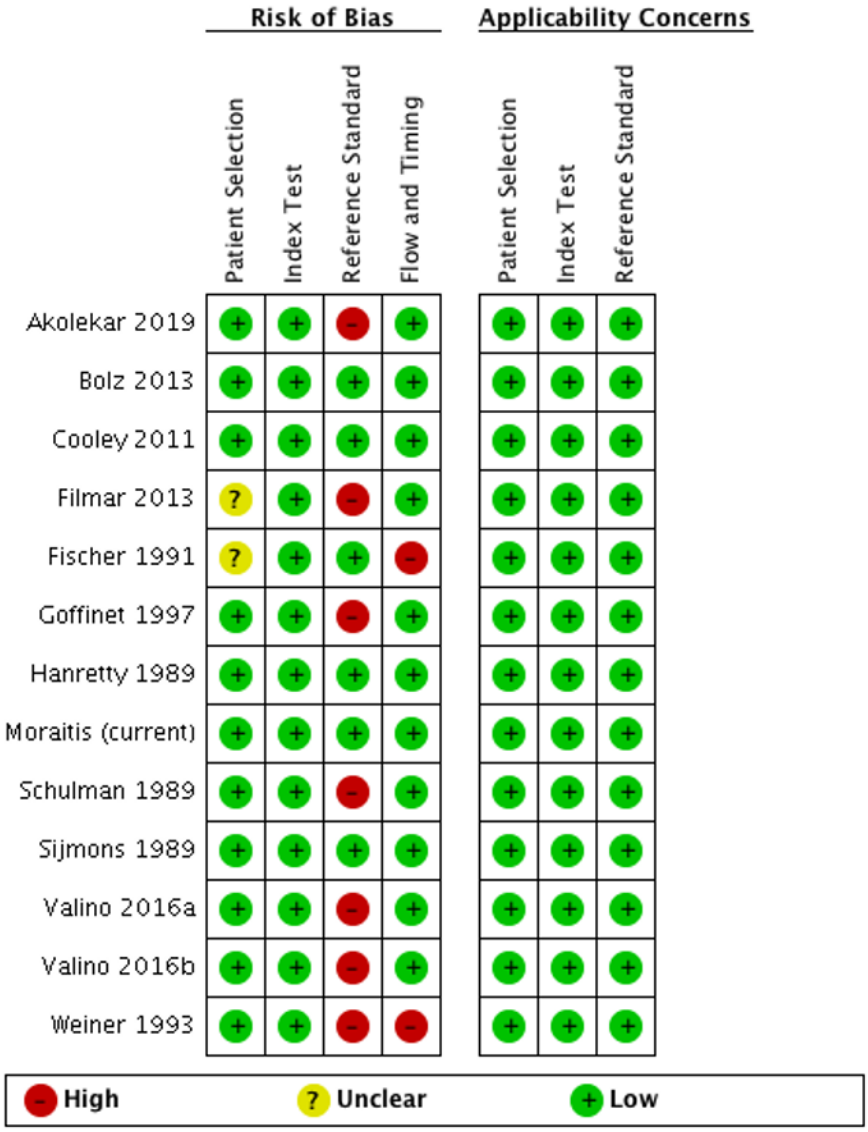
Table 4. Characteristics of studies included in the meta-analysis of UA Doppler

First Author (Year)	Type of Study, Setting	Number of fetuses and selection (All singleton, non anomalous unless otherwise stated)	Index test	Gestational age at ultrasound	Reference standard	Gestational age (ga) at delivery	Other comments
Akolekar 2019	Prospective cohort, 2 NHS Hospital, UK Between March 2014 and September 2018 (potential overlap with Valino studies)	N= 47,211 Universal, >36 weeks.	PI >90th centile. Not blinded.	Between 35+6 and 37+6 weeks.	Severe APO (composite of stillbirth, neonatal deaths and HIE grade 2 or 3), perinatal hypoxia (cord artery PH <7.0, 5-minute Apgar score <7, NICU admission), CS for fetal compromise, SGA <3rd centile.	Median ga at delivery 40.0 (39.0-40.9) weeks.	Nulliparous: 45.4% for those with no adverse outcome, 58.5% for those with adverse outcome.
Bolz 2013	Prospective cohort, Single Hospital, Germany	N=514 Low risk, term, cephalic only. Excluded maternal disease, SGA, RFM.	PI>1.2 Blinded UA Doppler.	Within 1 week from delivery. Mean ga 39+2 weeks.	Neonatal acidosis (cord arterial PH <7.10)	Mean ga 40+1 weeks	Nulliparity: Not reported. IOL: Not reported.
Cooley 2011	Prospective cohort, Single Hospital, Ireland	N=810 Mixed risk, nulliparous only. Only included Caucasian aged 18-40 years.	PI>95th centile UA blinded but EFW not blinded.	Around 36 weeks (not specified)	Emergency CS, PIH, PET, preterm delivery (<37 weeks), SGA <10th centile, SGA <3rd centile, 5-minute Apgar score <7, Cord arterial PH <7.10, NICU admission, Stillbirth	Not reported	Nulliparity: All IOL: 22.4%.
Filmar 2013	Retrospective cohort, Single Hospital, New York, NY, USA	N=251 Mixed risk, EFW>10th centile.	S/D ratio >90th centile (persistent), Not blinded.	Mean ga 35.3 weeks for abnormal UA group. Mean	NICU admission, 5-minute Apgar score <7	Median ga 37 weeks for abnormal UA	Nulliparity: Not reported IOL: Not reported.

				ga 34.4 for control group.		group, 39 weeks for control group.	
Fischer 1991	Prospective cohort Single Hospital, Pennsylvania, USA	N= 75 Low risk, post dates >41 weeks. Excluded maternal disease, suspected IUGR.	S/D ratio >3.0 S/D ratio >2.4 Blinded UA Doppler.	Mean interval from scan to delivery 2 days	Composite perinatal outcome: 1) Non-reassuring intrapartum fetal heart rate. 2) Umbilical artery PH <7.15, or venous <7.2 3) 5-min Apgar score <7 4) meconium stained liquor, 5) NICU admission, 6) birthweight <10th centile.	Mean ga at delivery 292.2 days	Nulliparity: 57% IOL: Not reported
Goffinet 1996	Prospective cohort, 17 hospitals, France	N=1903 Low risk, excluded maternal disease, suspected IUGR	RI >90th centile Not blinded.	Between 28 and 34 weeks	PIH, PET, Intervention for fetal distress, 5-minute Apgar <7, NICU admission, birthweight <3rd centile, birthweight 3-10th centile	Mean ga 39.2 weeks for those with abnormal UA, 39.4 weeks for those with normal UA.	Nulliparous: 43.0% for those with abnormal UA, 45.3% for normal.
Hanretty 1989	Prospective cohort, Single Hospital, Glasgow, UK	N=395 Universal	AB ratio >95th centile. Blinded UA doppler	34-36 weeks	PIH, SGA <5th centile, 5-minute Apgar <6, NICU admission	Mean ga 38.9 weeks for those with abnormal UA, 39.5 for those with normal UA.	Nulliparity: Not reported IOL: Not reported.
Moraitis (unpublished)	Prospective cohort, Single Hospital, Cambridge, UK	N=3615 Universal, nulliparous only, >36 weeks	PI >90th centile Blinded.	Mean 36 weeks	NICU admission, metabolic acidosis (umbilical artery PH <7.10), 5-min Apgar score <7, composite neonatal morbidity (1 or more of the above), composite severe neonatal morbidity, SGA <10th centile, SGA <3rd centile	40.4 (39.3-41.1)	Nulliparity: All IOL: 36.1% for those with abnormal UA doppler, 33.1% for those with normal UA doppler.

Schulman 1989	Prospective cohort, Single Hospital, NY, USA	N=255 Mixed	S/D ratio >3 Not blinded.	Around 30 weeks	SGA <15th centile	Not reported	Nulliparity: Not reported IOL: Not reported.
Sijmons 1989	Prospective cohort Single Hospital, Netherlands	N=368 Mixed (randomly selected)	PI>95th centile Blinded UA doppler	At 28 and34 weeks	SGA <10th centile, SGA <3rd centile	Not reported	Nulliparity: Not reported IOL: Not reported.
Valino 2016a	Retrospective cohort, 3 NHS hospitals, South East England, UK May 2011- August 2014	N=8262 Universal	PI >95th centile PI >90th centile Not blinded	30+0- 34+6 weeks Mean 32.2 weeks	Term PET, term SGA <10th centile, Stillbirth, CS for fetal distress, Cord arterial PH <7.0, 5-minute Apgar score <7, NICU admission	Mean 40.0 weeks	Nulliparous: 49.2% IOL: 15.5%
Valino 2016b	Retrospective cohort, 2 NHS hospitals, South East England, UK February 2014-December 2014 (potential overlap with above)	N=3953 Universal	PI >95th centile Not blinded	35+0- 37+6 weeks Mean 36.1 weeks	PET, SGA <10th centile, CS for fetal distress, Cord arterial PH <7.0, 5-minute Apgar score <7, NICU admission	Mean 40.0 weeks	Nulliparous: 49.7% IOL: 19.1%
Weiner 1993	Prospective cohort, Single Hospital, Israel	N=142 Low risk, term only >41 weeks.	RI >95th centile. Not blinded	After 41 weeks	Composite adverse outcome:1) 5-minute Apgar <7, 2) NICU admission, 3) CS for fetal distress, SGA <5th centile	Mean 41.8 weeks	Nulliparous: n=43 IOL: Not reported.

Figure 9. Risk of bias graph of the studies included in the UA meta-analysis



Meta-analysis results

The summary results of the meta-analysis are presented in Table 5. The pattern of results was very similar to the POP study. A high resistance pattern of UA Doppler was associated with an increased risk of delivering an SGA infant or a severely SGA infant. However, the finding was not strongly predictive with positive LRs between 2.5 and 3.0. A high resistance pattern of UA Doppler was not associated with an increased risk of a range of indicators of neonatal morbidity. The summary ROC curves are presented in Figure 10. For some outcomes such as 5-minute Apgar score <7, caesarean section for fetal distress and pre-eclampsia (PET) the Rutter-Gatsonis model could not produce summary results despite an adequate number of studies. We additionally performed pooling of DORs for all the reported outcomes and illustrated the variation between studies using forest plots (Figure 11). Finally we used the Deeks' funnel plot asymmetry test to assess the risk of publication bias using the outcome of neonatal unit admission for the analysis. The test showed no evidence of publication bias (P=0.52; Figure 12)

Table 5. Summary results of meta-analysis of the UA Doppler.

Outcome	Number of studies	Number of patients	Summary Sensitivity (95% CI)	Summary Specificity (95% CI)	Summary Positive LR (95% CI)	Summary Negative LR (95% CI)
SGA <10th centile	8	19,203	21.7% (13.2-33.6)	91.8% (86.5-95.1)	2.65 (1.89-3.72)	0.85 (0.77-0.94)
SGA <3rd centile	5	53,907	25.4% (14.0-41.5%)	90.4% (78.6-96.1%)	2.65 (1.92-3.66)	0.83 (0.75-0.91)
NICU admission	8	66,253	13.6 (6.8-25.3)	89.9 (83.5-94.0)	1.35 (0.93-1.97)	0.96 (0.90-1.03)
Neonatal acidosis	5	9629	12.0% (5.3-25.0)	91.1% (81.0-96.1)	1.34 (0.86-2.08)	0.97 (0.91-1.02)
Severe APO*	4	58,866	9.3% (4.8-17.5)	88.3% (74.5-95.2)	0.80 (0.44-1.46)	1.03 (0.95-1.11)

SGA, Small for gestational age; LR, Likelihood ratio; CI, Confidence intervals; APO, Adverse pregnancy outcome

*The definition varied between studies and includes one or more of the following: stillbirth, neonatal death, hypoxic ischemic encephalopathy, inotrope support, or severe metabolic acidosis.

Figure 10. Summary ROC curves for the UA Doppler at predicting: A. NICU admission, B. Neonatal Metabolic acidosis, C. SGA (<10th centile), D. Severe SGA (<3rd centile).

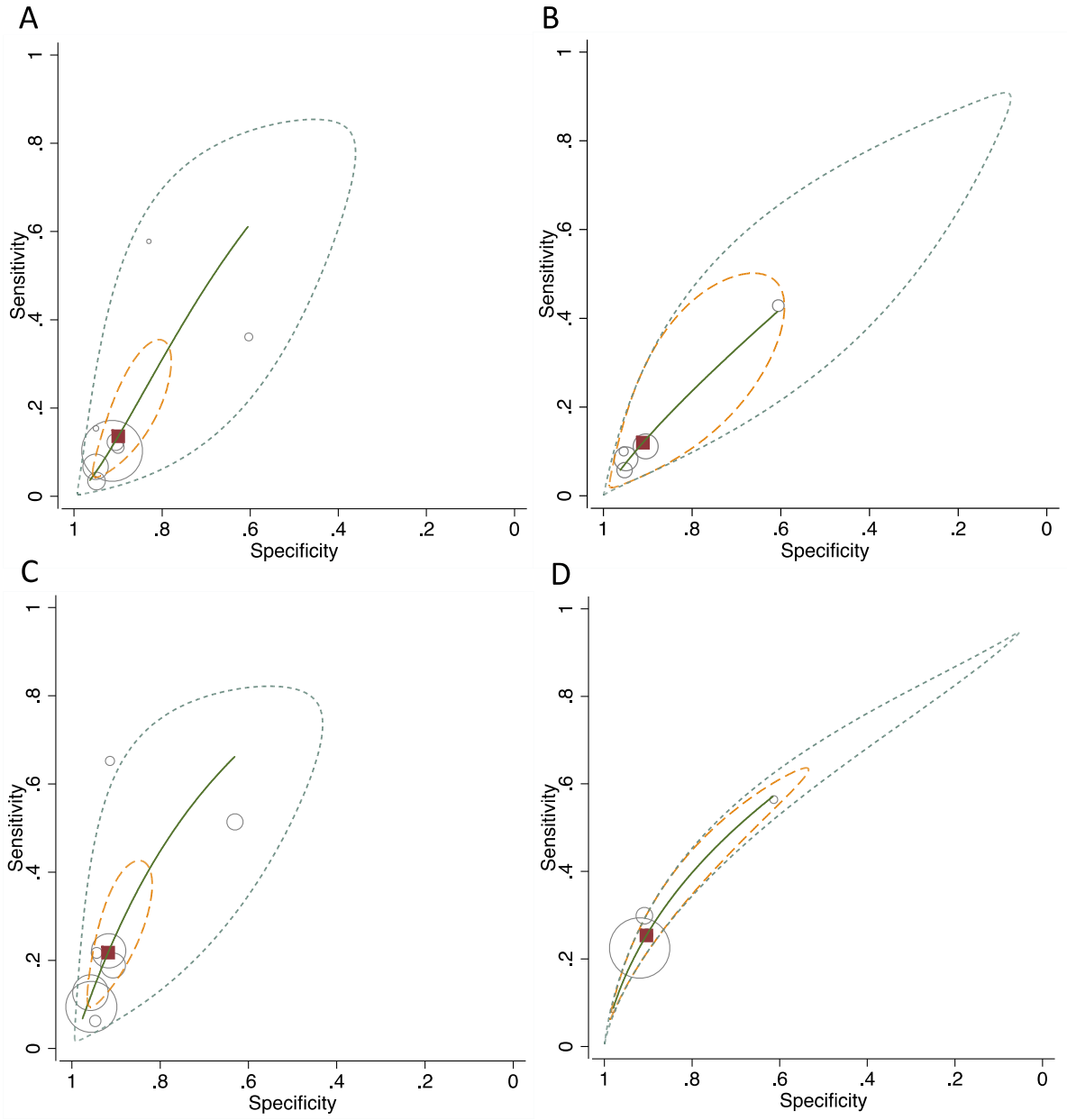
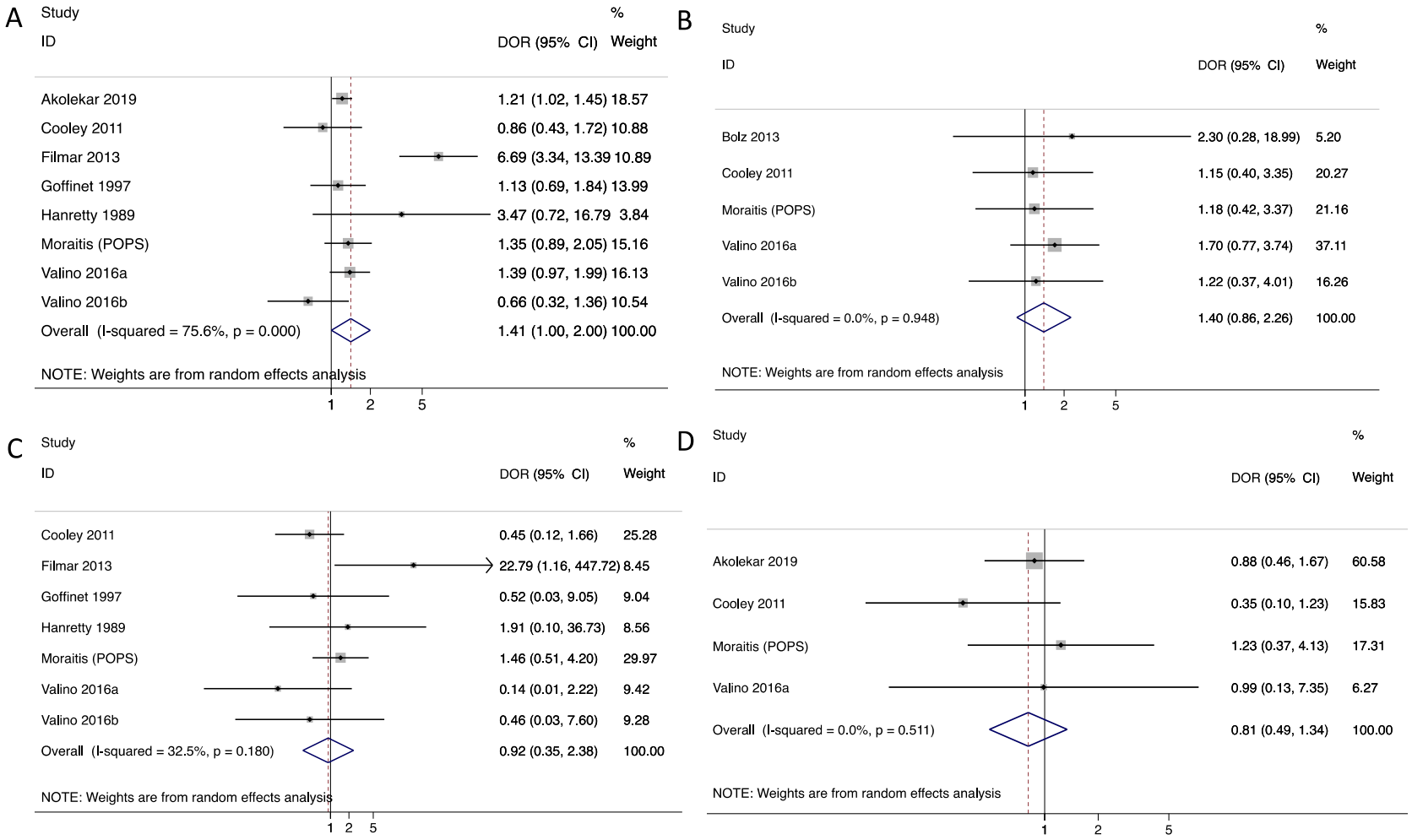
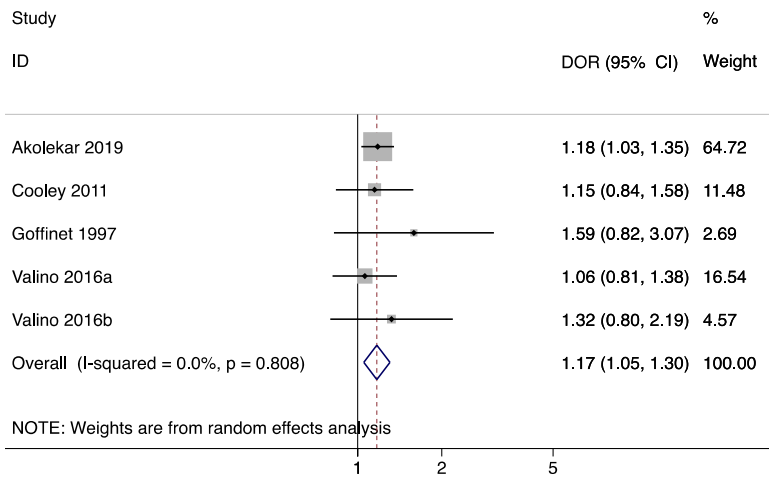


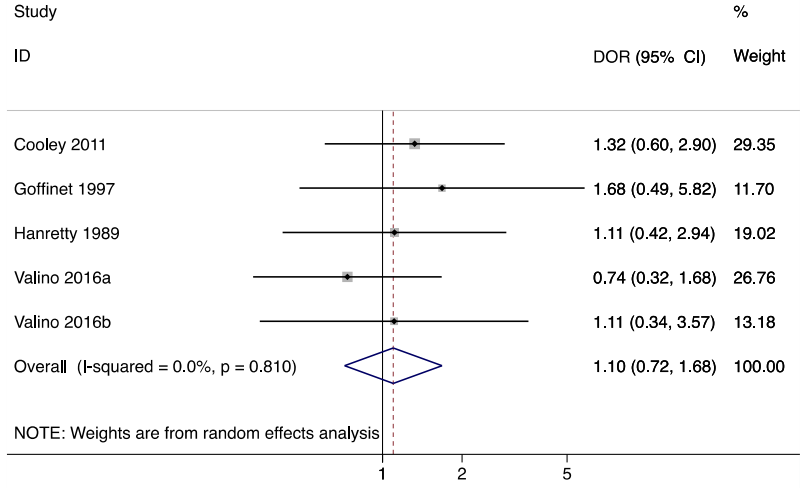
Figure 11. Meta-analysis of DORs of UA Doppler at predicting: A. NICU admission, B. Neonatal metabolic acidosis, C. 5-minute Apgar score <7, D. Severe adverse perinatal outcome, E. Caesarean section for fetal distress, F. Pre-eclampsia, G. SGA (<10th centile), H. Severe SGA (<3rd centile)



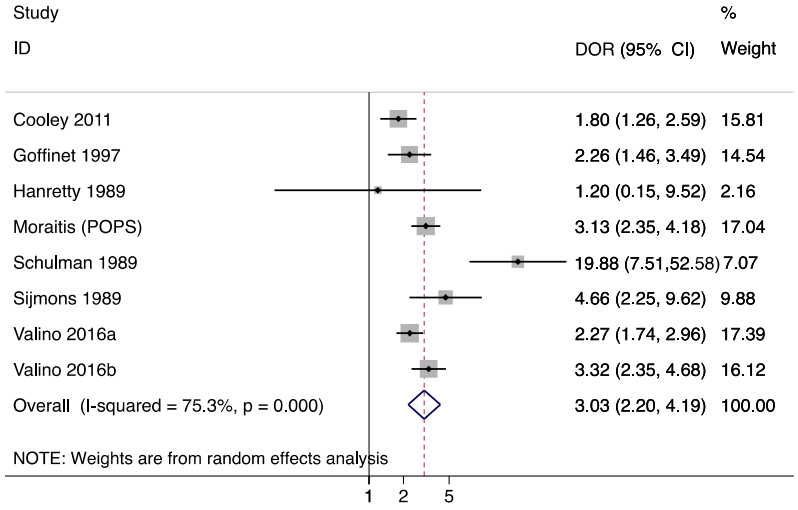
E


Study
ID
DOR (95% CI)
%
Weight
Akolekar 2019
1.18 (1.03, 1.35) 64.72
Cooley 2011
1.15 (0.84, 1.58) 11.48
Goffinet 1997
1.59 (0.82, 3.07) 2.69
Valino 2016a
1.06 (0.81, 1.38) 16.54
Valino 2016b
1.32 (0.80, 2.19) 4.57
Overall (I-squared = 0.0%, p = 0.808)
1.17 (1.05, 1.30) 100.00
NOTE: Weights are from random effects analysis
1
2
5

F


Study
ID
DOR (95% CI)
%
Weight
Cooley 2011
1.32 (0.60, 2.90) 29.35
Goffinet 1997
1.68 (0.49, 5.82) 11.70
Hanretty 1989
1.11 (0.42, 2.94) 19.02
Valino 2016a
0.74 (0.32, 1.68) 26.76
Valino 2016b
1.11 (0.34, 3.57) 13.18
Overall (I-squared = 0.0%, p = 0.810)
1.10 (0.72, 1.68) 100.00
NOTE: Weights are from random effects analysis
1
2
5

G


Study
ID
DOR (95% CI)
%
Weight
Cooley 2011
1.80 (1.26, 2.59) 15.81
Goffinet 1997
2.26 (1.46, 3.49) 14.54
Hanretty 1989
1.20 (0.15, 9.52) 2.16
Moraitis (POPS)
3.13 (2.35, 4.18) 17.04
Schulman 1989
19.88 (7.51,52.58) 7.07
Sijmons 1989
4.66 (2.25, 9.62) 9.88
Valino 2016a
2.27 (1.74, 2.96) 17.39
Valino 2016b
3.32 (2.35, 4.68) 16.12
Overall (I-squared = 75.3%, p = 0.000)
3.03 (2.20, 4.19) 100.00
NOTE: Weights are from random effects analysis
1
2
5

H

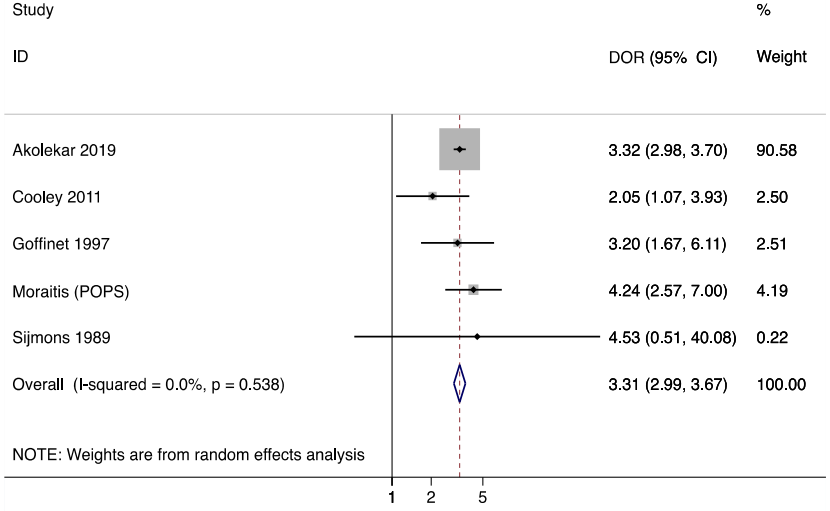
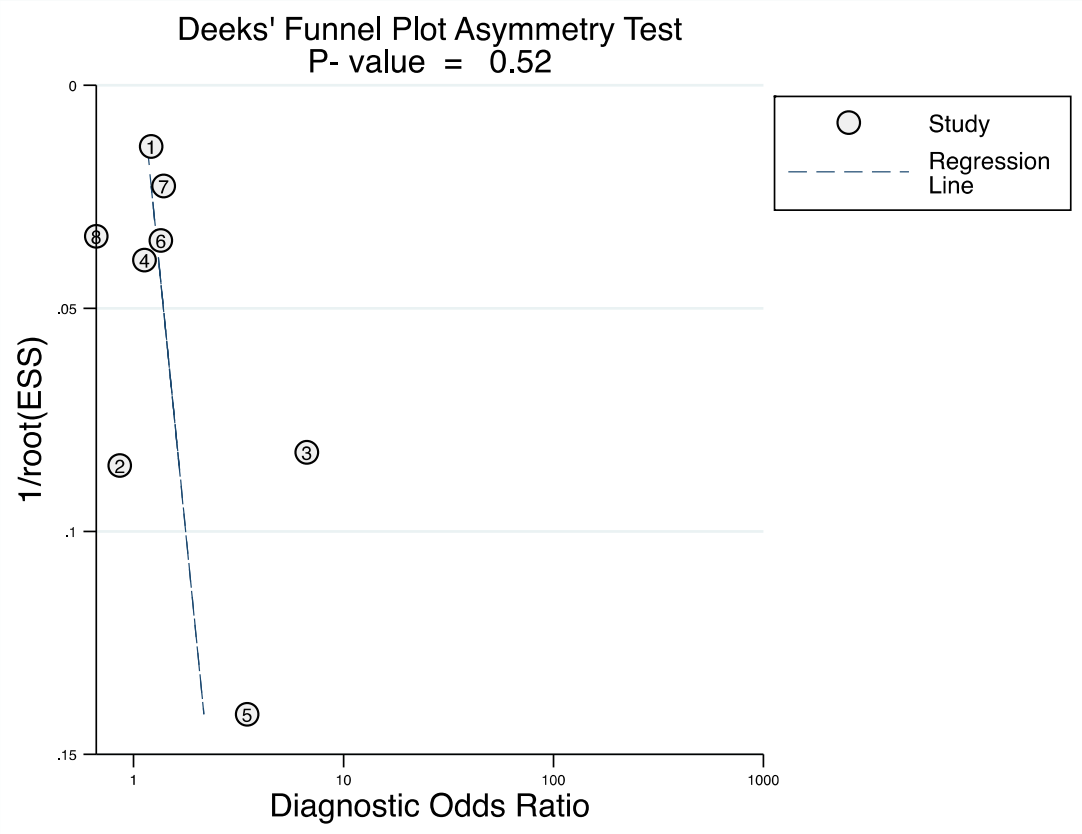

Study
ID
DOR (95% CI)
%
Weight
Akolekar 2019
3.32 (2.98, 3.70) 90.58
Cooley 2011
2.05 (1.07, 3.93) 2.50
Goffinet 1997
3.20 (1.67, 6.11) 2.51
Moraitis (POPS)
4.24 (2.57, 7.00) 4.19
Sijmons 1989
4.53 (0.51, 40.08) 0.22
Overall (I-squared = 0.0%, p = 0.538)
3.31 (2.99, 3.67) 100.00
NOTE: Weights are from random effects analysis
1
2
5

Figure 12.Deeks' funnel plot for publication bias for UA Doppler for the prediction of neonatal unit admission.



Discussion

The main finding of this chapter was that the umbilical artery Doppler has moderate predictive accuracy for detecting SGA and severely SGA infants. However, it did not predict neonatal morbidity at term. The results were very similar in both the POP study and the meta-analysis which included the POP study and other published studies. The only notable difference between the analysis of the POP study and the meta-analysis including the POP study is that the association in the former was slightly stronger for severe SGA. The outcome of SGA is used as a proxy for FGR. SGA is used a proxy for FGR but it is recognised that only a proportion of SGA infants are small due to FGR. As the threshold for defining SGA is lowered, the proportion of cases so defined which are truly FGR increases. Hence, the stronger association with severe SGA is most likely explained by a true association between high resistance patterns of UA Doppler and FGR.

The similar associations between the POP study and the meta-analysis is reassuring. Of all the studies evaluated, only the POP study blinded both the Doppler result and fetal biometry. The failure to blind studies could lead to bias. First, revealing the results could lead to interventions which then improve the outcome of the pregnancy. In this case, an investigation which is truly predictive for adverse outcome may not appear to be so when evaluated in a study where the result is revealed as knowledge of the result leads to interventions which prevent the adverse outcome. However, revealing the result could also lead to a non-informative test being wrongly identified as predictive of adverse outcome. The primary intervention following a concerning ultrasound finding is to deliver the baby which, if performed preterm or at early term, can cause iatrogenic morbidity. Hence a non-informative test could appear to be associated with adverse neonatal outcome when evaluated in a study where the result is revealed as revealing the result leads to interventions which cause iatrogenic morbidity. Moreover, if outcomes include events that are defined on the basis of the results of the diagnostic test being evaluated there is the risk of ascertainment bias. For example, if the presence of abnormal UA Doppler is used to define Caesarean section (CS) for fetal distress, there could be an association between the two because the test was being used to classify the outcome.

The lack of association between UA Doppler and adverse neonatal outcome is likely explained due to two reasons. First, the minority of term SGA infants have abnormal UA Doppler. This chapter showed that about 1 in 5 of the SGA infants born below the 10th birthweight centile and 1 in 4 of those born below the 3rd birthweight centile had abnormal UA Doppler. Second, only a small percentage of overall morbidity at term is associated with abnormal fetal growth. For example, previous studies of

perinatal death at term demonstrated that only 1 in 3 stillbirths at term are associated with abnormal fetal growth.[2] This association would likely be even weaker for other outcomes such as NICU admission which includes morbidity for various reasons not related to the fetal size such as neonatal infection. It is plausible that UA Doppler would be more strongly predictive of adverse neonatal outcome in fetuses which were actually SGA and this has been confirmed in a previous analysis of the POP study.[9]

Given that UA Doppler appears to be predictive of FGR in low risk women it might be regarded as surprising that the RCTs of its use as a screening test failed to demonstrate any benefit. However, a previous analysis of required sample sizes of screening and intervention to prevent stillbirth demonstrated that, even if a test had a positive LR of 5 for perinatal death, and was observed in 5% of women, and even if the test was coupled to an intervention that reduced the risk of perinatal death by 50%, an RCT of screen versus no screen would need to recruit >100,000 to achieve 90% power see Supplementary Figure 10 in Flenady et al 2016.[109] Thus, the Cochrane meta-analysis of low-risk pregnancies is significantly underpowered to identify a reduction in perinatal death.

In conclusion, a high resistance pattern of UA Doppler is somewhat predictive of the risk of delivering an SGA infant. The strength of prediction was similar using a blinded 36wkGA scan in unselected nulliparous women in the POP study as it was in a systematic review of the wider literature.

Chapter 5. Systematic review of the diagnostic effectiveness of universal ultrasonic screening using late pregnancy cerebro-placental ratio in the prediction of adverse perinatal outcome.

Introduction

The cerebroplacental ratio (CPR) is the ratio of middle cerebral artery (MCA) Doppler to umbilical artery (UA) Doppler. It is considered to be a marker of the "brain sparing" effect which is the redistribution of fetal circulation towards the brain and is a sign of fetal compromise. This effect is believed to be associated with placental dysfunction which reduces the amount of oxygen in fetal circulation. These reduced levels of oxygen activate the peripheral arterial chemoreceptors (PACs)[110] which in turn reduce the resistance to blood flow to the brain. Clinically, the increased vascular resistance in the placenta and the reduced vascular resistance in the fetal brain are measured using Doppler flow velocimetry in the UA and MCA respectively.

The CPR has been used for monitoring and management of pregnancies with fetal growth restriction (FGR). A recent systematic review showed that CPR appears to be useful in predicting perinatal death in pregnancies with suspected FGR.[16] However, the evidence is not clear whether we can use the CPR as part of universal third-trimester screening. Some studies have shown that the CPR can predict neonatal morbidity and intrapartum fetal compromise in normally grown fetuses.[111, 112] Contrary to those findings a recently published large study which performed the CPR as part of routine ultrasound at 36 weeks' gestation showed that the CPR has poor predictive accuracy for adverse pregnancy outcome in an unselected population.[97]

The aim of this chapter is to assess whether measurement of the CPR in the third-trimester predicts adverse pregnancy outcome in unselected, low and mixed-risk pregnancies.

Methods

Publication statement

This chapter was published as a chapter in the HTA report (Appendix 9C).[113] Some passages were quoted verbatim from this paper where I wrote the text for this publication. The figures and tables have been reproduced after gaining permission by the journal.

Sources for meta-analysis

A systematic search was performed using Medline, EMBASE, the Cochrane database of Systematic Reviews (CDSR) and Cochrane Central Register of Controlled Trials (CENTRAL) and ClinicalTrals.gov up to June 2019. No restrictions for language or geographic location were applied. The studies were identified using a combination of words related to "ultrasound", "pregnancy", "cerebroplacental", "cerebro-umbilical", "middle cerebral artery", and "fetal brain Doppler". We defined the cerebroplacental ratio as the ratio of middle cerebral artery (MCA) pulsatility index (PI) to the umbilical artery (UA) PI.

Study selection

Selection criteria included cohort or cross-sectional studies with singleton pregnancies where an ultrasound scan was performed ≥24wkGA. We included all studies where the ultrasound was performed as part of universal screening, studies that used low-risk populations only and studies with mixed-risk populations. We excluded studies that were focused on high risk patients such as FGR and studies that the ultrasound was performed during labour.

The CPR was commonly defined as the ratio of middle cerebral artery (MCA) pulsatility index (PI) to the umbilical artery (UA) PI. We included studies regardless of the threshold they used to define abnormality of the CPR and regardless of blinding of the result to the clinicians.

We included studies that reported the following outcomes: severe adverse perinatal outcome (which included stillbirth, neonatal death and hypoxic ischaemic encephalopathy); fetal growth abnormalities such as SGA (defined as birthweight <10th centile) and severe SGA (birthweight <3rd of <5th centile); adverse neonatal outcomes such as neonatal unit admission, 5-minute Apgar score <7, and neonatal metabolic acidosis (as defined in each study); Caesarean section or operative delivery (including both Caesarean section and instrumental delivery) for fetal compromise in labour. In cases of significant

population overlap between studies that reported the same outcomes we included the larger study in the meta-analysis.

Study quality assessment and statistical analysis

The literature search, study selection, and analysis ware performed independently by two authors (myself and Dexter Hayes) using Review Manager 5.3. Any differences were resolved in discussion with the senior author (GS). The methods employed are described in Chapter 2.

Results

Study characteristics and risk of bias assessment

The literature search flowchart is presented in Figure 13. We identified 16 studies[97, 111, 112, 114-126] that met our inclusion criteria involving 121,607 patients in total. The study characteristics are presented in Table 6. Four studies [97, 114, 115, 123](N= 85,059) included unselected pregnancies, seven studies[111, 112, 116, 117, 119, 122, 125] (N= 12,929) included only low-risk pregnancies and five studies[118, 120, 121, 124, 126] (N= 23,619) included mixed risk pregnancies. Nine studies (N= 87,208) were prospective and seven (N= 34,399) were retrospective. There was likely population overlap between the Akolekar 2015,[114] Akolekar 2019,[97] and Bakalis[115] studies. For the first two we reported different outcomes and for those outcomes that were the same we employed the data from the larger Akolekar 2019 study in the meta-analysis. The study published by Bakalis performed ultrasound at 32wkGA compared to the two Akolekar studies which performed ultrasound at around 36wkGA. There was also likely population overlap between the Khalil,[112] Monaghan[120] and Morales-Rosello[121] studies which reported different outcomes at the same tertiary maternity unit. Moreover, there was also likely population overlap between the Flatley,[118] Sabdia[124] and Twomey[126] studies. The study published by Twomey performed ultrasound at 32wkGA and the other two studies which performed ultrasound between 35 and 38 weeks reported different rates of nulliparity and different gestational age at delivery (Sabdia included preterm deliveries) which indicates that the potential population overlap was not significant. Finally, there was a complete population overlap between the studies published by Bligh but the two studies reported different outcomes.

The assessment of study quality was performed using the QUADAS-2 tool and is summarized in Figure 14. The main risk of bias was for reference standard due to the lack of blinding in the majority of studies. Only five studies [111, 116, 117, 122, 123](N=3079) blinded the results to the clinicians. The second more common risk of bias was for flow and timing due to the different gestational ages that the ultrasound was performed. Bakalis, Rial-Crestelo and Twomey performed ultrasound at around 32 to 33wkGA, and Prior (both studies) and Stumpfe performed the ultrasound prior to induction of labour (interval between ultrasound and delivery less than 72 hours). Hence, the results of the above studies might not be applicable to universal screening at 36wkGA. One study (Maged et al.) had unclear risk of selection bias as they did not specify if the selection of patients was consecutive or random.

Figure 13.Literature search PRISMA flow diagram for the systematic review on cerebro-placental ratio.

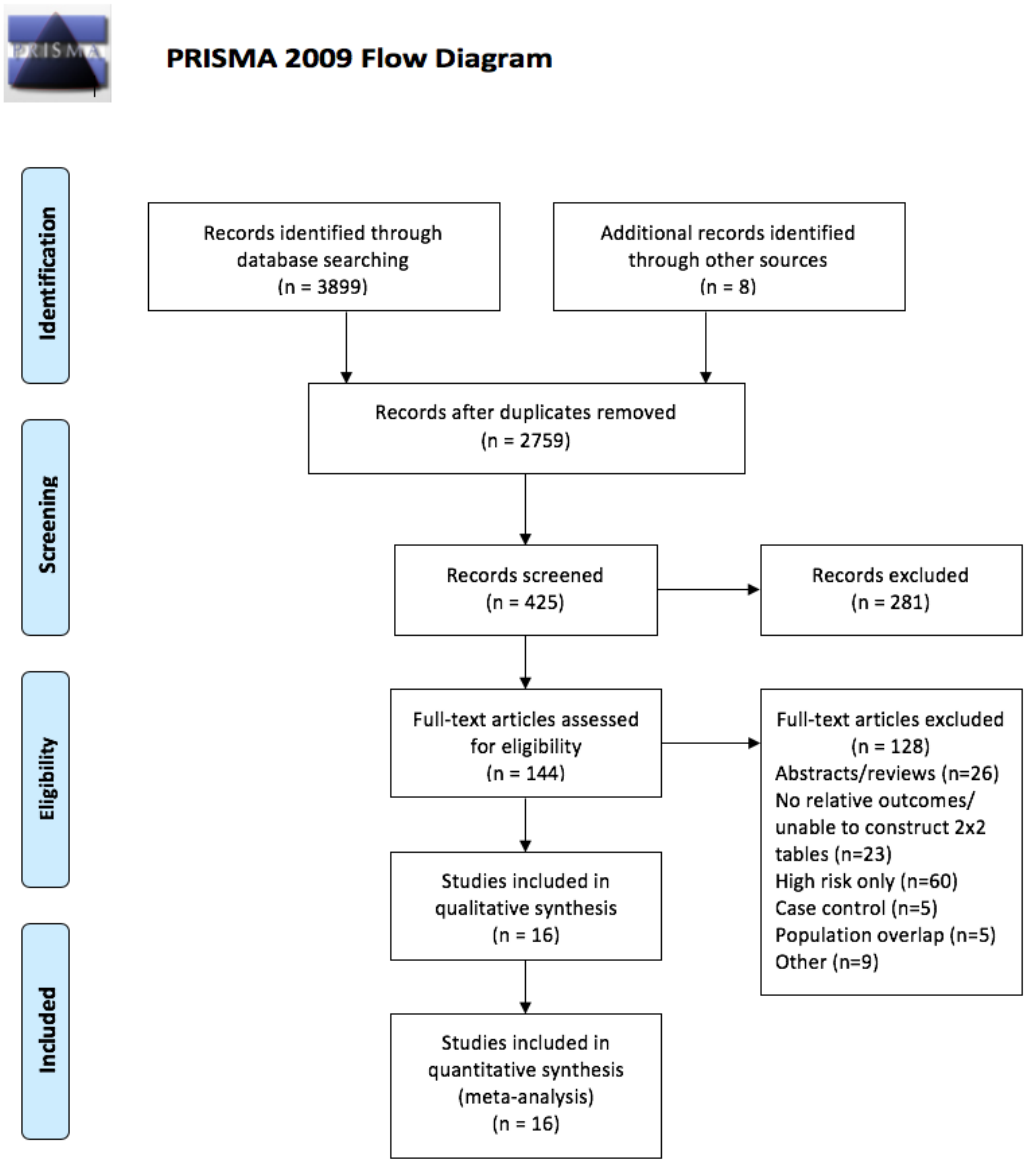


Table 6. Characteristics of the studies included in the meta-analysis of cerebroplacental ratio to predict adverse pregnancy outcome.

First Author (Year)	Type of Study, Setting	Number of fetuses and selection (All singleton, non-anomalous unless otherwise stated)	Index test CPR = MCA PI/ Umbilical Artery PI (unless otherwise stated)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Akolekar 2015	Prospective cohort. 2 NHS hospitals (King's College London, Medway Maritime Hospital), UK. (Between February 2014 and December 2014).	N= 6038. Universal screening.	CRP < 5th centile. Not blinded.	35+0 to 37+6 Median 36.1 (IQR 36.0-36.6)	Cord arterial PH <7.0, 5-min Apgar score <7, NICU admission.	Median 39.9 (IQR 39.0-40.7)	Nulliparous: 49.8% IOL: 20% overall.
Akolekar 2019	Prospective cohort, 2 NHS Hospitals (King's college, Medway Maritime Hospital), UK (Between March 2014 and	N= 47,211 Universal screening.	CRP < 10th centile. Not blinded.	Between 35+0 and 37+6 weeks.	Adverse perinatal outcome (composite of stillbirths, neonatal deaths and HIE grade 2 or 3), perinatal hypoxia (composite of cord artery PH <7.0 and	Median ga at delivery 40.0 (39.0-40.9) weeks.	Nulliparous: 45.4% for those with no adverse outcome, 58.5% for those with adverse outcome. IOL: Not reported.

	September 2018; Significant population overlap with Akolekar 2015 study)				venous <7.1, 5-minute Apgar score <7, NICU admission for >24 hours), CS for fetal compromise, SGA <3rd centile.		
Bakalis 2015	Prospective cohort. 3 NHS hospitals (KCL, UCL, Medway Maritime Hospital), UK (Between May 2011 to August 2014; likely population overlap with Akolekar 2015 and 2019 studies)	N= 30,780. Universal screening.	CRP < 5th centile. Not blinded.	30+0 to 34+6, Mean 32.3 (IQR 32.0-32.9)	Stillbirth; Emergency caesarean for fetal distress (ECFS), cord art PH <7.0; cord venous PH ,7.1; 5-min Apgar score <7; NNU admission; NICU admission.	Median 40 (IQR 39.0-40.9)	Nulliparous: 50.2% Further analysed in SGA vs. AGA and delivery < 2 weeks from scan vs. > 2 weeks from scan. IOL: 14.5% overall.
Bligh 2018 (A/UOG)	Prospective cohort, 1 hospital, Brisbane, Australia (May 2014 – August 2016)	N= 437 Low risk Uncomplicated, term only.	CPR <10th centile Blinded.	From 36+1 weeks forward. Within 2 weeks of delivery	CS for fetal distress. Composite adverse neonatal outcome (cord artery PH <7.10, 5-min Apgar <7, or NICU admission)	Median 40 (IQR 39.3-40.9)	Nulliparous: 87.4% IOL: Not reported.

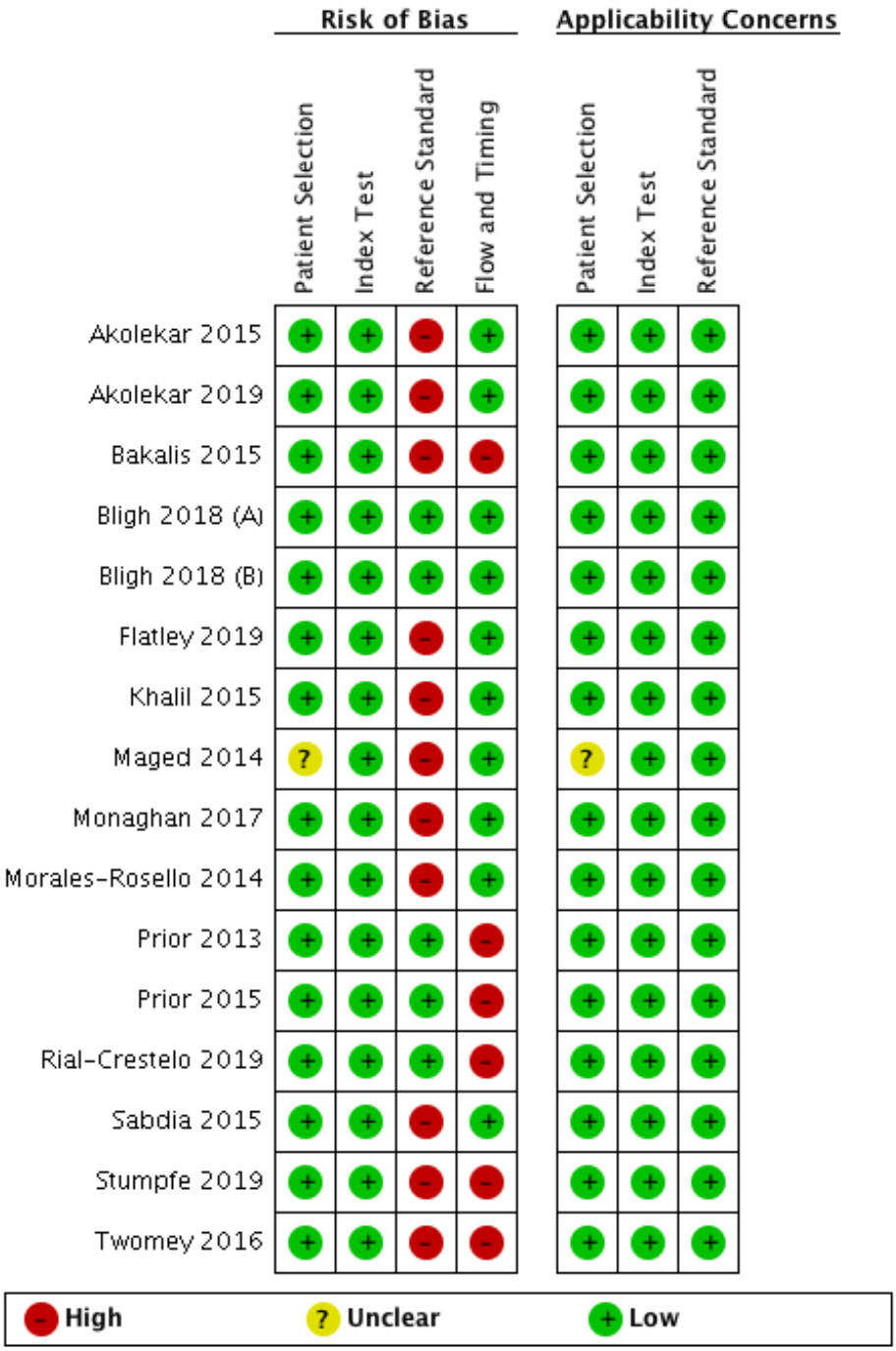
Bligh 2018 (B/FDT)	Prospective cohort, 1 hospital, Brisbane, Australia (May 2014 – August 2016)	N= 437 Low risk Uncomplicated, term only.	CPR <10th centile CPR <5th centile Blinded.	From 36 Within 2 weeks of delivery	SGA <10th centile SGA <5th centile	Median 40 (IQR 39.3-40.9)	Nulliparous: 87.4% IOL: Not reported.
Flatley 2019	Retrospective cohort, 1 hospital, Brisbane, Australia (2010-2015) (Likely some population overlap with Bligh 2018)	N= 2425 Mixed risk Excluded preterm delivery <37 weeks, maternal hypertension and diabetes mellitus.	CPR <10th centile. Not blinded.	Between 36-38 wks	Cord artery PH <7.00, 5-minute Apgar ≤3, NICU admission, perinatal death. Composite of all the above (SCNO) CS for fetal distress. SGA <10th centile, SGA <5th centile.	Term only, 54.5% of those with abnormal CPR delivered <39 wks, 36,4% of those with normal CPR	Nulliparous: 65.4% of those with abnormal CPR, 48.0% of those with normal CPR. IOL: 46.4% for those with abnormal CPR, 39.5% for those with normal CPR.
Khalil AJOG 2015	Retrospective cohort. 1 tertiary NHS hospital (St George's), UK (2000-2013)	N= 9772 Low risk. Term only. For the analysis of operative delivery for fetal distress , the patients that had elective CS were excluded.	CPR < 0.6765 MoM Not blinded.	Within 2 wks of delivery. Median 40.4 for those admitted to NNU, 40.4 wks for those not admitted.	NNU admission Operative delivery of fetal distress, (including instrumental delivery and CS),	Median 41.1 for both those admitted and those not admitted to NNU.	Nulliparous: 65.2% of those admitted to NNU, 54.6% for those not admitted to NNU. IOL: 44.1% for NNU 39.4% for no NNU.

Maged 2014	Prospective cohort 1 hospital, Cairo, Egypt	N= 100 Low risk. Included those delivered between 40-42 weeks. Excluded PPROM , APH, patients in labor and maternal HTN/DM.	CPR < 1.05 Not blinded.	37.8 weeks for those with adverse outcome, 39.5 weeks for those with normal.	C-Section for fetal distress (CSFD). Composite adverse pregnancy outcome defined as 1 or more of: CSFD, 5-min Apgar <7, MAS, NICU admission.	283.1 days for those with adverse outcome, 281.7 for those with normal outcome.	Nulliparous: Not reported. IOL: Not reported
Monaghan 2017	Retrospective cohort 1 NHS hospital (St George's), UK January 2008- June 2016 (Likely population overlap with Khalil 2015)	N= 7013 Mixed risk (had USS based on NHS indications). Only included those delivered after 36 weeks.	CPR <10th centile CPR <5th centile Not blinded	36.4 wks for all live births, 37 wks for perinatal deaths	Perinatal death	Median: 40.1 weeks for all live births, 39 weeks for perinatal deaths	Nulliparous: Not reported. IOL: Not reported.
Morales-Rosello 2014	Retrospective cohort 1 NHS hospital (St George's), UK, 2002-2012 (Likely population overlap with Khalil	N= 11,576 Mixed risk . Term only with USS within 14 days of delivery.	CPR <0.6765 MoM Not blinded	Mean: 40.1 +/-1.5 weeks.	SGA <10th centile.	Mean 40.8 +/-1.3	Nulliparous: Not reported. IOL: Not reported.

	2015 and Monaghan 2017)						
Prior 2013	Prospective cohort. 1 NHS hospital (Queen Charlotte's and Chelsea), UK. (March 2011-March 2014)	N= 400 Low risk. Term only. Recruited before active labor. Excluded PET, FGR, intrauterine infection.	CPR <10th centile Blinded.	Mean: 40 weeks + 2 days. (Range: 37+0 – 42+1)	CS for fetal compromise, 5-min Apgar <7, Cord arterial PH<7.20, NNU admission	Within 72 hours from scan	Nulliparous: 65.5% IOL: Not reported.
Prior 2015	Prospective cohort 1 tertiary NHS hospital (Chelsea), UK. (Likely population overlap with Prior 2013 study)	N= 775 Low risk Term only. Recruited before active labor or IOL (for postdates or social). Excluded SGA/FGR, PIH/PET, PPROM.	CRP <0.6765 MoM Blinded.	Median 41 weeks (range 37-42)	CS for fetal distress, 5-min Apgar score <7, cord arterial PH<7.20, NNU admission.	Within 72 hours from scan	Nulliparous: 80.8% IOL: Not reported.
Rial-Crestelo 2019	Prospective cohort, 1 hospital, Barcelona. January 2013- December 2016	N= 1030 Universal screening	CPR <10th centile Doppler blinded for those with EFW >10th centile.	Between 32+0 and 34+6 wks. Mean 33 wks	SGA <10th centile	Mean 40 weeks	Nulliparous: 70% of those born SGA, 54% of non-SGA. IOL: Not reported.
Sabdia 2015	Retrospective cohort	N= 1381 Mixed risk.	CPR < 10th centile (1.20). Not blinded.	Between 35 and 37 weeks	Operative delivery for fetal distress (CS or instrumental), 5 min	Median ga 36 wks for those with abnormal	Nulliparous: 53.9% of those with abnormal

	1 hospital, Brisbane, Australia (June 1998-November 2013)	Included cephalic with UA PI < 95th centile.			Apgar score <7, NICU admission.	CPR, 38 wks for normal CPR	CPR, 40.4% of those with normal CPR IOL: Not reported.
Stumpfe 2019	Retrospective cohort Single tertiary centre, Germany (January 2016- April 2017)	N= 1008 Low risk, Term only, excluded those in labour, elective CS, EFW <10th centile.	CPR <0.6765 MoM Not blinded.	Term , within 72 hours of delivery	CS for fetal distress, 5-min Apgar score <7, cord arterial PH <7.10	Term (not further specified)	Nulliparous: Not specified IOL: 42.4% overall.
Twomey 2016	Retrospective cohort. 1 l hospital, Brisbane, Australia. (January 2007-December 2013) (Population overlap with Sabdia 2015)	n =1224. Mixed risk. Excluded women that had elective caesarean section.	CPR <1. Not blinded.	30–34 wks. Median 32.1 wks.	CS for fetal compromise, Cord PH <7.0, 5-minute Apgar ≤3, NNU admission, SGA <10th centile, SGA <5th centile.	Mean ga 32 wks for those with CPR <1, 37 wks for those with CPR>1.	Nulliparous: 43.2% IOL: Not reported

Figure 14. Risk of bias and applicability concerns using the QUADAS-2 tool for the studies included in the meta-analysis of cerebro-placental ratio.



Meta-analysis results

The summary results for the diagnostic accuracy of CPR at predicting adverse pregnancy outcomes are presented in Table 7. Overall, the strongest associations were with the risk of delivering an SGA or severely SGA infant and the positive LRs were in the region of 3.5 to 4.0, which was stronger than for UA on its own. Moreover, unlike the UA Doppler in the previous chapter, a low CPR was associated with a statistically significantly increased risk of neonatal morbidity. However, the strength of prediction was weak, with positive LRs between 1.5 and 3.0.

The summary ROC curves are presented in Figure 15. Generally, the larger studies reported lower sensitivities and higher specificities for all the outcomes. We also present the pooling of the DORs in Figure 16. These demonstrate that for many of the outcomes there was a very high level of heterogeneity between the studies.

Finally we used the Deeks' funnel plot asymmetry test to assess the risk of publication bias using the outcome of neonatal unit admission for the analysis. The test showed no significant risk of publication bias (P=0.28; Figure 17)

Table 7. Diagnostic accuracy of CPR in predicting adverse pregnancy outcome.

Outcome	Studies	Patients	Summary sensitivity (95% CI)	Summary specificity (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)
Neonatal unit admission	9	52,554	22.9% (10.5-42.9%)	89.1% (82.1-93.5%)	2.10 (1.60-3.68)	0.86 (0.74-1.01)
5-minute Apgar score <7	8	35,586	13.5% (8.8-20.2%)	92.1% (90.0-93.8%)	1.71 (1.22-2.40)	0.94 (0.89-0.99)
Neonatal metabolic acidosis	7	16,321	10.9% (6.9-16.8%)	91.2% (87.9-93.6%)	1.24 (0.94-1.62)	0.98 (0.94-1.01)
Severe adverse perinatal outcome	4	87,429	18.6% (10.6-30.6%)	90.9% (87.4-93.5%)	2.04 (1.49-2.80)	0.90 (0.81-0.99)
SGA (<10th centile)	5	16,692	26.7% (18.0%-37.7%)	93.0% (86.9%-96.4%)	3.82 (1.68-8.71)	0.79 (0.67-0.92)
Severe SGA (<3rd or <5th centile)	4	51,297	32.3% (20.1-47.5%)	91.2% (84.3-95.3%)	3.70 (1.38-9.97)	0.74 (0.57-0.96)
C-Section for fetal distress	9	68,506	25.9% (14.9-41.2%)	90.6% (87.6-92.9%)	2.75 (1.96-3.88)	0.82 (0.70-0.96)
Operative delivery for fetal distress	5	12,162	19.4% (13.2-27.6%)	92.6% (90.1-94.5%)	2.63 (1.81-3.83)	0.87 (0.80-0.94)

Figure 15. Summary ROC curves for the diagnostic performance of abnormal cerebroplacental ratio at predicting adverse pregnancy outcomes. A. Neonatal unit admission; B. 5-minute Apgar score <7; C. Neonatal metabolic acidosis; D. Severe adverse perinatal outcome (including stillbirth, neonatal death and hypoxic ischaemic encephalopathy); E. SGA (birthweight <10th centile); F. Severe SGA (<3rd or <5th centile); G. Caesarean section for fetal distress; H. Operative delivery for fetal distress (including both caesarean section and instrumental delivery)

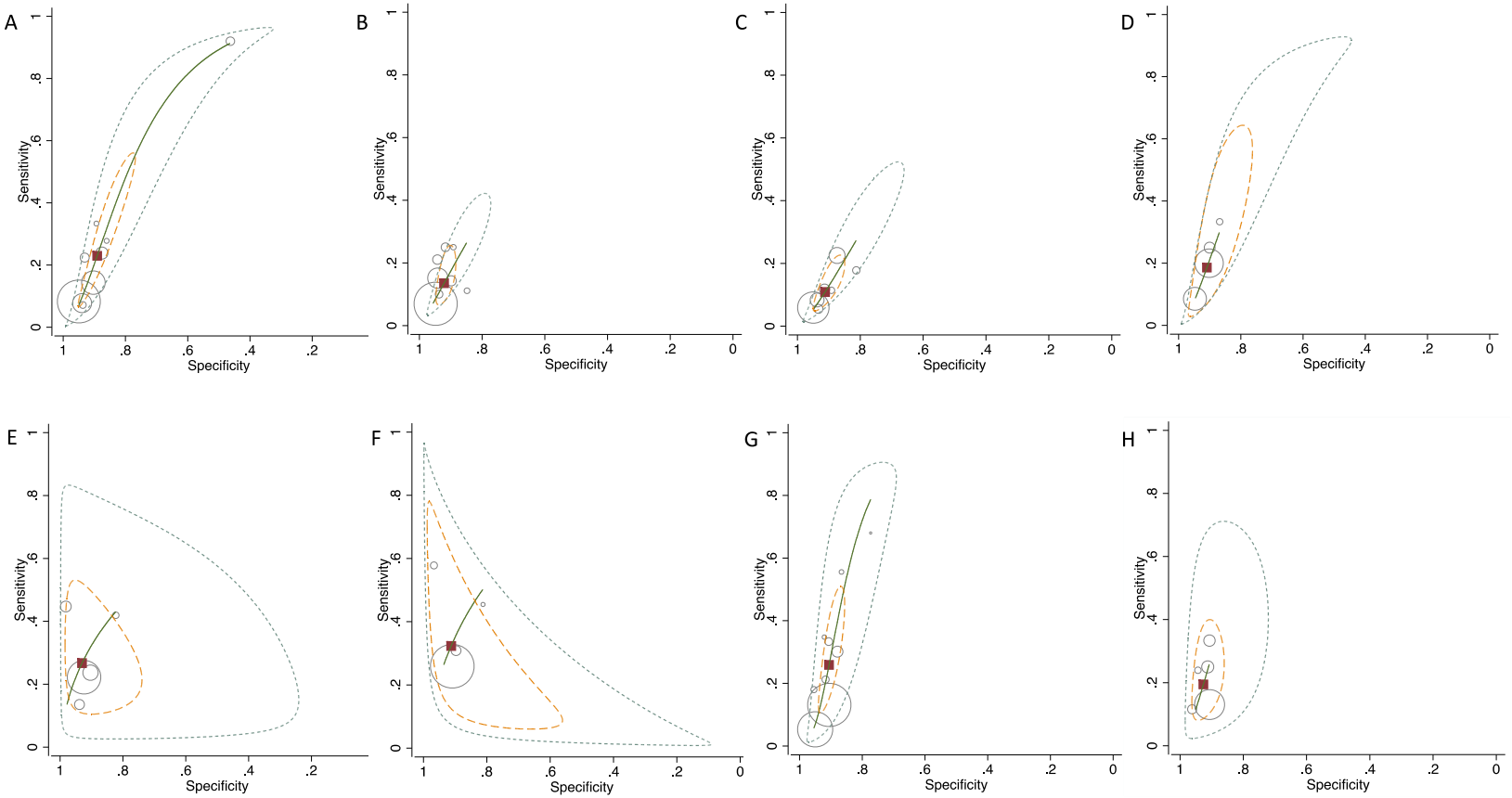
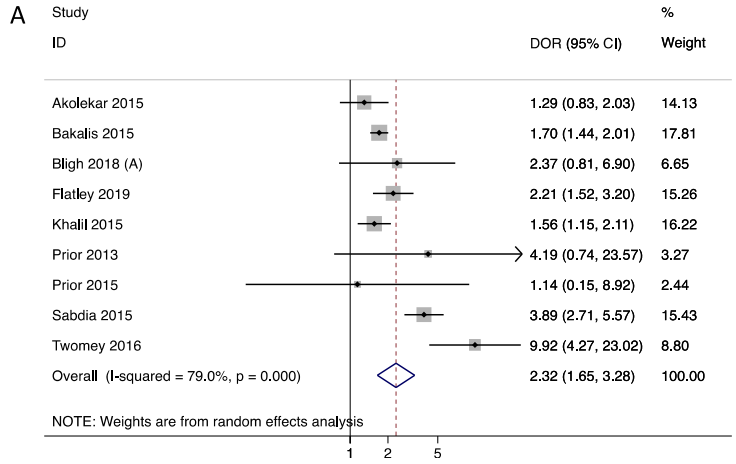
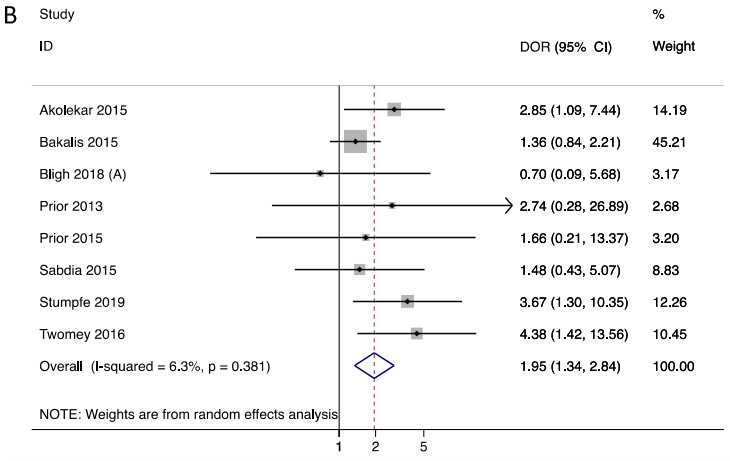
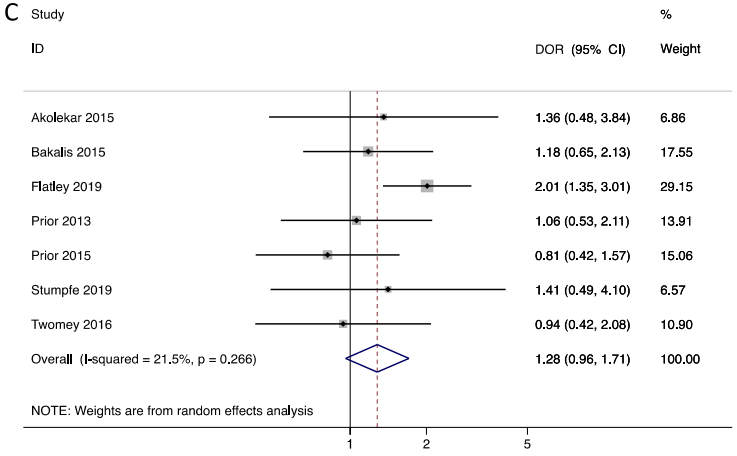
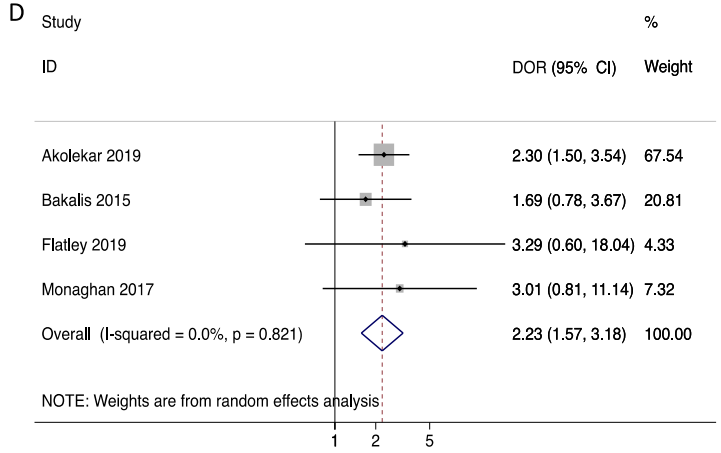


Figure 16. DORs for the diagnostic performance of abnormal cerebroplacental ratio at predicting adverse pregnancy outcomes: A. Neonatal unit admission; B. 5-minute Apgar score <7; C. Neonatal metabolic acidosis; D. Severe adverse perinatal outcome (including stillbirth, neonatal death and hypoxic ischaemic encephalopathy); E. SGA (birthweight <10th centile); F. Severe SGA (<3rd or <5th centile); G. Caesarean section for fetal distress; H. Operative delivery for fetal distress (including both caesarean section and instrumental delivery)

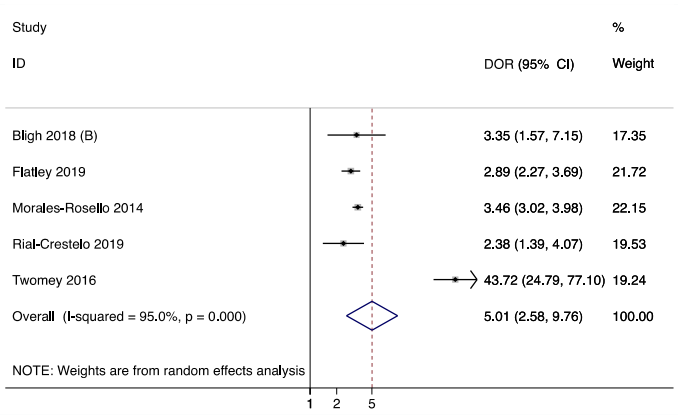




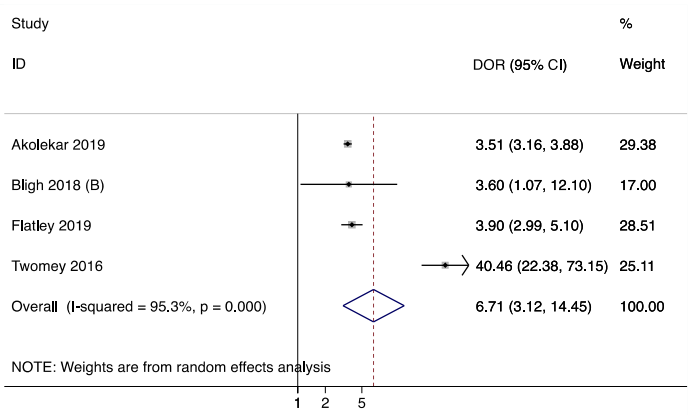




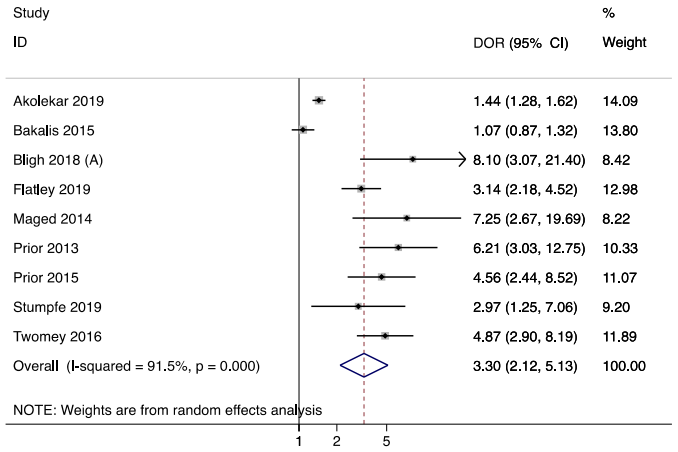
E


Study
ID
DOR (95% CI)
%
Weight
Bligh 2018 (B) 3.35 (1.57, 7.15) 17.35
Flatley 2019 2.89 (2.27, 3.69) 21.72
Morales-Rosello 2014 3.46 (3.02, 3.98) 22.15
Rial-Crestelo 2019 2.38 (1.39, 4.07) 19.53
Twomey 2016 43.72 (24.79, 77.10) 19.24
Overall (I-squared = 95.0%, p = 0.000) 5.01 (2.58, 9.76) 100.00
NOTE: Weights are from random effects analysis
1
2
5

F


Study
ID
DOR (95% CI)
%
Weight
Akolekar 2019 3.51 (3.16, 3.88) 29.38
Bligh 2018 (B) 3.60 (1.07, 12.10) 17.00
Flatley 2019 3.90 (2.99, 5.10) 28.51
Twomey 2016 40.46 (22.38, 73.15) 25.11
Overall (I-squared = 95.3%, p = 0.000) 6.71 (3.12, 14.45) 100.00
NOTE: Weights are from random effects analysis
1
2
5

G


Study
ID
DOR (95% CI)
%
Weight
Akolekar 2019 1.44 (1.28, 1.62) 14.09
Bakalis 2015 1.07 (0.87, 1.32) 13.80
Bligh 2018 (A) 8.10 (3.07, 21.40) 8.42
Flatley 2019 3.14 (2.18, 4.52) 12.98
Maged 2014 7.25 (2.67, 19.69) 8.22
Prior 2013 6.21 (3.03, 12.75) 10.33
Prior 2015 4.56 (2.44, 8.52) 11.07
Stumpfe 2019 2.97 (1.25, 7.06) 9.20
Twomey 2016 4.87 (2.90, 8.19) 11.89
Overall (I-squared = 91.5%, p = 0.000) 3.30 (2.12, 5.13) 100.00
NOTE: Weights are from random effects analysis
1
2
5

H

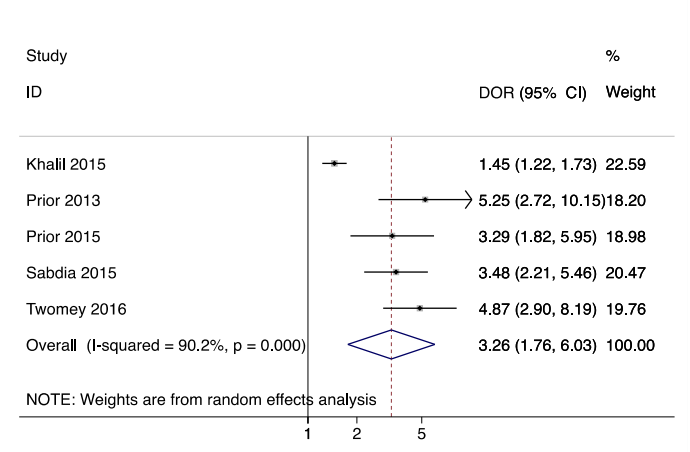
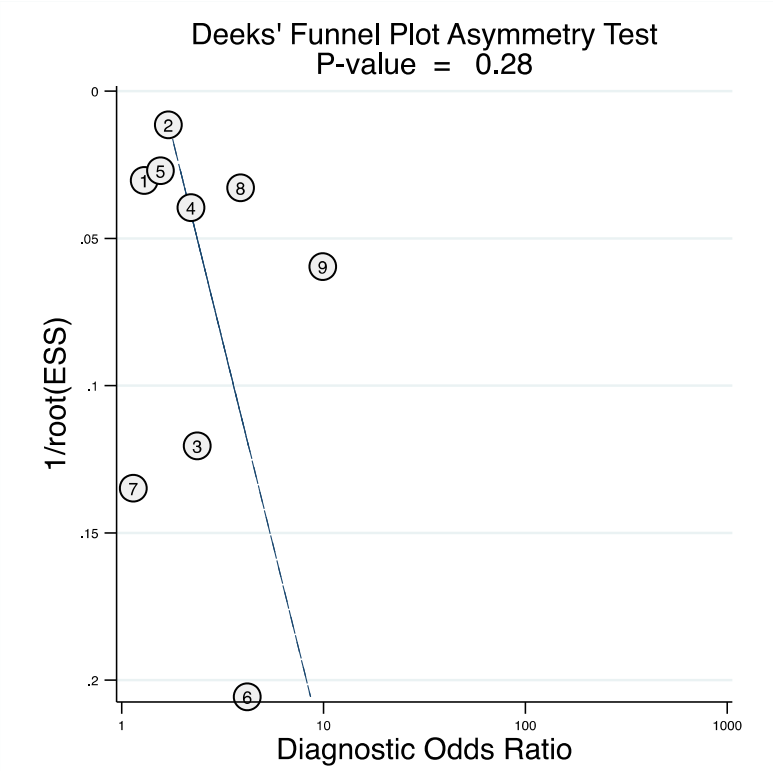

Study
ID
DOR (95% CI)
%
Weight
Khalil 2015 1.45 (1.22, 1.73) 22.59
Prior 2013 5.25 (2.72, 10.15) 18.20
Prior 2015 3.29 (1.82, 5.95) 18.98
Sabdia 2015 3.48 (2.21, 5.46) 20.47
Twomey 2016 4.87 (2.90, 8.19) 19.76
Overall (I-squared = 90.2%, p = 0.000) 3.26 (1.76, 6.03) 100.00
NOTE: Weights are from random effects analysis
1
2
5

Figure 17. Deeks' funnel plot for publication bias for cerebroplacental ratio for the prediction of neonatal unit admission.



Discussion.

The meta-analysis demonstrated that the CPR performed slightly better than UA Doppler in identifying pregnancies at an increased risk of adverse outcome. In the case of SGA, the positive LRs were in the region of 3.5 to 4.0 compared with 2.5 to 3.0 for UA Doppler. Moreover, unlike UA Doppler, CPR was predictive, albeit weakly, of neonatal morbidity (positive LRs of <2.0).

With more than 100,000 patients included in this analysis, this is the largest systematic review to date on the prognostic accuracy of CPR and the only one that was conducted in unselected and low-risk populations. However, despite the size of the study there was significant heterogeneity in relation to both fetal weight and neonatal morbidity. This heterogeneity could be explained mainly due to two factors. First, only five studies blinded the ultrasound result to the clinicians and some of them did not blind other parts of the scan such as the fetal size. This could potentially explain the lack of association between CPR and neonatal metabolic acidosis since knowing the ultrasound result could affect the clinical decision making and expedite delivery prior to severe fetal compromise. Second, the timing of the ultrasound. Twomey et al performed a non-blinded ultrasound at 32wkGA in a mixed-risk population and the study showed significant association with NICU admission, SGA and severe SGA. Since it is well known that CPR is more strongly associated with early FGR,[127] the iatrogenic delivery of those infants at preterm gestations would increase the need for neonatal unit support. We did not observe this strong association in the Bakalis et al study which also offered ultrasound at 32 wkGA likely because the scan was offered as part or routine ultrasound screening and not because of a clinical indication.

This study shows that the association between CPR and SGA indicates that the ratio is likely to predict FGR. However, it is only weakly predictive of adverse obstetric and perinatal outcome. Our findings contradict the previously published systematic review[128] which concluded that CPR at term has a strong association with adverse obstetric and perinatal outcomes. We believe this is because the systematic review by Dunn et al included mostly studies done in high-risk populations, did not include some large, recently published studies which offered ultrasound as part of universal screening (Akolekar[97], Bakalis[115]) and did not produce any pooled analysis. Another systematic review[129] that was focused in pregnancies with FGR showed that the CPR was strongly predictive of perinatal death but only weakly predictive of markers of neonatal morbidity such as neonatal metabolic acidosis and low 5-minute Apgar score with positive LRs of 1.6 (95 % CI 1.3-2.0) and 1.9 (95% CI 1.5-2.4) respectively. Since, we would expect that the association between CPR and adverse neonatal outcome

would be stronger in growth restricted fetuses, we believe that our corresponding positive LRs of 1.2 and 1.7 are likely to be true findings in unselected pregnancies.

In this systematic review I found that the CPR would be abnormal in about 1 in 4 SGA infants and 1 in 3 severe SGA infants at term. Since abnormal fetal growth accounts for about 1 in 3 stillbirths at term,[2] pregnancies with abnormal CPR need to be monitored closely. However, an optimal screening policy will likely need to use a combination of ultrasound findings and biomarkers. We have previously shown that combining ultrasonic measurement of fetal size with an elevated sFLT1/PlGF ratio could increase the predictive accuracy for maternal pre-eclampsia or perinatal morbidity and mortality at term with a sensitivity of 38% and a positive LR of 17.5.[130] A recently published study showed that a serum metabolite ratio could further increase the predictive accuracy for term FGR to an AUC of 0.78 compared to 0.64 for the sFLT1/PlGF ratio.[131]

In conclusion, the CPR is clearly associated with the delivery of SGA infants. However, it is unlikely that a policy of early term delivery based on the CPR alone would reduce overall neonatal morbidity. Future research is needed on examining the use of a combination of ultrasound and biomarkers to reduce neonatal morbidity at term.

Chapter 6. Systematic review of the diagnostic effectiveness of universal ultrasonic screening using severe oligohydramnios in the prediction of adverse perinatal outcome.

Introduction

The amniotic fluid is necessary for fetal development and the measurement of amniotic fluid volume has become a standard marker for the assessment of fetal wellbeing in the third trimester. Decreased amniotic fluid is called oligohydramnios and increased amniotic fluid is called polyhydramnios. Fetal urine is the main source of amniotic fluid in the second half of pregnancy and fetuses with no kidneys (renal agenesis) produce no amniotic fluid in the second and third trimester. However, congenital anomalies account only for a minority of cases with oligohydramnios. Oligohydramnios is commonly associated with rupture of the fetal membranes in which amniotic fluid is lost vaginally and can lead to other problems such as poor fetal lung development since the amniotic fluid helps expand and develop fetal lungs in utero. In these cases the ultrasonic assessment of the fetal bladder (filling and emptying) can confirm the production of amniotic fluid by the fetus. However, fetal distress can also be a cause of oligohydramnios through reduced fetal urine production. Fetal stress – such as arterial hypoxaemia – results in activation of a number of compensatory responses which include the increased release or arginine vasopressin (aka anti-diuretic hormone) which has a direct effect on the kidney [110] . Fetal hypoxia leads to changes in fetal circulatory distribution which increases blood supply to the vital organs (heart and brain) but reduces blood flow to other organs including the kidneys. The combination of increased arginine vasopressin and reduced renal blood flow will reduce fetal urine output and lead to oligohydramnios.

The most common markers for assessing of amniotic fluid volume are the amniotic fluid index (AFI , the sum of the four deepest pockets of amniotic fluid in four quadrants of the uterus)[132] and the single deepest pocket (SDP). Severe oligohydramnios is commonly defined as AFI<5cm or SDP<2cm. The aim of this chapter was to assess the diagnostic effectiveness of severe oligohydramnios in predicting adverse pregnancy outcomes at or near term.

Methods

Publication statement

This chapter was published as a chapter in the HTA report (Appendix 9C).[113] Some passages were quoted verbatim from this paper where I wrote the text for this publication. The figures and tables have been reproduced after gaining permission by the journal.

Sources for meta-analysis

We identified a previous systematic review[133] which was published in 2014 and included source material from publications up to 2011. However, the review did not limit searches to low or mixed risk pregnancies. We updated the systematic review including studies published from 01//01/2011 up to June 2019. An update search was done in November 2020. The systematic search was performed using Medline, EMBASE, the Cochrane database of Systematic Reviews (CDSR) and Cochrane Central Register of Controlled Trials (CENTRAL). No restrictions for language or geographic location were applied. The studies were identified using a combination of words related to "ultrasound", "pregnancy", "amniotic fluid volume", "AFI", "oligohydramnios", and "single deepest pocket". The exact search strategy is described in Appendix 4.

Study selection

Selection criteria included cohort or cross-sectional studies with singleton pregnancies where an ultrasound scan was performed ≥24wkGA. We included all studies where the ultrasound was performed as part of universal screening, studies that used low-risk populations only and studies with mixed-risk populations. These criteria were applied to the studies included in the previously published review and to the studies published subsequent to that review. We excluded studies that were focused in high risk patients such as FGR, studies which included pregnancies with preterm premature rupture of membranes, and studies where the ultrasound was performed intrapartum. We included studies that reported the outcomes described in chapter 2.

Study quality assessment and statistical analysis

The literature search, study selection, and analysis ware performed independently by two authors (AM and DW) using Review Manager 5.3. Any differences were resolved in discussion with the senior author (GS). The methods employed are described in Chapter 2.

Results

The literature search flowchart is presented in Figure 18. We identified 14 studies[134-147] that met our inclusion criteria involving 109,679 patients in total. The study characteristics are presented in Table 8. Two studies[137, 138] (N= 30,555) included unselected pregnancies, ten studies [134-136, 140-145, 147](N= 61,047) included low-risk pregnancies only and two studies[139, 146] (N= 18,077) included mixed risk pregnancies. Six studies [135, 138, 139, 141, 142, 144](N= 5740) were prospective, six [134, 137, 140, 143, 145, 146](N= 97,022) were retrospective, one[136] (N=260) was cross-sectional and one[147] (N= 6657) was done as part of a clinical trial.

The assessment of study quality was performed using the QUADAS-2 tool and is summarized in Figure 19. The main risk of bias was for reference standard due to the lack of blinding in the majority of studies. Only two studies [141, 144](N=1892) blinded the results to the clinicians, one of which blinded only the AFI result and not the other aspects of the ultrasound. The second more common risk of bias was for flow and timing. Two studies[135, 145] performed ultrasound prior to induction of labour or within 4 days from delivery. Two other studies[137, 142] did not report the gestational age at either ultrasound or delivery. Hence, these results may not be applicable for universal third trimester screening at 36wkGA. Two studies had unclear risk of selection bias[139, 146] as they did not report how they selected their patients and one study[136] had high applicability concerns for patient selection as they included prolonged (>41 weeks's gestation) pregnancies only.

The summary results for the diagnostic accuracy of oligohydramnios at predicting adverse pregnancy outcomes are presented in Table 9. The most reported outcomes were neonatal unit admission and Caesarean section for fetal distress (11 and 10 studies respectively). The stronger statistically significant association was with SGA <10th centile with positive LR of 2.8 (Table 9). There were also statistically significant associations with NICU admission and Caesarean section for fetal distress with positive LRs of 1.7 and 2.2 respectively. The positive LR for neonatal death was 3.7 but because of the small number of events the confidence intervals were very large and include unity. The summary ROC curves are presented in Figure 20. Generally, the larger studies reported lower sensitivities and higher specificities for all the outcomes. Figure 21 illustrates forest plots of DORs. Finally we used the Deeks' funnel plot asymmetry test to assess the risk of publication bias using the outcome of neonatal unit admission for the analysis (Figure 22). The test showed no evidence of publication bias (P=0.54).

Figure 18. PRISMA flowchart for the systematic review for severe oligohydramnios

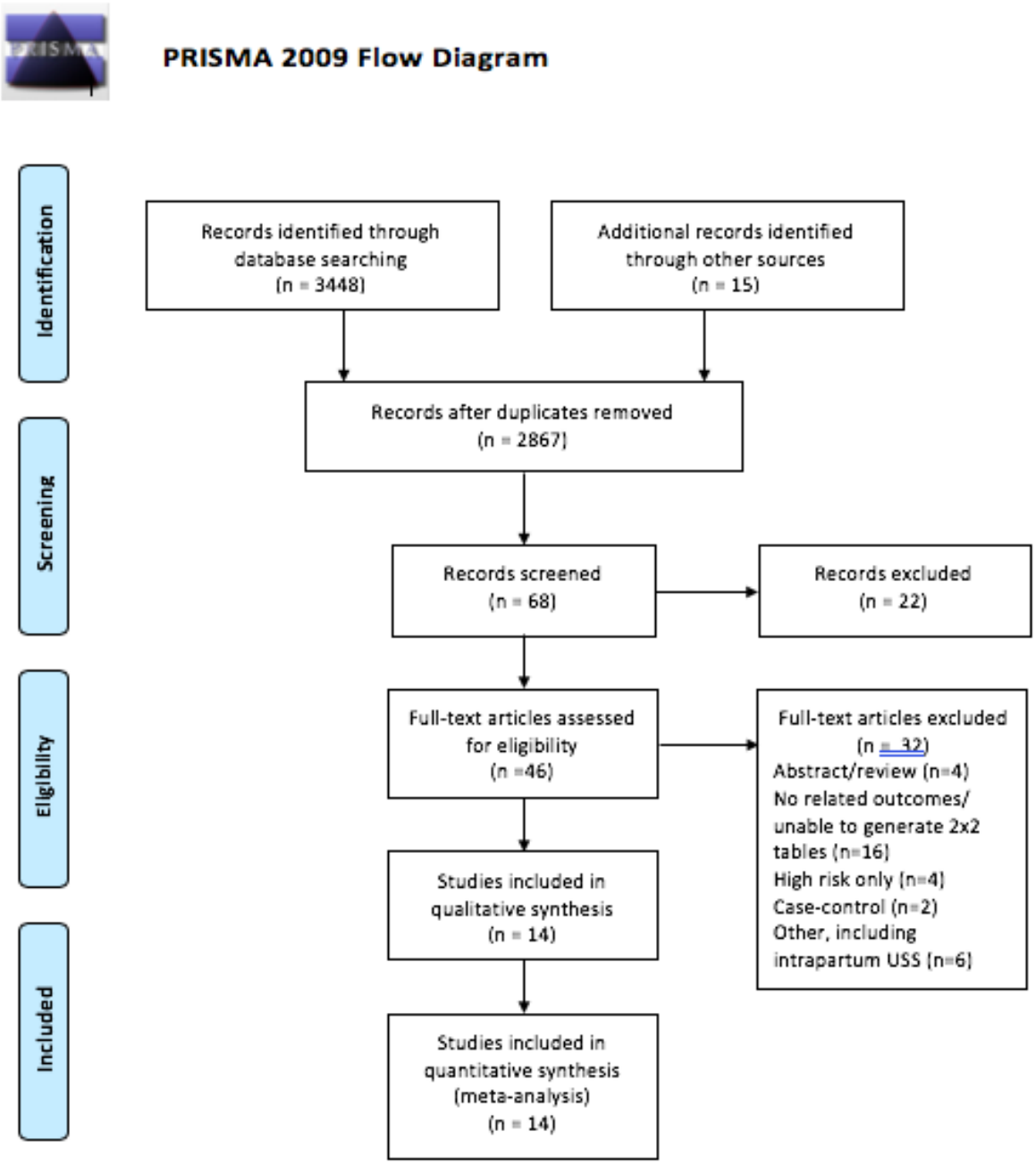


Table 8. Characteristics of the studies included in the meta-analysis for severe oligohydramnios

First Author (Year)	Type of Study, Setting	Number of fetuses and selection (All singleton, non anomalous unless otherwise stated)	Index test	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Ashwal 2014	Retrospective cohort Single University hospital, Israel	N=23,267 Low risk Term only. Excluded pregnancies with hypertensive disorders, diabetes, AFI >25cm, and EFW <10th centile.	AFI <5cm Not blinded	Within 1 week from delivery	C-Section for fetal distress (CSFD), operative vaginal delivery for fetal distress, 5-min Apgar <7, umbilical artery pH < 7.10, NICU admission, need for intubation, meconium aspiration syndrome (MAS) or HIE. Also stillbirth, neonatal death, IVH, meconium amniotic fluid (not MAS).	39+8 +/- 1.1 for isolated oligohydramnios; 39.3 +/- 1.1 for normal AFI	Nulliparous: N= 442 (44.8%) for isolated oligohydramnios, N=6,848 (30.7%) for normal AFI IOL: N= 273 (27.7%) for oligo, N= 824 (3.7%) for normal.
Ghosh 2002	Prospective cohort, Single hospital, Sweden	N= 333 Low risk, Term only, in early labour or prior to IOL	AFI <5cm Not blinded	In early labour or before IOL	Operative delivery for fetal distress, C-Section for fetal distress, 5-min Apgar <7, cord arterial PH <7.10, NICU admission.	Mean GA 283 days for those with AFI <5cm, 280 days for AFI >5cm	Nulliparous: 26/49 of those with AFI <5cm, 134 for those with AFI >5cm.
Hassan 2005	Cross-sectional,	N= 260 Low risk, Postdates (after 41+0).	AFI <6cm Not blinded	After 41+0	Neonatal death, caesarean section, meconium stained amniotic fluid.	After 41+0	Nulliparous: 34% of low AFI, 19.7% of those with normal. IOL: Not specified.

	Single hospital, Pakistan						
Hsieh 1998	Retrospective cohort, Single hospital Taiwan	N=27,506 Universal Excluded those with AFI>24cm, PPROM.	AFI <5cm Not blinded	Not specified	Stillbirth, SGA <10th centile, 5-min Apgar <7, NICU admission, Neonatal death.	Not specified	Nulliparous: Not specified IOL: Not specified.
Locatelli 2004	Prospective cohort Single hospital, Italy	N= 3049 Universal Routine scan at 40 weeks. Excluded those with PPROM and those with other indications for USS.	AFI <5cm Not blinded	40 weeks	Meconium stained amniotic fluid, CS for fetal distress, SGA <10th centile, Apgar score <7, Cord arterial PH <7.0.	40+0 – 41+6 weeks	Nulliparous: 72% for those with low AFI, 58% for those with normal. IOL: 83% for those with low AFI, 25% for those with normal
Megha 2013	Prospective cohort Single centre, India	N=200 Mixed. Selection not specified.	AFI <5cm Blinded	34-41 weeks Within 7 days of delivery	C-Section for fetal distress, meconium stained fluid, 5-min Apgar score <7, cord arterial PH <7.10. Admission to NICU for >48 hours.	Not specified. 56% of those with low AFI delivered <37 weeks vs. 34.3% with normal AFI	Nulliparous: 68% of those with low AFI, 58.9% of those with normal. IOL: 72% of those with low AFI, 51% of those with normal.

Melamed 2011	Matched cohort (3:1) Single hospital, Israel	N= 432 Low risk. Excluded pregnancies with PET/DM/GDM, EFW <10th centile, abnormal umbilical artery doppler, and PROM.	AFI <5cm Not blinded	GA at initial USS: 33.9 for low AFI , 33.9 for normal. GA at last scan not reported.	C-Section for fetal distress, meconium stained fluid, preterm delivery (<37 weeks), admission to NICU.	37.3 +/-1.6 for cases, 39.1 +/-1.8 for controls	Nulliparous: 62 (57.4%) of cases, 186 (57.4% of controls) IOL: 54 (50%) of cases, 31 (9.6%) of controls.
Morris 2003	Prospective cohort, Single Hospital, Oxford, UK	N= 1584 Low risk, Term only (>40 weeks). Excluded non-vertex and those with clinically required ultrasound.	AFI <5cm SDP <2cm Not blinded	At or after 40 weeks (59% at 40 wks)	C-Section for fetal distress, NICU admission, 5 min Apgar score <7	At or after 40 weeks (615 at 41weeks)	Nulliparous: 778 (49.1%) IOL: 643 (40.6%)
Myles 2002	Prospective cohort, Single hospital Florida, USA	N= 266 Low risk Term only. Excluded non-vertex, SROM, polyhydramnios, and any pregnancies with fetal or maternal complications.	AFI <5cm SDP <2.5cm Not blinded	Between 37+0 and 41+6 (Not specified)	C-Section for fetal distress, NICU admission, Meconium stained amniotic fluid.	Not specified.	Nulliparous: Not specified IOL: Not specified.

Naveiro-Fuentes 2015	Retrospective cohort Single hospital, Spain	N= 27,708 Low risk, Term only. Routine antenatal scan at 39 weeks. Excluded pregnancies with maternal or fetal pathology including suspected IUGR.	AFI <5cm Not blinded	39 weeks	C-Section for fetal distress, instrumental delivery for fetal distress, meconium stained fluid, small for gestational age (<10th centile), 5-min Apgar score <7, Admission to NICU, umbilical artery pH < 7.10.	279 +/- 7.3 days for those with oligohydramnios, 278.2 +/- 7.5 for normal	Nulliparous: 65.1%) of those with low AFI. IOL: Not reported.
Quinones 2012	Prospective cohort, 2 centres, Pennsylvania, USA	N= 308 Low risk Between 37-40 weeks Excluded pregnancies with maternal or obstetric complications (including suspected FGR).	AFI <5cm AFI <8cm AFI <10cm SDP <2cm	37-40 weeks (Mean 38.1 +/- 0.9 weeks)	Fetal vulnerability index (FVI) which is defined as 1 or more of the following: 5 min Apgar <3, umbilical cord PH <7.0, intrapartum fetal death, neonatal seizures, intubation in the absence of meconium, or NICU admission for >24 hours.	Mean ga 39.9 +/- 0.8	Nulliparous: 50%
Rainford 2001	Retrospective cohort, Single hospital, USA	N=232 Low risk Term only. Excluded those with any maternal or fetal complications.	AFI <5cm Not blinded	Within 4 days of delivery	Operative delivery for fetal distress, NICU admission, 5-min Apgar score <7, meconium stained amniotic fluid.	Mean ga 40.1 for those with oligohydramnios, 40.9 for normal AFI.	Nulliparous: 17% for low AFI, 20% for normal AFI. IOL: 98% of those with low AFI , 51% of those with normal AFI.

Shanks 2011	Retrospective cohort Single centre, USA	N= 17,877 Mixed risk Selection criteria not specified	AFI <5cm AFI <5th centile Not blinded	Mean 34.38 +/- 3.04 weeks	NICU admission	Mean 38.27 +/- 2.86	Nulliparous: n=7069 (39.5%)
Zhang 2004	Clinical trial (USS screening vs. no screening). For this study data used by the screening group.	N=6657 in the low risk group. They all had 2 research scans at 15-22 weeks and 31-35 weeks. Excluded multiple pregnancies and those with any maternal or fetal conditions.	AFI <5cm Not blinded	31-35 weeks	CS for fetal distress, 5-min Apgar score <7, NICU admission, perinatal mortality	Mean ga 39.6 weeks for those with oligohysramnios, 39.8 for those with normal AFI	Nulliparous: 53% of oligohydramnios cases, 45% of normal AFI IOL: Not specified.

Figure 19. Risk of bias graph of included studies for systematic review of severe oligohydramnios.

Table 9. Summary diagnostic performance of low AFI (<5cm) at predicting adverse pregnancy outcome.

Pregnancy outcome	Studies	Patients	Summary sensitivity (95% CI)	Summary specificity (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)
NICU admission	11	106,072	10.9% (6.3-18.3%)	93.7% (88.4-96.6%)	1.73 (1.15-2.60)	0.95 (0.91-0.99)
5-minute Apgar <7	9	90,536	9.9% (5.8-16.4%)	94.4% (89.0-97.2%)	1.77 (0.91-3.44)	0.95 (0.90-1.01)
Neonatal metabolic acidosis	5	54,557	9.8% (6.1-15.5%)	92.1% (87.1-95.2%)	1.24 (0.87-1.77)	0.98 (0.95-1.01)
Caesarean section for fetal distress	10	63,706	18.7% (9.6-33.2%)	91.6% (86.1-95.1%)	2.24 (1.80-2.78)	0.89 (0.80-0.98)
SGA	4	58,463	10.6% (4.4-23.6%)	96.2% (89.4-98.7%)	2.79 (1.42-5.46)	0.93 (0.86-1.00)
Neonatal death	4	57,640	12.8% (0.4-83.2%)	96.6% (87.5-99.1%)	3.73 (0.29-48.8)	0.90 (0.59-1.38)

Figure 20. Summary ROC curves for AFI <5cm at predicting adverse pregnancy outcome. A. NICU admission; B. 5-minute Apgar score <7; C. Neonatal metabolic acidosis; D. Caesarean section for fetal distress; E. SGA (<10th centile); F. Neonatal death

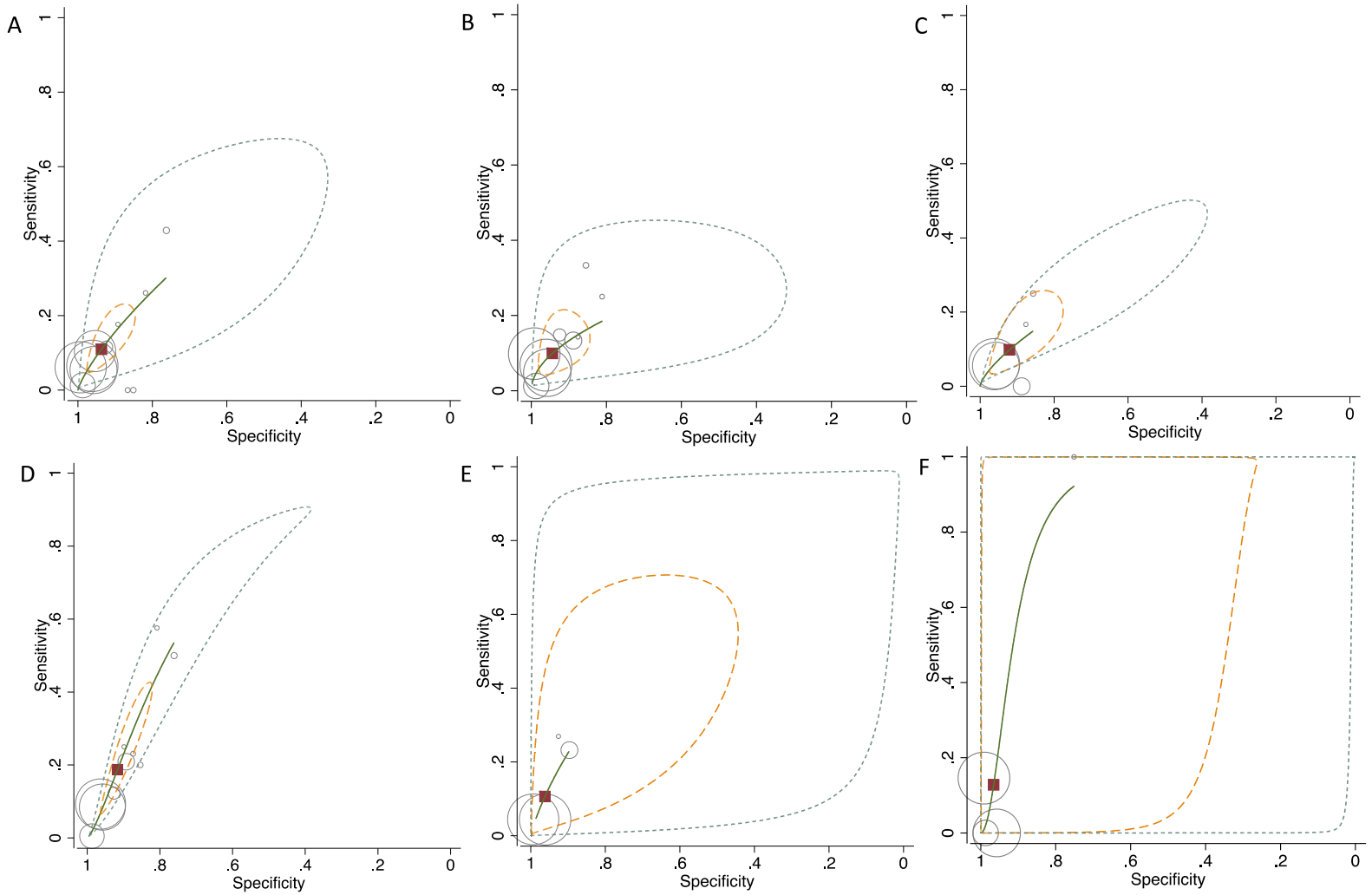
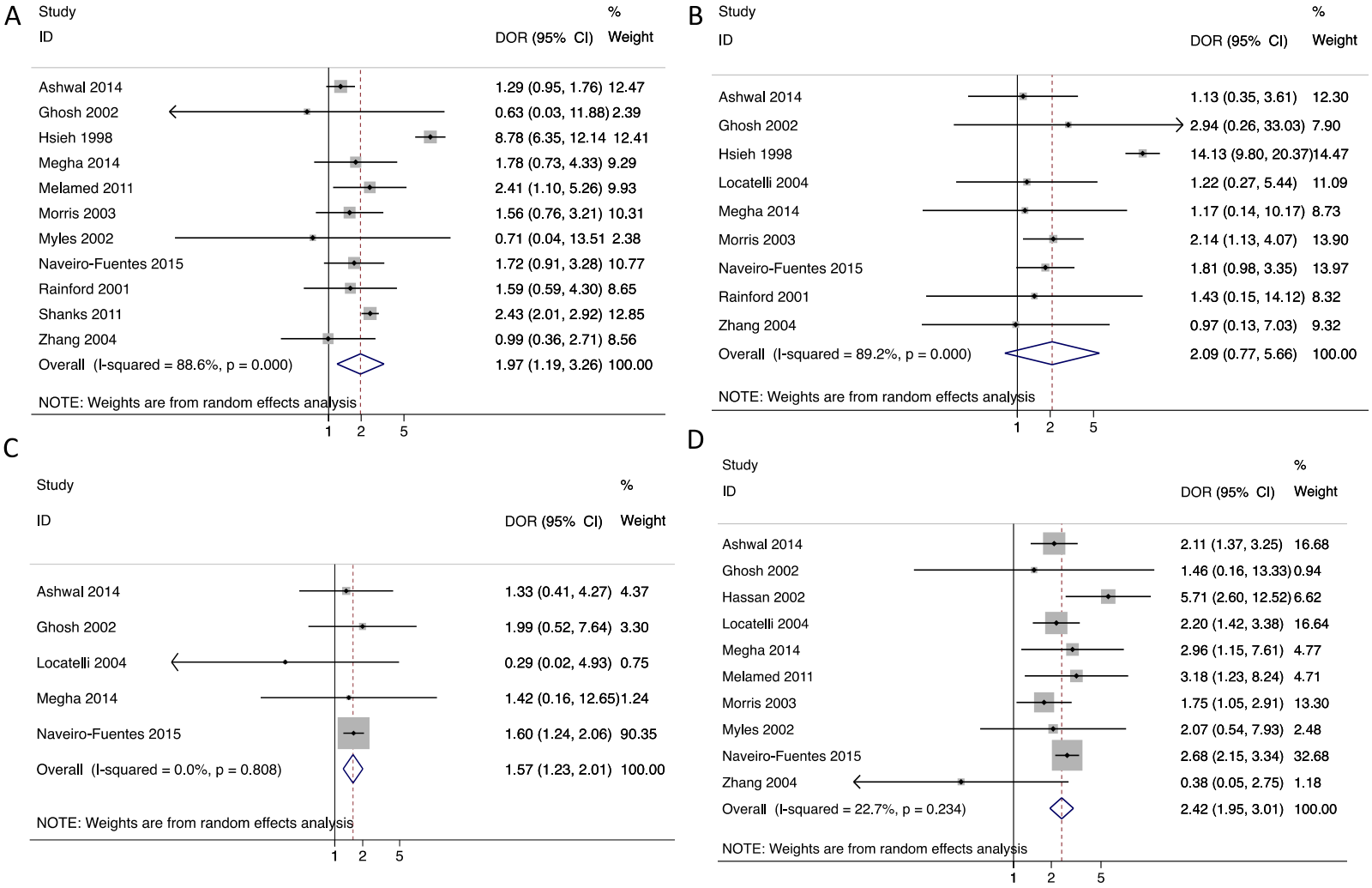
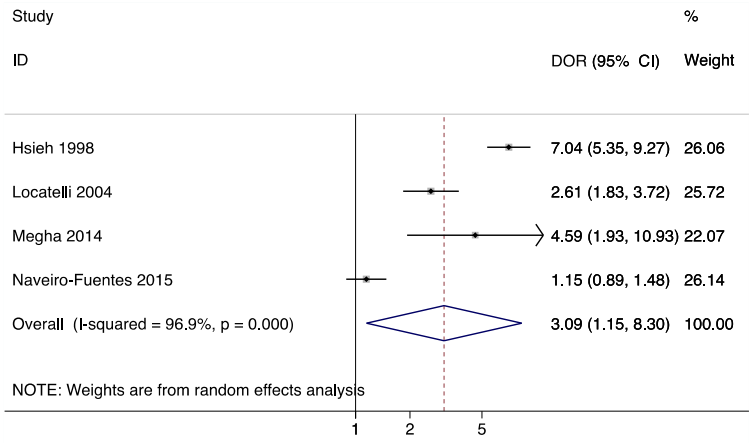


Figure 21. Meta-analysis of DORs for AFI <5cm at predicting adverse pregnancy outcome: A. NICU admission; B. 5-minute Apgar score <7; C. Neonatal metabolic acidosis; D. Caesarean section for fetal distress; E. SGA (<10th centile); F. Neonatal death.



E


Study
ID
DOR (95% CI)
%
Weight
Hsieh 1998
7.04 (5.35, 9.27)
26.06
Locatelli 2004
2.61 (1.83, 3.72)
25.72
Megha 2014
4.59 (1.93, 10.93)
22.07
Naveiro-Fuentes 2015
1.15 (0.89, 1.48)
26.14
Overall (I-squared = 96.9%, p = 0.000)
3.09 (1.15, 8.30)
100.00
NOTE: Weights are from random effects analysis
1
2
5

F

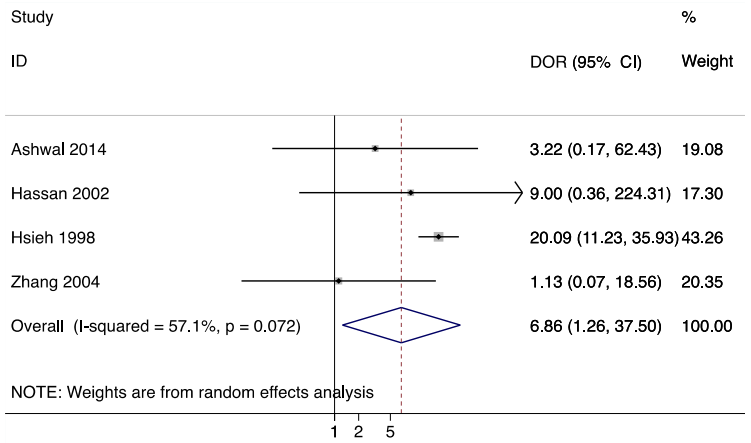
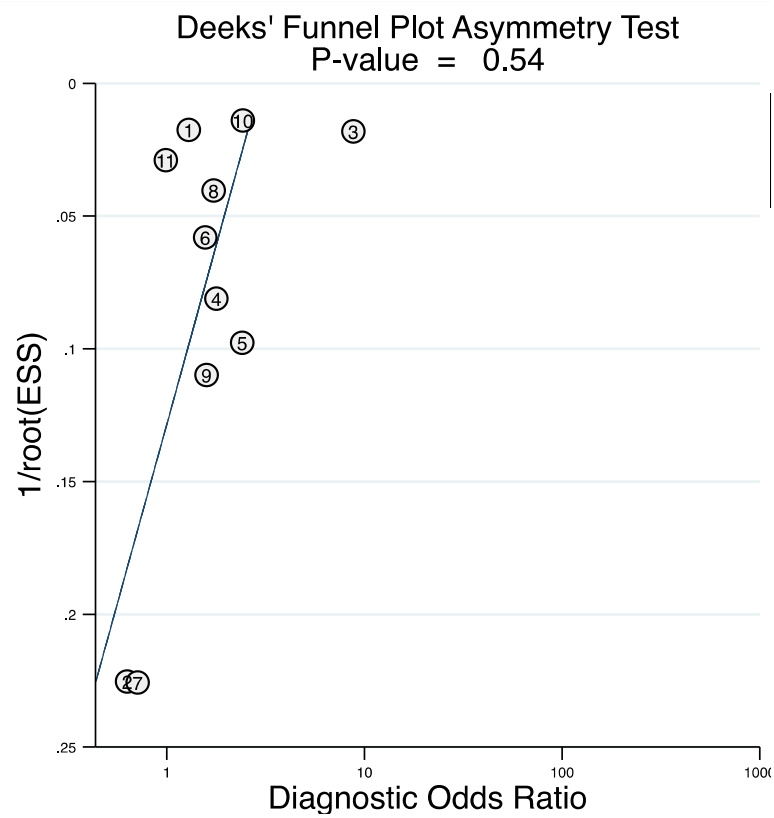

Study
ID
DOR (95% CI)
%
Weight
Ashwal 2014
3.22 (0.17, 62.43)
19.08
Hassan 2002
9.00 (0.36, 224.31)
17.30
Hsieh 1998
20.09 (11.23, 35.93)
43.26
Zhang 2004
1.13 (0.07, 18.56)
20.35
Overall (I-squared = 57.1%, p = 0.072)
6.86 (1.26, 37.50)
100.00
NOTE: Weights are from random effects analysis
1
2
5

Figure 22. Deeks' funnel plot for publication bias for severe oligohydramnios for the prediction of neonatal unit admission.



Discussion

The key finding of this meta-analysis was that severe oligohydramnios had a positive LR for SGA of between 2.5 and 3.0. Although, there was significant heterogeneity between the studies, 3 out of 4 studies showed strong association with SGA. The only study[143] that showed no correlation with SGA had excluded all those fetuses with EFW < 10th centile at the time of the scan. Moreover, this study performed the ultrasound at 39 weeks gestation by when it's likely that most previously diagnosed SGA babies would have already been delivered. Finally, the outcome of SGA, when measured in percentiles for the gestational age, is unlikely to have been affected by clinical decision making. Thus, it is likely that the association between oligohydramnios and SGA is true.

The associations with admission to NICU, emergency Caesarean section for fetal distress, and neonatal death are more difficult to interpret. These outcomes can significantly be affected by the timing of delivery as shown in the Hsieh study[137] which had significant weight in all those meta-analyses. In this study, out of the 245 cases with oligohydramnios 32% were delivered preterm and 64% were delivered by Caesarean section. Since the study was not blinded, it is likely that revealing the result to the clinicians contributed to the decision to deliver the fetus early for suspected fetal distress. Preterm delivery is a major cause for NICU admission and neonatal death and revealing the results of the scan could explain both associations. In the case of Caesarean delivery for fetal distress, revealing the result that there is severe oligohydramnios was likely used as an indication (in whole or in part) to perform a Caesarean section for suspected fetal distress which would explain the high rates of Caesarean sections in this group.

There were important differences between our findings and the previously published systematic review by Morris et al. [133] This meta-analysis included 43 studies, the majority of which were conducted in high-risk populations. They showed that the risk of SGA was 6-times higher in pregnancies with oligohydramnios which is double the risk in our study. They also showed a 9-times increased risk of neonatal death which is about double compared to our study. The results are likely explained by the differences in the populations included in each meta-analysis.

In conclusion, this analysis confirms that severe oligohydramnios is associated with SGA. This can confidently be stated as there was an association with SGA which is much less likely to arise from biases. However, the association between oligohydramnios and neonatal morbidity is less clear.

Despite the association with SGA, the positive LR was not very high and its capacity to act as a screening test in unselected nulliparous women at 36wkGA is limited.

Chapter 7. Systematic review of the diagnostic effectiveness of universal ultrasonic screening using borderline oligohydramnios in the prediction of adverse perinatal outcome.

Introduction

In the preceding chapter, we assessed the association between severe oligohydramnios and the risk of adverse pregnancy outcome. Although associated with the risk of SGA, the finding was not strongly predictive of SGA and associations with neonatal morbidity were difficult to assess as >95% of the patients included in the meta-analysis participated in studies where the ultrasound scan was revealed. The aim of this element of the work was to determine the association between borderline oligohydramnios and adverse pregnancy outcome. First, we aimed to determine whether there was indeed a gradient in the strength of association comparing severe and borderline. Second, we were able to analyse previously unpublished data which were obtained from the POP study, where the ultrasonic finding of borderline oligohydramnios was blinded to the clinicians. This allowed us to address the true association between the finding and the risk of adverse outcome avoiding associated biases, for example, treatment paradox and ascertainment bias.

Whereas severe oligohydramnios is defined as AFI <5cm, borderline oligohydramnios can be defined as 5cm to 8cm or 5cm to 10cm. In order to establish the predictive associations, we analysed unpublished data from the POP study (described above and below) and a systematic review of other studies of diagnostic effectiveness.

Methods

Publication statement

This chapter was published as a chapter in the HTA report (Appendix 9C).[113] Some passages were quoted verbatim from this paper where I wrote the text for this publication. The figures and tables have been reproduced after gaining permission by the journal.

Analysis of data from the Pregnancy Outcome Prediction study

In the systematic review we included unpublished data from a prospective cohort study, the Pregnancy Outcome Prediction (POP) study, as described in Chapter 2. For the present analysis, women who delivered prior to their 36wkGA scan appointment were excluded. Screen positive was defined as an Amniotic Fluid Index (AFI) between 5 and 8 cm and screen negative as an AFI between 8 and 24 cm. The definition of outcome data has previously been described.[9]

Sources for meta-analysis

We searched Medline, EMBASE, the Cochrane database of Systematic Reviews (CDSR) and Cochrane Central Register of Controlled Trials (CENTRAL) from inception to June 2019. An update search was done in November 2020. The studies were identified using a combination of words related to "ultrasound", "pregnancy", "amniotic fluid index", "AFI", "liquor volume", and "prenatal diagnosis". No restrictions for language or geographic location were applied. The exact search strategy is described in Appendix 5.

Study selection

Selection criteria included cohort or cross-sectional studies with singleton pregnancies where an ultrasound scan was performed ≥24wkGA. We included studies that used a matched design based on the ultrasound finding (borderline oligohydramnios versus normal AFI) but excluded case-control studies (matched on outcome). We included all studies where the ultrasound was performed as part of universal screening (i.e. ultrasound was offered to women regardless of indication), studies that were performed in low-risk populations (i.e. those that excluded pregnancies with any maternal or fetal complication) and studies with mixed risk population (i.e. those that did not specify the indication for the ultrasound). We included studies defining borderline oligohydramnios as either an AFI of 5-8 cm or 5-10 cm and included studies both where the result was revealed (i.e. the result of the scan was reported to the clinician) and those where it was not revealed (clinicians masked to result). We excluded studies that were focused only on high risk populations, e.g. pregnancies known to be

complicated by FGR, and those where the scan was performed during labour. We included studies that reported the outcomes described in chapter 2.

Study quality assessment and statistical analysis

The literature search, study selection, and analysis ware performed independently by two authors (myself and Ilianna Armata) using Review Manager 5.3. Any differences were resolved in discussion with the senior author (GS). The methods employed are described in Chapter 2.

Results

The POP study

Initially we analysed the previously unpublished data from the POP study. Applying the inclusion criteria described above yielded a total of 3387 women with a blinded scan at 36wkGA out of the 4512 women recruited (Figure 23) and 108 (3.2%) of these women had borderline oligohydramnios (AFI 5-8 cm). The maternal age, socio-economic deprivation, ethnicity, BMI, and rates of alcohol consumption and smoking were similar between the two groups (Appendix 5, Table 16). Moreover, the groups had similar rates of pre-existing hypertension and pre-eclampsia. The median birthweight was 200g lower in the cases of borderline oligohydramnios with a small difference in the gestational age at delivery. The rates of IOL were similar in both groups but women with borderline oligohydramnios had higher rates of spontaneous vaginal delivery. The screening performance of borderline AFI in the POP study is presented in Table 10. Borderline AFI was associated with an increased risk of delivering a severely SGA infant but was not associated with SGA or an increased risk of a range of indicators of neonatal morbidity in the POP study.

Meta-analysis

The literature search flowchart is presented in Figure 24. We identified 11 studies[148-157] (including the POP study) that met our inclusion criteria involving 37,848 patients in total. The study characteristics are presented in Table 11. Only the POP study (N=3387) included unselected pregnancies, three studies[150, 156, 157] (N=1890) included only low-risk pregnancies and seven studies[148, 149, 151-155] (N=32,571) included mixed risk pregnancies. Two studies[156] (N=3817) were prospective and nine studies[148-155, 157] (N=34,031) were retrospective. Seven studies[150, 152-156] (N=36,293) defined borderline oligohydramnios as between 5 and 8 cm and four studies[148, 149, 151] (N=1555) as between 5 and 10 cm. The majority of patients in all the studies delivered at term. However, four studies[148, 151, 154, 156] reported a significantly higher rate of preterm delivery for those with borderline oligohydramnios.

The assessment of study quality was performed using the QUADAS-2 tool and is summarized in Figure 25. The main risk of bias was lack of blinding of the ultrasound result (which we defined as high risk for reference standard) which affected all studies except the POP study. We classified one study[152] as high risk for selection bias as they used only low risk patients for their comparison group and two studies[148, 149] as unclear risk of selection bias as they did not specify if they enrolled a consecutive or random sample of patients. Moreover, we classified five studies[148, 151, 153, 155, 157] as having an unclear risk of flow and timing because they did not report the gestational age at ultrasound or delivery.

The summary diagnostic performance of borderline oligohydramnios at predicting adverse pregnancy outcome is presented in Table 12. The meta-analysis demonstrated a statistically significant association between borderline oligohydramnios and all of the outcomes, and the strongest association was with delivery of an SGA infant (positive LR = 2.6). The summary ROC curves are presented in Figure 26. Forrest plots of the DORs demonstrated heterogeneity which was statistically significant for SGA and NICU admission (Figure 27). Two studies (POP and Petrozella et al) reported SGA below the 3rd centile and three studies reported perinatal death. However, we could not generate summary results for outcomes that were reported in less than four studies. Finally we used the Deeks' funnel plot asymmetry test to assess the risk of publication bias using the outcome of SGA <10th centile for the analysis (Figure 28). The test showed no evidence of publication bias (P=0.33).

Figure 23. POP study inclusion flowchart for the analysis of borderline oligohydramnios

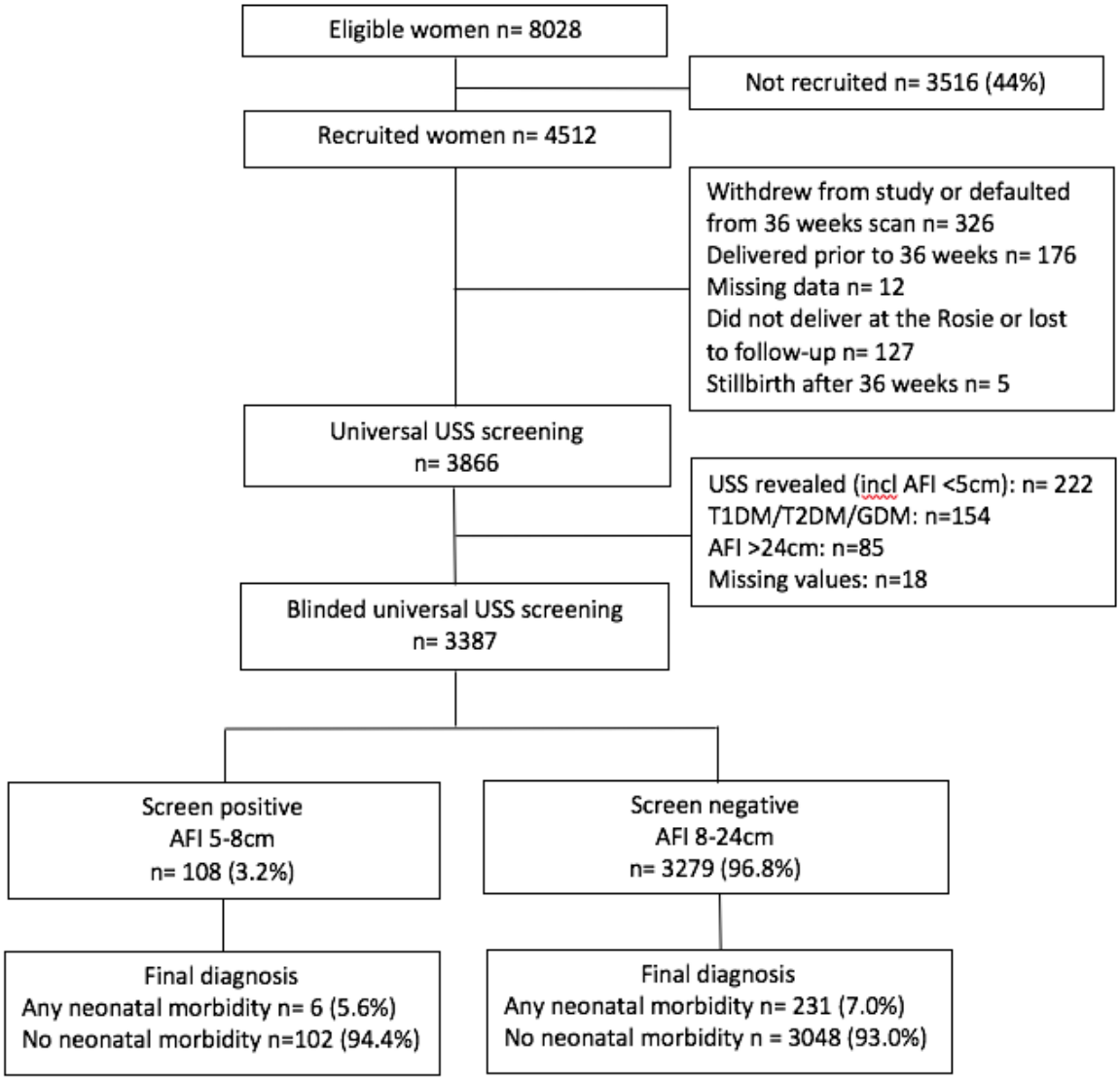


Table 10. Diagnostic performance of borderline low AFI (5-8cm) at predicting adverse pregnancy outcome at term in the POP study (N=3387).

Outcome	True Positive / False Positive	True Negative / False Negative	Sensitivity (95% CI)	Specificity (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)
SGA <10th centile	10/98	2969/310	3.1% (1.2-5.0)	96.8% (96.2-97.4)	0.98 (0.52-1.86)	1.00 (0.98-1.02)
SGA <3rd centile	6/102	3212/67	8.2% (1.9-14.5)	96.9% (96.3-97.5)	2.67 (1.21-5.88)	0.95 (0.88-1.01)
Any neonatal morbidity[a]	6/102	3048/231	2.5% (0.5-4.5)	96.8% (96.1-97.4)	0.78 (0.35-1.76)	1.01 (0.99-1.03)
NICU admission	6/102	3084/195	3.0% (0.6-5.3)	96.8% (96.2-97.2)	0.93 (0.41-2.10)	1.00 (0.98-1.03)
5-min Apgar <7	0/108	3251/28	N/A	96.8% (96.2-97.4)	N/A	N/A
Metabolic acidosis	0/108	3245/34	N/A	96.8% (96.1-97.3)	N/A	N/A
Severe neonatal morbidity[b]	1/107	3256/23	4.2% (0.5-27.4)	96.8% (96.2-97.4)	1.31 (0.18-9.38)	0.99 (0.91-1.08)

[a] One or more of the following: 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), NICU admission. [b] 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).

Figure 24. PRISMA flowchart for the systematic review for borderline oligohydramnios

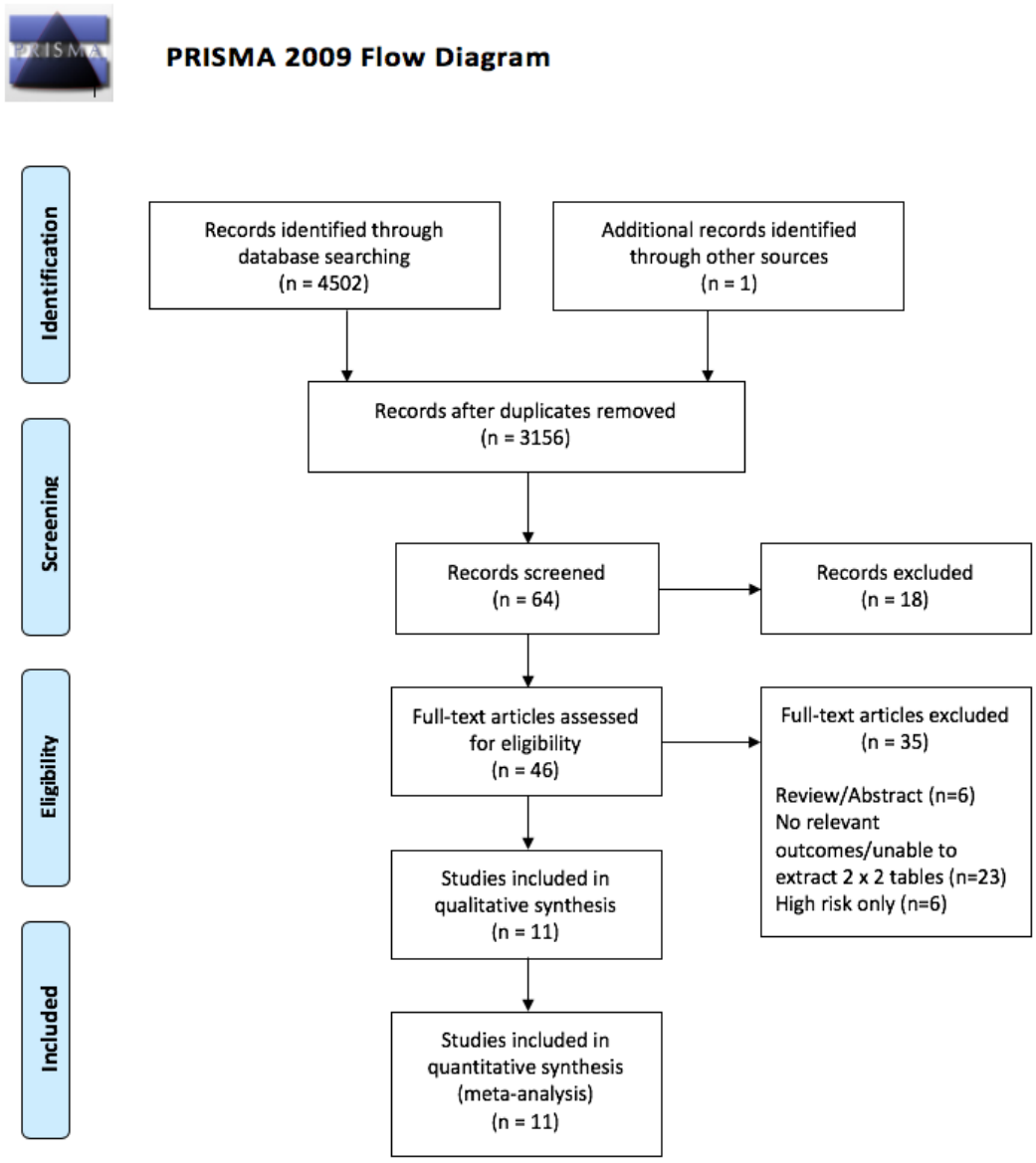


Table 11. Characteristics of the studies included in the meta-analysis for borderline oligohydramnios

First Author (Year)	Type of Study, Setting	Population & selection (Singletons only unless otherwise specified)	Index test	Gestational age at ultrasound	Reference standard	Gestational age at delivery (Mean unless otherwise specified)	Other comments
Asgharnia 2013	Retrospective cohort, Single hospital, Iran	N= 235 Mixed risk. Pregnancies >28 wks, Excluded PPROM, uterine anomalies, vaginal bleeding.	5 <AFI<10cm Not blinded	>28 weeks (mean ga not reported)	RDS, 5-minute Apgar score <7, NICU, IUGR, SGA <10th centile.	Mean GA not reported Preterm: BAFI 40.4% normal AFI 14.9%	Nulliparous: BAFI 68.1%, normal AFI 58.2% IOL: BAFI 22.3%, normal AFI 10.6%
Banks, 1999	Retrospective cohort, Single hospital, USA	N= 214 Mixed risk Pregnancies with antepartum testing within 1 week of delivery.	5cm <AFI <10cm Not blinded	Not reported	Intrapartum fetal distress, Meconium stained amniotic fluid, SGA <10th centile.	Not reported	Nulliparous: Not reported IOL: Not reported
Choi 2016	Retrospective cohort Single Hospital, South Korea	n=721 Low risk Uncomplicated, term pregnancies only. Excluded SROM, elective CS, breech presentation, pre-eclampsia, and other maternal disease.	5.1≤ AFI ≤ 8.0 cm	Within 1 week of delivery	Meconium stained amniotic fluid, C-Section for fetal distress, 5-min Apgar score <7, NICU admission, SGA <10th centile	BAFI: 39.2 wks Normal AFI: 39.4 wks	Nulliparous: BAFI 66.1%, normal AFI 57.3% IOL: BAFI 60.7%, normal AFI 27.4%
Gumus, 2007	Retrospective cohort Single hospital, Turkey	n= 367 Mixed risk Excluded PROM, uterine anomalies, PV bleeding	5cm <AFI< 10cm	Not reported	Intrapartum fetal distress, meconium stained amniotic fluid, SGA <10th centile), NICU admission, RDS	BAFI 37.7 wks for Normal AFI 38.3 wks Preterm: BAFI 18.9% Normal AFI 9.7%	IOL: BAFI 73.3% Normal AFI 54.5%
Jamal 2016	Matched cohort (matched 1:1),	n=128 Mixed risk	5.1≤ AFI ≤ 8.0	37-40 weeks	Meconium stained amniotic fluid, 5-min	BAFI (median): 37 wks +5 days	Nulliparous: Not reported

	Single hospital, Iran	Term only, Excluded PPROM, anomalies, maternal medical diseases, contraindications for vaginal delivery		within 1 wk of delivery	Apgar score <7, umbilical artery pH <7.0, NICU admission, SGA <10th centile.	Normal AFI: 38wks +6 days	IOL: Not reported
Kwon 2006	Retrospective cohort, Single hospital, South Korea	n= 3740 Mixed risk Excluded fetal malformations, SROM preeclampsia, chromosomal anomalies, AFI >25cm	5.1≤ AFI ≤ 8.0	Within 2 weeks of delivery	Perinatal death, NICU admission, CS for fetal distress, 5-min Apgar score <7, SGA <10th centile.	BAFI: 36.3 weeks normal AFI: 38.0 weeks.	Nulliparous: Not reported. IOL: Not reported.
Moraitis (current paper)	Prospective cohort, Single centre, Cambridge, UK	N= 3387 Nulliparous only, Universal screening	5cm <AFI< 8cm Blinded	36 weeks	NICU admission, metabolic acidosis, 5-min Apgar score <7, composite morbidity (all above), composite severe morbidity,		Nulliparous only.
Petrozella, 2011	Retrospective cohort Regional hospitals, USA	n= 27,601 Mixed risk Those that received USS between 24-34 weeks. Excluded AFI>24cm, SROM	5cm <AFI< 8cm	24+0 to 33+6 weeks. Mean ga 29.2wks	CS for fetal distress, SGA <10th centile, SGA <3rd centile Neonatal death	BAFI 37.1 weeks Normal AFI 39.2 weeks Preterm: BAFI 37%, normal AFI 8%	Nulliparous: Not reported. IOL: Not reported.
Rutherford, 1987	Retrospective cohort Single hospital, USA	n= 286 Mixed risk Those who had antepartum surveillance. Excluded PPROM,	5cm <AFI< 8cm	Not reported	Meconium, CS for fetal distress, 5-minute Apgar score <7	Not reported	Nulliparous: Not reported. IOL: Not reported.
Sahin, 2018	Prospective (matched 1:3)	n= 430 Low risk	5cm <AFI ≤8cm	Between 34+0 and 36+6 weeks	5-minute Apgar <7, CS forfetal distress, RDS, meconium stained AF,	BAFI: 37.5 wks Normal AFI: 38.6wks.	Nulliparous: Not reported.

	Singleton hospital, Turkey	Excluded maternal disease, IUGR chromosomal/ fetal abnormalities, SROM, abnormal Doppler.		Mean 35,4 weeks	meconium aspiration syndrome, NICU, neonatal death	Preterm: BAFI 15.9%, normal AFI 8,4%	IOL: BAFI 34.6%, normal AFI 23.8%
Wood 2014	Retrospective cohort (matched 1:3) 2 hospitals, USA	n= 739 Low risk Exclusion criteria: AFI ≤5 cm, PPROM, preeclampsia	5cm <AFI ≤10cm	Not reported	CS for fetal distress, SGA, meconium stained amniotic fluid, 5-min Apgar score <7, NICU admission, preterm delivery	BAFI: 38.3 wks normal AFI: 38.9 wks	Nulliparous: Not reported IOL: Not reported.

Figure 25. Risk of bias graph of included studies for systematic review of borderline oligohydramnios.

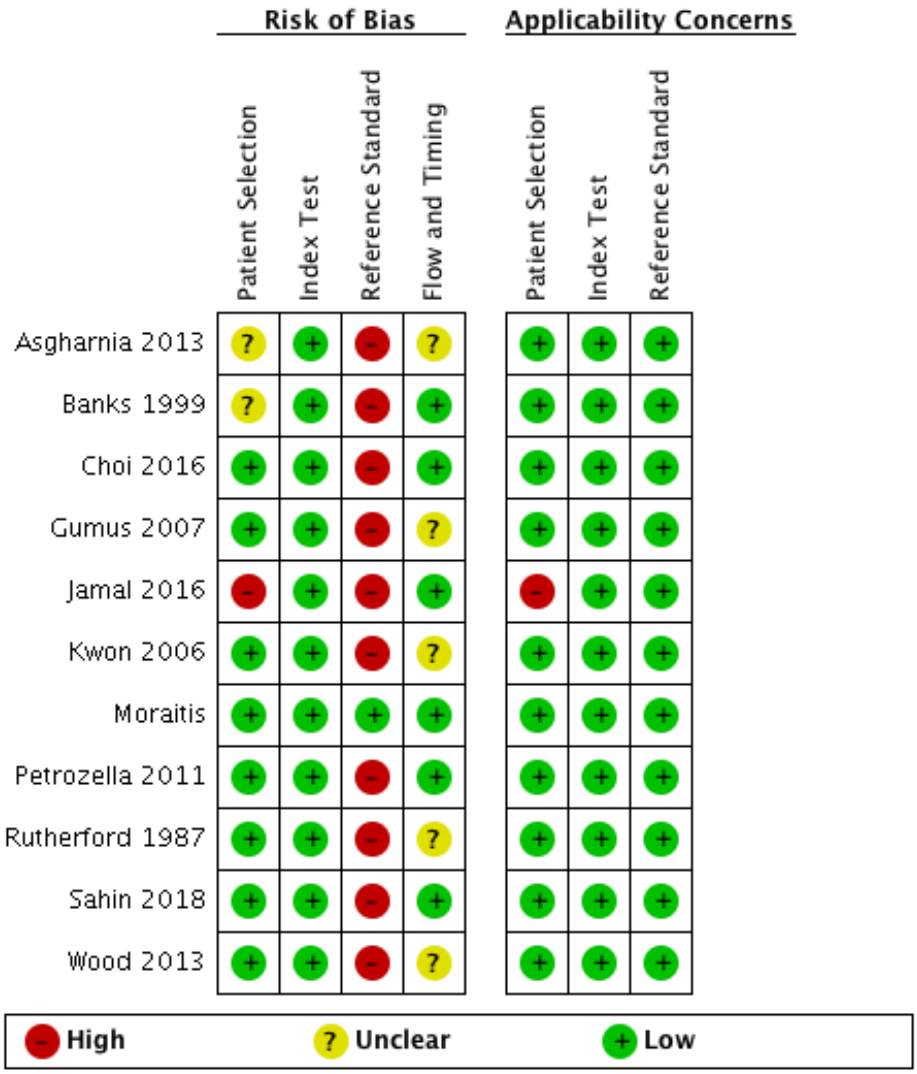


Table 12. Summary diagnostic performance of borderline low AFI to predict adverse pregnancy outcome.

Outcome	Studies	Patients	Summary sensitivity (95% CI)	Summary specificity (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)
SGA <10th centile	9	37,132	31.6% (13.0-58.7%)	87.9% (71.9-95.3%)	2.60 (1.83-3.69)	0.78 (0.61-0.99)
NICU admission	8	9,747	34.8% (15.9-60.1%)	82.6% (69.1-91.0%)	2.00 (1.41-2.85)	0.79 (0.61-1.02)
5-minute Apgar score <7	8	9,666	34.0% (17.4-55.8%)	82.0% (68.8-90.4%)	1.89 (1.47-2.42)	0.80 (0.66-0.98)
C-Section for fetal distress	6	33,517	21.2% (7.5-47.2%)	90.0% (74.5-96.5%)	2.13 (1.56-2.90)	0.87 (0.75-1.02)
Meconium amniotic fluid	7	2,885	42.1% (28.7-56.9%)	74.9% (67.7-81.0%)	1.68 (1.24-2.28)	0.77 (0.62-0.96)

SGA, Small for gestational age; LR, Likelihood ratio; CI, Confidence intervals

Figure 26. Summary ROC curves of borderline AFI at predicting: A. SGA <10th centile, B. NICU admission, C. 5-minute Apgar score <7, D. Caesarean section for fetal distress.

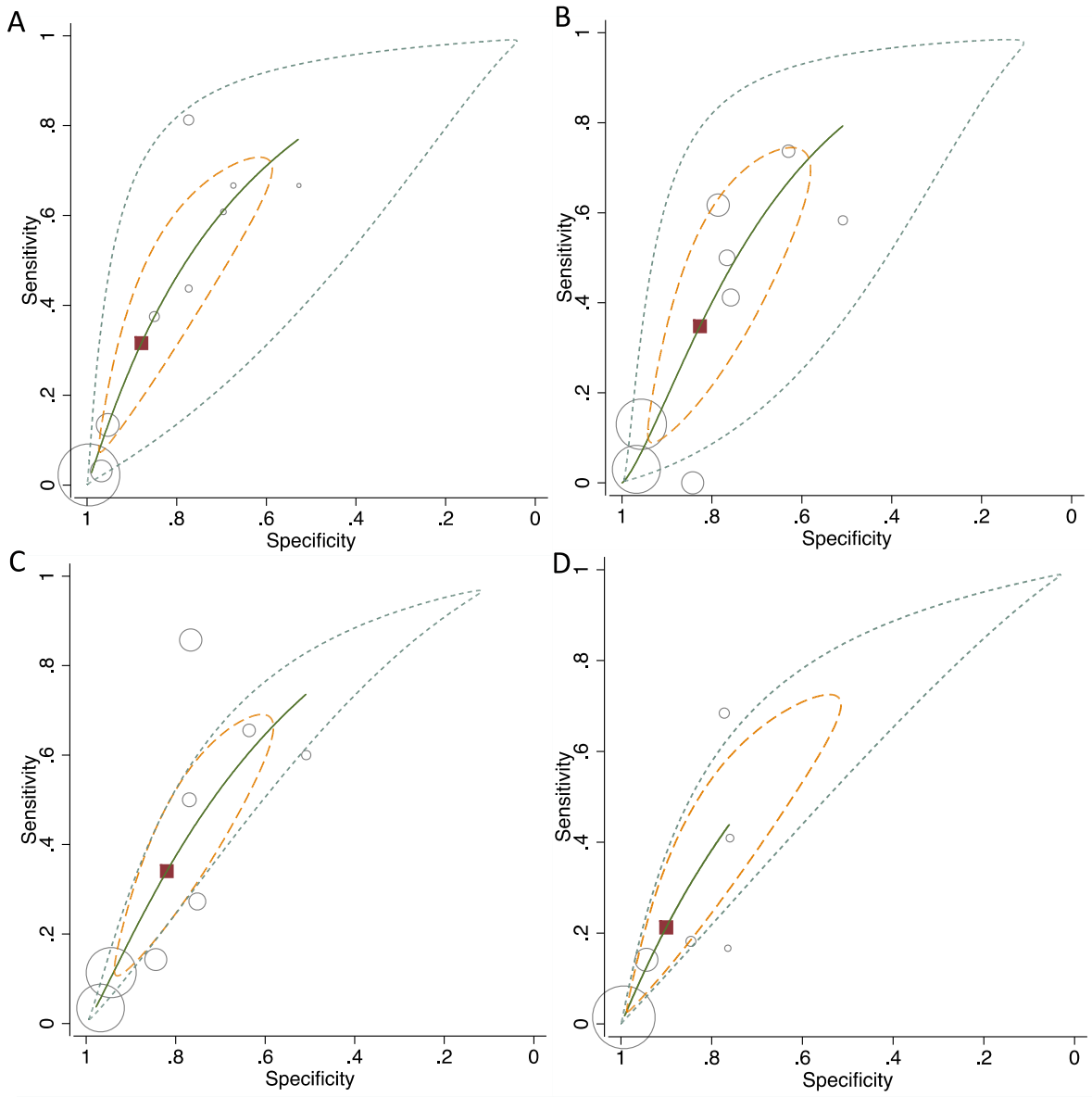
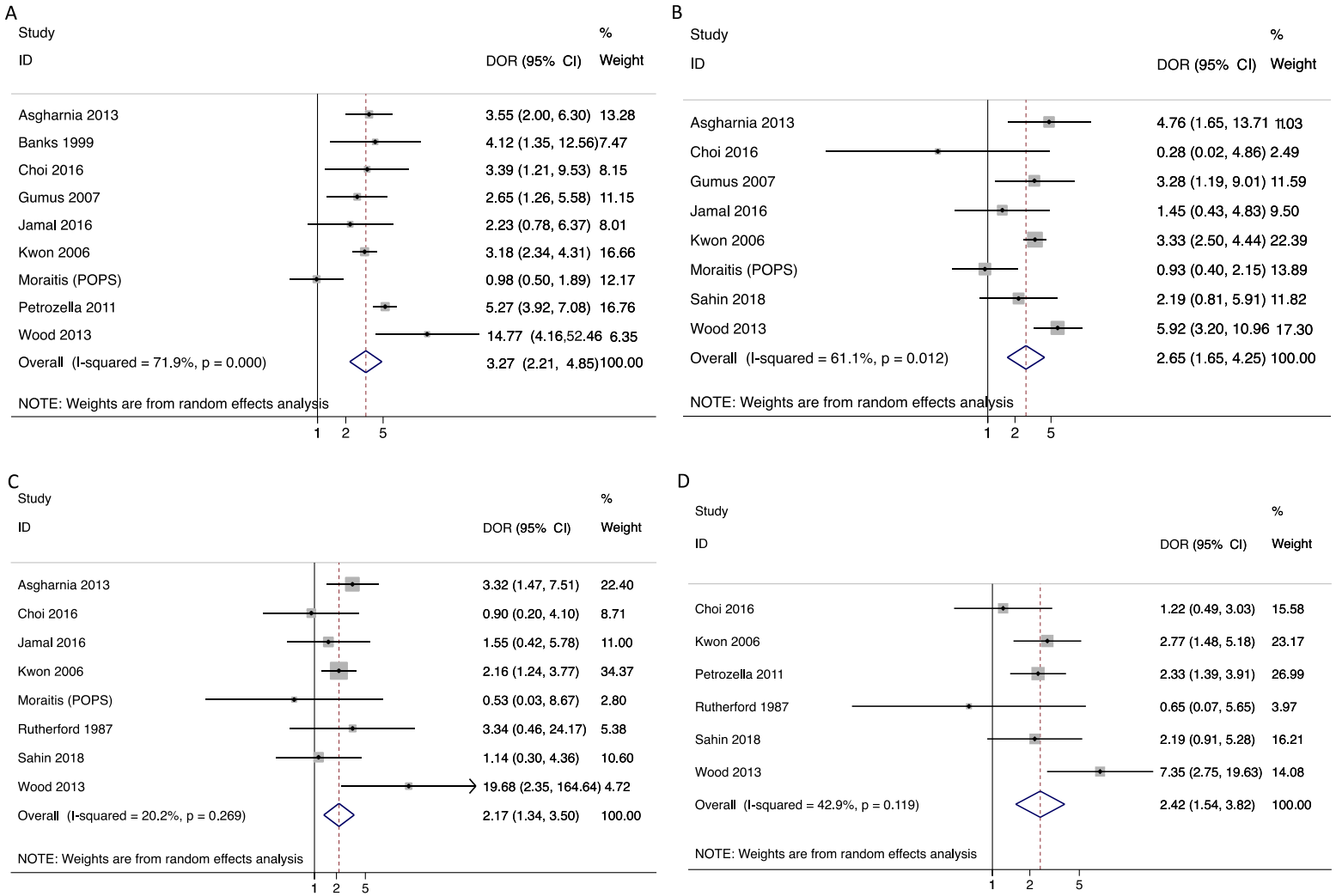
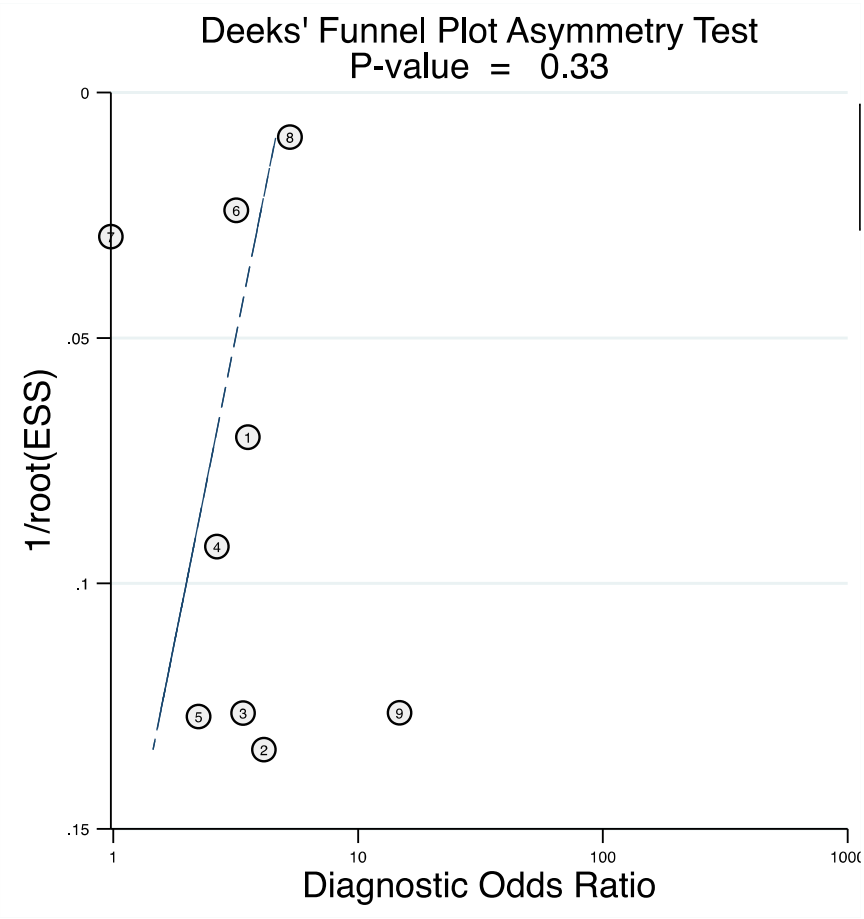


Figure 27. DORs of borderline AFI at predicting A. SGA <10th centile, B. NICU admission, C. 5-minute Apgar score <7, D. Caesarean section for fetal distress.



DOR = Diagnostic odds ratio

Figure 28. Deeks' funnel plot for publication bias for borderline oligohydramnios for the prediction of SGA <10th centile.



Discussion

The main finding of the present study is that borderline oligohydramnios is moderately predictive of SGA. This was observed in the meta-analysis of multiple studies of variable quality. There was also a comparable association between borderline oligohydramnios and severe SGA in the only study where the scan result was blinded, the POP study.

The observation that borderline oligohydramnios was associated only with severe SGA in the POP study is of interest. One possible explanation for this is that the scan result was not revealed, hence, the finding did not lead to changes in clinical management. The success of the blinding of the result is evidenced by the fact that borderline oligohydramnios was not associated with increased rates of IOL. in the POP study. A previous RCT of routine early term induction versus expectant management for suspected SGA infants (<10th centile) demonstrated that early delivery was associated with a significantly decreased the risk of delivering a severely SGA baby (<3rd percentile).[3] A possible explanation for the POP study association with severe SGA and the meta-analysis association with all SGA is that a finding of borderline oligohydramnios may have led to increased rates of early delivery in studies where the result was revealed, whereas the lack of intervention in the POP study led to growth restricted fetuses becoming progressively smaller for gestational age as the pregnancy advanced.

The other major difference between the meta-analysis and the POP study may also relate to the lack of blinding in the other studies. Borderline oligohydramnios was associated with increased rates of neonatal morbidity in the meta-analysis but none of the outcomes of neonatal morbidity were associated with this finding in the POP study. However, the confidence intervals were wide and one explanation could be the lower statistical power of the POP study. However, plotting the DORs demonstrates that, in relation to NICU admission, the 95% CI observed in the POP study excluded the point estimate of the meta-analysis. This result could also be explained by the absence of blinding in the other studies. If the scan result is revealed the only disease modifying intervention available in late pregnancy is early delivery, and this could be late preterm or early term. It is well recognized that both are associated with increased rates of neonatal morbidity and NICU admission. Hence, the association between borderline oligohydramnios and neonatal morbidity in the meta-analysis could be because the finding led to iatrogenic prematurity and the absence of the finding in the POP study could be due to the lack of this effect. Assessment of individual studies in the meta-analysis is consistent with this interpretation. Gumus et al.[151] reported higher rates of IOL in women with

borderline oligohydramnios which was associated with higher rates of preterm and early term delivery, and higher rates of NICU admission. Similarly, Asgharnia et al.[148] offered screening after 28 weeks, found that those with borderline oligohydramnios had a rate of preterm delivery of 40.4% (compared to 14.9% for those with normal AFI) and this is the likely explanation for the strong association between borderline oligohydramnios and NICU admission. This association was not found in studies that offered ultrasound later in pregnancy such as those by Sahin et al.[156]

In conclusion, we provide strong evidence that borderline oligohydramnios is associated with an increased risk of delivering an SGA infant. However, when the finding of borderline oligohydramnios is revealed to clinicians, it may lead to increased risks of neonatal morbidity through earlier delivery. Given that the strength of prediction of SGA was not strong and that revealing the result may have led to increased risks of neonatal morbidity, the observed association with SGA does not necessarily mean that screening unselected nulliparous women near term with this method will result in better clinical outcomes.

Chapter 8. Screening for breech presentation (clinical and health economic case)

Introduction

Undiagnosed breech presentation in labour is associated with neonatal morbidity and mortality and represents a challenge for obstetric management.[20] Fetal presentation is routinely assessed by palpation of the maternal abdomen by a midwife, obstetrician or general practitioner. Abdominal palpation detect between 57 and 70% of breech presentations, and depends on the skill and experience of the practitioner.[23] In contrast, ultrasound examination could accurately and quickly identify fetal presentation. Effective interventions exist for the care of women who have breech presentation diagnosed near term. The Royal College of Obstetricians and Gynaecologists recommends "that all women with an uncomplicated breech presentation at term should be offered External Cephalic Version (ECV)".[24] The rationale for this is to reduce the incidence of breech presentation at term and avoid the risks of vaginal breech birth or Caesarean section. The success rate of ECV is considered to be approximately 50%,[158] but it differs greatly between nulliparous and parous women (34% and 66% respectively).[159] Should ECV be declined, or fail, generally women are offered delivery by planned (elective) caesarean section, as there is level 1 evidence of reduced risk of perinatal death and severe morbidity compared with attempting vaginal breech birth.[20]

I used data from the Pregnancy Outcome Prediction (POP) study to analyse the outcomes for pregnant nulliparous women with breech presentation in the study. I collaborated with health economists (David Wastlund and Ed Wilson) to perform a cost-effectiveness analysis of universal ultrasound as a screening test for breech presentation.

Methods

Publication statement

This chapter was published as an individual paper in Plos Medicine (Appendix 8D).[160] Some passages (in the methods and results sections) were quoted verbatim from this paper. The figures and tables have been reproduced under open access licence.

Study design

The POP study is described in detail in Chapter 2. If the fetus was in a breech presentation at 36wkGA, women were counselled by a member of the medical team. In line with NICE guidelines, external cephalic version (ECV) was routinely offered unless there was a clinical indication which contra-indicated the procedure, e.g. reduced amniotic fluid volume (AFI <5cm).[161] ECV was performed by one of five obstetric consultants in the unit between 36-38 wkGA, patients were scanned before the procedure to confirm presentation and it was performed with ultrasound assessment; 0.25mg terbutaline SC was given prior to the procedure at the discretion of the clinician. If women refused ECV or the procedure failed, the options of vaginal breech delivery and elective caesarean section were discussed and documented. The local guideline for management of breech presentation, including selection criteria for vaginal breech delivery, was based upon recommendations from the RCOG.[18] We extracted information about ECV from case records that were individually reviewed by research midwives. Finally, we obtained delivery related information from our hospital electronic database (Protos; iSoft, Banbury, UK).

Statistical analysis

The statistical methods for analysing the cohort are described in Chapter 2.

Economic model and analysis

To evaluate the cost-effectiveness of routinely offering late pregnancy presentation scan, a decision-tree simulation model was constructed using R (version 3.4.1).[162-165] The time horizon of the economic analysis was from the ultrasound scan (36wkGA) to infant lifetime, and costs were from the perspective of the English NHS. Costs for modes of delivery were obtained from NHS reference costs;[166] since these do not list a separate cost for vaginal breech delivery, we assumed that the cost ratio between vaginal breech and elective Caesarean section deliveries was the same as in another study.[167]

The population of interest is unselected nulliparous women The model compares the outcomes at birth for two strategies: 'universal ultrasound' and 'selective ultrasound' (Figure 29). For universal ultrasound we assumed that all breech presentations at the time of scanning would be detected (i.e. assumed 100% sensitivity and specificity for the test). For selective ultrasound, the breech presentation was diagnosed either clinically (by abdominal palpation followed by ultrasound for confirmation) or as an incidental finding during a scan for a different indication. These assumptions were based upon current practice and derived from the POP study. The model input are presented in Appendix 6, Table 17.

The end-state of the decision-tree was the mode of delivery, which was either vaginal, elective Caesarean section (ELCS), or emergency Caesarean section (EMCS). Delivery could be either cephalic or breech. Emergency Caesarean section could be either due to previously undiagnosed breech presentation, or for other reasons. All cases of breech could spontaneously revert to cephalic presentation. However, we assumed the probability of this to be lower if ECV had been attempted and failed.[168] If ECV was successful, a reversion back to breech presentation was possible. It is currently unclear whether the probability of mode of delivery varies depending on whether cephalic presentation is the result of successful ECV or spontaneous reversion,[19, 169-172] but we assumed that the probabilities differed.

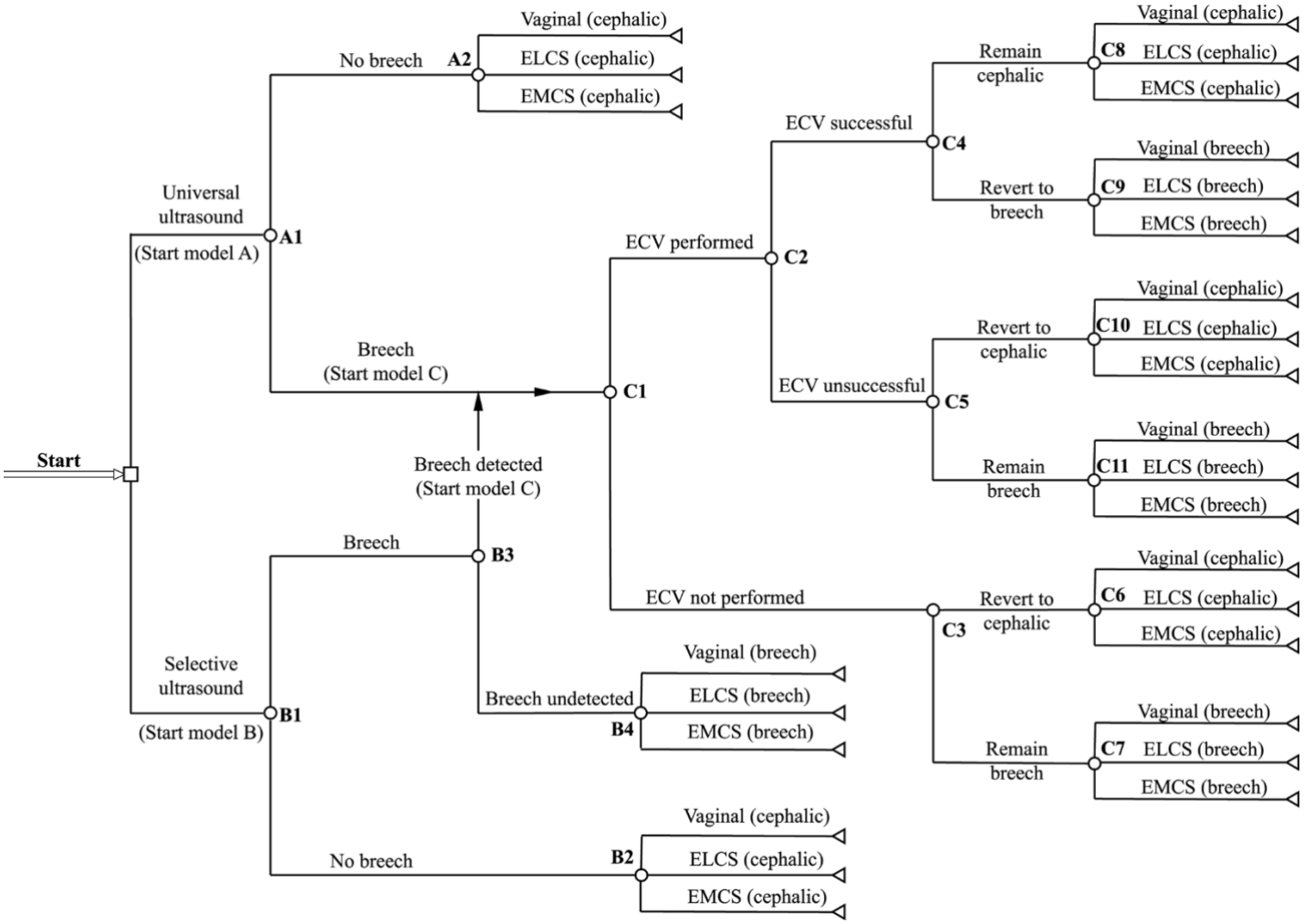
Long-term health outcomes were modelled based upon the mortality risk associated with each mode of delivery (MOD). The risk of neonatal mortality was taken from the RCOG guidelines. For breech presentation, these risks were 0.05% for delivery through ELCS, and 0.20% for vaginal delivery. The risk of neonatal mortality for cephalic presentation with vaginal delivery was 0.10%.[18] There were no randomized clinical trials that allowed us to compare the outcomes of ELCS vs. vaginal delivery for uncomplicated pregnancies with cephalic presentation, however, most observational studies found no significant difference in neonatal mortality and serious morbidity between the two modes.[173-175] For this reason, we assumed the mortality risk for cephalic vaginal and ELCS deliveries to be identical. We also assumed that emergency Caesarean section (EMCS) would have the same mortality rate as ELCS, both for cephalic and breech deliveries. Studies have found that the mode of delivery for breech presentation affects the risk of serious neonatal morbidity in the short term, but not in the long term.[18, 20, 176] For this reason, we focused the economic analysis on the effect from mortality only. The

average lifetime QALYs per member of the UK population was estimated using data on quality of life from Euroqol, weighted by longevity indexes from ONS.[177, 178] Using the annual discount rate of 3.5% as recommended by NICE, the net present value for the average lifetime QALYs at birth was 24.3.[179]

The model was probabilistic, capturing how uncertainty in the input parameters affected the outputs by allowing each parameter to vary according to its distribution. Binary and multivariable outcomes were modelled using the beta and the Dirichlet distributions, respectively.[180] Probabilities of events were calculated from the POP study and presented in Table 17. On top of the probabilistic sensitivity analysis, the sensitivity of individual parameters was also explored through one-way sensitivity analyses modifying probabilities by +/- 1 percentage point, and costs by +/- £10, to see which parameters had the greatest impact on cost-effectiveness estimates.

Total costs depended on the distribution of mode of delivery, the number of expected mortalities, and the cost of ultrasound scanning and ECV. Nationwide costs for each screening strategy were calculated for 585,489 deliveries, i.e. the number of births in England 2016-17, assuming 92% occur after 36wkGA.[9, 181] Model parameters were sampled from their respective distributions in a Probabilistic Sensitivity Analysis (PSA) of 100,000 simulations for each strategy. To determine cost-effectiveness, we used two different willingness-to-pay thresholds, £20,000 and £30, 000.[179]

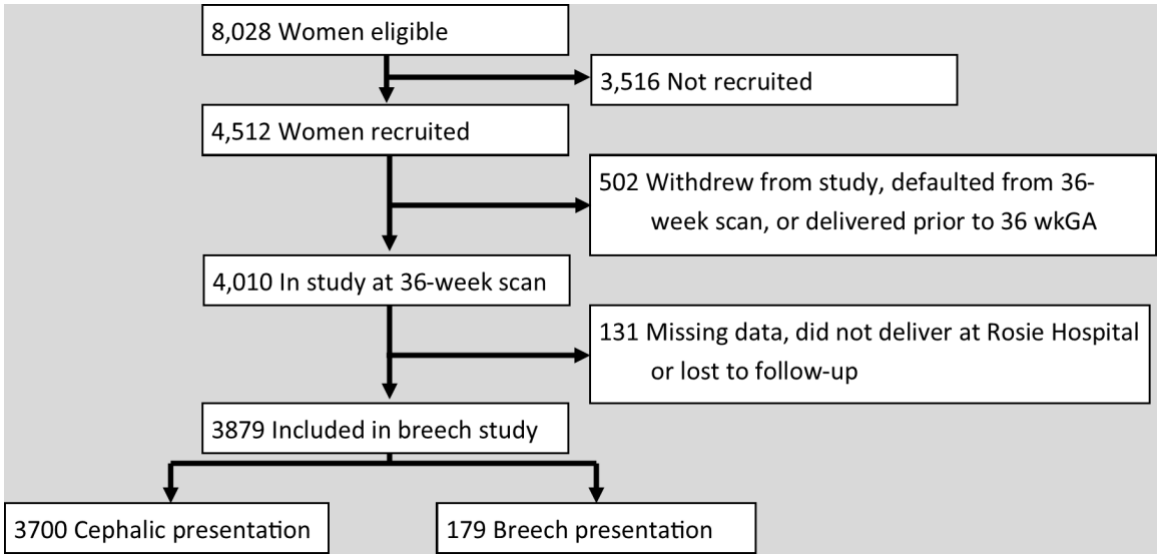
Figure 29. Simulation model structure for screening for breech. 'Universal ultrasound' strategy starts in Model A, and patients with breech presentation enter Model C. 'Selective ultrasound', i.e. no routine ultrasound, starts in Model B, and only those with a detected breech presentation enters Model C.



Results

Recruitment to the POP study cohort is shown in Figure 30 and has been previously described.[182] Information about presentation at the 36-week scan was available for 3879 women who delivered at the Rosie Hospital, Cambridge, UK; 179 of these had a breech presentation.

Figure 30. Patient recruitment. Schedule of patient recruitment in the POP study, shown by fetal presentation.



We compared maternal and fetal characteristics of the 179 women with breech presentation at 36 weeks to the women with a cephalic presentation (Appendix 6; Table 18). Women diagnosed with breech presentation were on average a year older than women with a cephalic presentation, but other maternal characteristics did not differ. The babies of women diagnosed breech were smaller and born earlier but their birth weight centile and the proportions of SGA or LGA were not markedly different. There were no differences in maternal BMI between the groups. As expected, women with breech presentation were more likely to deliver by elective or emergency Caesarean section.

Breech presentation was suspected before the 36wkGA scan for 79 (44.1%) of the women with breech presentation through abdominal palpation by the midwife or doctor; out of these, 27 had a clinically indicated scan between 32-36 weeks in which the presentation was reported. For 96 women, the breech presentation was unsuspected before the 36-week scan. Information on suspected breech position was missing for 4 women. There were no differences in BMI between the 79 women with suspected breech and the 96 women misdiagnosed as cephalic prior to the scan (median BMI was 24 in both groups, Wilcoxon rank sum test p=0.31).

Mode of delivery by external cephalic version (ECV) status is shown in Table 13. ECV was performed for 84 women, declined by 45 women, and unsuitable for 23; contraindications included low AFI at screening (18 women), uterine abnormalities (2), and other reasons (3). For 25 women, an ECV was never performed despite consent; 17 babies turned spontaneously, 6 had reduced AFI on the day of the ECV, and 2 went into labour before ECV. When performed, ECV was successful for 12 women; in one case, the baby later reverted to breech presentation before delivery. Information on ECV uptake was missing for 2 women.

Table 13. Mode of delivery by presentation and response to ECV for POP study participants with breech presentation at 36-week scan (n = 179).

ECV status	Vaginal	ELCS	EMCS	Total
ECV successful	8	1	3	12
ECV unsuccessful	0	54	18	72
ECV not offered *	1	17	5	23
ECV discussed but declined	1	32	12	45
ECV accepted but not performed †	9	5	11	25
Missing	0	1	1	2
Total	19	110	50	179

ELCS = Elective caesarean section; EMCS = Emergency caesarean section

* 18 women were contraindicated due to low AFI at screening, 2 for uterine abnormalities, and 3 for other reasons

† 17 babies turned spontaneously, 6 had reduced AFI on the day of the ECV, and 2 went into labour before ECV

The results from the economic analysis are presented in Table 14. On average, universal ultrasound resulted in an absolute decrease in breech deliveries by 0.39%. It also led to fewer vaginal breech deliveries (absolute decrease by 1.04%), and overall EMCS deliveries (0.72%) than selective ultrasound, but increased overall deliveries through ELCS (1.51%). Resulting from the more favourable distribution of mode of delivery, the average risk of mortality fell by 0.0013%. On average, 40 women had to be scanned to identify one previously unsuspected breech presentation (95% Credibility Interval (CrI): 33 to 49); across England, this would mean that 14,826 (95% CrI: 12.048 – 17,883) unidentified breech presentations could be avoided annually.

Table 14. Simulated cost and mode of delivery distribution for universal ultrasound and no ultrasound

	Universal ultrasound	Selective ultrasound	Difference (per patient)	Difference (Total population)
Total cost	2956.59	2949.30	7.29	4,268,004
Screening cost	20.70	0.43	20.27	11,867,159
ECV cost	6.52	2.94	3.57	2,093,048
Delivery cost	2927.78	2944.31	-16.53	-9,679,396
Mortality cost	1.59	1.62	-0.02	-12,806
Vaginal cephalic	0.6850	0.6826	0.0024	1,399
ELCS cephalic	0.0442	0.0441	0.0001	84
EMCS cephalic	0.2321	0.2305	0.0016	918
Vaginal breech	0.0007	0.0110	-0.0104	-6,061
ELCS breech	0.0273	0.0123	0.0150	8,774
EMCS breech	0.0107	0.0194	-0.0087	-5,115
Total mortality	0.000982	0.000995	-0.000013	-7.89
Total QALY	24.27615	24.27582	0.000327	191.73

Costs (£) are presented per patient, except in column for 'Total population' (n = 585,489).
CV = Cephalic vaginal; ECV = External cephalic version; ELCS = Elective caesarean section; EMCS = Emergency caesarean section; QALY = Quality-adjusted life years; VB = Vaginal breech.

The expected per person cost of universal ultrasound was £2,957 (95% Credibility Interval (CrI): £2,922 - £2,991), compared to £2,949 (95% CrI: £2,915 - £2,984) from selective ultrasound, a cost increase of £7.29 (95% CrI: 2.41 – 11.61). Across England, this means that universal ultrasound would cost £4.27M more annually than current practice. The increase stems from higher costs of ultrasound scan (£20.3 per person) and ECV (£3.6 per person), but is partly offset by the lower delivery costs (-£16.5 per person). The simulation shows that universal ultrasound would on average increase the number of total ELCS deliveries by 8,858 (95% CrI: 7,662 – 10,068), but decrease the number of EMCS and vaginal breech deliveries by 4,196 (95% CrI: 2,779 – 5,603) and 6,061 (95% CrI: 6,617 – 8,670) per year, respectively.

The long-term health outcomes are presented in Table 14. Nationwide, universal ultrasound would be expected to lower mortality by 7.89 cases annually (95% CrI: 3.71, 12.7). After discounting, this means that universal ultrasound would be expected to yield 192 QALYs annually (95% CrI: 90, 308). The cost-effectiveness of universal ultrasound depends on the value assigned to these QALYs. The incremental cost-effectiveness ratio was £23,611 (95% CrI: 8,184, 44,851), which is of borderline cost-effectiveness (given NICE's willingness to pay of £20,000 to £30,000).[179] The number needed to scan per prevented mortality was 74,204 (95% CrI: 46,124 – 157,642).

One-way sensitivity analysis showed that the probability parameter with the greatest impact upon the cost-effectiveness of universal ultrasound was the prevalence of breech: increasing this parameter by 1 percentage point was associated with a relative reduction of costs for universal ultrasound by £3.07. The results were less sensitive to the ECV success rate, an increase by 1 percentage point led to a relative reduction in the cost of universal ultrasound by £0.12. The most important cost parameter was the unit cost of ultrasound scan, an increase in this parameter by £10 led to a relative increase for universal ultrasound by £9.79. Keeping all other parameters equal, universal ultrasound would be cost-effective if ultrasound scanning could be provided for less than £19.80 or £23.10 per mother, for a willingness-to-pay threshold of £20,000 or £30,000, respectively. For universal ultrasound to be cost-saving, scans would need to cost less than £12.90 per mother.

Discussion

This is the first study that was published to assess routine ultrasound screening for presentation in the third trimester. The incidence of breech presentation was 4.6%, and for more than half of these, it had not previously been diagnosed by clinical assessment. The majority of these women were delivered by planned Caesarean section. Some women were delivered by emergency Caesarean section because they went into labour before their elective procedure, and a small number had a cephalic vaginal delivery following either spontaneous or. ECV. No woman in the cohort had a vaginal breech delivery or experienced an intrapartum Caesarean for undiagnosed breech.

Universal late pregnancy presentation scan would decrease the number of fetal mortalities associated with breech presentation. The cost-effectiveness analysis showed, and that this is of borderline cost-effectiveness, costing an estimated £23,611 per QALY gained, and that this is mainly driven by the cost of the scan itself and the maximum cost at which it would be cost-effective was £19.80. These unit costs may be feasible if assessment of presentation could be performed as part of a routine antenatal visit using low-cost portable ultrasound systems. Training midwives to perform presentation scans would likely require a few sessions and could be done at fraction of the costs associated with the training of a specialised sonographer. If universal ultrasound could be provided for less than £12.90 per scan, the policy would also be cost-saving.

Our analysis shows that introducing a policy of universal late pregnancy ultrasound screening would decrease the number of emergency Caesarean sections by about 5000 annually in the UK. However, it will increase the total number of Caesarean sections due to the increase of the elective Caesarean section by slightly less than 9000. Evidence suggests that Caesarean delivery may have long-term consequences on the health of the child (increased risk of asthma and obesity), the mother (reduced risk of pelvic organ prolapse and increased risk of subfertility) and future pregnancies (increased risk of placenta previa and stillbirth).[183, 184] There is no evidence that these are related to the type of the Caesarean section (elective vs emergency). [183, 184]

Fetal presentation was revealed to all women in the POP study. Consequently, this study cannot say what would have happened without routine screening. However, revealing the result was appropriate since there is level 1 evidence that planned caesarean delivery reduces the risk of perinatal morbidity and mortality at term.[185] However, less than half of all breech presentations in the POP study were

detected clinically. It is unclear whether the detection rates were affected by midwives knowing that the women were part of the POP study and hence would receive an ultrasound scan at 36wkGA. It is also unclear whether malpresentation would have been detected clinically later in pregnancy if the women had no presentation scan.

The prevalence of breech presentation in this analysis was 4.6% which is higher than the 3-4% that is often reported in literature.[18] However, this study report the prevalence of breech presentation at 36 weeks' gestation rather than at delivery. Considering that some fetuses will revert spontaneously to cephalic and some will be cephalic after successful ECV our finding is consistent with the literature. The ECV success rate in the POP study was considerably lower than reported elsewhere in the literature. This could be partly because the cohort consisted of nulliparous women, who have higher rates of ECV failure than parous women.[159, 186, 187] It is also possible that the real world ECV success rate is lower than in the literature due to publication bias. However, sensitivity analysis indicates that the impact from an increased ECV success rate would be modest (an increase in ECV success rate by 10 percentage points lowers the incremental cost of universal ultrasound by £0.91 per patient).

In conclusion this analysis showed that introducing a policy of universal late pregnancy ultrasound for fetal presentation would virtually eliminate undiagnosed breech presentation at term. It would also likely reduce perinatal morbidity and mortality. Introducing presentation scan into routine care into the routine 36 weeks appointment, for example, by midwives using a portable ultrasound system, it is likely to be cost-effective.

Chapter 9. Discussion

Overall conclusions

This thesis aimed to address the question whether we should offer a third-trimester ultrasound to all pregnant women regardless of their risk profile. More specifically, it was the first study to assess ultrasonic screening for fetal presentation, the largest study to assess universal screening using fetal macrosomia, and the first study to assess the effectiveness of screening using markers such as liquor volume and Doppler in a low-risk population. The main conclusions of the thesis are presented below:

First, there is a strong clinical and health economic case for implementing late pregnancy ultrasound screening to assess fetal presentation. It would virtually eliminate undiagnosed breech presentation, and prevent about 5000 emergency Caesarean sections and 8 perinatal deaths annually in the UK. The policy would be cost-effective at about £20 and cost-saving at about £13 per scan.

Second, universal ultrasound screening for macrosomia increased the detection of LGA fetuses but was only weakly predictive of shoulder dystocia. Despite the large number of studies in the analysis, there was not enough evidence to assess the effect of screening on neonatal morbidity. We recommend caution prior to introducing universal ultrasound screening for macrosomia as it would increase the rates of intervention without necessarily reducing neonatal morbidity.

Third, umbilical artery Doppler, CPR, severe oligohydramnios, and borderline oligohydramnios were all weakly predictive of the risk of delivering an SGA infant but either non-predictive or weakly predictive of the risk of neonatal morbidity. They should not be used alone to screen for neonatal morbidity, however a positive result would justify further fetal monitoring due to the association of all above markers with SGA.

Fourth, the quality of the studies included in the meta-analyses was variable and most studies did not blind the result of the index test. This could introduce bias either by not seeing associations where true associations exist (e.g. through treatment paradox) or by seeing associations where no true associations exist (e.g. through ascertainment bias or iatrogenic harm). Hence, much of the existing literature informing the utility of ultrasound as a screening test in late pregnancy has a profound limitation in addressing the research question. Only two studies (the POP study and the Genesis study in Ireland) performed universal late pregnancy ultrasound in unselected nulliparous women and

blinded the results to the clinicians. However, the latter has not published widely on the results. The results of the POP study in relation to both SGA and LGA (outcomes which are objectively defined and less prone to biases) were comparable to the summary estimates across all studies.

Clinical interpretation

Screening for breech presentation

This study makes a strong case for universal third trimester ultrasound screening for breech presentation. Our original publication in Plos Medicine[160] (Appendix 10) was the first study to assess universal ultrasound screening for presentation. The incidence of breech diagnosis was 4.6% and for about half of them the breech was not diagnosed prior to the scan. However, after the universal scan there were no undiagnosed breech presentations at delivery. It showed that the introduction of a policy of universal screening for presentation would prevent about 5000 emergency Caesarean sections which have higher risk of complications for both the mother and the fetus compared to elective Caesarean sections. Finally, it showed that ECV would have a minimal effect on the overall incidence of breech presentation at delivery (about 0.4%). However, in our study ECV was attempted in about half of the eligible cases with low rates of success (14.3%) which could be associated to the nulliparous population of the study.

Since that publication a large cohort study was published in Oxford, UK[188] which included both nulliparous and multiparous women. The Salim et al. study confirmed that the introduction of universal screening reduced the rate of undiagnosed breech presentation by about 80% from 22.3% to 4.7%. This remaining percentage of undiagnosed breech presentations was mostly due to multiparous women and only 4 nulliparous women reverted from cephalic to breech between the time of scan and delivery. It could also be explained by the earlier timing of scan starting at 35 weeks' gestation which makes reversion more likely. Finally, despite their higher uptake of ECV (>80%) and higher success rates (47.2%) the overall rate of breech presentation at delivery was only reduced by 0.1% from 2.7% to 2.6% which confirms that the main benefit of this policy would be through reducing the number of emergency Caesareans and undiagnosed vaginal breech deliveries rather than increasing the number of cephalic vaginal deliveries through ECV.

A policy of universal ultrasound screening for presentation could be implemented relatively quickly and cheaply. There is no need for long specialised training in ultrasonography to diagnose a breech presentation and it could be done by midwives after a few training sessions. Moreover, it does not require expensive equipment and it can be done with low-cost portable devices at an outpatient setting. There will be practically no false positives or false negatives at the time of the scan and only a small number of fetuses will revert to breech between the timing of the scan and delivery. Our

economic analysis shows that the policy would be cost-effective at £20 and cost-saving at £13 which we consider feasible.

After our publication, there was public interest and media attention on the findings which also led to MP questions in the UK parliament about the implementation of the policy. Finally, we have made a proposal to the National Screening Committee which is currently under consideration.

Screening for fetal biometry, liquor volume and Doppler.

Contrary to our findings about screening for breech presentation, the findings about screening for macrosomia, liquor volume, umbilical artery Doppler and CPR do not justify the introduction of policy of universal third trimester screening using those markers.

Our study on universal screening for macrosomia was published in Plos Medicine[47] (Appendix 8). It showed that universal screening for macrosomia identified slightly more than half LGA babies and 2 in 3 severely LGA babies at delivery with high diagnostic accuracy (positive likelihood ratios between 7 and 12). However, it identified only 1 in 5 cases of shoulder dystocia. This finding is not surprising given that the majority of cases of shoulder dystocia occur in appropriately sized babies.[90] Interestingly, two blinded studies, the POP[85] and Genesis[62] studies, reported no association between diagnosis of a LGA fetus and shoulder dystocia. Revealing a scan result may lead to ascertainment bias with an over reporting of shoulder dystocia as clinicians may have a lower threshold to initiate manoeuvres for shoulder dystocia or to document that a delayed delivery is due to shoulder dystocia when the baby was suspected to be LGA. Only the POP study reported on outcomes such as Apgar score, metabolic acidosis, neonatal jaundice and neonatal hypoglycaemia with very weak associations.

Moreover, a screening policy for macrosomia will need to be coupled with a successful intervention. A randomised control trial indicated that routine induction of labour in the presence of suspected macrosomia may reduce the risk of shoulder dystocia by about 70%[5]. However, in this study the incidence of shoulder dystocia in the control group was 4% which is unusually high and could possibly be explained by the fact that the study was not conducted in a low-risk population. Another large randomised trial, the Big Baby Trial trial (Induction of labour for predicted macrosomia: the Big Baby trial"; ISRCTN1822989) has not reported the results yet.

Our initial plan for analysis did not include screening for SGA. This was because another systematic review which was performed by one of our co-applicants which was later published.[29]. This review identified the POP study[9] as the only one that was done antenatally and blinded the results to clinicians. For SGA <10th centile it reported a sensitivity of 74% for a specificity of 88%. In the POP study, the sensitivity was 57% for a specificity of 90%. For severe SGA (birth weight <3rd percentile) it reported a sensitivity of 66% for a specificity of 87%. In the POP study, the sensitivity was 77% for a specificity of 87%. The findings are surprising since it would be expected that the prediction would be higher for a more severe outcome, as was observed in the POP study. This could reflect the differences in the studies included.

A policy of induction of labour for suspected SGA has some potential risks due to the false positives (incorrectly identified as small). This was evident in France.[189] where after the introduction of routine ultrasonography there were higher rates of adverse outcomes in the false positive women compared to those correctly identified as having an appropriately grown baby. These adverse outcomes include a six-fold increased risk of neonatal resuscitation and a two-fold increase in admission to NICU. This was likely secondary to a four-fold increase in delivery at preterm gestations in the group of women who were incorrectly classified as having an SGA baby. Using routine ultrasonography to detect SGA fetuses improves the sensitivity of detection of SGA at the expense of reduced specificity. In the POP study when selective ultrasound was used, 2% of women were incorrectly classified as SGA but this increased to 10% when routine ultrasonography was used. This equates to two additional false positives for every one additional true positive result with the use of universal scanning.

Ultrasonic markers such as the umbilical artery Doppler, the CPR, and the amniotic fluid volume have been studied extensively in high-risk populations such as fetuses with FGR. We studied in detail all those markers in this study using both systematic review and by analysing original data of the POP study (Appendix 9).[95] As described above all those markers could be moderately predictive of SGA and severe SGA but non- or weakly predictive of neonatal morbidity. Moreover, there are no clinical trials to assess the effectiveness of delivery based on the abnormality of those markers in low-risk populations. Hence, universal screening using those markers is unlikely to reduce neonatal morbidity.

Additional cost-effectiveness and value of information analysis

As part of the larger HTA project which we later published[113], we planned to use the results generated from the meta-analyses to produce cost-effectiveness analyses for universal third trimester screening in collaboration with the health economists David Wastlund and Ed Wilson. The findings for the cost-effectiveness of universal screening for presentation are presented in chapter 8. I contributed to the two following publications[160, 190] mainly by helping design the decision models, filling the inputs and writing the clinical part of the draft.

Cost-effectiveness of late pregnancy screening for macrosomia

In the Wastlund et al paper[191] we sought to identify the most cost-effective policy for detection and management of fetal macrosomia in late-stage pregnancy. We compared long-term maternal-fetal health and cost outcomes for two strategies for detection (universal ultrasound scanning versus selective ultrasound scanning), combined with three strategies for management (planned Caesarean section versus induction of labour versus expectant management) of suspected fetal macrosomia.

The model structure for detection and management for macrosomia is shown in Appendix 9E (Figure 1 of the paper). Four different screening statuses were possible: true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN). The likelihood of each state was driven by the sensitivity and specificity of the test used for detection, as well as the prevalence of macrosomia. Suspected macrosomia was managed according to the pre-determined management strategy: planned Caesarean section, induction of labour, or expectant management. If macrosomia was not suspected, it was assumed vaginal delivery would be attempted, with a risk of emergency Caesarean section.

Five fetal delivery outcomes were possible: No complications, Respiratory distress, Shoulder dystocia, Other acidosis (i.e. acidosis not induced by shoulder dystocia), and death. Their respective likelihood were affected by both screening and management strategies (see below). The outcomes from the fetal delivery outcomes were then extrapolated into long-term costs and quality-adjusted life years (QALYs).

We showed that the least expensive option was selective ultrasound with expectant management, and the most, universal ultrasound with planned CS. The least effective option (in terms of QALYs

gained) was universal ultrasound with planned CS, and the most, universal ultrasound with induction of labour. Assuming a willingness to pay threshold of £20,000 per QALY (as per standard in the NHS), universal US plus induction of labour was expected to yield marginally greater QALYs (+0.002) at an added cost of £113 which yields an ICER of £52,719. This was above the threshold and is therefore not cost-effective.

Late pregnancy ultrasound to screen for and manage potential birth complications in nulliparous women: a cost-effectiveness and value of information analysis

In the paper published by Wilson et al[190] we estimate the cost-effectiveness of various late-pregnancy screening and management strategies, and predict the return on investment from a future clinical trial.

We compared three screening options, 'selective', 'universal breech' and 'universal'. 'Selective' screening means only those mothers who are clinically indicated (current practice in the NHS). The 'universal breech' scanning strategy offers all mothers a simple presentation-only scan, whereas scans assessing other features, such as biometry, would be as per current clinical indications. 'Universal' screening is defined as all mothers receiving an ultrasound scan including fetal biometry. Findings from a presentation scan can be either cephalic or breech, and fetal size could be either appropriate, small or large for gestational age (AGA, SGA and LGA, respectively). If a breech presentation is identified, all mothers are offered ECV unless contraindicated. If this is declined or unsuccessful, an elective Caesarean section may be scheduled. If LGA is detected, the mother may be offered either IOL or expectant management. If SGA is detected, all mothers are offered IOL. We therefore compared six alternative screening and management policies comprising three possible screening modes and two alternative management plans, summarised in Appendix 9F (Table 1 in the paper). We created decision trees and used short term outcomes for each condition (breech, LGA and SGA) as described in the paper. In this analysis we also included long-term outcomes such as special educational needs.

We showed that given a willingness to pay of £20,000 per QALY gained, the most cost-effective strategy is a routine presentation-only scan for all women. Universal ultrasound screening for fetal size is unlikely to be cost-effective. Value of information analysis showed that the greatest uncertainty was around the costs associated with IOL and that a randomised controlled trial with an endpoint of stillbirth is extremely unlikely to be a value for money investment.

Patient and public involvement

In the HTA project we sought to assess views of recently delivered and currently pregnant women on universal ultrasound screening in late pregnancy which we have published in the HTA report.[113]. The aims of this section were threefold. First, to assess the knowledge of pregnant women on the current antenatal care pathway for low-risk pregnancies. Second, to assess their understanding of the potential benefits and drawbacks of third trimester screening. Third, to estimate their willingness to participate in a future randomised clinical trial, examine which trial design they would prefer to participate in, and calculate the expected recruitment rate.

For these above aims we conducted a survey and ran focus groups. For both aims we collaborated with the NIHR Cambridge BRC Communications and PPI Department of Cambridge University Hospitals NHS Foundation Trust (CUHFT).

Survey

The objective of the survey was to meet the requirements of aims 1 and 3 by involving a large and representative number of women. We planned to recruit low risk nulliparous women after their ultrasound scan at 12 or 20 weeks' gestation, given that the scan confirms a viable pregnancy. The questionnaire was approved by all the collaborators of the study and tested by the PPI office in CUHFT to ensure it was understood by the women. We received feedback from five anonymous individuals and modified our form accordingly. The final version of the questionnaire in Appendix 13. In brief, this questionnaire included three parts. The first two questions were about their knowledge of the current antenatal care and their willingness to have an additional ultrasound scan in the third trimester. The second part included three questions about potential participation in a future randomised trial. We discussed two possible trial designs. The first study (study A) would randomize low risk women to have a scan at 36 weeks' gestation or not (current standard of care). The ultrasound results would be revealed to their clinical care team and their management would be affected accordingly. In the second study (study B) all women would have an ultrasound at 36 weeks' gestation. If there was a major problem (eg breech presentation or very small amount of fluid around the baby) the result would be revealed to the care team. In all other cases the result would be blinded to the mothers and the clinicians. Finally, we included some questions on women's demographics, such as age, ethnicity, and education to ensure that the sample of women was diverse. All the replies were anonymised.

We collected 100 replies from pregnant women in the Rosie hospital, Cambridge. We present the results in Table 16. The respondents were diverse regarding their age group, ethnicity and education level. The majority (85%) was aware that low risk pregnancies are not been offered routine ultrasound in the third trimester and 84% would like to have a routine third trimester scan. Regarding participation in a future clinical trial, 76% would agree or strongly agree to participate in study A and 66% in study B. When asked which study they would prefer to participate in, out of the 65 women that replied this question, 10 (15.4%) preferred study A, 23 (35.4%) study B, and 32 (49.2%) would be happy to participate in either study.

Focus group

The second part this section was to run a focus group in which we could discuss the qualitative aspects of all the above aims. We planned to recruit women that have recently delivered (within the last two years), and discuss in detail the benefits and potential risks of third trimester screening. For the advertisement we used the mailing list of the PPI office, personal contact by midwives, and social media. The focus group discussion was run by myself and Amanda Stranks, the Lead of the PPI office in CUHFT.

Eight women showed an initial interest in participating in our focus groups. Due to difficulties with childcare four of the women could not participate in a focus group in one of multiple suggested dates. We managed to run one focus group with four participants. The participant characteristics are as below (quoted verbatim form the HTA report):

A: One previous delivery, low risk, she was measuring slightly small on symphysis-fundal height (2cm below the appropriate for the gestational age) but had no extra scans. Normal uncomplicated delivery of 2.49kg baby at 40wkGA. Her motivation for participation was whether she needed a third scan. She also mentioned that her husband is French where they all have a third trimester scan and she wanted to know why this is not the policy in the UK.

B: Two previous deliveries (4 and 2 years old), both low risk. The first baby was born in the birth centre, for the second she had induction of labour for postdates. Both deliveries were uncomplicated. Her motivation for participation was that four of her friends had stillbirths at term in the last few years which she found very stressful as she was planning for a third pregnancy.

C: One previous delivery, initially high risk due to low BMI, had growth scans at 32 and 36 weeks (both normal). Then discharged to midwifery care. Delivered in the midwifery unit without complications. Her motivation for participating was whether she needed all these scans as it was difficult to attend due to work.

D: One previous delivery, initially low risk. Due to low PAPP-A she had close monitoring during pregnancy. She had IOL at 37wkGA for suspected FGR. She delivered vaginally a 2.1Kg baby (2nd centile) who stayed in NICU for 3 days. Her motivation for participation was whether this could have been missed if the PAPP-A was not marginally abnormal in the first trimester.

We initially discussed their opinion on the current screening schedule and whether they would want an additional ultrasound scan in the third trimester. Two participants (A and B) thought that this is not enough and there is long period after 20wkGA that they don't know about the fetal wellbeing. They both believed that an additional scan would make them feel more reassured. One participant (C) considered herself low risk (despite her low BMI) and found it difficult to attend the additional scans that she was offered. Finally, the fourth participant thought that the schedule was about right and she wanted to have more evidence that the additional scans would be beneficial before introducing them.

We then discussed about potential diagnoses such as breech presentation, SGA and LGA. The management in each case was explained and the statistics regarding the risks and benefits. We also discussed a large study from France which showed that universal screening could cause harm. In the case of breech presentation all participants said that they would definitely want to know and they would all opt for external cephalic version in case of diagnosis. In the cases of SGA and LGA one participant (B) said that she would definitely want to know and that she would opt for IOL if she was diagnosed with either SGA and LGA. Two participants (A and D) said that they would still want to have the scan but were not sure about induction of labour and they would like to have further conversation with the doctors. One participant (C) said that she was sceptical about the potential misdiagnosis and hesitant about the management.

Finally we discussed about participation in a future trial. All women would be happy to participate in a future trial. When we specifically discussed the two potential study designs as above they all preferred study B (screening all women and randomizing to blind or not the result). This was because they would be reassured about the baby's presentation and that a diagnosis of a severe problem would be revealed. The main suggestions about blinding were that we had to make clear which

conditions would be revealed and which would not. Additionally they wanted us to explain clearly that we are not withholding information from them but we simply collect more of it, and that they would receive the normal care in case they were randomized in the control group. When we discussed about the timing of the consent they would all be happy to be approached in the first or second trimester. However, they would prefer to have a second discussion about the randomization at 36 wkGA because they would have forgotten the details of the consent form at 12 or 20 wkGA and they would prefer to have a longer conversation at that point.

Future research

Feasibility of a randomised clinical trial of screening with ultrasound and intervention

We considered all the above findings from the meta-analyses, cost-effectiveness and VOI analyses, and patient perspectives to assess potential designs for a randomised controlled trial. The results were published in Smith et al[113]**.** The size and, consequently, cost of the trial would depend both on the design of the trial and the primary outcomes selected.

There are two possible approaches to trial design. First, a "screen versus no screen" design in which women might be randomised (a) to be screened, with the offer of intervention if they screen positive, or (b) to receive routine care. The result of this trial design is a simple comparison between the two groups The drawback of this design is that in the event of a negative result, it is impossible to determine whether the trial failed because of the screening or the intervention. The second approach is a "screen all" design, in which the whole of the population is screened and high risk women are randomised to intervention or routine care (masking the result in the latter group). The advantages of the second approach are that the number of women who need to be recruited is substantially reduced and that the same trial can assess both the diagnostic effectiveness of the screening test and the clinical effectiveness of the intervention. The two approaches are illustrated in Appendix 8, Figure 33.

The choice of primary outcome would greatly influence the power calculations for a randomised trial (Appendix 8, Table 19). Using perinatal mortality would require at least 200.000 participants, assuming that the test has high predictive accuracy (positive LR) and that the intervention would reduce the outcome by 50%. Using neonatal morbidity would require close to 40.000 participants. Using SGA as the primary outcome would reduce the number of participants required but would include a lot of babies constitutionally small, therefore not at higher risk. All power calculations are performed for $P<0.05$ (two-sided) with 90% power to detect the effect.

Overall, the conclusion was that using ultrasound alone as screening test in a randomised trial would make this trial too large and too expensive to conduct.

Combining ultrasound with biochemical markers

A large number of biochemical markers have been assessed for the prediction of SGA and stillbirth in all trimesters. These included markers that are currently used for Down's syndrome screening, such as the pregnancy-associated plasma protein A (PAPP-A), angiogenic factors, such as soluble fms-like tyrosine kinase-1 (sFLT1), soluble endoglin (sENG), and placental growth factor (PlGF), and hormonal factors such as the insulin-like growth factor binding protein (IGFBP). A detailed review for all those markers was published by Gaccioli et al.[192].

A Cochrane systematic review[29] assessed the predictive accuracy of four biomarkers including human placental lactogen (hPL), oestriol, PlGF and uric acid, and compared it with ultrasound for the prediction of SGA or stillbirth. It showed that PlGF had high predictive accuracy for identifying stillbirth (DOR ~49), and PlGF and hPL had moderate predictive accuracy for SGA. However, most of the studies included were done in high-risk populations and none was done exclusively in a low-risk population.

A study that was published in the POP cohort[130] using ultrasound in combination with the sFLT1/PlGF ratio at 28 and 36 weeks' gestation. At 28 weeks, the combination of ultrasonic small for gestational age and an elevated sFLT1/PlGF ratio which was observed in 1% of pregnancies had high predictive accuracy (positive LR ~40) for the preterm delivery of a small for gestational age infant. At 36 weeks, the combination of ultrasonic small for gestational age and an elevated sFLT1/PlGF ratio was observed in 3% of the cases and the positive LR for delivery of a small for gestational age infant associated with maternal pre-eclampsia or perinatal morbidity or mortality was ~17.

Finally, study by Sovio et al[131] was done in the POP cohort and additionally in the Born in Bradford (BiB) cohort using ultrahigh performance liquid chromatography-tandem mass spectroscopy (UPLC-MS/MS) metabolomics on maternal serum at 12, 20 and 28 weeks' gestation. It demonstrated that a ratio of the products of the relative concentrations of two positively associated metabolites (1-(1-enyl-stearoyl)-2-oleoyl-GPC (P-18:0/18:1) and 1,5-anhydroglucitol) to the product of the relative concentrations of two negatively associated metabolites (5α-androstan-3α,17α-diol disulfate and N1,N12-diacetylspermine) predicted FGR at term. The ratio had approximately double the discrimination as compared to the sFLT1:PlGF ratio (AUC 0.78 versus 0.64, P = 0.0001).

Overall, the above show that combining ultrasound with biochemical markers could make a trial of screening and intervention feasible and likely effective as it would identify fetuses that are at higher risk of an adverse outcome compared to those identified by ultrasound alone.

References

1. National Institute for Health and Care Excellence. NICE guidline: Antenatal Care. London; 2008.

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

3. Boers KE, Vijgen SM, Bijlenga D, van der Post JA, Bekedam DJ, Kwee A, et al. Induction versus expectant monitoring for intrauterine growth restriction at term: randomised equivalence trial (DIGITAT). BMJ (Clinical research ed). 2010 Dec 21;341:c7087.

4. Middleton P, Shepherd E, Crowther CA. Induction of labour for improving birth outcomes for women at or beyond term. The Cochrane database of systematic reviews. 2018 May 9;5:Cd004945.

5. Boulvain M, Senat MV, Perrotin F, Winer N, Beucher G, Subtil D, et al. Induction of labour versus expectant management for large-for-date fetuses: a randomised controlled trial. Lancet (London, England). 2015 Jun 27;385(9987):2600-5.

6. 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. American journal of obstetrics and gynecology. 1985 Feb 1;151(3):333-7.

7. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. Radiology. 1991 Oct;181(1):129-33.

8. Kiserud T, Piaggio G, Carroli G, Widmer M, Carvalho J, Neerup Jensen L, et al. The World Health Organization Fetal Growth Charts: A Multinational Longitudinal Study of Ultrasound Biometric Measurements and Estimated Fetal Weight. PLoS medicine. 2017 Jan;14(1):e1002220.

9. 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 (London, England). 2015 Nov 21;386(10008):2089-97.

10. Hoffman C, Galan HL. Assessing the 'at-risk' fetus: Doppler ultrasound. Curr Opin Obstet Gynecol. 2009 Apr;21(2):161-6.

11. Krebs C, Macara LM, Leiser R, Bowman AW, Greer IA, Kingdom JC. Intrauterine growth restriction with absent end-diastolic flow velocity in the umbilical artery is associated with maldevelopment of the placental terminal villous tree. American journal of obstetrics and gynecology. 1996 Dec;175(6):1534-42.

12. Morris RK, Malin G, Robson SC, Kleijnen J, Zamora J, Khan KS. Fetal umbilical artery Doppler to predict compromise of fetal/neonatal wellbeing in a high-risk population: systematic review and

bivariate meta-analysis. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2011 Feb;37(2):135-42.

13. Alfirevic Z, Stampalija T, Dowswell T. Fetal and umbilical Doppler ultrasound in high-risk pregnancies. The Cochrane database of systematic reviews. 2017 Jun 13;6:Cd007529.

14. Alfirevic Z, Stampalija T, Medley N. Fetal and umbilical Doppler ultrasound in normal pregnancy. The Cochrane database of systematic reviews. 2015 Apr 15(4):Cd001450.

15. Smith GC. Researching new methods of screening for adverse pregnancy outcome: lessons from pre-eclampsia. PLoS medicine. 2012;9(7):e1001274.

16. Conde-Agudelo A, Villar J, Kennedy SH, Papageorghiou AT. Predictive accuracy of cerebroplacental ratio for adverse perinatal and neurodevelopmental outcomes in suspected fetal growth restriction: systematic review and meta-analysis. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2018 Oct;52(4):430-41.

17. Hamza A, Herr D, Solomayer EF, Meyberg-Solomayer G. Polyhydramnios: Causes, Diagnosis and Therapy. Geburtshilfe und Frauenheilkunde. 2013 Dec;73(12):1241-6.

18. Impey LWM MD, Griffiths M, Penna LK on behalf of the Royal College of Obstetricians and Gynaecologists. Management of Breech Presentation. BJOG : an international journal of obstetrics and gynaecology. 2017;124:e151-e77.

19. Impey LWM MD, Griffiths M, Penna LK on behalf of the Royal College of Obstetricians and Gynaecologists. External Cephalic Version and Reducing the Incidence of Term Breech Presentation. BJOG : an international journal of obstetrics and gynaecology. 2017;124:e178-e92.

20. Hannah ME, Hannah WJ, Hewson SA, Hodnett ED, Saigal S, Willan AR. Planned caesarean section versus planned vaginal birth for breech presentation at term: a randomised multicentre trial. Term Breech Trial Collaborative Group. Lancet (London, England). 2000 Oct 21;356(9239):1375-83.

21. Hunter LA. Vaginal breech birth: can we move beyond the Term Breech Trial? Journal of midwifery & women's health. 2014 May-Jun;59(3):320-7.

22. Leung WC, Pun TC, Wong WM. Undiagnosed breech revisited. BJOG : an international journal of obstetrics and gynaecology. 1999 Jul;106(7):638-41.

23. Nassar N, Roberts CL, Cameron CA, Olive EC. Diagnostic accuracy of clinical examination for detection of non-cephalic presentation in late pregnancy: cross sectional analytic study. BMJ (Clinical research ed). 2006 Sep 16;333(7568):578-80.

24. Impey LWM MD, Griffiths M, Penna LK on behalf of the Royal College of Obstetriciansand Gynaecologists. External Cephalic Version and Reducing the Incidence of Term Breech Presentation. BJOG: An International Journal of Obstetrics & Gynaecology. 2017;124(7):e178-e92.

25. Morrison JJ, Rennie JM, Milton PJ. Neonatal respiratory morbidity and mode of delivery at term: influence of timing of elective caesarean section. British journal of obstetrics and gynaecology. 1995 Feb;102(2):101-6.

26. MacKay DF, Smith GC, Dobbie R, Pell JP. Gestational age at delivery and special educational need: retrospective cohort study of 407,503 schoolchildren. PLoS medicine. 2010 Jun 8;7(6):e1000289.

27. Bricker L, Medley N, Pratt JJ. Routine ultrasound in late pregnancy (after 24 weeks' gestation). The Cochrane database of systematic reviews. 2015 Jun 29(6):Cd001451.

28. Smith G. A critical review of the Cochrane meta-analysis of routine late-pregnancy ultrasound. BJOG : an international journal of obstetrics and gynaecology. 2021 Jan;128(2):207-13.

29. Heazell AE, Hayes DJ, Whitworth M, Takwoingi Y, Bayliss SE, Davenport C. Biochemical tests of placental function versus ultrasound assessment of fetal size for stillbirth and small-for-gestational-age infants. The Cochrane database of systematic reviews. 2019 May 14;5:Cd012245.

30. Brouwers L, van der Meiden-van Roest AJ, Savelkoul C, Vogelvang TE, Lely AT, Franx A, et al. Recurrence of pre-eclampsia and the risk of future hypertension and cardiovascular disease: a systematic review and meta-analysis. BJOG : an international journal of obstetrics and gynaecology. 2018 Dec;125(13):1642-54.

31. Phillips C, Velji Z, Hanly C, Metcalfe A. Risk of recurrent spontaneous preterm birth: a systematic review and meta-analysis. BMJ Open. 2017 Jul 5;7(6):e015402.

32. Lamont K, Scott NW, Jones GT, Bhattacharya S. Risk of recurrent stillbirth: systematic review and meta-analysis. BMJ (Clinical research ed). 2015 Jun 24;350:h3080.

33. Kinzler WL, Kaminsky L. Fetal growth restriction and subsequent pregnancy risks. Semin Perinatol. 2007 Jun;31(3):126-34.

34. Brocklehurst P, Hardy P, Hollowell J, Linsell L, Macfarlane A, McCourt C, et al. 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 (Clinical research ed). 2011 Nov 23;343:d7400.

35. Group SCRNW. Association between stillbirth and risk factors known at pregnancy confirmation. Jama. 2011 Dec 14;306(22):2469-79.

36. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011 Oct 18;155(8):529-36.

37. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. Journal of clinical epidemiology. 1991;44(8):763-70.

38. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Statistics in medicine. 2001 Oct 15;20(19):2865-84.

39. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of clinical epidemiology. 2005 Oct;58(10):982-90.

40. Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. BMJ (Clinical research ed). 2001 Jul 21;323(7305):157-62.

41. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. Journal of clinical epidemiology. 2005 Sep;58(9):882-93.

42. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS medicine. 2009 Jul 21;6(7):e1000097.

43. Freeman JV, Cole TJ, Chinn S, Jones PR, White EM, Preece MA. Cross sectional stature and weight reference curves for the UK, 1990. Archives of disease in childhood. 1995 Jul;73(1):17-24.

44. 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 Jun;200(6):672.e1-4.

45. Campbell S, Wilkin D. Ultrasonic measurement of fetal abdomen circumference in the estimation of fetal weight. British journal of obstetrics and gynaecology. 1975 Sep;82(9):689-97.

46. Boulvain M, Irion O, Dowswell T, Thornton JG. Induction of labour at or near term for suspected fetal macrosomia. The Cochrane database of systematic reviews. 2016 May 22(5):Cd000938.

47. Moraitis AA, Shreeve N, Sovio U, Brocklehurst P, Heazell AEP, Thornton JG, et al. Universal third-trimester ultrasonic screening using fetal macrosomia in the prediction of adverse perinatal outcome: A systematic review and meta-analysis of diagnostic test accuracy. PLoS medicine. 2020 Oct;17(10):e1003190.

48. Aviram A, Yogev Y, Ashwal E, Hiersch L, Hadar E, Gabbay-Benziv R. Prediction of large for gestational age by various sonographic fetal weight estimation formulas-which should we use? Journal of Perinatology. 2017;37(5):513-7.

49. Balsyte D, Schaffer L, Burkhardt T, Wisser J, Kurmanavicius J. Sonographic prediction of macrosomia cannot be improved by combination with pregnancy-specific characteristics. Ultrasound in Obstetrics & Gynecology. 2009;33(4):453-8.

50. Benacerraf BR, Gelman R, Frigoletto Jr FD. Sonographically estimated fetal weight: Accuracy and limitation. American journal of obstetrics and gynecology. 1988;159(5):1118-21.

51. Ben-Haroush A, Yogev Y, Hod M, Bar J. Predictive value of a single early fetal weight estimate in normal pregnancies. European Journal of Obstetrics Gynecology and Reproductive Biology. 2007 February;130(2):187-92.

52. Ben-Haroush A, Melamed N, Mashiach R, Meizner I, Yogev Y. Use of the amniotic fluid index combined with estimated fetal weight within 10 days of delivery for prediction of macrosomia at birth. Journal of Ultrasound in Medicine. 2008 July;27(7):1029-32.

53. Benson CB, Coughlin BF, Doubilet PM. Amniotic fluid volume in large-for-gestational-age fetuses of nondiabetic mothers. Journal of Ultrasound in Medicine. 1991;10(3):149-51.

54. Burkhardt T, Schmidt M, Kurmanavicius J, Zimmermann R, Schaffer L. Evaluation of fetal anthropometric measures to predict the risk for shoulder dystocia. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2014 Jan;43(1):77-82.

55. Chauhan SP, Parker D, Shields D, Sanderson M, Cole JH, Scardo JA. Sonographic estimate of birth weight among high-risk patients: feasibility and factors influencing accuracy. American journal of obstetrics and gynecology. 2006 Aug;195(2):601-6.

56. Chervenak JL, Divon MY, Hirsch J, Girz BA, Langer O. Macrosomia in the postdate pregnancy: Is routine ultrasonographic screening indicated? American journal of obstetrics and gynecology. 1989;161(3):753-6.

57. Cohen JM, Hutcheon JA, Kramer MS, Joseph KS, Abenhaim H, Platt RW. Influence of ultrasound-to-delivery interval and maternal-fetal characteristics on validity of estimated fetal weight. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2010 Apr;35(4):434-41.

58. Crimmins S, Mo C, Nassar Y, Kopelman JN, Turan OM. Polyhydramnios or Excessive Fetal Growth Are Markers for Abnormal Perinatal Outcome in Euglycemic Pregnancies. American Journal of Perinatology. 2018 01 Jan;35(2):140-5.

59. Cromi A, Ghezzi F, Di Naro E, Siesto G, Bergamini V, Raio L. Large cross-sectional area of the umbilical cord as a predictor of fetal macrosomia. Ultrasound in Obstetrics & Gynecology. 2007;30(6):861-6.

60. De Reu PAOM, Smits LJM, Oosterbaan HP, Nijhuis JG. Value of a single early third trimester fetal biometry for the prediction of birth weight deviations in a low risk population. Journal of Perinatal Medicine. 2008 01 Jul;36(4):324-9.

61. Freire DMC, Cecatti JG, Paiva CSM. Correlation between estimated fetal weight by ultrasound and neonatal weight. [Portuguese]. Revista Brasileira de Ginecologia e Obstetricia. 2010 Jan.;32(1):4-10.

62. Galvin DM, Burke N, Burke G, Breathnach F, McAuliffe F, Morrison J, et al. 94: Accuracy of prenatal detection of macrosomia >4,000g and outcomes in the absence of intervention: results of the prospective multicenter genesis study. American journal of obstetrics and gynecology. 2017 2017/01/01/;216(1, Supplement):S68.

63. Gilby JR, Williams MC, Spellacy WN. Fetal abdominal circumference measurements of 35 and 38 cm as predictors of macrosomia. A risk factor for shoulder dystocia. Journal of Reproductive Medicine. 2000;45(11):936-8.

64. Hasenoehrl G, Pohlhammer A, Gruber R, Staudach A, Steiner H. Fetal weight estimation by 2D and 3D ultrasound: Comparison of six formulas. Ultraschall in der Medizin. 2009;30(6):585-90.

65. Hendrix NW, Grady CS, Chauhan SP. Clinical vs. sonographic estimate of birth weight in term parturients. A randomized clinical trial. Journal of Reproductive Medicine. 2000;45(4):317-22.

66. Henrichs C, Magann EF, Brantley KL, Crews JH, Sanderson M, Chauhan SP. Detecting fetal macrosomia with abdominal circumference alone. Journal of Reproductive Medicine. 2003;48(5):339-42.

67. Humphries J, Reynolds D, Bell-Scarbrough L, Lynn N, Scardo JA, Chauhan SP. Sonographic estimate of birth weight: relative accuracy of sonographers versus maternal-fetal medicine specialists. Journal of Maternal-Fetal & Neonatal Medicine. 2002;11(2):108-12.

68. Kayem G, Grange G, Breart G, Goffinet F. Comparison of fundal height measurement and sonographically measured fetal abdominal circumference in the prediction of high and low birth weight at term. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2009 Nov;34(5):566-71.

69. Kehl S, Brade J, Schmidt U, Berlit S, Bohlmann MK, Sutterlin M, et al. Role of fetal abdominal circumference as a prognostic parameter of perinatal complications. Archives of Gynecology & Obstetrics. 2011;284(6):1345-9.

70. Khan N, Ciobanu A, Karampitsakos T, Akolekar R, Nicolaides KH. Prediction of large for gestational age neonates by routine third trimester ultrasound. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2019;25.

71. Levine AB, Lockwood CJ, Brown B, Lapinski R, Berkowitz RL. Sonographic diagnosis of the large for gestational age fetus at term: does it make a difference? Obstetrics & Gynecology. 1992;79(1):55-8.

72. Melamed N, Yogev Y, Meizner I, Mashiach R, Pardo J, Ben-Haroush A. Prediction of fetal macrosomia: effect of sonographic fetal weight-estimation model and threshold used. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2011 Jul;38(1):74-81.

73. Miller JM, Jr., Korndorffer FA, 3rd, Gabert HA. Fetal weight estimates in late pregnancy with emphasis on macrosomia. Journal of Clinical Ultrasound. 1986;14(6):437-42.

74. Miller Jr JM, Brown HL, Khawli OF, Pastorek IJG, Gabert HA. Ultrasonographic identification of the macrosomic fetus. American journal of obstetrics and gynecology. 1988;159(5):1110-4.

75. Nahum GG, Pham KQ, McHugh JP. Ultrasonic prediction of term birth weight in Hispanic women. Accuracy in an outpatient clinic. Journal of Reproductive Medicine. 2003;48(1):13-22.

76. Nahum GG, Stanislaw H. A computerized method for accurately predicting fetal macrosomia up to 11 weeks before delivery. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2007;133(2):148-56.

77. Nicod AC, Hohlfeld P, Vial Y. Performance of ultrasound estimation of fetal weight in fetuses weighing <= 2000 g and more than 4000 g. [French]. Revue Medicale Suisse. 2012 24 Oct;8(359):2022-7.

78. O'Reilly-Green CP, Divon MY. Receiver operating characteristic curves of sonographic estimated fetal weight for prediction of macrosomia in prolonged pregnancies. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 1997 Jun;9(6):403-8.

79. Pates JA, McIntire DD, Casey BM, Leveno KJ. Predicting macrosomia. Journal of Ultrasound in Medicine. 2008;27(1):39-43.

80. Peregrine E, O'Brien P, Jauniaux E. Clinical and ultrasound estimation of birth weight prior to induction of labor at term. Ultrasound in Obstetrics & Gynecology. 2007;29(3):304-9.

81. Pollack RN, Hauer-Pollack G, Divon MY. Macrosomia in postdates pregnancies: the accuracy of routine ultrasonographic screening. American Journal of Obstetrics & Gynecology. 1992;167(1):7-11.

82. Rossavik IK, Joslin GL. Macrosomatia and ultrasonography: what is the problem? Southern Medical Journal. 1993;86(10):1129-32.

83. Sapir A, Khayyat I, Drukker L, Rabinowitz R, Samueloff A, Sela HY. Ultrasound predication of shoulder dystocia in low risk term singleton deliveries. American journal of obstetrics and gynecology. 2017 January;216 (1 Supplement 1):S221.

84. Smith GC, Smith MF, McNay MB, Fleming JE. The relation between fetal abdominal circumference and birthweight: findings in 3512 pregnancies. British journal of obstetrics and gynaecology. 1997 Feb;104(2):186-90.

85. Sovio U, Moraitis AA, Wong HS, Smith GCS. Universal vs selective ultrasonography to screen for large-for-gestational-age infants and associated morbidity. Ultrasound in obstetrics & gynecology

: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2018 Jun;51(6):783-91.

86. Sritippayawan S, Anansakunwat W, Suthantikorn C. The accuracy of gestation-adjusted projection method in estimating birth weight by sonographic fetal measurements in the third trimester. Journal of the Medical Association of Thailand. 2007 June;90(6):1058-67.

87. Sylvestre G, Divon MY, Onyeije C, Fisher M. Diagnosis of macrosomia in the postdates population: combining sonographic estimates of fetal weight with glucose challenge testing. Journal of Maternal-Fetal Medicine. 2000;9(5):287-90.

88. Weiner Z, Ben-Shlomo I, Beck-Fruchter R, Goldberg Y, Shalev E. Clinical and ultrasonographic weight estimation in large for gestational age fetus. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2002;105(1):20-4.

89. Burke N, Burke G, Breathnach F, McAuliffe F, Morrison JJ, Turner M, et al. Prediction of cesarean delivery in the term nulliparous woman: results from the prospective, multicenter Genesis study. American journal of obstetrics and gynecology. 2017 Jun;216(6):598.e1-.e11.

90. Ouzounian JG. Shoulder Dystocia: Incidence and Risk Factors. Clin Obstet Gynecol. 2016 Dec;59(4):791-4.

91. Little SE, Edlow AG, Thomas AM, Smith NA. Estimated fetal weight by ultrasound: a modifiable risk factor for cesarean delivery? American journal of obstetrics and gynecology. 2012 Oct;207(4):309.e1-6.

92. 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? American journal of obstetrics and gynecology. 2009 Mar;200(3):340.e1-3.

93. Parry S, Severs CP, Sehdev HM, Macones GA, White LM, Morgan MA. Ultrasonographic prediction of fetal macrosomia. Association with cesarean delivery. The Journal of reproductive medicine. 2000 Jan;45(1):17-22.

94. Smith G. Should we implement universal screening with late pregnancy ultrasound to prevent stillbirth? BJOG : an international journal of obstetrics and gynaecology. 2018 Jan;125(2):101-3.

95. Moraitis AA, Bainton T, Sovio U, Brocklehurst P, Heazell AE, Thornton JG, et al. Fetal umbilical artery Doppler as a tool for universal third trimester screening: A systematic review and meta-analysis of diagnostic test accuracy. Placenta. 2021 May;108:47-54.

96. 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 Nov 19;8:51.

97. Akolekar R, Ciobanu A, Zingler E, Syngelaki A, Nicolaides KH. Routine assessment of cerebroplacental ratio at 35-37 weeks' gestation in the prediction of adverse perinatal outcome. American Journal of Obstetrics and Gynecology. 2019.

98. Bolz N, Kalache KD, Proquitte H, Slowinski T, Hartung JP, Henrich W, et al. Value of Doppler sonography near term: can umbilical and uterine artery indices in low-risk pregnancies predict perinatal outcome? Journal of Perinatal Medicine. 2013;41(2):165-70.

99. Cooley SM, Donnelly JC, Walsh T, MacMahon C, Gillan J, Geary MP. The impact of umbilical and uterine artery Doppler indices on antenatal course, labor and delivery in a low-risk primigravid population. Journal of Perinatal Medicine. 2011;39(2):143-9.

100. Filmar G, Panagopoulos G, Minior V, Barnhard Y, Divon MY. Elevated umbilical artery systolic/diastolic ratio in the absence of fetal growth restriction. Archives of gynecology and obstetrics. 2013 August;288(2):279-85.

101. Fischer RL, Kuhlman KA, Depp R, Wapner RJ. Doppler evaluation of umbilical and uterine-arcuate arteries in the postdates pregnancy. Obstetrics & Gynecology. 1991;78(3 Pt 1):363-8.

102. Goffinet F, Paris J, Heim N, Nisand I, Breart G. Predictive value of Doppler umbilical artery velocimetry in a low risk population with normal fetal biometry. A prospective study of 2016 women. European Journal of Obstetrics Gynecology and Reproductive Biology. 1997 January;71(1):11-9.

103. Hanretty KP, Primrose MH, Neilson JP, Whittle MJ. Pregnancy screening by Doppler uteroplacental and umbilical artery waveforms. British journal of obstetrics and gynaecology. 1989 Oct;96(10):1163-7.

104. Schulman H, Winter D, Farmakides G, Ducey J, Guzman E, Coury A, et al. Pregnancy surveillance with Doppler velocimetry of uterine and umbilical arteries. American Journal of Obstetrics & Gynecology. 1989;160(1):192-6.

105. Sijmons EA, Reuwer PJ, van Beek E, Bruinse HW. The validity of screening for small-for-gestational-age and low-weight-for-length infants by Doppler ultrasound. British Journal of Obstetrics & Gynaecology. 1989;96(5):557-61.

106. Valino N, Giunta G, Gallo DM, Akolekar R, Nicolaides KH. Biophysical and biochemical markers at 30-34 weeks' gestation in the prediction of adverse perinatal outcome. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2016 01 Feb;47(2):194-202.

107. Valino N, Giunta G, Gallo DM, Akolekar R, Nicolaides KH. Biophysical and biochemical markers at 35-37 weeks' gestation in the prediction of adverse perinatal outcome. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2016 01 Feb;47(2):203-9.

108. Weiner Z, Reichler A, Zlozover M, Mendelson A, Thaler I. The value of Doppler ultrasonography in prolonged pregnancies. European Journal of Obstetrics Gynecology and Reproductive Biology. 1993;48(2):93-7.

109. Flenady V, Wojcieszek AM, Middleton P, Ellwood D, Erwich JJ, Coory M, et al. Stillbirths: recall to action in high-income countries. Lancet (London, England). 2016 Feb 13;387(10019):691-702.

110. Giussani DA. The fetal brain sparing response to hypoxia: physiological mechanisms. J Physiol. 2016 Mar 1;594(5):1215-30.

111. Prior T, Mullins E, Bennett P, Kumar S. Prediction of intrapartum fetal compromise using the cerebroumbilical ratio: a prospective observational study. American Journal of Obstetrics & Gynecology. 2013;208(2):124.e1-6.

112. Khalil AA, Morales-Rosello J, Morlando M, Hannan H, Bhide A, Papageorghiou A, et al. Is fetal cerebroplacental ratio an independent predictor of intrapartum fetal compromise and neonatal unit admission? American Journal of Obstetrics & Gynecology. 2015;213(1):54.e1-10.

113. Smith GC, Moraitis AA, Wastlund D, Thornton JG, Papageorghiou A, Sanders J, et al. Universal late pregnancy ultrasound screening to predict adverse outcomes in nulliparous women: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2021 Feb;25(15):1-190.

114. Akolekar R, Syngelaki A, Gallo DM, Poon LC, Nicolaides KH. Umbilical and fetal middle cerebral artery Doppler at 35-37 weeks' gestation in the prediction of adverse perinatal outcome. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2015 01 Jul;46(1):82-92.

115. Bakalis S, Akolekar R, Gallo DM, Poon LC, Nicolaides KH. Umbilical and fetal middle cerebral artery Doppler at 30-34 weeks' gestation in the prediction of adverse perinatal outcome. Ultrasound in Obstetrics & Gynecology. 2015;45(4):409-20.

116. Bligh LN, Alsolai AA, Greer RM, Kumar S. Cerebroplacental ratio thresholds measured within 2 weeks before birth and risk of Cesarean section for intrapartum fetal compromise and adverse neonatal outcome. Ultrasound in Obstetrics & Gynecology. 2018;52(3):340-6.

117. Bligh LN, Al Solai A, Greer RM, Kumar S. Diagnostic Performance of Cerebroplacental Ratio Thresholds at Term for Prediction of Low Birthweight and Adverse Intrapartum and Neonatal Outcomes in a Term, Low-Risk Population. Fetal Diagnosis & Therapy. 2018;43(3):191-8.

118. Flatley C, Kumar S. Is the fetal cerebroplacental ratio better that the estimated fetal weight in predicting adverse perinatal outcomes in a low risk cohort? Journal of Maternal-Fetal and Neonatal Medicine. 2019 18 Jul;32(14):2380-6.

119. Maged AM, Abdelhafez A, Al Mostafa W, Elsherbiny W. Fetal middle cerebral and umbilical artery Doppler after 40 weeks gestational age. Journal of Maternal-Fetal & Neonatal Medicine. 2014;27(18):1880-5.

120. Monaghan C, Binder J, Thilaganathan B, Morales-Rosello J, Khalil A. Perinatal Loss at Term: The Role of Uteroplacental and Fetal Doppler Assessment. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2017 Apr 24.

121. Morales-Rosello J, Khalil A, Morlando M, Papageorghiou A, Bhide A, Thilaganathan B. Changes in fetal Doppler indices as a marker of failure to reach growth potential at term. Ultrasound in Obstetrics & Gynecology. 2014;43(3):303-10.

122. Prior T, Paramasivam G, Bennett P, Kumar S. Are fetuses that fail to achieve their growth potential at increased risk of intrapartum compromise? Ultrasound in Obstetrics & Gynecology. 2015;46(4):460-4.

123. Rial-Crestelo M, Martinez-Portilla RJ, Cancemi A, Caradeux J, Fernandez L, Peguero A, et al. Added value of cerebro-placental ratio and uterine artery Doppler at routine third trimester screening as a predictor of SGA and FGR in non-selected pregnancies. Journal of Maternal-Fetal and Neonatal Medicine. 2019 03 Aug;32(15):2554-60.

124. Sabdia S, Greer RM, Prior T, Kumar S. Predicting intrapartum fetal compromise using the fetal cerebro-umbilical ratio. Placenta. 2015;36(5):594-8.

125. Stumpfe FM, Kehl S, Pretscher J, Baier F, Bayer CM, Schwenke E, et al. Correlation of short-term variation and Doppler parameters with adverse perinatal outcome in low-risk fetuses at term. Archives of gynecology and obstetrics. 2019 05 Feb;299(2):411-20.

126. Twomey S, Flatley C, Kumar S. The association between a low cerebro-umbilical ratio at 30-34 weeks gestation, increased intrapartum operative intervention and adverse perinatal outcomes. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2016;203:89-93.

127. Figueras F, Gratacos E. An integrated approach to fetal growth restriction. Best practice & research Clinical obstetrics & gynaecology. 2017 Jan;38:48-58.

128. Dunn L, Sherrell H, Kumar S. Review: Systematic review of the utility of the fetal cerebroplacental ratio measured at term for the prediction of adverse perinatal outcome. Placenta. 2017 Jun;54:68-75.

129. Conde-Agudelo A, Villar J, Kennedy SH, Papageorghiou AT. Predictive accuracy of cerebroplacental ratio for adverse perinatal and neurodevelopmental outcomes in suspected fetal growth restriction: systematic review and meta-analysis. Ultrasound in Obstetrics & Gynecology. 2018;52(4):430-41.

130. Gaccioli F, Sovio U, Cook E, Hund M, Charnock-Jones DS, Smith GCS. Screening for fetal growth restriction using ultrasound and the sFLT1/PlGF ratio in nulliparous women: a prospective cohort study. Lancet Child Adolesc Health. 2018 Aug;2(8):569-81.

131. Sovio U, Goulding N, McBride N, Cook E, Gaccioli F, Charnock-Jones DS, et al. A maternal serum metabolite ratio predicts fetal growth restriction at term. Nature medicine. 2020 Mar;26(3):348-53.

132. Phelan JP, Ahn MO, Smith CV, Rutherford SE, Anderson E. Amniotic fluid index measurements during pregnancy. The Journal of reproductive medicine. 1987 Aug;32(8):601-4.

133. Morris RK, Meller CH, Tamblyn J, Malin GM, Riley RD, Kilby MD, et al. Association and prediction of amniotic fluid measurements for adverse pregnancy outcome: systematic review and meta-analysis. BJOG : an international journal of obstetrics and gynaecology. 2014 May;121(6):686-99.

134. Ashwal E, Hiersch L, Melamed N, Aviram A, Wiznitzer A, Yogev Y. The association between isolated oligohydramnios at term and pregnancy outcome. Archives of gynecology and obstetrics. 2014 Nov;290(5):875-81.

135. Ghosh G, Marsal K, Gudmundsson S. Amniotic fluid index in low-risk pregnancy as an admission test to the labor ward. Acta obstetricia et gynecologica Scandinavica. 2002 Sep;81(9):852-5.

136. Hassan AA. The role of amniotic fluid index in the management of postdate pregnancy. Journal of the College of Physicians and Surgeons--Pakistan : JCPSP. 2005 Feb;15(2):85-8.

137. Hsieh TT, Hung TH, Chen KC, Hsieh CC, Lo LM, Chiu TH. Perinatal outcome of oligohydramnios without associated premature rupture of membranes and fetal anomalies. Gynecologic and obstetric investigation. 1998;45(4):232-6.

138. Locatelli A, Vergani P, Toso L, Verderio M, Pezzullo JC, Ghidini A. Perinatal outcome associated with oligohydramnios in uncomplicated term pregnancies. Archives of gynecology and obstetrics. 2004 Jan;269(2):130-3.

139. Megha B, Indu C. Correlation of amniotic fluid index with perinatal outcome. Journal of Obstetrics and Gynecology of India. 2014 February;64(1):32-5.

140. Melamed N, Pardo J, Milstein R, Chen R, Hod M, Yogev Y. Perinatal outcome in pregnancies complicated by isolated oligohydramnios diagnosed before 37 weeks of gestation. American journal of obstetrics and gynecology. 2011 Sep;205(3):241.e1-6.

141. Morris JM, Thompson K, Smithey J, Gaffney G, Cooke I, Chamberlain P, et al. The usefulness of ultrasound assessment of amniotic fluid in predicting adverse outcome in prolonged pregnancy: a prospective blinded observational study. BJOG : an international journal of obstetrics and gynaecology. 2003 Nov;110(11):989-94.

142. Myles TD, Santolaya-Forgas J. Normal ultrasonic evaluation of amniotic fluid in low-risk patients at term. The Journal of reproductive medicine. 2002 Aug;47(8):621-4.

143. Naveiro-Fuentes M, Prieto AP, Ruiz RS, Badillo MPC, Ventoso FM, Vallejo JLG. Perinatal outcomes with isolated oligohydramnios at term pregnancy. Journal of Perinatal Medicine. 2015(pagination).

144. Quinones JN, Odibo AO, Stringer M, Rochon ML, Macones GA. Determining a threshold for amniotic fluid as a predictor of perinatal outcome at term. Journal of Maternal-Fetal & Neonatal Medicine. 2012;25(8):1319-23.

145. Rainford M, Adair R, Scialli AR, Ghidini A, Spong CY. Amniotic fluid index in the uncomplicated term pregnancy. Prediction of outcome. The Journal of reproductive medicine. 2001 Jun;46(6):589-92.

146. Shanks A, Tuuli M, Schaecher C, Odibo AO, Rampersad R. Assessing the optimal definition of oligohydramnios associated with adverse neonatal outcomes. Journal of Ultrasound in Medicine. 2011;30(3):303-7.

147. Zhang J, Troendle J, Meikle S, Klebanoff MA, Rayburn WF. Isolated oligohydramnios is not associated with adverse perinatal outcomes. BJOG : an international journal of obstetrics and gynaecology. 2004 Mar;111(3):220-5.

148. Asgharnia M, Faraji R, Salamat F, Ashrafkhani B, Dalil Heirati SF, Naimian S. Perinatal outcomes of pregnancies with borderline versus normal amniotic fluid index. Iranian Journal of Reproductive Medicine. 2013;11(9):705-10.

149. Banks EH, Miller DA. Perinatal risks associated with borderline amniotic fluid index. American Journal of Obstetrics & Gynecology. 1999;180(6 Pt 1):1461-3.

150. Choi SR. Borderline amniotic fluid index and perinatal outcomes in the uncomplicated term pregnancy. Journal of Maternal-Fetal & Neonatal Medicine. 2016;29(3):457-60.

151. Gumus, II, Koktener A, Turhan NO. Perinatal outcomes of pregnancies with borderline amniotic fluid index. Archives of gynecology and obstetrics. 2007 Jul;276(1):17-9.

152. Jamal A, Kazemi M, Marsoosi V, Eslamian L. Adverse perinatal outcomes in borderline amniotic fluid index. International Journal of Reproductive Biomedicine. 2016;14(11):705-8.

153. Kwon JY, Kwon HS, Kim YH, Park YW. Abnormal Doppler velocimetry is related to adverse perinatal outcome for borderline amniotic fluid index during third trimester. Journal of Obstetrics & Gynaecology Research. 2006;32(6):545-9.

154. Petrozella LN, Dashe JS, McIntire DD, Leveno KJ. Clinical significance of borderline amniotic fluid index and oligohydramnios in preterm pregnancy. Obstetrics & Gynecology. 2011;117(2 Pt 1):338-42.

155. Rutherford SE, Phelan JP, Smith CV, Jacobs N. The four-quadrant assessment of amniotic fluid volume: an adjunct to antepartum fetal heart rate testing. Obstetrics & Gynecology. 1987;70(3 Pt 1):353-6.

156. Sahin E, Madendag Y, Tayyar AT, Sahin ME, Col Madendag I, Acmaz G, et al. Perinatal outcomes in uncomplicated late preterm pregnancies with borderline oligohydramnios. Journal of Maternal-Fetal & Neonatal Medicine. 2018;31(23):3085-8.

157. Wood SL, Newton JM, Wang L, Lesser K. Borderline amniotic fluid index and its relation to fetal intolerance of labor: a 2-center retrospective cohort study. Journal of Ultrasound in Medicine. 2014;33(4):705-11.

158. Hofmeyr GJ, Kulier R, West HM. External cephalic version for breech presentation at term. The Cochrane database of systematic reviews. 2015 Apr 1;2015(4):Cd000083.

159. Beuckens A, Rijnders M, Verburgt-Doeleman GH, Rijninks-van Driel GC, Thorpe J, Hutton EK. An observational study of the success and complications of 2546 external cephalic versions in low-risk pregnant women performed by trained midwives. BJOG : an international journal of obstetrics and gynaecology. 2016 Feb;123(3):415-23.

160. Wastlund D, Moraitis AA, Dacey A, Sovio U, Wilson ECF, Smith GCS. Screening for breech presentation using universal late-pregnancy ultrasonography: A prospective cohort study and cost effectiveness analysis. PLoS medicine. 2019 Apr;16(4):e1002778.

161. National Institute for Health and Care Excellence. Quality standard: Antenatal care. In: National Institute for Health and Care Excellence, editor.; 2012.

162. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.

163. Martin AD, Quinn KM, Park JH. MCMCpack: Markov Chain Monte Carlo in R. Journal of Statistical Software. 2011;42(9):22.

164. Fan FY. FinCal: Time Value of Money, Time Series Analysis and Computational Finance. R package version 063; 2016.

165. Baio G, Heath A. BCEA: Bayesian Cost Effectiveness Analysis. 2016.

166. Department of Health. NHS Reference Costs 2015-16. In: Health Do, editor.; 2016.

167. Palencia R, Gafni A, Hannah ME, Ross S, Willan AR, Hewson S, et al. The costs of planned cesarean versus planned vaginal birth in the Term Breech Trial. Cmaj. 2006 Apr 11;174(8):1109-13.

168. Ben-Meir A, Elram T, Tsafrir A, Elchalal U, Ezra Y. The incidence of spontaneous version after failed external cephalic version. American journal of obstetrics and gynecology. 2007 Feb;196(2):157.e1-3.

169. Nassar N, Roberts CL, Barratt A, Bell JC, Olive EC, Peat B. Systematic review of adverse outcomes of external cephalic version and persisting breech presentation at term. Paediatric and perinatal epidemiology. 2006 Mar;20(2):163-71.

170. Rosman AN, Vlemmix F, Ensing S, Opmeer BC, Te Hoven S, Velzel J, et al. Mode of childbirth and neonatal outcome after external cephalic version: A prospective cohort study. Midwifery. 2016 Aug;39:44-8.

171. Krueger S, Simioni J, Griffith LE, Hutton EK. Labour Outcomes After Successful External Cephalic Version Compared With Spontaneous Cephalic Version. Journal of obstetrics and gynaecology Canada : JOGC = Journal d'obstetrique et gynecologie du Canada : JOGC. 2017 Aug 02.

172. Hants Y, Kabiri D, Elchalal U, Arbel-Alon S, Drukker L, Ezra Y. Induction of labor at term following external cephalic version in nulliparous women is associated with an increased risk of cesarean delivery. Archives of gynecology and obstetrics. 2015 Aug;292(2):313-9.

173. Lavender T, Hofmeyr GJ, Neilson JP, Kingdon C, Gyte GM. Caesarean section for non-medical reasons at term. The Cochrane database of systematic reviews. 2012 Mar 14(3):Cd004660.

174. National Institute of Health. NIH State-of-the-Science Conference Statement on cesarean delivery on maternal request. NIH consensus and state-of-the-science statements. 2006 Mar 27-29;23(1):1-29.

175. American College of Obstetricians and Gynecologists. ACOG committee opinion no. 559: Cesarean delivery on maternal request. Obstetrics and gynecology. 2013 Apr;121(4):904-7.

176. Whyte H, Hannah ME, Saigal S, Hannah WJ, Hewson S, Amankwah K, et al. Outcomes of children at 2 years after planned cesarean birth versus planned vaginal birth for breech presentation at term: the International Randomized Term Breech Trial. American journal of obstetrics and gynecology. 2004 Sep;191(3):864-71.

177. Szende A, Janssen B, Cabasés J. Self-Reported Population Health: An International Perspective based on EQ-5D. Springer Dordrecht Heidelberg New York London; 2014.

178. Office for National Statistics. National Life Tables, United Kingdom, 1980-82 to 2014-16. Office for National Statistics; 2017.

179. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal 2013. In: Excellence NIfHaC, editor.; 2013.

180. Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.

181. NHS Digital. NHS Maternity Statistics 2016-17: Summary report. NHS Digital - Health and Social Care Information Centre; 2017.

182. Gaccioli F, Lager S, Sovio U, Charnock-Jones S, Smith GC. The pregnancy outcome prediction (POP) study: Investigating the relationship between serial prenatal ultrasonography, biomarkers, placental phenotype and adverse pregnancy outcomes. Placenta. 2017;59(Suppl 1):S17-S25.

183. Keag OE, Norman JE, Stock SJ. Long-term risks and benefits associated with cesarean delivery for mother, baby, and subsequent pregnancies: Systematic review and meta-analysis. PLoS medicine. 2018 Jan;15(1):e1002494.

184. Moraitis AA, Oliver-Williams C, Wood AM, Fleming M, Pell JP, Smith G. Previous caesarean delivery and the risk of unexplained stillbirth: retrospective cohort study and meta-analysis. BJOG : an international journal of obstetrics and gynaecology. 2015 Oct;122(11):1467-74.

185. Hofmeyr GJ, Hannah M, Lawrie TA. Planned caesarean section for term breech delivery. The Cochrane database of systematic reviews. 2015 Jul 21(7):Cd000166.

186. Lau TK, Lo KW, Wan D, Rogers MS. Predictors of successful external cephalic version at term: a prospective study. BJOG : an international journal of obstetrics and gynaecology. 1997 Jul;104(7):798-802.

187. Cho LY, Lau WL, Lo TK, Tang HHT, Leung WC. Predictors of successful outcomes after external cephalic version in singleton term breech pregnancies: A nine-year historical cohort study. Hong Kong Medical Journal. 2012 February;18(1):11-9.

188. Salim I, Staines-Urias E, Mathewlynn S, Drukker L, Vatish M, Impey L. The impact of a routine late third trimester growth scan on the incidence, diagnosis, and management of breech presentation in Oxfordshire, UK: A cohort study. PLoS medicine. 2021 Jan;18(1):e1003503.

189. Monier I, Blondel B, Ego A, Kaminiski M, Goffinet F, Zeitlin J. Poor effectiveness of antenatal detection of fetal growth restriction and consequences for obstetric management and neonatal outcomes: a French national study. BJOG : an international journal of obstetrics and gynaecology. 2015 Mar;122(4):518-27.

190. Wilson ECF, Wastlund D, Moraitis AA, Smith GCS. Late Pregnancy Ultrasound to Screen for and Manage Potential Birth Complications in Nulliparous Women: A Cost-Effectiveness and Value of Information Analysis. Value Health. 2021 Apr;24(4):513-21.

191. Wastlund D, Moraitis AA, Thornton JG, Sanders J, White IR, Brocklehurst P, et al. The cost-effectiveness of universal late-pregnancy screening for macrosomia in nulliparous women: a decision-analysis. BJOG : an international journal of obstetrics and gynaecology. 2019 May 8.

192. Gaccioli F, Aye I, Sovio U, Charnock-Jones DS, Smith GCS. Screening for fetal growth restriction using fetal biometry combined with maternal biomarkers. American journal of obstetrics and gynecology. 2018 Feb;218(2s):S725-s37.

193. James M, Kevin H, B. R, J. R. A decision analytical cost analysis of offering ECV in a UK district general hospital. BMC Health Services Research. 2001;1(6).

194. Curtis L, Burns A. Unit Costs of Health and Social Care 2017. Personal Social Services Research Unit; 2017.

Appendix 1. Supporting data for the systematic review of the diagnostic effectiveness of universal ultrasonic screening using macrosomia in the prediction of adverse perinatal outcome.

Literature search strategy for Medline and Embase (from inception to May 2020)

1. exp fetus echography/
2. ultrasonography, prenatal.mp.
3. exp ultrasound/
4. ultraso*.mp.
5. sonograph*.mp.

6. exp biometry/
7. USS.mp.
8. estimated fetal weight.mp.
9. EFW.mp.
10. abdominal circumference.mp.
11. AC.mp.

12. exp macrosomia/
13. macrosomi*.mp.
14. exp fetus weight/
15. fetal weight.mp.
16. exp birth weight/
17. birthweight.mp.
18. large for gestational age.mp.
19. LGA.mp.
20. large fetus.mp.
21. exp brachial plexus injury/ or brachial plexus injury.mp.
22. exp shoulder dystocia/ or shoulder dystocia.mp.

23. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
24. 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22
25. 23 and 24

26. exp pregnancy/
27. 25 and 26

Appendix 2. Supporting data for the systematic review of the diagnostic effectiveness of universal ultrasonic screening using late pregnancy umbilical artery Doppler flow velocimetry in the prediction of adverse perinatal outcome.

Literature search strategy for Medline and Embase (from inception to the 19/03/2019)

1. exp pregnant woman/
2. exp pregnancy/
3. pregnan*.mp.

4. exp prenatal diagnosis/
5. exp fetus echography/
6. exp Doppler ultrasonography/

7. arterial doppler.mp.
8. doppler velocimetry.mp.
9. doppler ultraso*.mp.
10. umbilical arter*.mp.

11. 1 or 2 or 3
12. 4 or 5 or 6
13. 7 or 8 or 9 or 10
14. 11 and 12
15. 13 and 14

Table 15. Maternal characteristics and birth outcomes of POP study for the analysis of UA Doppler.

Characteristic	Umbilical artery PI >90th centile (N=346)	Umbilical artery PI <90th centile (N=3269)	P Value	Overall baseline characteristics (N=3615)
Maternal characteristics				
Age, years	29.7 (26.2-32.7)	30.3 (26.8-33.3)	*0.05*	30.2 (26.7-33.3)
Deprivation quartile				
1 (lowest)	97 (28.0)	784 (24.0)	*0.14*	881 (24.4)
2	73 (21.1)	776 (23.7)		849 (23.5)
3	92 (26.6)	773 (23.7)		865 (23.9)
4 (highest)	71 (20.5)	799 (24.4)		870 (24.1)
Missing	13 (3.7)	137 (4.2)		150 (4.2)
White ethnicity	324 (93.6)	3036 (92.9)	*0.53*	3360 (93.0)
Missing	6 (1.7)	56 (1.7)		62 (1.7)
Married	229 (66.2)	2238 (68.5)	*0.39*	2467 (68.2)
Smoker	24 (6.9)	152 (4.7)	*0.06*	176 (4.9)
Any alcohol consumption	13 (3.8)	155 (4.7)	*0.40*	168 (4.7)
Missing	0 (0)	1(0)		1 (0)
BMI, kg/m^2	24.3 (21.7-28.1)	24.0 (21.8-27.2)	*0.44*	24.0 (21.8-27.3)
≥1 previous miscarriage	34 (9.8)	331 (10.1)	*0.86*	365 (10.1)
Chronic hypertension	25 (7.3)	161 (4.9)	*0.06*	

Pre-eclampsia	29 (8.4)	204 (6.2)	*0.12*	233 (6.5)
Missing	0(0)	2(0.1)		2 (0.1)
Diabetes				
Type 1 or type 2 DM	2 (0.6)	10 (0.3)	*0.14*	12 (0.3)
Gestational DM	20 (5.8)	124 (3.8)		144 (4.0)
Birth outcomes				
Birth weight, g	3263 (2970-3560)	3470 (3170-3770)	*<0.001*	3445 (3150-3750)
Gestational age, weeks	40.4 (39.3 – 41.1)	40.4 (39.4- 41.3)	*0.74*	40.4 (39.4- 41.3)
<37	3 (0.9)	34 (1.0)	*0.19**	37 (1.0)
37	22 (6.4)	133 (4.1)		155 (4.3)
38	35 (10.1)	360 (11.0)		395 (10.9)
39	71 (20.5)	641 (19.6)		712 (19.7)
40	92 (26.6)	1001 (30.6)		1093 (30.2)
41	102 (29.5)	909 (27.8)		1011 (30.0)
≥ 42	21 (6.1)	191 (5.8)		212 (5.9)
Induction of labor	125 (36.1)	1081 (33.1)	*0.25*	1206 (33.4)
Mode of delivery				
Spontaneous vaginal	178 (51.5)	1662 (50.8)	*0.20*	1840 (50.9)
Assisted vaginal	86 (24.9)	821 (25.1)		907 (25.1)
Intrapartum cesarean	54 (15.6)	601 (18.4)		655 (18.1)
Pre-labor cesarean	27 (7.8)	176 (5.4)		203 (5.6)
Missing	1 (0.3)	9 (0.3)		10 (0.3)

Appendix 3. Supporting data for the systematic review of the diagnostic effectiveness of universal ultrasonic screening using late pregnancy cerebro-placental ratio in the prediction of adverse perinatal outcome.

Literature search strategy for Medline and Embase (from inception to the 30/05/2019)

1. exp pregnant woman/
2. exp pregnancy/
3. pregnan*.mp.

4. exp fetus echography/
5. exp prenatal diagnosis/
6. exp Doppler ultrasonography/
7. exp fetus monitoring/
8. ultraso*.mp.

9. exp middle cerebral artery/
10. middle cerebral artery.mp.
11. uteroplacental.mp.
12. utero-placental.mp.
13. cerebroplacental.mp.
14. cerebro-placental.mp.
15. cerebroumbilical.mp.
16. cerebro-umbilical.mp.
17. fetal brain doppler.mp.
18. fetal cerebral doppler.mp.

19. 1 or 2 or 3
20. 4 or 5 or 6 or 7 or 8
21. 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18
22. 19 and 20
23. 21 and 22

Appendix 4. Supporting data for the systematic review of the diagnostic effectiveness of universal ultrasonic screening using severe oligohydramnios in the prediction of adverse perinatal outcome.

Literature search strategy for Medline and Embase (from 01/01/2011 to 10/11/2020)

1. exp Pregnant Women/
2. limit 1 to yr="2011 -Current"
3. exp Pregnancy Trimester/
4. limit 3 to yr="2011 -Current"
5. pregnan*.mp.
6. limit 5 to yr="2011 -Current"
7. exp Prenatal Diagnosis/
8. limit 7 to yr="2011 -Current"
9. exp Ultrasonography, Prenatal/
10. limit 9 to yr="2011 -Current"
11. exp Amniotic Fluid/
12. limit 11 to yr="2011 -Current"
13. exp Oligohydramnios/
14. limit 13 to yr="2011 -Current"
15. oligohydramnio*.mp.
16. limit 15 to yr="2011 -Current"
17. exp Polyhydramnios/
18. limit 17 to yr="2011 -Current"
19. polyhydramnio*.mp.
20. limit 19 to yr="2011 -Current"
21. amniotic fluid index.mp.
22. limit 21 to yr="2011 -Current"
23. AFI.mp.
24. limit 23 to yr="2011 -Current"
25. maximum pool depth.mp.
26. limit 25 to yr="2011 -Current"
27. MPD.mp.
28. limit 27 to yr="2011 -Current"
29. single deepest pocket.mp.
30. limit 29 to yr="2011 -Current"
31. SDP.mp.
32. limit 31 to yr="2011 -Current"
33. largest vertical pocket.mp.
34. limit 33 to yr="2011 -Current"
35. LVP.mp.
36. limit 35 to yr="2011 -Current"
37. maximum vertical pocket.mp.
38. limit 37 to yr="2011 -Current"
39. MVP.mp.
40. limit 39 to yr="2011 -Current"
41. amniotic fluid volume.mp.

42. limit 41 to yr="2011 -Current"
43. anhydramnios.mp.
44. limit 43 to yr="2011 -Current"
45. liquor volume.mp.
46. limit 45 to yr="2011 -Current"
47. quadrants.mp.
48. limit 47 to yr="2011 -Current"
49. biophysical profile.mp.
50. limit 49 to yr="2011 -Current"
51. BPP.mp.
52. limit 51 to yr="2011 -Current"
53. 2 or 4 or 6
54. 8 or 10 or 12 or 14 or 16 or 18 or 20
55. 22 or 24 or 26 or 28 or 30 or 32 or 34 or 36 or 38 or 40 or 42 or 44 or 46 or 48 or 50 or 52
56. 53 and 54 and 55
57. 8 or 10
58. 12 or 14 or 16 or 18 or 20 or 22 or 24 or 26 or 28 or 30 or 32 or 34 or 36 or 38 or 40 or 42 or 44 or 46 or 48 or 50 or 52
59. 53 and 57 and 58

Appendix 5. Supporting data for the systematic review of the diagnostic effectiveness of universal ultrasonic screening using borderline oligohydramnios in the prediction of adverse perinatal outcome.

Literature search strategy for Medline and Embase (from inception to 10/11/2020)

1. exp Pregnant Women/
2. exp pregnancy/
3. pregnan$.mp.

4. exp oligohydramnios/
5. oligohydramnio$.mp.
6. exp Amniotic Fluid/
7. amniotic fluid index.mp.
8. AFI.mp.
9. liquor volume.mp.

10. low.mp.
11. borderline.mp.
12. decreased.mp.

13. perinatal.mp.
14. peripartum.mp.
15. fetal.mp.

16. 1 or 2 or 3
17. 4 or 5 or 6 or 7 or 8 or 9
18. 13 or 14 or 15

19. 16 and 17 and 18
20. 10 or 11 or 12
21. 19 and 20

Table 16. Patient characteristics and birth outcomes of POP study for the analysis of borderline oligohydramnios.

Characteristic	Borderline AFI 5-8cm (N= 108)	Normal AFI 8-24cm (N= 3279)	P Value	Overall baseline characteristics (N= 3387)
Maternal characteristics				
Age, years	30.1 (26.7-33.2)	30.3 (26.2-33.7)	*0.60*	30.1 (26.7-33.2)
Deprivation quartile				
1 (lowest)	29 (26.9)	808 (24.6)	*0.53*	837 (24.7)
2	28 (25.9)	769 (23.5)		797 (23.5)
3	23 (21.3)	776 (23.7)		799 (23.6)
4 (highest)	25 (23.2)	783 (23.9)		808 (23.9)
Missing	3 (2.8)	143 (4.4)		146 (4.3)
White ethnicity	96 (88.9)	3052 (93.1)	*0.16*	3148 (92.9)
Missing	3 (2.8)	54 (1.7)		57 (1.7)
Married	81 (75.0)	2222 (67.8)	*0.11*	2303 (68.0)
Smoker	3 (2.8)	164 (5.0)	*0.29*	167 (4.9)
Any alcohol consumption	1 (0.9)	154 (4.7)	*0.06*	155 (4.6)
Missing	0 (0.0)	1(0.0)		1 (0.0)
BMI, kg/m^2	23.4 (21.6-26.5)	23.9 (21.8-27.1)	*0.19*	23.9 (21.8-27.0)
≥1 previous miscarriage	8 (7.4)	327 (10.0)	*0.38*	335 (9.9)
Chronic hypertension	4 (3.7)	164 (5.0)	*0.54*	
Pre-eclampsia	9 (8.3)	201 (6.1)	*0.35*	210 (6.2)

	Missing	0(0)	2(0.1)		2 (0.1)
Birth outcomes					
Birth weight, g		3260 (3005-3520)	3460 (3150-3770)	*<0.001*	3450 (3150-3760)
Gestational age, weeks		40.0 (38.8 – 40.9)	40.4 (39.6- 41.3)	*<0.001*	40.4 (39.6- 41.3)
Induction of labor		41 (38.0)	1016 (31.0)	*0.12*	1057 (31.2)
Mode of delivery					
	Spontaneous vaginal	70 (64.8)	1685 (51.4)	*0.04*	1755 (51.8)
	Assisted vaginal	19 (17.6)	832 (25.4)		851 (25.1)
	Intrapartum Caesarean	13 (12.0)	596 (18.2)		609 (18.0)
	Pre-labor Caesarean	6 (5.6)	157 (4.8)		163 (4.8)
	Missing	0 (0.0)	9 (0.3)		9 (0.3)

Appendix 6. Additional data for screening for breech presentation

Table 17. Inputs for costs and probabilities for the economic model

Costs	**Costs**				**Source**	
Ultrasound scanning	20.7				Expert opinion *	
ECV	297.4				James et al. (2001)[193] †	
CV delivery	2297.3				NHS Reference costs 2015-16[166] ‡	
Elective caesarean delivery	3438.1				NHS Reference costs 2015-16[166] ‡	
Emergency caesarean delivery	4553.4				NHS Reference costs 2015-16[166] ‡	
VB delivery	3999.7				Expert opinion *	
Probabilities	**Alpha**	**Beta**	**Mean**	**Node**	**Source**	
Breech prevalence at ~36wkGA	179	3700	0.046	A1 & B1	POP study	
ECV attempted	84	93	0.475	C1	POP study	
Detection without ultrasound	79	96	0.451	B3	POP study	
Successful ECV	12	72	0.143	C2	POP study	
SRC (ECV not attempted)	21	72	0.226	C3	POP study	
SRB	1	11	0.083	C4	POP study	
SRC (failed ECV)	3	127	0.023	C5	Ben-Meir et al.[168] §	
Mode of delivery	**CV**	**ELCS**	**EMCS**	**VB**	**Node**	**Source**
No breech	2813	141	735	0	A2 & B2	POP study
Cephalic (successful ECV)	8	0	3	0	C8	POP study
Cephalic (spontaneous reversion)	11	1	9	0	C6 & C10	POP study
Breech (ECV not attempted)	0	52	20	0	C7	POP study
Breech (Unsuccessful ECV)	0	54	18	0	C11	POP study
Breech (spontaneous reversion)	0	0	15	11	C9	Leung et al.[22]
Undetected breech	0	0	15	11	B4	Leung et al.[22]

Costs given per unit/episode. For probabilities, Alpha represent case of event and Beta case of no event. Mode of delivery shows input values for Dirichlet distribution.

CV = Cephalic Vaginal; ELCS = Elective caesarean section; EMCS = Emergency caesarean section; IDR = Incidental detection rate; SRB = Spontaneous reversion to breech; SRC = Spontaneous reversion to cephalic; VB = Vaginal breech

* Details on how this value was estimated is provided as supporting information, S1 Text.

† Cost for ECV (high staff cost), converted to 2017 price level using the Hospital & Community Health Services (HCHS) index.[194]

‡ Weighted average of all complication levels (Total HRG's)

§ Due to the small sample size for these parameters in the POP study, the model used inputs for mode of delivery for undetected breech instead.

Table 18. Characteristics and delivery outcomes in the POP study by presentation at 36 weeks.

Characteristics	Breech (N=179)	Cephalic (N=3,700)	P-value
Maternal			
Age (years)	31 (28 - 34)	30 (27 - 33)	0.002
Age stopped FTE (years)	21 (18 - 23)	21 (18 - 23)	0.19
Missing	5 (3%)	105 (3%)	
Racial ancestry			
White European	172 (96%)	3437 (93%)	0.38
Missing	0 (0%)	66 (2%)	
Alcohol consumption	7 (4%)	172 (5%)	0.65
Missing	0 (0%)	1 (<0.1%)	
Smoker	4 (2%)	179 (5%)	0.11
BMI, kg/m2	24 (22 - 27)	24 (22 - 27)	0.69
Missing	0 (0%)	1 (<0.1%)	
Deprivation quartile			0.08
1 (lowest)	46 (26%)	899 (24%)	
2	53 (30%)	873 (24%)	
3	39 (22%)	886 (24%)	
4 (highest)	33 (18%)	892 (24%)	
Missing	8 (4%)	150 (4%)	
Fetal or neonatal			
Female sex	96 (54%)	1841 (50%)	0.31
Missing	0 (0%)	1 (<0.1%)	
Birth weight (grams)	3310 (2995 – 3560)	3445 (3145 – 3750)	<0.001
Gestational age (weeks)	39.1 (38.7 – 39.7)	40.4 (39.4 – 41.3)	<0.001
Birth weight centile	49 (25 – 70)	44 (24 – 66)	0.22
Birth weight centile category			0.32
SGA	12 (7%)	332 (9%)	
AGA	158 (88%)	3199 (86%)	
LGA	9 (5%)	168 (5%)	
Missing	0 (0%)	1 (<0.1%)	
Mode of delivery			<0.001
Spontaneous vaginal cephalic	11 (6.1%)	1885 (50.9%)	
Instrumental vaginal cephalic	8 (4.5%)	928 (25.1%)	
Elective caesarean section	110 (61.5%)	141 (3.8%)	
Emergency caesarean section	50 (27.9%)	735 (19.9%)	
Missing	0 (0%)	11 (0.3%)	

Statistics are presented as n (%) for binary outcomes, and median (inter-quartile range) for continuous variables. The "Missing" category was not included in statistical tests. For variables without a "Missing" category, data were 100% complete. P-values are reported for the difference between groups using the two-sample Wilcox rank-sum test for continuous variables and the Pearson Chi-square test for categorical variables, with trend test as appropriate (i.e. for deprivation quartile and birth weight centile category).
FTE= full time education; BMI= body mass index; SGA, AGA and LGA denotes small, average and large for gestational age, respectively.

Appendix 7. Questionnaire for attitudes towards universal ultrasound screening in late pregnancy.

Thank you for taking the time to read the background of our research project and considering the following five questions.

Background

As part of routine NHS care all pregnant women are offered two scans. The first scan is usually done at about 12 weeks. This scan dates the pregnancy, checks for twins and contributes to screening for Down's syndrome. The second scan is usually performed at around 20 weeks. This scan looks for some physical abnormalities and can often check to see if the baby is a boy or girl. Healthy women with an uncomplicated pregnancy are NOT routinely scanned after 20 weeks but a scan may be suggested if their doctor or midwife has concerns.

We want to carry out research to find out whether offering all women expecting their first baby a third scan at around 36 weeks would result in better outcomes for babies. By this we mean fewer babies having to be admitted to special baby units because they are born unwell, fewer babies being born who are smaller than expected and the worst outcome of all which is when a baby dies before he or she is born, a stillbirth. The reason for having a scan at 36 weeks would be to check the baby is growing normally, check the placenta (the baby's life line to the mother) is still healthy and check if the baby is head down, which is the correct position for birth.

Research is needed because while having a third scan at 36 weeks as part of normal care may be useful in some cases, it may not always give accurate information and could therefore be harmful. For example, there might be a difference of up to 10% between the weight of the baby as calculated during the scan and the actual weight, which can be up to 1 pound (lb) difference (equivalent to about 450 grams) for large babies. Similarly, the scan may suggest a baby is not growing well when in fact the baby is perfectly healthy. This can lead to unnecessary and potentially harmful interventions such as delivering the baby earlier than needed, which can increase the risk of the baby being admitted to special care. We would like to plan a study that women would be happy to join. For this reason your views are important, and will help us decide on the design a future research project on whether we should be offering women scans in late pregnancy.

1. Were you aware that women whose pregnancies are straight-forward are NOT routinely scanned after 20 weeks? (circle one)

A) Yes, I was aware that healthy women are NOT routinely scanned after 20 weeks.

B) No, I thought all women have a scan after 20 weeks.

2. How much do you agree/disagree with the following statement?

"I would like to have the option of a scan at around 36 weeks as part of my routine NHS care". Circle one.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
(don't want scan)				(do want scan)

3. Imagine that today you are asked to be in a research study. This study is called "A". If you agreed to take part you would be randomly put into one of two groups. One group would have a scan at 36 weeks and the other group would not have a scan at 36 weeks (i.e the current standard of care). That is, you would agree to take part in the research and, after you had consented, you would find out whether or not you were one of the women selected to have a routine scan at 36 weeks.

How much do you agree/disagree with the following statement? "I would be likely to agree to take part in such a research project".

Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree

(wouldn't want to take part) (would take part)

4. Now imagine that you are asked to be in study (B) where you would definitely have a scan at 36 weeks. All women would be told whether their baby was head first or bottom first and if there was a major obvious problem (e.g. very small amount of fluid around the baby). However, in this new study you would also be randomly put into one of two groups. In this study other information from the scan (such as the estimated size of the baby – the part that may suggest you should be delivered early) would only be told to women and the midwives and doctors looking after women in one of the groups. If you were in this group, the care you received might change in the light of knowing your scan results (such as being required to deliver in the consultant-led unit and not in the midwife-led unit). If you were in the other group the midwives and doctors and you would not be told this extra information and you will receive the standard care.

How much do you agree/disagree with the following statement? "I would be likely to agree to take part in such a research project".

Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree

(wouldn't want to take part) (would take part)

5. If you are happy to participate in one of the above research projects which one would you prefer?

A. The study in which you may or may not have an additional scan at 36 weeks (depending on which group you were randomly put in). For women who have a scan the results will be revealed to you and your midwife or doctor.

B. The study in which all women have an additional scan at 36 weeks. If there is any major problem (as described above) the results will be revealed to you and your midwife and doctor. If there is not a major problem the results might or might not be revealed (depending on which group you were randomly put in).

C. I will be happy to participate in either study.

About you

Age (circle one): <20 20-24 25-29 30-34 35-39 40+

Ethnicity: ……

Age stopped full time education (circle one): <18 18-21 22-24 25+

Have you been told that you are going to have extra NHS scans anyway? YES NO

Have you had a previous birth (births include stillbirths but not miscarriages)? YES NO

Appendix 8. Additional data for future randomised controlled trial.

Figure 31. Flowcharts of possible trial designs.

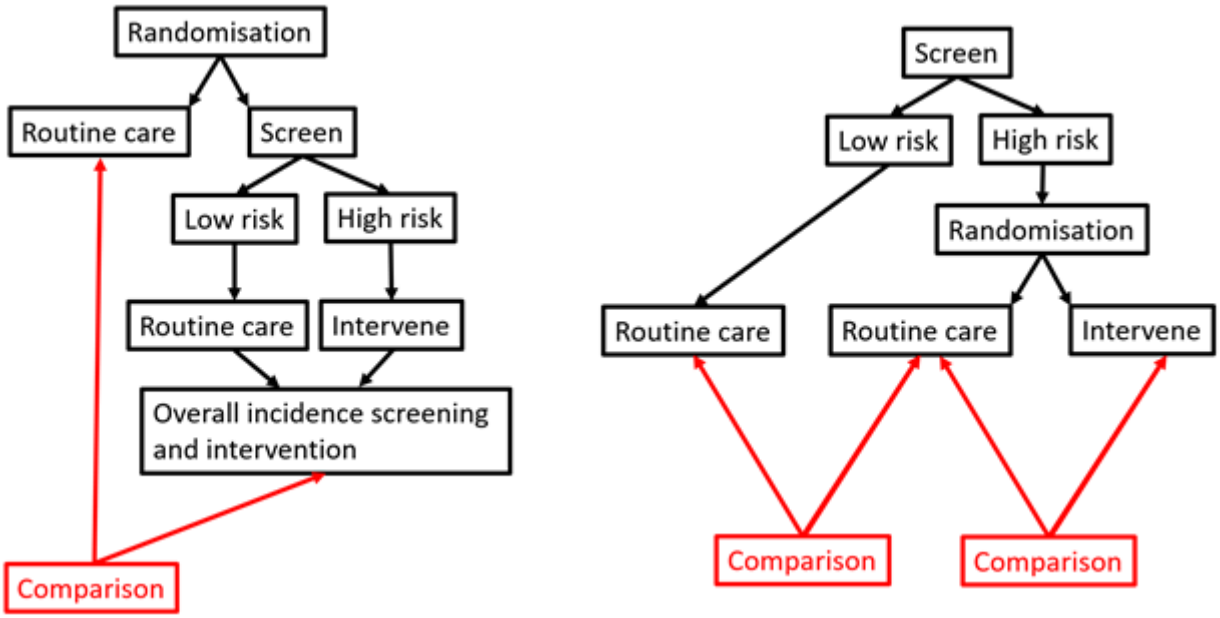


Table 19. Sample size calculations for different outcomes, screening tests and trial designs.

Outcome	Screening test	SPR	PPV	Sample size			Reference
					Screen all, randomise high risk		
				Screen vs. no screen	Number needed to screen	Number of high risk women	
Perinatal death (background = 0.2%)							
	LR+ = 2	10%	0.4%	1,488,448	234,740	23,474	
	LR+ = 3	10%	0.6%	644,156	156,260	15,626	
	LR+ = 5	10%	1.0%	219,382	93,460	9,346	
	LR+ = 2	5%	0.4%	6,110,172	469,480	23,474	
	LR+ = 3	5%	0.6%	2,680,882	312,520	15,626	
	LR+ = 5	5%	1.0%	940,096	186,920	9,346	
	LR+ = 10	5%	2.0%	219,382	92,760	4,638	
Any neonatal morbidity							
	EFW <10th	14%	10.3%	36,910	6,014	842	Sovio et al 2015
	EFW< 10th + ACGV	4.3%	15.7%	172,522	12,279	528	Sovio et al 2015
Severe neonatal morbidity							
	EFW <10th	14%	1.07%	422,336	63,743	8,924	Sovio et al 2015
	EFW< 10th + ACGV	4.3%	2.33%	965,714	93,256	4,010	Sovio et al 2015

Complicated SGA						
EFW <10th	14%	7.5%	13,920	8,457	1,184	Gaccioli et al 2018
EFW< 10th + ACGV	4.3%	11.2%	73,538	17,860	768	Gaccioli et al 2018
Delphi	11.3%	8.5%	16,952	9,168	1,036	Gaccioli et al 2018

SPR = screen positive rate, PPV = positive predictive value, EFW = estimated fetal weight, ACGV = abdominal circumference growth velocity in the lowest decile (see Sovio et al 2015). Delphi = fulfilled definition of late FGR using criteria of Gordjin et al 2016 (except MCA Doppler not included). Neonatal morbidity and severe neonatal morbidity are defined in Sovio et al 2015 and complicated SGA is defined in Gaccioli et al 2018 (in brief = delivery of a baby with a birth weight <10th percentile where either the mother had a diagnosis of preeclampsia or the baby experienced neonatal morbidity

Appendix 9. List of publications

Publications that include core parts of this thesis

A. **Moraitis AA**, Shreeve N, Sovio U, Brocklehurst P, Heazell AEP, Thornton JG, Robson SC, Papageorghiou A, Smith GCS. *Universal third-trimester ultrasonic screening using fetal macrosomia in the prediction of adverse perinatal outcome: A systematic review and meta-analysis of diagnostic test accuracy.* PLoS medicine. 2020 Oct;17(10):e1003190. (attached)
B. **Moraitis AA**, Bainton T, Sovio U, Brocklehurst P, Heazell AE, Thornton JG, Robson SC, Papageorghiou A, Smith GCS. *Fetal umbilical artery Doppler as a tool for universal third trimester screening: A systematic review and meta-analysis of diagnostic test accuracy.* Placenta. 2021 May;108:47-54. (attached)
C. Smith GCS, **Moraitis AA**, Wastlund D, Thornton JG, Papageorghiou A, Sanders J, Heazell AE P, Robson SC, Sovio U, Brocklehurst P & Wilson ECF. *Universal late pregnancy ultrasound screening to predict adverse outcomes in nulliparous women: a systematic review and cost-effectiveness analysis.* Health Technol Assess. 2021 Feb;25(15):1-190. (not attached due to size)
D. Wastlund D, **Moraitis AA**, Dacey A, Sovio U, Wilson ECF, Smith GCS. *Screening for breech presentation using universal late-pregnancy ultrasonography: A prospective cohort study and cost effectiveness analysis.* PLoS medicine. 2019 Apr;16(4):e1002778. (attached)

Additional publications that I have contributed towards and have used data generated by this thesis

E. Wastlund D, **Moraitis AA**, Thornton JG, Sanders J, White IR, Brocklehurst P, Smith GCS, Wilson ECF. *The cost-effectiveness of universal late-pregnancy screening for macrosomia in nulliparous women: a decision-analysis*. BJOG : an international journal of obstetrics and gynaecology. 2019;126(10):1243-1250. (attached)
F. Wilson ECF, Wastlund D, **Moraitis AA**, Smith GCS. Late Pregnancy Ultrasound to Screen for and Manage Potential Birth Complications in Nulliparous Women: A Cost-Effectiveness and Value of Information Analysis. Value Health. 2021 Apr;24(4):513-21. (attached)

PLOS MEDICINE

RESEARCH ARTICLE

Universal third-trimester ultrasonic screening using fetal macrosomia in the prediction of adverse perinatal outcome: A systematic review and meta-analysis of diagnostic test accuracy

Alexandros A. Moraitis[1], Norman Shreeve[1], Ulla Sovio[1], Peter Brocklehurst[2], Alexander E. P. Heazell[3,4], Jim G. Thornton[5], Stephen C. Robson[6], Aris Papageorghiou[7], Gordon C. Smith[1]*

1 Department of Obstetrics and Gynaecology, University of Cambridge, NIHR Cambridge Comprehensive Biomedical Research Centre, Cambridge, United Kingdom, **2** Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, United Kingdom, **3** Maternal and Fetal Health Research Centre, School of Medical Sciences, Faculty of Biological, Medical and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom, **4** St. Mary's Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, United Kingdom, **5** Division of Child Health, Obstetrics and Gynaecology, School of Medicine, University of Nottingham, Nottingham, United Kingdom, **6** Reproductive and Vascular Biology Group, The Medical School, University of Newcastle, Newcastle, United Kingdom, **7** Nuffield Department of Obstetrics and Gynaecology, Oxford, United Kingdom

* gcss2@cam.ac.uk

Citation: Moraitis AA, Shreeve N, Sovio U, Brocklehurst P, Heazell AEP, Thornton JG, et al. (2020) Universal third-trimester ultrasonic screening using fetal macrosomia in the prediction of adverse perinatal outcome: A systematic review and meta-analysis of diagnostic test accuracy. PLoS Med 17(10): e1003190. https://doi.org/10.1371/journal.pmed.1003190

Academic Editor: Eva Pajkrt, The University of Amsterdam, NETHERLANDS

Received: January 10, 2020

Accepted: September 9, 2020

Published: October 13, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pmed.1003190

Data Availability Statement: All relevant data are within the manuscript and its Supporting

Abstract

Background

The effectiveness of screening for macrosomia is not well established. One of the critical elements of an effective screening program is the diagnostic accuracy of a test at predicting the condition. The objective of this study is to investigate the diagnostic effectiveness of universal ultrasonic fetal biometry in predicting the delivery of a macrosomic infant, shoulder dystocia, and associated neonatal morbidity in low- and mixed-risk populations.

Methods and findings

We conducted a predefined literature search in Medline, Excerpta Medica database (EMBASE), the Cochrane library and ClinicalTrials.gov from inception to May 2020. No language restrictions were applied. We included studies where the ultrasound was performed as part of universal screening and those that included low- and mixed-risk pregnancies and excluded studies confined to high risk pregnancies. We used the estimated fetal weight (EFW) (multiple formulas and thresholds) and the abdominal circumference (AC) to define suspected large for gestational age (LGA). Adverse perinatal outcomes included macrosomia (multiple thresholds), shoulder dystocia, and other markers of neonatal morbidity. The risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. Meta-analysis was carried out using the hierarchical summary receiver

Information files. All the studies included in the meta-analysis are publicly available.

Funding: This study was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme, grant number 15/105/01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: AAM, NS, PB, JGT, and SCR have no competing interests to declare. US reports grants from NIHR Cambridge Biomedical Research Centre during the conduct of the study. GCS is a member of the Editorial Board of PLOS Medicine. GCS reports grants and personal fees from GlaxoSmithKline Research and Development, Ltd., grants from Sera Prognostics, Inc., nonfinancial support from Illumina, Inc., and personal fees from Roche Diagnostics, Ltd., outside the submitted work. In addition, GCS and US have a patent in preparation for a novel predictive test for fetal size pending. AP reports personal fees from educational events/lectures, clinical services in the private sector and from Consultancy via Oxford University Innovation, royalties from published works, and editorial work for UOG and BJOG, outside the submitted work.

Abbreviations: AC, abdominal circumference; BPD, Biparietal diameter; CDSR, Cochrane Database of Systematic Reviews; CENTRAL, Cochrane Central Register of Controlled Trials; CI, confidence interval; DOR, diagnostic odds ratio; EFW, estimated fetal weight; EMBASE, Excerpta Medica database; FL, femur length; HC, head circumference; HSROC, hierarchical summary receiver–operating characteristics; IOL, induction of labor; LGA, large for gestational age; LR, likelihood ratio; NICE, National Institute for Health and Care Excellence; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; QUADAS 2, Quality Assessment of Diagnostic Accuracy Studies; RCT, randomized controlled trial; ROC, receiver operating characteristic; SGA, small for gestational age; wkGA, weeks' gestation.

operating characteristic (ROC) and the bivariate logit-normal (Reitsma) models. We identified 41 studies that met our inclusion criteria involving 112,034 patients in total. These included 11 prospective cohort studies ($N = 9986$), one randomized controlled trial (RCT) ($N = 367$), and 29 retrospective cohort studies ($N = 101,681$). The quality of the studies was variable, and only three studies blinded the ultrasound findings to the clinicians. Both EFW >4,000 g (or 90th centile for the gestational age) and AC >36 cm (or 90th centile) had >50% sensitivity for predicting macrosomia (birthweight above 4,000 g or 90th centile) at birth with positive likelihood ratios (LRs) of 8.74 (95% confidence interval [CI] 6.84–11.17) and 7.56 (95% CI 5.85–9.77), respectively. There was significant heterogeneity at predicting macrosomia, which could reflect the different study designs, the characteristics of the included populations, and differences in the formulas used. An EFW >4,000 g (or 90th centile) had 22% sensitivity at predicting shoulder dystocia with a positive likelihood ratio of 2.12 (95% CI 1.34–3.35). There was insufficient data to analyze other markers of neonatal morbidity.

Conclusions

In this study, we found that suspected LGA is strongly predictive of the risk of delivering a large infant in low- and mixed-risk populations. However, it is only weakly (albeit statistically significantly) predictive of the risk of shoulder dystocia. There was insufficient data to analyze other markers of neonatal morbidity.

Author summary

Why was this study done?

- There is a debate regarding introducing universal third-trimester screening for macrosomia. An effective screening program requires two elements: an effective test at predicting a condition and an effective intervention.
- There is evidence that early-term induction of labor (IOL) could reduce the rates of shoulder dystocia. However, there is no high-quality evidence regarding the diagnostic effectiveness of fetal biometry at predicting macrosomia and associated morbidity.

What did the researchers do and find?

- We searched more than 10,000 titles and identified 41 studies including 112,034 patients that offered third-trimester ultrasounds for the prediction of macrosomia as part of universal ultrasound screening or were done in low- and mixed-risk populations. The quality of the studies was variable, and only three studies blinded the ultrasound findings to the clinicians.
- We found that the two most common ultrasound markers, the estimated fetal weight (EFW) and the abdominal circumference (AC), could predict the majority of macrosomic infants at birth (sensitivity >50%) with high diagnostic performance (positive LRs between 7 and 10).

- However, the EFW could only predict about 1 in 5 cases of shoulder dystocia (22% sensitivity) with low diagnostic performance (positive likelihood ratio of about 2). There was insufficient data to analyze other markers of neonatal morbidity.

What do these findings mean?

- Universal third-trimester ultrasound screening will identify more pregnancies with macrosomia. However, it will not have a clinically significant effect at predicting shoulder dystocia. There is not enough evidence on the effect of ultrasound screening on neonatal morbidity.
- We recommend caution prior to introducing universal third-trimester screening for macrosomia, as it would increase the rates of intervention, with potential iatrogenic harm, without clear evidence that it would reduce neonatal morbidity.

Introduction

Macrosomia is usually defined as birthweight >4,000 g or >90th centile for sex and gestational age. Macrosomic birth weight is associated with the risk of adverse outcomes, including perinatal death [1] and injuries related to traumatic delivery [2]. Ultrasonic estimated fetal weight (EFW) was first described in 1975 [3]. The equation for EFW that is in most widespread use was published by Hadlock and colleagues in 1985 [4], and the distribution of EFW in relation to week of gestation was published in 1991 [5]. Hence, the diagnostic tools to identify small for gestational age (SGA) and large for gestational age (LGA) fetuses have been available for many years. One of the main complications associated with macrosomia is shoulder dystocia, and a Cochrane review of four randomized controlled trials (RCTs) including 1,190 women demonstrated that routine induction of labor (IOL) for suspected LGA may prevent this outcome [6]. However, it remains unclear whether screening and intervention for suspected LGA results in better outcomes.

An RCT of IOL in women with an ultrasonically suspected LGA infant is in progress in the United Kingdom (The Big Baby trial, ISRCTN18229892). However, the women recruited to this trial will have been scanned because they were high risk for some reason, as the National Institute for Health and Care Excellence (NICE) has recommended that women should not be routinely scanned in late pregnancy [7]. Although the trial will confirm whether IOL is effective in high-risk women, it will not determine whether screening women without risk factors and intervening results in net benefit. It is often the case that screening and intervention programs that work well in high-risk groups do not work as well in low-risk populations, and one explanation for this can be that the screening test is less informative in low- and mixed-risk populations due to the lower prior risk of disease. In this study, we sought to quantify the diagnostic effectiveness of screening for fetal macrosomia and associated complications using universal ultrasonic fetal biometry in late pregnancy.

Methods

Sources

The protocol for this review was prospectively written and registered with PROSPERO (the International Prospective Register of Systematic Reviews), and the registration number was

CRD42017064093. We searched the literature systematically using the Cochrane Database of Systematic Reviews (CDSR), Cochrane Central Register of Controlled Trials (CENTRAL), Medline, EMBASE, and ClinicalTrials.gov from inception to August 2019. An update search was done on May 28, 2020. We applied no restrictions on the language of the report or the location of the study. The studies were identified using a combination of words related to "ultrasound," "pregnancy," "estimated fetal weight," "EFW," "birthweight," "macrosomia," "large for gestational age," "shoulder dystocia," and "brachial plexus injury." The exact search strategy is presented in S1 Text.

Study selection

We set out to include cohort studies where an ultrasound scan was performed ≥24 weeks' gestation (wkGA), excluding multiple pregnancies. We included studies of low-risk populations, universal screening, and mixed-risk populations (i.e., included both high-risk and low-risk pregnancies). Studies that included only high-risk women, such as patients with preexisting or gestational diabetes, and those in which the ultrasound was performed during labor were excluded. Studies were not selected on the basis of the definition of the index test, i.e., the formula and the threshold used. Finally, we included both blinded and unblinded studies.

Index tests and outcomes

For the purposes of the meta-analysis, we defined suspected LGA as a fetus with an EFW >4,000 g or >90th centile or with an abdominal circumference (AC) >36 cm or >90th centile. However, we have also documented other thresholds used. The outcomes studied included macrosomic birth weight (>4,000 g or >90th centile) and severe macrosomic birth weight (>4,500 g or >97th centile); shoulder dystocia; and perinatal morbidity (neonatal unit admission, 5-minute Apgar score of six or less, metabolic acidosis, neonatal hypoglycaemia, and neonatal jaundice).

Quality assessment

Two authors (AAM and NS) independently performed the literature search, using the software package Review Manager 5.3. Any differences were addressed in consultation with the senior author (GCS). The Quality Assessment of Diagnostic Accuracy Studies (QUADAS 2) tool was used to assess the risk of biases, following the *Cochrane Handbook of Diagnostic Test Accuracy Studies* [8]. The QUADAS 2 tool was employed to assess potential biases in patient selection, index test, reference standard, and flow and timing. In relation to flow and timing, we assessed the risk from the perspective of universal ultrasound screening near term (i.e., around 36 wkGA). Flow and timing are based on the timing of the ultrasound scan, the timing of delivery, and the length of the interval between scan and delivery. A standardized data extraction form was employed to obtain information on the characteristics of the study (publication year, location, setting, study design, blinding), the participants (inclusion and exclusion rules and number, including inclusion or exclusion of women with diabetes, either preexisting or gestational), the index test (range of wkGA when the scan was conducted, the EFW equation employed, and the threshold for screen positive), reference standard (outcome, wkGA at delivery, and the scan-to-delivery interval).

Data extraction and synthesis

Sensitivity, specificity, positive and negative likelihood ratios (LRs) [9] were calculated from standard two-by-two tables, which had been extracted for each study by tabulating each of the

different definitions of screen positive with each of the different outcomes studied. The "hierarchical summary receiver–operating characteristics" (HSROC) model of Rutter and Gatsonis [10] was utilized for data synthesis. This method allows the results of studies to be combined despite variation between studies in the threshold employed for screen positive. The bivariate logit-normal (Reitsma) model [11] was used to calculate average estimates of sensitivity and specificity and respective variances, at a specific threshold, in analyses in which data were available from at least four studies. We also used meta-analysis to obtain a summary of the diagnostic odds ratios (DORs) [12]. Publication bias was assessed using the Deeks' funnel plot asymmetry test when data was available from a sufficient number of studies. Significant asymmetry was assumed at $P < 0.05$ [13]. Statistical analyses were performed using STATA version 14 (StataCorp LP, College Station, Texas), specifically, its METANDI, METAN, and MIDAS packages. Analysis and reporting was performed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (S1 PRISMA Checklist) [14].

Results

Study characteristics

Fig 1 is the literature search PRISMA flowchart. Out of 9,811 unique titles and 72 full paper reviews, we identified 41 studies [15–55] fulfilling the inclusion criteria, including a total of 112,034 participants. The study characteristics are presented in S1 Table. Six studies [18,27,33,36,37,52] (N = 53,935) included unselected pregnancies, nine [23,29,31–33,35,43,45,53,54] (N = 6,436) were confined to low-risk pregnancies, and 26 [15–17, 9–22,24–26,28,30,34,38–42,44,46–51,55] (N = 51,663) recruited pregnancies at mixed risk. The list of the excluded studies and the reasons for the exclusion are presented in S2 Table.

Quality assessment

The risk of bias, as assessed by the QUADAS-2 tool, is summarized in Fig 2 and presented in detail in S1 Fig. The Galvin 2017 study [29] was published as an abstract; hence, we used a different study from the same cohort (GENESIS study) [56] to assess the risk of bias. Two of the included studies [51,52] have been authored by some of the coauthors of this paper. We used the same criteria for the quality assessment and analysis. Only three studies—Sovio 2018 [52] (Pregnancy Outcome Prediction study), Galvin 2017 [29] (GENESIS study), and Peregrine 2007 [47]—blinded the results to the clinicians. Hence, the large majority of studies were at risk of bias in relation to the reference standard. The second most common risk of bias was in relation to flow and timing, as six studies [19,24,36,39,47,55] performed the ultrasound either prior to IOL or less than 72 hours before delivery, resulting in a very short interval between the scan and delivery. Conversely, two studies [18,27] had a very long interval (ultrasound <33 wkGA). Two studies [17,20] did not present data on the gestational age at delivery. Finally, three studies [23,48,54] were confined to pregnancies progressing beyond 41 wkGA and were classified as having "high applicability concerns due to patient selection".

Meta-analysis results

Full details of the summary diagnostic performance are presented in Table 1. In summary, both definitions of ultrasonically suspected macrosomia (i.e., either EFW >4,000 g or >90th percentile) had >50% sensitivity for predicting LGA at birth. Many associations were similar regardless of the formula employed, but the positive LRs for the Hadlock formulae (ranging between 7.5 and 12) tended to be higher than for the Shepard formula (around 5). The

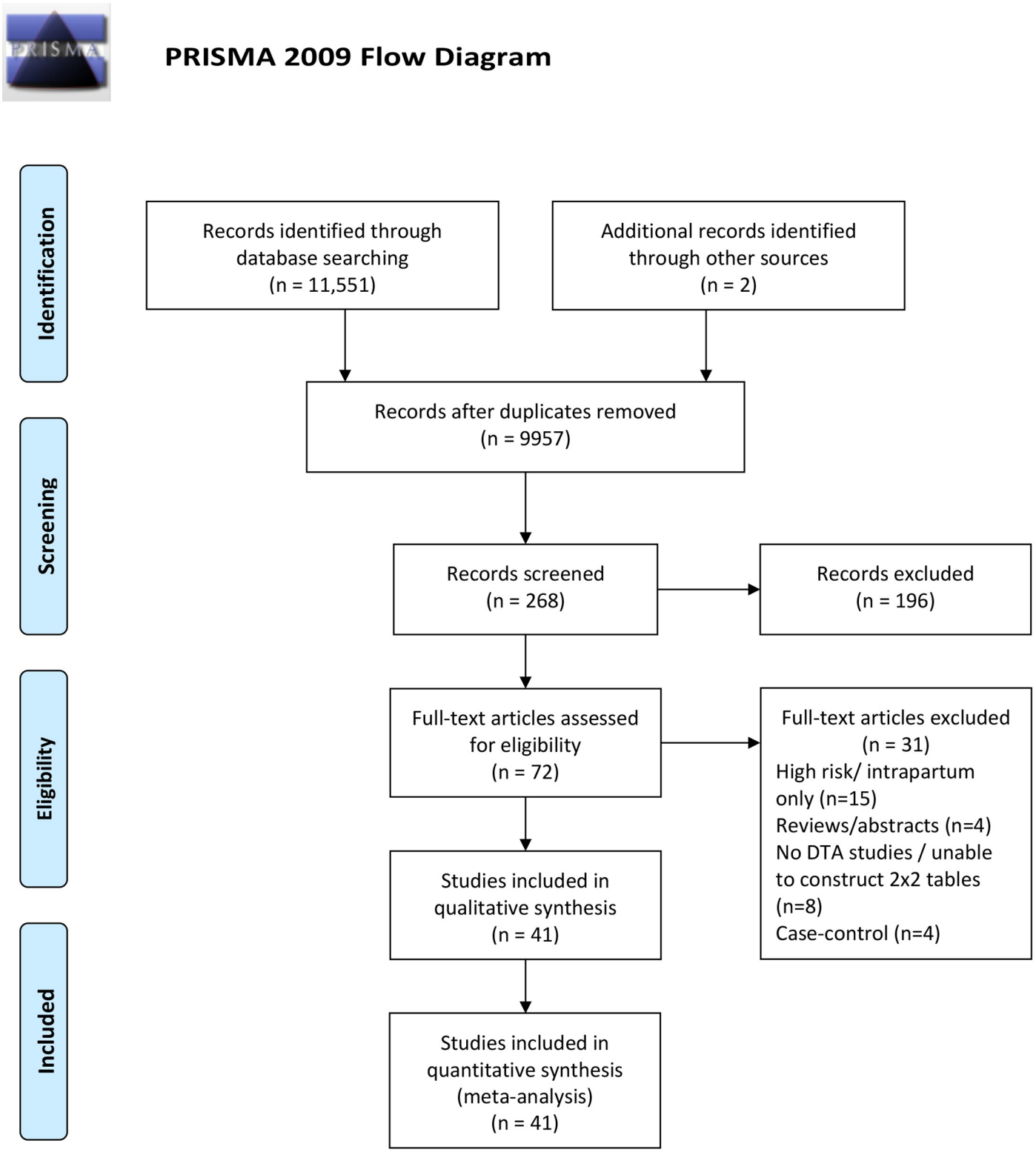


Fig 1. PRISMA flow diagram.

https://doi.org/10.1371/journal.pmed.1003190.g001

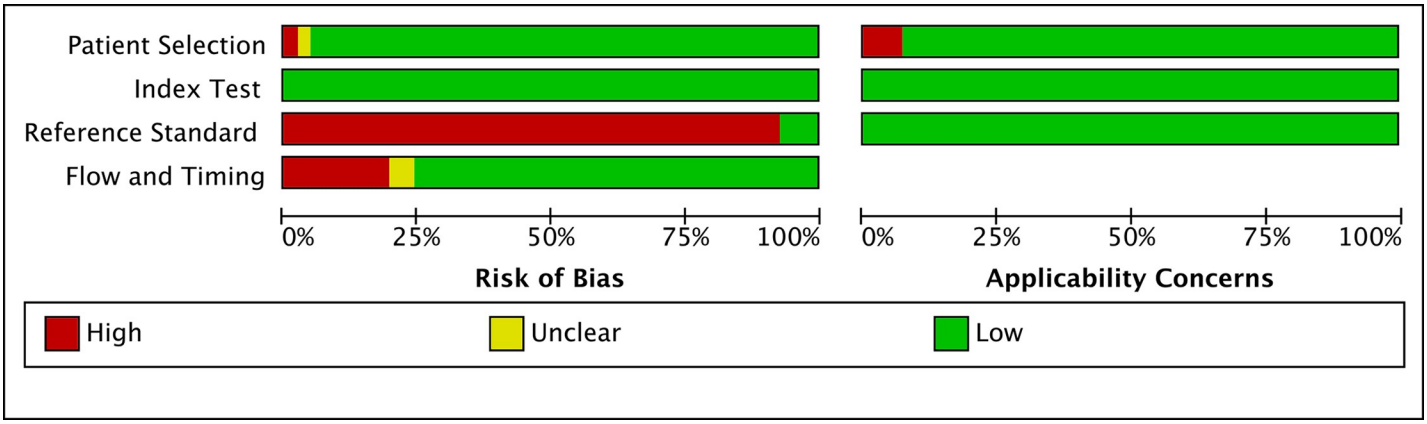


Fig 2. Summary of bias assessment using the QUADAS-2 tool of the studies included in the meta-analysis. QUADAS 2, Quality Assessment of Diagnostic Accuracy Studies.

https://doi.org/10.1371/journal.pmed.1003190.g002

Table 1. Summary diagnostic performance of suspected LGA to predict adverse perinatal outcome.

Diagnostic test	Studies	Patients	Summary sensitivity	Summary specificity	Positive LR (95% CI)	Negative LR (95% CI)
			(95% CI)	(95% CI)		
Outcome: Birthweight >4,000 g (or 90th centile)						
EFW (any) >4,000 g (or 90th centile)	30	80,045	53.2%	93.9%	8.74	0.50
			(47.2%–59.1%)	(91.9%–95.5%)	(6.84–11.17)	(0.44–0.56)
EFW (Hadlock-AC/FL/HC/BPD)	9	22,073	63.1%	94.3%	11.13	0.39
			(49.1%–75.2%)	(90.9%–96.5%)	(8.24–15.04)	(0.28–0.55)
EFW (Hadlock- AC/FL/BPD)	10	17,110	55.1%	92.9%	7.77	0.48
			(44.1%–65.7%)	(89.7%–95.2%)	(5.55–10.89)	(0.38–0.61)
EFW (Hadlock- AC/FL/HC)	7	60,648	55.2%	94.9%	11.84	0.47
			(45.7%–64.2%)	(92.4%–96.6%)	(7.46–15.74)	(0.39–0.58)
EFW (Hadlock- AC/FL)	9	16,736	60.5%	92.0%	7.54	0.43
			(50.7%–69.5%)	(89.4%–93.7%)	(6.13–9.29)	(0.34–0.54)
EFW (Hadlock- AC/BPD)	6	13,617	62.9%	93.7%	9.99	0.40
			(36.1%–83.5%)	(85.9%–97.3%)	(6.40–15.58)	(0.21–0.75)
EFW (Shepard)	7	14,060	73.7%	85.1%	4.96	0.31
			(54.4%–86.9%)	(76.5%–90.9%)	(3.29–7.48)	(0.17–0.56)
AC >36cm (or 90th centile)	5	10,543	57.8%	92.3%	7.56	0.46
			(39.6%–74.2%)	(88.7%–94.9%)	(5.85–9.77)	(0.30–0.68)
Outcome: Birthweight >4,500 g (or 97th centile)						
EFW (any) >4,000 g (or 90th centile)	5	51,686	67.5%	89.7%	6.58	0.36
			(47.8%–82.6%)	(79.1%–95.3%)	(2.78–15.58)	(0.20–0.65)
Outcome: Shoulder dystocia						
EFW (any) >4,000 g (or 90th centile)	6	26,264	22.0%	89.6%	2.12	0.87
			(9.9%–42.0%)	(80.8%–94.6%)	(1.34–3.35)	(0.74–1.02)

Abbreviations: AC, abdominal circumference; BPD, Biparietal diameter; CI, confidence interval; EFW, estimated fetal weight; FL, femur length; HC, head circumference; LR, likelihood ratio

https://doi.org/10.1371/journal.pmed.1003190.t001

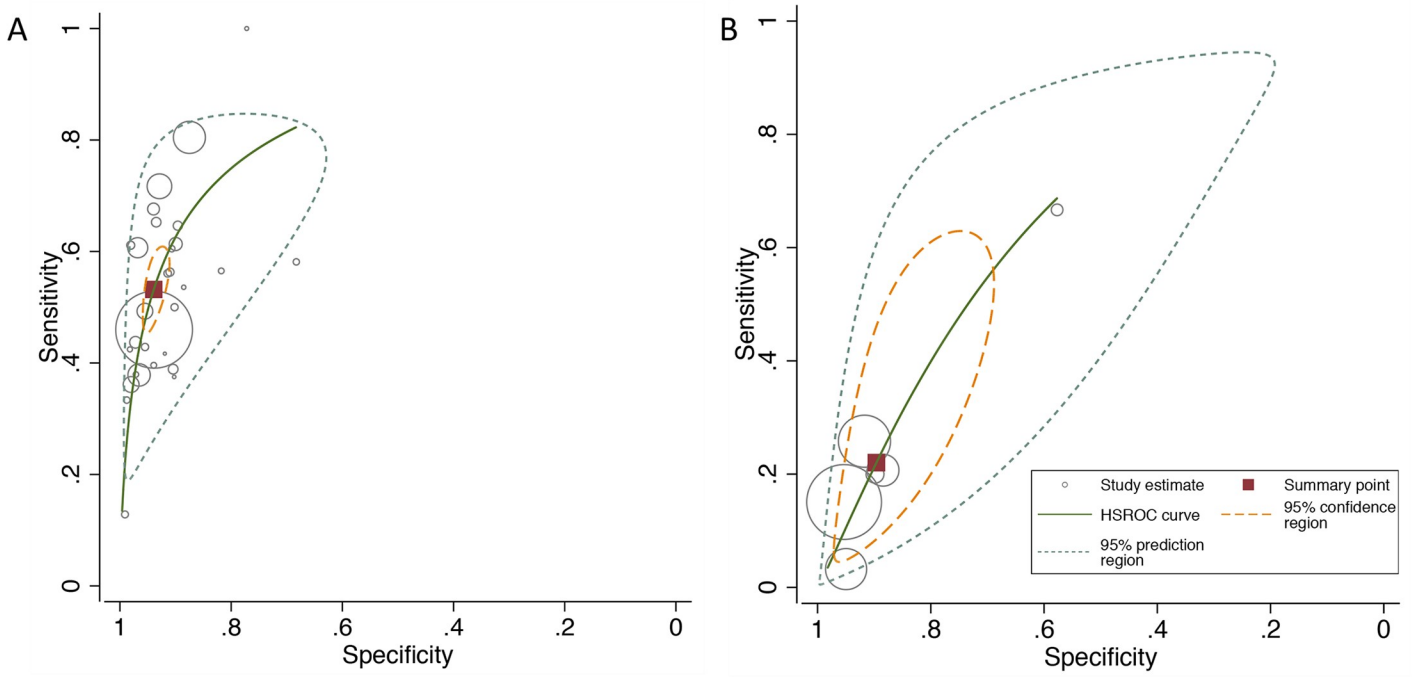


Fig 3. Summary ROC curves for the diagnostic performance of EFW >4,000 g (or 90th centile) at predicting (A) macrosomia at birth (birthweight above 4,000 g or above the 90th centile) and (B) shoulder dystocia. EFW, estimated fetal weight.

https://doi.org/10.1371/journal.pmed.1003190.g003

performance of definitions using just the AC was similar to using an ultrasonic EFW. The sensitivity for predicting severe macrosomia at birth of suspected LGA was around 70%. However, macrosomia (EFW >4,000 g or >90th centile) had a lower (22%) sensitivity for predicting shoulder dystocia, although the association was statistically significant and the positive LR was approximately 2.

Fig 3 has summary ROC curves for shoulder dystocia and macrosomia. For the prediction of macrosomia at birth, most of the large studies were close to the point estimate, and only a few small studies were outside the prediction intervals. For shoulder dystocia, most studies reported sensitivities below 30%, and only one study [55] reported a sensitivity of >50%. However, in this study, the total number of shoulder dystocia cases was very small ($n = 3$). Fig 4 and Fig 5 present graphs of the pooling of DORs for macrosomia and shoulder dystocia, respectively. There was significant heterogeneity for the prediction of macrosomia but not for the prediction of shoulder dystocia.

Only three studies—Crimmins 2018 [25], Galvin 2017 [29], and Sovio 2018 [52]—reported neonatal unit admission as an outcome, and a meta-analysis was not feasible. However, none of the studies reported statistically significant results with positive LRs of 0.73 (95% confidence interval [CI] 0.36–1.48), 1.39 (95% CI 0.97–2.00), and 1.33 (95% CI 0.80–2.22), respectively. Only the Sovio 2018 [52] study reported on 5-minute Apgar score of less than 7 and neonatal metabolic acidosis with positive LRs of 1.94 (95% CI 0.66–5.75) and 1.08 (95% CI 0.28–4.18), respectively. Moreover, the Sovio 2018 study was the only one that reported on neonatal hypoglycaemia and neonatal jaundice with positive LRs of 1.9 (95% CI 1.1–3.4) and 1.2 (95% CI 0.6–2.4), respectively.

The analysis demonstrated no significant evidence of publication bias ($P = 0.57$) when evaluated using Deeks' funnel plot asymmetry test (S2 Fig).

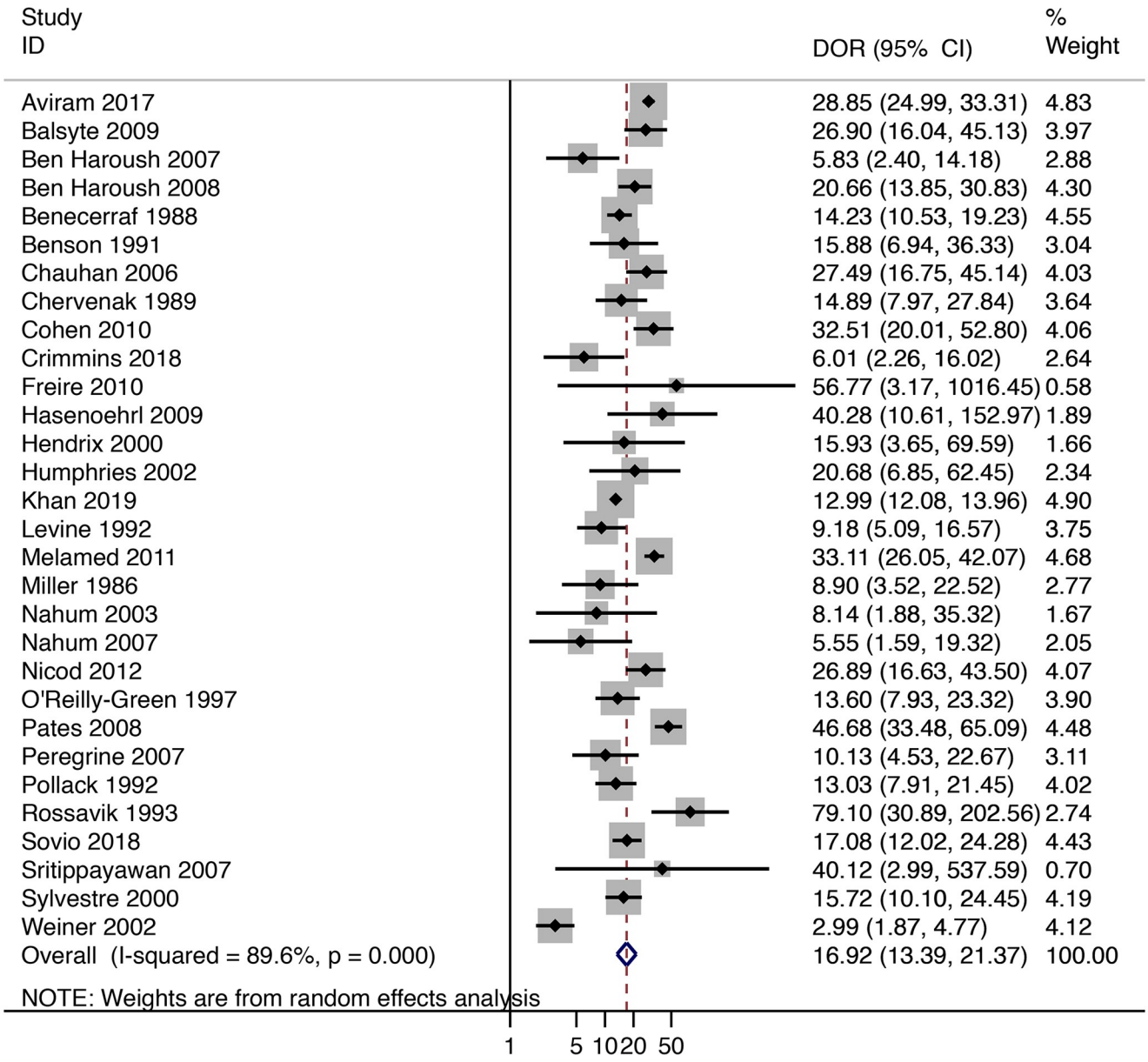


Fig 4. Diagnostic performance of EFW >4,000 g (or 90th centile) at predicting macrosomia at birth (birthweight above 4,000 g or above the 90th centile). EFW, estimated fetal weight.

https://doi.org/10.1371/journal.pmed.1003190.g004

Discussion

The main conclusion of this analysis is that an ultrasonic EFW indicating an increased risk of a large baby was strongly associated with delivering a macrosomic infant, but it was only weakly associated with the risk of shoulder dystocia. When the EFW was calculated using the widely employed Hadlock method, the positive LRs for macrosomia were in the region of 7 to 12, whereas they were approximately 2 in relation to the risk of shoulder dystocia.

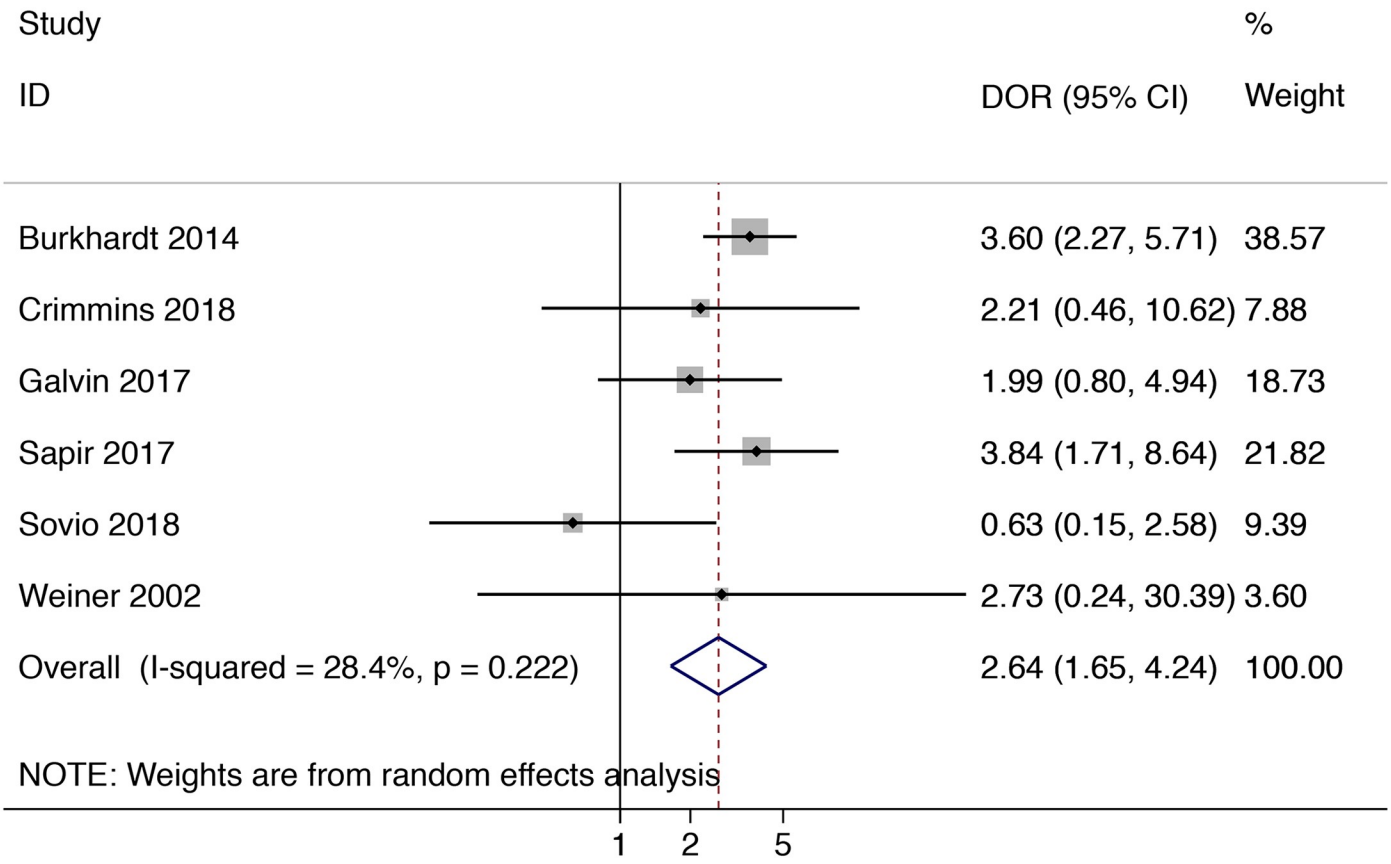


Fig 5. Diagnostic performance of EFW >4,000 g (or 90th centile) at predicting shoulder dystocia. EFW, estimated fetal weight.

https://doi.org/10.1371/journal.pmed.1003190.g005

This is the largest systematic review on the prediction of macrosomia and the only study that was focused on low- and mixed-risk populations from the perspective of using third-trimester ultrasound as routine screening in all pregnancies. We reported on multiple ultrasound markers and formulas. Moreover, we also reported on the prediction of shoulder dystocia, which is a major perinatal complication, the prevention of which would be a major aim of the routine ultrasound screening. The main limitation of this study is that there was significant heterogeneity between the studies in the ability to predict a macrosomic infant, as the forest plot of DORs indicates. The source of this heterogeneity is unclear, but it could relate to differences in the quality of the performance of the diagnostic test, such as the quality of the imaging equipment, the skill and training of sonographers, and the characteristics of the population. Finally, despite the large amount of studies included, only three studies [25, 29, 52] reported any outcomes of neonatal morbidity, and a meta-analysis was not feasible.

In the current study, we incorporated previously published data from the POP study (Sovio 2018) [52], which included nulliparous women who had a research scan at 36 wkGA, which was blinded in most cases to the clinicians. We found that the DOR (95% CI) from the POP study was very similar to the summary DOR derived from all of the other studies, which suggests that the results from the POP study are likely to be generalizable. The POP study was one of only a few identified that blinded the ultrasound result. Another blinded study, conducted in seven centers across Ireland between 2012 and 2015, the GENESIS study (Galvin 2017) [29], was a prospective cohort study of 2,772 nulliparous pregnant women. The results of the GENESIS study have only

been published in conference proceedings [29] and include the outcome of shoulder dystocia but not macrosomia. Interestingly, neither the POP study nor the GENESIS study observed a statistically significant association between ultrasonic LGA and shoulder dystocia. When blinded and unblinded studies were combined, the meta-analysis demonstrated that ultrasound may be predictive of shoulder dystocia, albeit weakly. However, the associations observed in the other studies may be due to ascertainment bias. Specifically, if the fetus is suspected to be large on the basis of the EFW, the staff attending the delivery may have a lower threshold for using maneuvers for shoulder dystocia in the event of any delay. They may also be more likely to document a given delay as being due to shoulder dystocia. Hence, unblinded studies could result in stronger associations with shoulder dystocia through ascertainment bias. The fact that ultrasonic EFW is relatively poor as a predictor of shoulder dystocia is not unexpected, given that the actual birth weight of the baby is also not strongly predictive of the outcome: the majority of cases of shoulder dystocia involve a normal birth weight infant [57].

Finally, ultrasonic suspicion of a large baby is a clinical situation where there is evidence that knowledge of the scan result may itself cause complications. Multiple studies have demonstrated that women who have a false positive diagnosis of fetal macrosomia based on EFW are more likely to be delivered by emergency caesarean section [58,59]. This finding underlines the potential for harm caused by screening low-risk women. Research studies in which the results of the scan are revealed could lead to associations with adverse outcomes that were caused by an iatrogenic harm from a false positive result. Conversely, analysis of studies in which the scan was revealed may fail to show true associations with adverse outcome as knowledge of the scan result led to interventions that mitigated the risk.

We conclude that ultrasonically suspected LGA in the general population has quite good diagnostic effectiveness for macrosomic birth weight. However, it is not strongly predictive of the risk of associated complications, such as shoulder dystocia. Similar observations have been made in relation to ultrasonically suspected SGA [60, 61]. That study indicated that reduced fetal abdominal growth velocity helped discriminate between healthy SGA babies and those that were at increased risk of complications. Interestingly, the analogous finding is also true in LGA babies, in whom the combination of LGA and accelerated abdominal growth velocity was associated with the risk of neonatal morbidity [52]. We believe that future studies should address the other factors which help differentiate those suspected LGA fetuses which are at the greatest risk of complications.

Supporting information

S1 PRISMA Checklist. PRISMA checklist.
(DOC)

S1 Text. Literature search strategy for Medline and EMBASE (from inception to May 2020).
(DOCX)

S1 Table. Characteristics of the studies included in the meta-analysis.
(DOCX)

S2 Table. List of studies excluded from the meta-analysis and reason for exclusion.
(DOCX)

S1 Fig. Risk of bias assessment using the QUADAS 2 tool. QUADAS 2, Quality Assessment of Diagnostic Accuracy Studies
(TIFF)

S2 Fig. Deeks' funnel plot asymmetry test.
(TIFF)

Acknowledgments

The views expressed here are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

Author Contributions

Conceptualization: Alexandros A. Moraitis, Peter Brocklehurst, Alexander E. P. Heazell, Jim G. Thornton, Stephen C. Robson, Aris Papageorghiou, Gordon C. Smith.

Data curation: Alexandros A. Moraitis, Norman Shreeve, Ulla Sovio.

Formal analysis: Alexandros A. Moraitis, Norman Shreeve, Ulla Sovio, Gordon C. Smith.

Funding acquisition: Peter Brocklehurst, Jim G. Thornton, Stephen C. Robson, Aris Papageorghiou, Gordon C. Smith.

Investigation: Alexandros A. Moraitis.

Methodology: Alexandros A. Moraitis, Ulla Sovio, Peter Brocklehurst, Gordon C. Smith.

Supervision: Gordon C. Smith.

Visualization: Alexandros A. Moraitis.

Writing – original draft: Alexandros A. Moraitis, Gordon C. Smith.

Writing – review & editing: Alexandros A. Moraitis, Norman Shreeve, Ulla Sovio, Peter Brocklehurst, Alexander E. P. Heazell, Jim G. Thornton, Stephen C. Robson, Aris Papageorghiou, Gordon C. Smith.

References

1. 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. Epub 2014/07/09. doi: 10.1097/aog.0000000000000388. PMID: 25004344.
2. 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. Epub 2009/04/21. doi: 10.1016/j.ajog.2009.02.035. PMID: 19376489.
3. Campbell S, Wilkin D. Ultrasonic measurement of fetal abdomen circumference in the estimation of fetal weight. British journal of obstetrics and gynaecology. 1975; 82(9):689–97. Epub 1975/09/01. doi: 10.1111/j.1471-0528.1975.tb00708.x. PMID: 1101942.
4. 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. American journal of obstetrics and gynecology. 1985; 151(3):333–7. Epub 1985/02/01. https://doi.org/10.1016/0002-9378(85)90298-4 PMID: 3881966.
5. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. Radiology. 1991; 181(1):129–33. Epub 1991/10/01. doi: 10.1148/radiology.181.1.1887021. PMID: 1887021.
6. Boulvain M, Irion O, Dowswell T, Thornton JG. Induction of labour at or near term for suspected fetal macrosomia. Cochrane Database Syst Rev. 2016;(5):Cd000938. Epub 2016/05/23. doi: 10.1002/14651858.CD000938.pub2. PMID: 27208913.
7. National Collaborating Centre for Ws, Children's H. National Institute for Health and Clinical Excellence: Guidance. Antenatal Care: Routine Care for the Healthy Pregnant Woman. London: RCOG Press National Collaborating Centre for Women's and Children's Health.; 2008.

8. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011; 155 (8):529–36. Epub 2011/10/19. doi: 10.7326/0003-4819-155-8-201110180-00009. PMID: 22007046.

9. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol. 1991; 44(8):763–70. Epub 1991/01/01. https://doi.org/10.1016/0895-4356(91)90128-v PMID: 1941027.

10. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Statistics in medicine. 2001; 20(19):2865–84. Epub 2001/09/25. https://doi.org/10.1002/sim.942 PMID: 11568945.

11. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005; 58(10):982–90. Epub 2005/09/20. doi: 10.1016/j.jclinepi.2005.02.022. PMID: 16168343.

12. Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. BMJ (Clinical research ed). 2001; 323(7305):157–62. Epub 2001/07/21. https://doi.org/10.1136/bmj.323.7305.157 PMID: 11463691; PubMed Central PMCID: PMC1120791.

13. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol. 2005; 58 (9):882–93. Epub 2005/08/09. doi: 10.1016/j.jclinepi.2005.01.016. PMID: 16085191.

14. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009; 6(7):e1000097. Epub 2009/07/22. doi: 10.1371/journal.pmed.1000097. PMID: 19621072; PubMed Central PMCID: PMC2707599.

15. Aviram A, Yogev Y, Ashwal E, Hiersch L, Hadar E, Gabbay-Benziv R. Prediction of large for gestational age by various sonographic fetal weight estimation formulas-which should we use? Journal of Perinatology. 2017; 37(5):513–7. https://doi.org/10.1038/jp.2017.5 PMID: 28151496.

16. Balsyte D, Schaffer L, Burkhardt T, Wisser J, Kurmanavicius J. Sonographic prediction of macrosomia cannot be improved by combination with pregnancy-specific characteristics. Ultrasound in Obstetrics & Gynecology. 2009; 33(4):453–8. https://doi.org/10.1002/uog.6282 PMID: 19266500.

17. Benacerraf BR, Gelman R, Frigoletto FD Jr. Sonographically estimated fetal weight: Accuracy and limitation. American Journal of Obstetrics and Gynecology. 1988; 159(5):1118–21. https://doi.org/10.1016/0002-9378(88)90425-5 PMID: 3055999.

18. Ben-Haroush A, Yogev Y, Hod M, Bar J. Predictive value of a single early fetal weight estimate in normal pregnancies. European Journal of Obstetrics Gynecology and Reproductive Biology. 2007; 130 (2):187–92. PMID: 46097242.

19. Ben-Haroush A, Melamed N, Mashiach R, Meizner I, Yogev Y. Use of the amniotic fluid index combined with estimated fetal weight within 10 days of delivery for prediction of macrosomia at birth. Journal of Ultrasound in Medicine. 2008; 27(7):1029–32. https://doi.org/10.7863/jum.2008.27.7.1029 PMID: 18577666.

20. Benson CB, Coughlin BF, Doubilet PM. Amniotic fluid volume in large-for-gestational-age fetuses of nondiabetic mothers. Journal of Ultrasound in Medicine. 1991; 10(3):149–51. https://doi.org/10.7863/jum.1991.10.3.149 PMID: 2027186.

21. Burkhardt T, Schmidt M, Kurmanavicius J, Zimmermann R, Schaffer L. Evaluation of fetal anthropometric measures to predict the risk for shoulder dystocia. Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2014; 43(1):77–82. PMID: 373970964.

22. Chauhan SP, Parker D, Shields D, Sanderson M, Cole JH, Scardo JA. Sonographic estimate of birth weight among high-risk patients: feasibility and factors influencing accuracy. American journal of obstetrics and gynecology. 2006; 195(2):601–6. Epub 2006/06/27. doi: 10.1016/j.ajog.2006.04.012. PMID: 16796980.

23. Chervenak JL, Divon MY, Hirsch J, Girz BA, Langer O. Macrosomia in the postdate pregnancy: Is routine ultrasonographic screening indicated? American Journal of Obstetrics and Gynecology. 1989; 161 (3):753–6. https://doi.org/10.1016/0002-9378(89)90395-5 PMID: 2675605.

24. Cohen JM, Hutcheon JA, Kramer MS, Joseph KS, Abenhaim H, Platt RW. Influence of ultrasound-to-delivery interval and maternal-fetal characteristics on validity of estimated fetal weight. Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2010; 35(4):434–41. PMID: 359676712.

25. Crimmins S, Mo C, Nassar Y, Kopelman JN, Turan OM. Polyhydramnios or Excessive Fetal Growth Are Markers for Abnormal Perinatal Outcome in Euglycemic Pregnancies. American Journal of Perinatology. 2018; 35(2):140–5. https://doi.org/10.1055/s-0037-1606186 PMID: 28838004.

26. Cromi A, Ghezzi F, Di Naro E, Siesto G, Bergamini V, Raio L. Large cross-sectional area of the umbilical cord as a predictor of fetal macrosomia. Ultrasound in Obstetrics & Gynecology. 2007; 30(6):861–6. https://doi.org/10.1002/uog.5183 PMID: 17960667.

27. De Reu PAOM, Smits LJM, Oosterbaan HP, Nijhuis JG. Value of a single early third trimester fetal biometry for the prediction of birth weight deviations in a low risk population. Journal of Perinatal Medicine. 2008; 36(4):324–9. https://doi.org/10.1515/JPM.2008.057 PMID: 18598122

28. Freire DMC, Cecatti JG, Paiva CSM. Correlation between estimated fetal weight by ultrasound and neonatal weight. [Portuguese]. Revista Brasileira de Ginecologia e Obstetricia. 2010; 32(1):4–10. https://doi.org/10.1590/s0100-72032010000100002 PMID: 20209256.

29. Galvin DM, Burke N, Burke G, Breathnach F, McAuliffe F, Morrison J, et al. 94: Accuracy of prenatal detection of macrosomia >4,000g and outcomes in the absence of intervention: results of the prospective multicenter genesis study. American Journal of Obstetrics and Gynecology. 2017; 216(1, Supplement):S68. https://doi.org/10.1016/j.ajog.2016.11.983.

30. Gilby JR, Williams MC, Spellacy WN. Fetal abdominal circumference measurements of 35 and 38 cm as predictors of macrosomia. A risk factor for shoulder dystocia. Journal of Reproductive Medicine. 2000; 45(11):936–8. PMID: 11127108.

31. Hasenoehrl G, Pohlhammer A, Gruber R, Staudach A, Steiner H. Fetal weight estimation by 2D and 3D ultrasound: Comparison of six formulas. Ultraschall in der Medizin. 2009; 30(6):585–90. https://doi.org/10.1055/s-0028-1109185 PMID: 19544232.

32. Hendrix NW, Grady CS, Chauhan SP. Clinical vs. sonographic estimate of birth weight in term parturients. A randomized clinical trial. Journal of Reproductive Medicine. 2000; 45(4):317–22. PMID: 10804488.

33. Henrichs C, Magann EF, Brantley KL, Crews JH, Sanderson M, Chauhan SP. Detecting fetal macrosomia with abdominal circumference alone. Journal of Reproductive Medicine. 2003; 48(5):339–42. PMID: 12815906.

34. Humphries J, Reynolds D, Bell-Scarbrough L, Lynn N, Scardo JA, Chauhan SP. Sonographic estimate of birth weight: relative accuracy of sonographers versus maternal-fetal medicine specialists. Journal of Maternal-Fetal & Neonatal Medicine. 2002; 11(2):108–12. https://doi.org/10.1080/jmf.11.2.108.112 PMID: 12375539.

35. Kayem G, Grange G, Breart G, Goffinet F. Comparison of fundal height measurement and sonographically measured fetal abdominal circumference in the prediction of high and low birth weight at term. Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2009; 34(5):566–71. Epub 2009/07/08. doi: 10.1002/uog.6378. PMID: 19582801.

36. Kehl S, Brade J, Schmidt U, Berlit S, Bohlmann MK, Sutterlin M, et al. Role of fetal abdominal circumference as a prognostic parameter of perinatal complications. Archives of Gynecology & Obstetrics. 2011; 284(6):1345–9. https://doi.org/10.1007/s00404-011-1888-8 PMID: 21431328.

37. Khan N, Ciobanu A, Karampitsakos T, Akolekar R, Nicolaides KH. Prediction of large for gestational age neonates by routine third trimester ultrasound. Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2019; 25. PMID: 628504046.

38. Levine AB, Lockwood CJ, Brown B, Lapinski R, Berkowitz RL. Sonographic diagnosis of the large for gestational age fetus at term: does it make a difference? Obstetrics & Gynecology. 1992; 79(1):55–8. PMID: 1727587.

39. Melamed N, Yogev Y, Meizner I, Mashiach R, Pardo J, Ben-Haroush A. Prediction of fetal macrosomia: effect of sonographic fetal weight-estimation model and threshold used. Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2011; 38(1):74–81. PMID: 362515594.

40. Miller JM Jr., Korndorffer FA 3rd, Gabert HA. Fetal weight estimates in late pregnancy with emphasis on macrosomia. Journal of Clinical Ultrasound. 1986; 14(6):437–42. https://doi.org/10.1002/jcu.1870140606 PMID: 3091643.

41. Miller JM Jr, Brown HL, Khawli OF, Pastorek IJG, Gabert HA. Ultrasonographic identification of the macrosomic fetus. American Journal of Obstetrics and Gynecology. 1988; 159(5):1110–4. https://doi.org/10.1016/0002-9378(88)90423-1 PMID: 3055998.

42. Nahum GG, Pham KQ, McHugh JP. Ultrasonic prediction of term birth weight in Hispanic women. Accuracy in an outpatient clinic. Journal of Reproductive Medicine. 2003; 48(1):13–22. PMID: 12611089.

43. Nahum GG, Stanislaw H. A computerized method for accurately predicting fetal macrosomia up to 11 weeks before delivery. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2007; 133(2):148–56. https://doi.org/10.1016/j.ejogrb.2006.08.011 PMID: 17010500.

44. Nicod AC, Hohlfeld P, Vial Y. Performance of ultrasound estimation of fetal weight in fetuses weighing < = 2000 g and more than 4000 g. [French]. Revue Medicale Suisse. 2012; 8(359):2022–7. PMID: 23167077.

45. O'Reilly-Green CP, Divon MY. Receiver operating characteristic curves of sonographic estimated fetal weight for prediction of macrosomia in prolonged pregnancies. Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 1997; 9 (6):403–8. Epub 1997/06/01. doi: 10.1046/j.1469-0705.1997.09060403.x. PMID: 9239826.

46. Pates JA, McIntire DD, Casey BM, Leveno KJ. Predicting macrosomia. Journal of Ultrasound in Medicine. 2008; 27(1):39–43. https://doi.org/10.7863/jum.2008.27.1.39 PMID: 18096729.

47. Peregrine E, O'Brien P, Jauniaux E. Clinical and ultrasound estimation of birth weight prior to induction of labor at term. Ultrasound in Obstetrics & Gynecology. 2007; 29(3):304–9. https://doi.org/10.1002/uog.3949 PMID: 17290365.

48. Pollack RN, Hauer-Pollack G, Divon MY. Macrosomia in postdates pregnancies: the accuracy of routine ultrasonographic screening. American Journal of Obstetrics & Gynecology. 1992; 167(1):7–11. https://doi.org/10.1016/s0002-9378(11)91615-9 PMID: 1442959.

49. Rossavik IK, Joslin GL. Macrosomatia and ultrasonography: what is the problem? Southern Medical Journal. 1993; 86(10):1129–32. https://doi.org/10.1097/00007611-199310000-00010 PMID: 8211330.

50. Sapir A, Khayyat I, Drukker L, Rabinowitz R, Samueloff A, Sela HY. Ultrasound predication of shoulder dystocia in low risk term singleton deliveries. American Journal of Obstetrics and Gynecology. 2017; 216 (1 Supplement 1):S221. PMID: 614090070.

51. Smith GC, Smith MF, McNay MB, Fleming JE. The relation between fetal abdominal circumference and birthweight: findings in 3512 pregnancies. British journal of obstetrics and gynaecology. 1997; 104 (2):186–90. Epub 1997/02/01. https://doi.org/10.1111/j.1471-0528.1997.tb11042.x PMID: 9070136.

52. Sovio U, Moraitis AA, Wong HS, Smith GCS. Universal vs selective ultrasonography to screen for large-for-gestational-age infants and associated morbidity. Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2018; 51(6):783–91. Epub 2017/04/21. doi: 10.1002/uog.17491. PMID: 28425156.

53. Sritippayawan S, Anansakunwat W, Suthantikorn C. The accuracy of gestation-adjusted projection method in estimating birth weight by sonographic fetal measurements in the third trimester. Journal of the Medical Association of Thailand. 2007; 90(6):1058–67. PMID: 17624197.

54. Sylvestre G, Divon MY, Onyeije C, Fisher M. Diagnosis of macrosomia in the postdates population: combining sonographic estimates of fetal weight with glucose challenge testing. Journal of Maternal-Fetal Medicine. 2000; 9(5):287–90. https://doi.org/10.1002/1520-6661(200009/10)9:5<287::AID-MFM6>3.0.CO;2-1 PMID: 11132584.

55. Weiner Z, Ben-Shlomo I, Beck-Fruchter R, Goldberg Y, Shalev E. Clinical and ultrasonographic weight estimation in large for gestational age fetus. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2002; 105(1):20–4. https://doi.org/10.1016/s0301-2115(02)00140-9 PMID: 12270559.

56. Burke N, Burke G, Breathnach F, McAuliffe F, Morrison JJ, Turner M, et al. Prediction of cesarean delivery in the term nulliparous woman: results from the prospective, multicenter Genesis study. American journal of obstetrics and gynecology. 2017; 216(6):598.e1–.e11. Epub 2017/02/19. doi: 10.1016/j.ajog.2017.02.017. PMID: 28213060.

57. Ouzounian JG. Shoulder Dystocia: Incidence and Risk Factors. Clin Obstet Gynecol. 2016; 59(4):791–4. Epub 2016/10/21. doi: 10.1097/GRF.0000000000000227. PMID: 27662540.

58. Little SE, Edlow AG, Thomas AM, Smith NA. Estimated fetal weight by ultrasound: a modifiable risk factor for cesarean delivery? American journal of obstetrics and gynecology. 2012; 207(4):309.e1–6. Epub 2012/08/21. doi: 10.1016/j.ajog.2012.06.065. PMID: 22902073.

59. 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? American journal of obstetrics and gynecology. 2009; 200(3):340.e1–3. Epub 2009/03/04. doi: 10.1016/j.ajog.2008.12.043. PMID: 19254597.

60. 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 (London, England). 2015; 386(10008):2089–97. Epub 2015/09/12. doi: 10.1016/s0140-6736(15)00131-2. PMID: 26360240; PubMed Central PMCID: PMC4655320.

61. Heazell AE, Hayes DJ, Whitworth M, Takwoingi Y, Bayliss SE, Davenport C. Biochemical tests of placental function versus ultrasound assessment of fetal size for stillbirth and small-for-gestational-age infants. Cochrane Database Syst Rev. 2019; 5:Cd012245. Epub 2019/05/16. doi: 10.1002/14651858.CD012245.pub2. PMID: 31087568; PubMed Central PMCID: PMC6515632.

Placenta 108 (2021) 47–54

Contents lists available at ScienceDirect

Placenta

journal homepage: http://www.elsevier.com/locate/placenta

Review

Fetal umbilical artery Doppler as a tool for universal third trimester screening: A systematic review and meta-analysis of diagnostic test accuracy

Alexandros A. Moraitis [a], Thomas Bainton [a], Ulla Sovio [a], Peter Brocklehurst [b], Alexander EP. Heazell [c,d], Jim G. Thornton [e], Stephen C. Robson [f], Aris Papageorghiou [g], Gordon CS. Smith [a,*]

[a] Department of Obstetrics and Gynaecology, University of Cambridge; NIHR Cambridge Comprehensive Biomedical Research Centre, CB2 2SW, United Kingdom
[b] Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, United Kingdom
[c] Maternal and Fetal Health Research Centre, School of Medical Sciences, Faculty of Biological, Medical and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, M13 9WL, United Kingdom
[d] St. Mary's Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, M13 9WL, United Kingdom
[e] Division of Child Health, Obstetrics and Gynaecology, School of Medicine, University of Nottingham, Nottingham, United Kingdom
[f] Reproductive and Vascular Biology Group, The Medical School, University of Newcastle, Newcastle, United Kingdom
[g] Nuffield Department of Obstetrics and Gynaecology, Oxford, United Kingdom

ARTICLE INFO

Keywords:
Fetal growth restriction
Neonatal morbidity
Small for gestational age
Ultrasound
Screening
Umbilical artery Doppler

ABSTRACT

The objective of this study was to investigate the accuracy of universal third trimester umbilical artery (UA) Doppler to predict adverse pregnancy outcome at term. We searched Medline, EMBASE, the Cochrane library and ClinicalTrials.gov from inception to October 2020 and we also analyzed previously unpublished data from a prospective cohort study of nulliparous women, the Pregnancy Outcome Prediction (POP) study. We included studies that performed a third-trimester ultrasound scan in unselected, low or mixed risk populations, excluding studies which only included high risk pregnancies. Meta-analysis was performed using the hierarchal summary receiver operating characteristic curve (HSROC) analysis and bivariate logit-normal models. We identified 13 studies (including the POP study) involving 67,764 pregnancies which met our inclusion criteria. The overall quality was variable and only six studies (N = 5777 patients) blinded clinicians to the UA Doppler result. The summary sensitivity and positive likelihood ratio (LR) for small for gestational age (SGA; birthweight <10th centile) were 21.7% (95% CI 13.2–33.6%) and 2.65 (95% CI 1.89–3.72) respectively. The summary positive LR for NICU admission and metabolic acidosis were 1.35 (95% CI 0.93–1.97) and 1.34 (95% CI 0.86–2.08) respectively. The results were similar in the POP study: associations with SGA (positive LR 2.66 [95% CI 2.11–3.36]) and severe SGA (birthweight <3rd centile; positive LR 3.27 [95% CI 2.29–4.68]) but no statistically significant association with neonatal morbidity. We conclude that third trimester UA Doppler has moderate predictive accuracy for small for gestational age but not for indicators of neonatal morbidity in unselected and low risk pregnancies.

1. Introduction

Measurement of the umbilical artery (UA) Doppler has been used to monitor high risk pregnancies, including those with suspected fetal growth restriction (FGR). A systematic review of diagnostic test accuracy has shown that the UA Doppler can be useful at predicting perinatal mortality and risk of compromise in high-risk pregnancies [1]. A Cochrane review of 19 randomized controlled trials (RCTs) of 10,667 women has shown that UA Doppler ultrasound in high-risk pregnancies reduces the number of perinatal deaths (risk ratio [RR] 0.71, 95% confidence interval [CI] 0.52 to 0.98) and the number of Cesarean sections (RR 0.90, 95% CI 0.84 to 0.97) [2]. However, a similar Cochrane

* Corresponding author. Department of Obstetrics and Gynaecology, University of Cambridge, The Rosie Hospital, Cambridge, CB2 2SW, United Kingdom.
E-mail address: gcss2@cam.ac.uk (G.CS. Smith).

https://doi.org/10.1016/j.placenta.2021.03.011
Received 10 January 2021; Received in revised form 11 March 2021; Accepted 17 March 2021
Available online 22 March 2021

review of RCTs in low risk pregnancies was unable to demonstrate any difference in perinatal deaths between pregnancies that were offered UA Doppler and those that had no Doppler (risk ratio 0.80, 95% confidence interval 0.35 to 1.83) [3]. This latter review included 5 studies that compared routine Doppler vs no Doppler, none of the studies blinded the results to the clinicians, but despite this there was no consistent management plan for the abnormal findings. Critically, the meta-analysis only included data from a total of 14,185 women whereas this study design would require more than 100,000 women to assess rare outcomes such as perinatal death [4]. The conclusion was that there is no evidence that the use of routine UA Doppler ultrasound benefits either the mother or the baby.

In order for a large screening programme to be implemented it needs to meet two prerequisites; first, an index test that can accurately predict adverse outcome and, second, a safe and effective intervention. In this context we have suggested that ultrasound screening at 36 weeks' gestation with a policy of induction of labor (IOL) for those screened positive at 37 weeks could have the potential to be safe and effective [5]. We have previously shown that universal ultrasound screening using biometry and fetal growth velocity can triple the detection of SGA infants and identify infants at increased risk of neonatal morbidity [6]. However, there is no clear evidence of the effectiveness of UA Doppler as part of universal ultrasound screening.

The aim of this study is to assess the accuracy of third trimester UA Doppler to predict adverse pregnancy outcome at term. We conducted a systematic review and meta-analysis focusing in low and mixed risk populations. In the above analysis we also included unpublished data from a large prospective cohort study, the Pregnancy Outcome Prediction (POP) study [6].

2. Methods

2.1. Search strategy

The protocol for the review was designed a priori and registered with the PROSPERO International Prospective Register of Systematic Reviews (Registration number: CRD42017064093). This review is part of a larger project on the value of the universal third trimester screening which was funded by NIHR Health Tachnology Assessment (HTA 15/105/01). A systematic search was performed using Medline, EMBASE, the Cochrane database of Systematic Reviews (CDSR), Cochrane Central Register of Controlled Trials (CENTRAL), and ClinicalTrials.gov from inception to October 2020. The studies were identified using a combination of words related to "ultrasound", "Doppler", "umbilical artery", "pregnancy" and "prenatal diagnosis". No restrictions for language or geographic location were applied. The exact search strategy is presented in the Supplementary file.

2.2. Study selection

We included cohort and cross-sectional studies as well as randomized clinical trials with singleton pregnancies which had an ultrasound done after 24 weeks' gestation. Case-control studies were excluded. We included all studies in which the ultrasound was done as part of universal ultrasound screening (the ultrasound was offered to all women regardless of indication), studies that were done in low-risk populations (those that excluded pregnancies with any maternal or fetal complication) and studies with mixed risk population (the ultrasound was offered selectively based on current clinical indications including low risk indications such as presentation). We excluded studies that were focused only on high risk populations such as pregnancies with FGR as well as studies that performed the ultrasound intrapartum.

We included all reported indices of umbilical artery Doppler such as the Pulsatility Index (PI), Resistance Index (RI) or the systolic to diastolic ratio (S/D ratio), as well as all reported cut-off values. We included studies that reported the following outcomes: Small for gestational age (SGA; defined as birthweight <10th centile), severe SGA (birthweight <3rd centile), neonatal unit admission, 5-min Apgar score <7, neonatal metabolic acidosis (most commonly defined as cord arterial PH < 7.1), Cesarean section for fetal distress, and pre-eclampisa (PET). We also defined as severe adverse perinatal outcome one or more of the following: stillbirth, neonatal death, hypoxic ischemic encephalopathy, inotrope support, or severe metabolic acidosis (defined as a cord arterial pH < 7.0 and a base deficit of >12 mEq/L). Finally, we included studies regardless of blinding of the ultrasound to the clinicians but this was reported in the study characteristics.

2.3. Study quality assessment

The literature search, study selection, and analysis were performed independently by two authors (AM and TB) using Review Manager 5.3. Any differences were resolved in discussion with the senior author (GS). The risk of bias in each included study was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS 2) tool as explained in the Cochrane Handbook of Diagnostic Test Accuracy Studies [7]. We used a pre-designed data extraction form to extract information on study characteristics (year of publication, country, setting, study design, blinding), patient characteristics (inclusion and exclusion criteria, sample size), the index test (gestation at scan, Doppler indices and cut-off values used), reference standard (pregnancy outcome, gestation at delivery, and interval from scan to delivery).

2.4. Analysis of data from the POP study

In the systematic review we included unpublished data from a prospective cohort study, the Pregnancy Outcome Prediction study (POPS), which was conducted at the Rosie Hospital, Cambridge (UK) between 2008 and 2012 and which has previously been described in detail [6]. In brief, the study included nulliparous women only, and all women who agreed to participate had two research ultrasound scans at 28 and 36 weeks' gestational age (wkGA). The ultrasound results were blinded to the women and the clinicians except in certain circumstances based on the protocol (detailed in Pasupathy et al., 2008) [8], but all the women who had a result revealed were excluded from this analysis. About 40% of the women had clinically indicated ultrasound scans in the third trimester based on the local and national guidelines and those results were revealed to the clinicians. In the present analysis we included women that 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.

Screen positive was defined as an umbilical artery pulsatility index (PI) > 90th percentile [6]. The databases used, and the definition of the outcome data have been described before [6]. In brief, neonatal morbidity was defined as ≥1 of the following: a 5 min Apgar score less than 7, delivery with metabolic acidosis (defined as a cord arterial pH < 7.1 and a base deficit of >10 mEq/L) or admission to the neonatal unit at term (defined as admission <48 h after birth at ≥37 weeks gestational age and discharge ≥48 h after admission). Severe adverse perinatal outcome was defined as term perinatal death, hypoxic ischemic encephalopathy, use of inotropes, mechanical ventilation, or severe metabolic acidosis (defined as a cord arterial pH < 7.0 and a base deficit of >12 mEq/L). Small for gestational age (SGA) and severe SGA were defined as a birthweight <10th percentile and <3rd percentile respectively for sex and gestational age using a UK reference. [9].

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

2.5. Statistical analysis

From each study we extracted the 2 × 2 tables for all combinations of index tests and outcomes and we calculated the sensitivity, specificity, positive and negative likelihood ratios (LRs) respectively [10]. For the

data synthesis we used the hierarchical summary receiver – operator characteristics (HSROC) model of Rutter and Gatsonis [11] which allows for variation of cut-off points between studies. Whenever four or more studies were available, estimates of mean sensitivity and specificity and respective variances at a specific threshold were additionally generated using the bivariate logit-normal model [12]. Finally, we used pooling of the diagnostic odds ratios (DORs) [13]. For the assessment of publication bias we used the Deeks' funnel plot asymmetry test and plotted the inverse of the square root of the effective sample size against the DOR [14]. As this method requires a large number of studies, we used the most commonly reported outcome for the analysis. For the statistical analyses we used the METANDI, METAN and MIDAS packages from STATA version 14 (StataCorp LP, College Station, TX).

3. Results

3.1. Study selection and characteristics of included studies

The literature search PRISMA flowchart is presented in Fig. 1. Out of 6594 titles identified through database searching, 29 were eligible for full-text review. We identified 13 studies [15–25] including the POP study that met the inclusion criteria involving 67,764 participants in total. The study characteristics are presented in Supplementary Table 1. Five studies [15,21,24,25] (N = 63,436) included unselected pregnancies, four studies [16,19,20,26] (N = 2634) included only low-risk pregnancies and four studies [17,18,22,23](N = 1694) included mixed risk pregnancies. Three of the studies [15,24,25] that were done in the same hospitals might have a short period of patient overlap. Nine studies [16,17,19–23,26](N = 8097) were prospective and four [15,18,24,25] (N = 59,687) were retrospective. The differences between the studies included the gestational age at scan (ranging from 28 to 41 weeks' gestation), as well as the indices and the cut-off points used. The majority of patients from all of the studies delivered at term. The studies excluded from the meta-analysis and the reasons for the exclusion are presented in Supplementary Table 2.

3.2. Risk of bias of the included studies

The assessment of study quality is presented in Fig. 2. Overall the quality was variable. The main risk of bias was that only six studies [16,

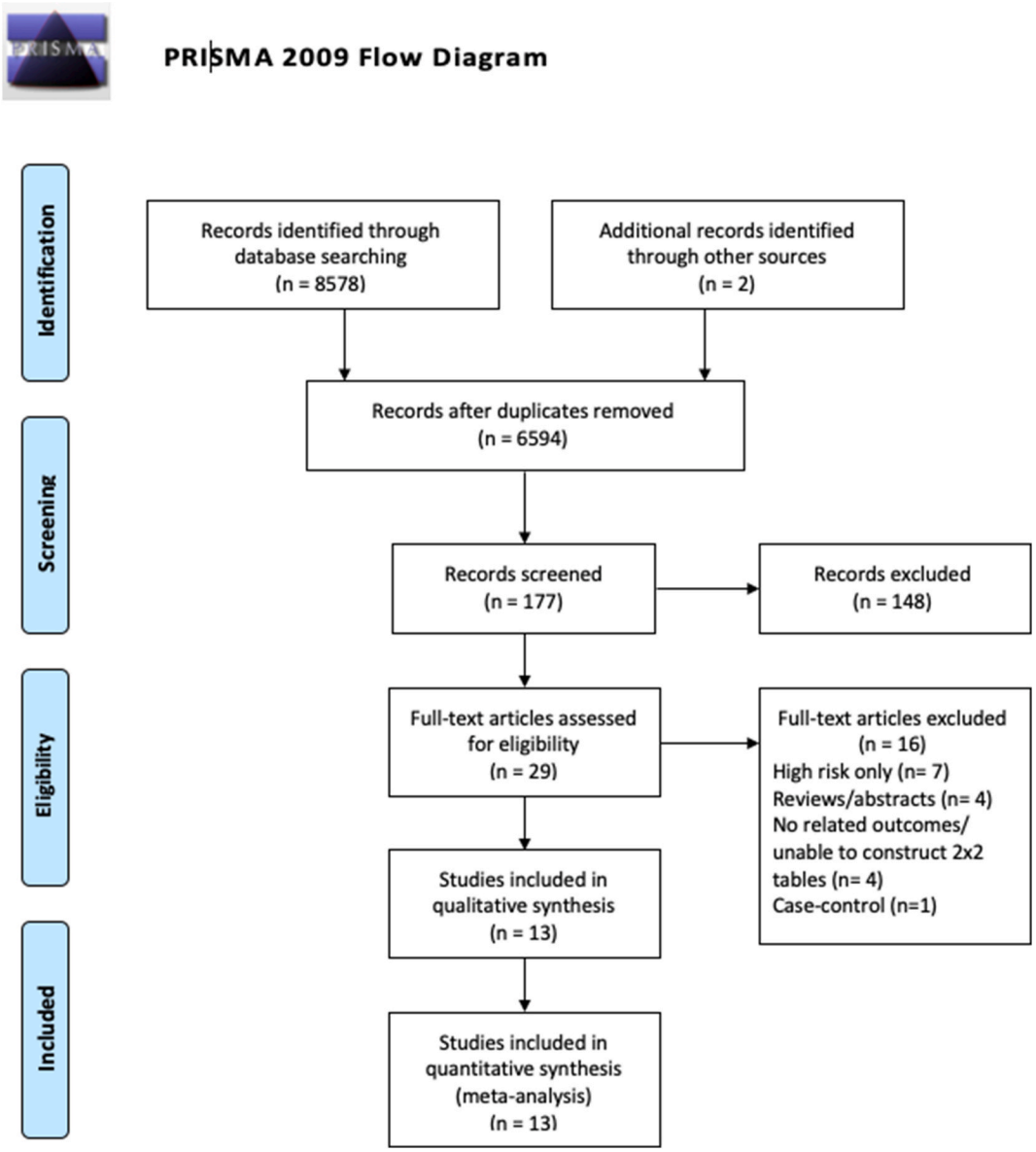


Fig. 1. Literature search PRISMA flow diagram.

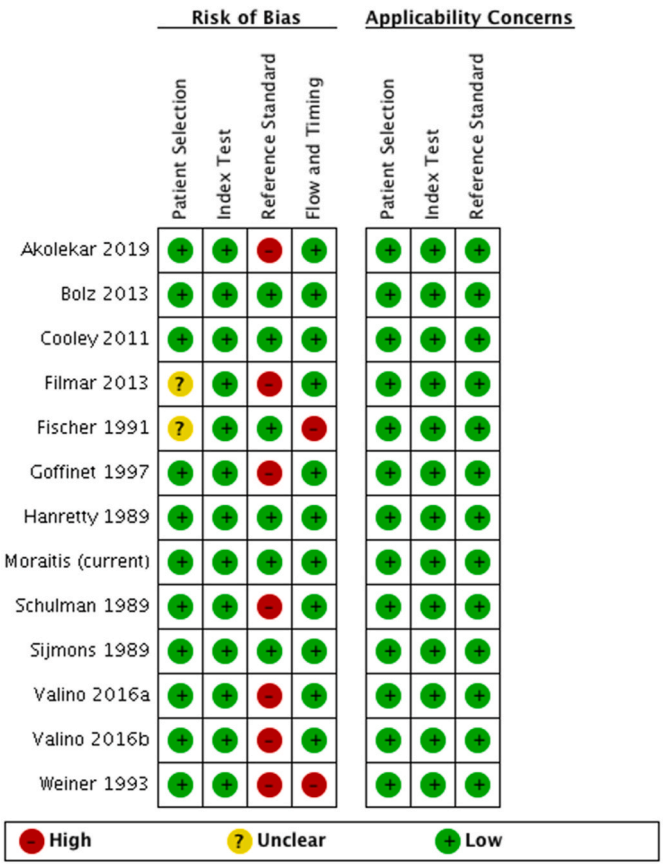


Fig. 2. Risk of bias graph of included studies.

17,19,21,23] (N = 5777) blinded the UA Doppler to the clinicians and only one study (POPS) blinded the whole ultrasound findings. Five of these six studies revealed other features of the scan result, such as fetal biometry meaning only the POP study blinded clinicians to both the utero-placental Doppler and fetal biometry. We classified two studies [19,26] as high risk for flow and timing as they included only pregnancies beyond 41 weeks's gestation. Finally we classified two studies [18,19] as unclear risk of patient selection bias as they did not clarify whether the recruitment was random or consecutive.

3.3. Synthesis of results

The summary results of the meta-analysis are presented in Table 1. A high resistance pattern of UA Doppler was associated with an increased risk of delivering an SGA infant or a severely SGA infant. However, the finding was not strongly predictive with positive LRs between 2.5 and 3.0. A high resistance pattern of UA Doppler was not associated with an increased risk of a range of indicators of neonatal morbidity. The summary ROC curves for those outcomes are presented in Fig. 3. Most of the large studies reported low sensitivities and high specifities for all the reported outcomes. However, some smaller studies reported higher sensitivities which resulted in large prediction intervals. For some outcomes such as 5-min Apgar score <7, Cesarean section for fetal distress, and pre-eclampsia (PET) the Rutter-Gatsonis model was unable to produce summary results despite an adequate number of studies. However, we performed pooling of DORs for all the reported outcomes and illustrated the variation between studies using forest plots (Fig. 4). The overall heterogeneity for most outcomes was low with notable exeption NICU admission (Fig. 4). This is likely due to the Filmar et al. [18] study reporting significantly higher DOR for NICU admission which could be related to the difference in mean gestation at delivery (37 weeks for those with abnormal UA Doppler and 39 weeks for those with normal).

We had planned to perform subgroup analyses for the different populations (universal, low-risk, and mixed-risk), gestational age at scan (<36 wkGA, ≥36 wkGA), blinding of results to the clinicians, and whether the study was prospective or retrospective. However, due to the small number of studies for each combination of subgroup and outcome separate meta-analyses per subgroup were not feasible. Nevertheless, we found that the three large studies that performed universal ultrasound scan at 36 weeks (Akolekar 2019, Valino 2016b and POPS) showed similar results (overlapping confidence intervals; Fig. 4) despite the first two studies being retrospective and not blinded, and POPS being prospective and blinded.

Finally, we used the Deeks' funnel plot asymmetry test to assess the risk of publication bias using the outcome of neonatal unit admission for the analysis (Supplementary Figure 1). The test suggested no evidence of publication bias (P = 0.52).

3.4. Analysis of the POP study

The analysis included 3615 women that met the inclusion criteria (Supplementary Figure 2). The maternal age, social status, ethnicity, BMI, and rates of alcohol consumption were similar between women with normal and high resistance UA at 36-weeks (Supplementary Table 2). The gestational age at delivery and rate of IOL were similar in both groups which can be attributed to the blinding of the ultrasound findings (Supplementary Table 3). The screening performance of UA PI > 90th centile is presented in Table 2. A high resistance pattern of UA Doppler was associated with an increased risk of delivering an SGA infant or a severely SGA infant and the association was stronger for the latter outcome. However, the finding was not strongly predictive with positive LRs between 2.5 and 3.5. A high resistance pattern of UA Doppler was not associated with an increased risk of the range of indicators of neonatal morbidity in the POP study, although severe adverse outcomes were uncommon.

Table 1
Summary diagnostic results of meta-analysis of the umbilical artery Doppler at predicting adverse pregnancy outcome.

Outcome	Number of studies	Number of patients	Summary Sensitivity (95% CI)	Summary Specificity (95% CI)	Summary Positive LR (95% CI)	Summary Negative LR (95% CI)	Diagnostic Odds Ratio (95% CI)
SGA <10th centile	8	19,203	21.7% (13.2–33.6)	91.8% (86.5–95.1)	2.65 (1.89–3.72)	0.85 (0.77–0.94)	3.03 (2.20–4.19)
SGA <3rd centile	5	53,907	25.4% (14.0–41.5%)	90.4% (78.6–96.1%)	2.65 (1.92–3.66)	0.83 (0.75–0.91)	3.31 (2.99–3.67)
NICU admission	8	66,253	13.6 (6.8–25.3)	89.9 (83.5–94.0)	1.35 (0.93–1.97)	0.96 (0.90–1.03)	1.41 (1.00–2.00)
Neonatal acidosis	5	9629	12.0% (5.3–25.0)	91.1% (81.0–96.1)	1.34 (0.86–2.08)	0.97 (0.91–1.02)	1.40 (0.86–2.26)
Severe APO[a]	4	58,866	9.3% (4.8–17.5)	88.3% (74.5–95.2)	0.80 (0.44–1.46)	1.03 (0.95–1.11)	0.81 (0.49–1.34)

SGA, Small for gestational age; LR, Likelihood ratio; CI, Confidence intervals; APO, Adverse pregnancy outcome.
[a] The definition varied between studies and includes one or more of the following: stillbirth, neonatal death, hypoxic ischemic encephalopathy, inotrope support, or severe metabolic acidosis.

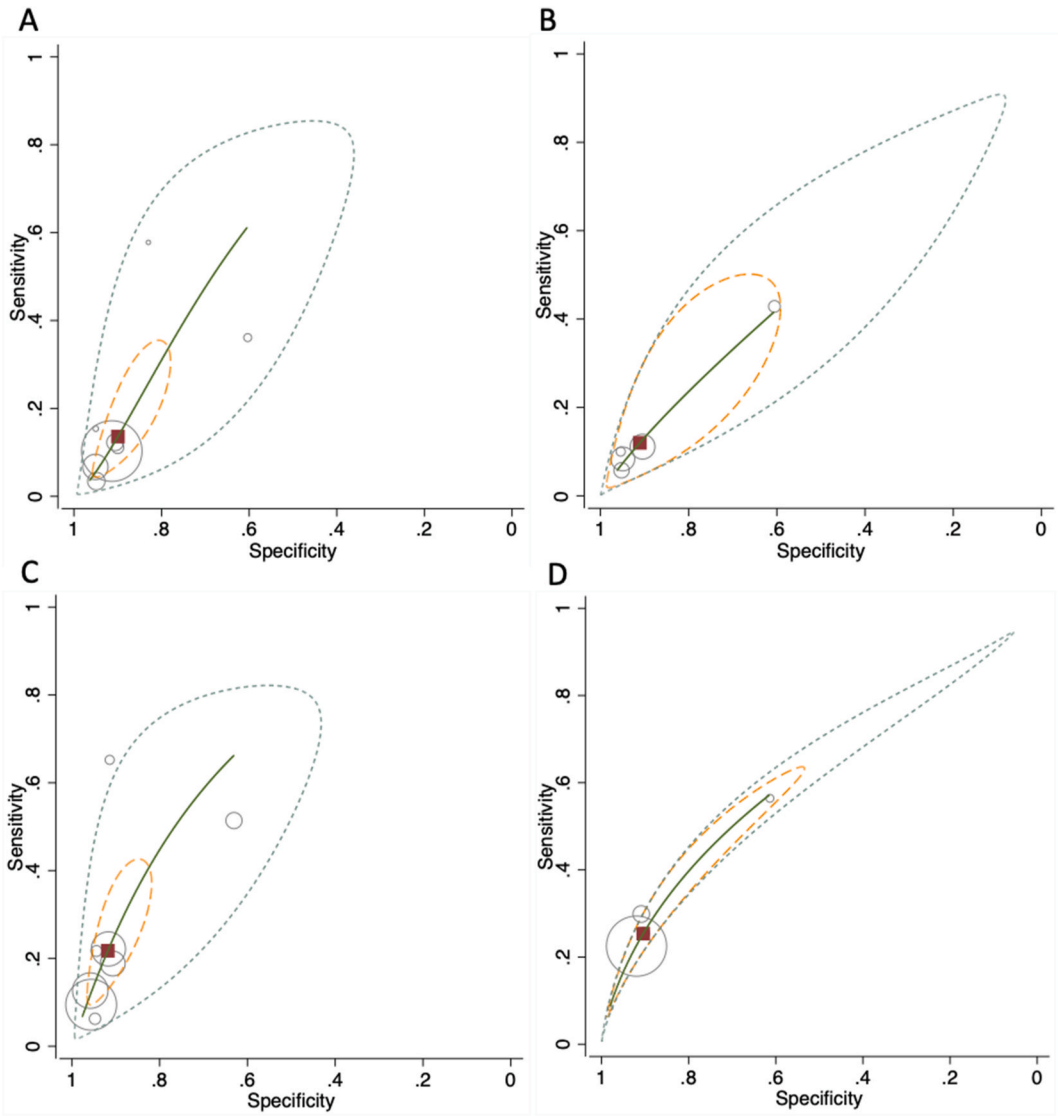


Fig. 3. Summary ROC curves for the elevated UA Doppler at predicting: A. NICU admission, B. Neonatal Metabolic acidosis, C. SGA (<10th centile), D. Severe SGA (<3rd centile).

4. Discussion

The main finding of this study was that in an unselected population a high resistance pattern of umbilical artery Doppler has moderate predictive accuracy for detecting SGA and severely SGA infants but it does not predict neonatal morbidity at term. The results were very similar in both the POP study and the meta-analysis. However, the association was slightly stronger for severe SGA in the POP study compared to the meta-analysis. The outcome of SGA is used as a proxy for FGR but it is recognised that only a proportion of SGA infants are small due to FGR. The lower the threshold for defining SGA, the higher the proportion of cases which are truly FGR. Hence, the stronger association with severe SGA is most likely explained by a true association between high resistance patterns of UA Doppler and FGR due to placental dysfunction.

The similar associations between the POP study and the meta-analysis suggestes that the findings were generalizable. Of all the studies evaluated, only the POP study blinded both the Doppler result and fetal biometry. The failure to blind studies could lead to bias. First, by revealing the results the clinicians might decide to intervene which then improves the outcome of the pregnancy. This could hide a true association between an investigation and an adverse outcome. However, revealing the result could also lead to a non-informative test being wrongly identified as predictive of adverse outcome. The decision to deliver the fetus preterm or at early term based on an abnormal result, can cause iatrogenic morbidity. Hence a non-informative test could appear to be associated with adverse neonatal outcome which was, in fact, iatrogenic. Moreover, there is the risk of ascertainment bias. For example, a study could show an association between an abnormal UA Doppler and Cesarean section for fetal distress because the clinicians had a lower threshold to deliver the infant by Cesarean section based on the prior knowledge of the abnormal ultrasound finding.

In this meta-analysis we found no association between abnormal UA Doppler and neonatal morbidity. This is likely explained by two reasons. First, abnormal UA Doppler is present in a minority of term SGA infants. This study showed that about 1 in 5 of the SGA infants born below the 10th birthweight centile and 1 in 4 of those born below the 3rd birthweight centile had abnormal UA Doppler. Second, only a small percentage of overall morbidity at term is associated with abnormal fetal growth. For example, previous studies on perinatal death at term demonstrated that only 1 in 3 stillbirths at term is associated with abnormal fetal growth [27]. This association would likely be even weaker for other outcomes such as NICU admission which includes morbidity for various reasons not related to the fetal size, such as neonatal infection. It is plausible that elevated UA Doppler would be

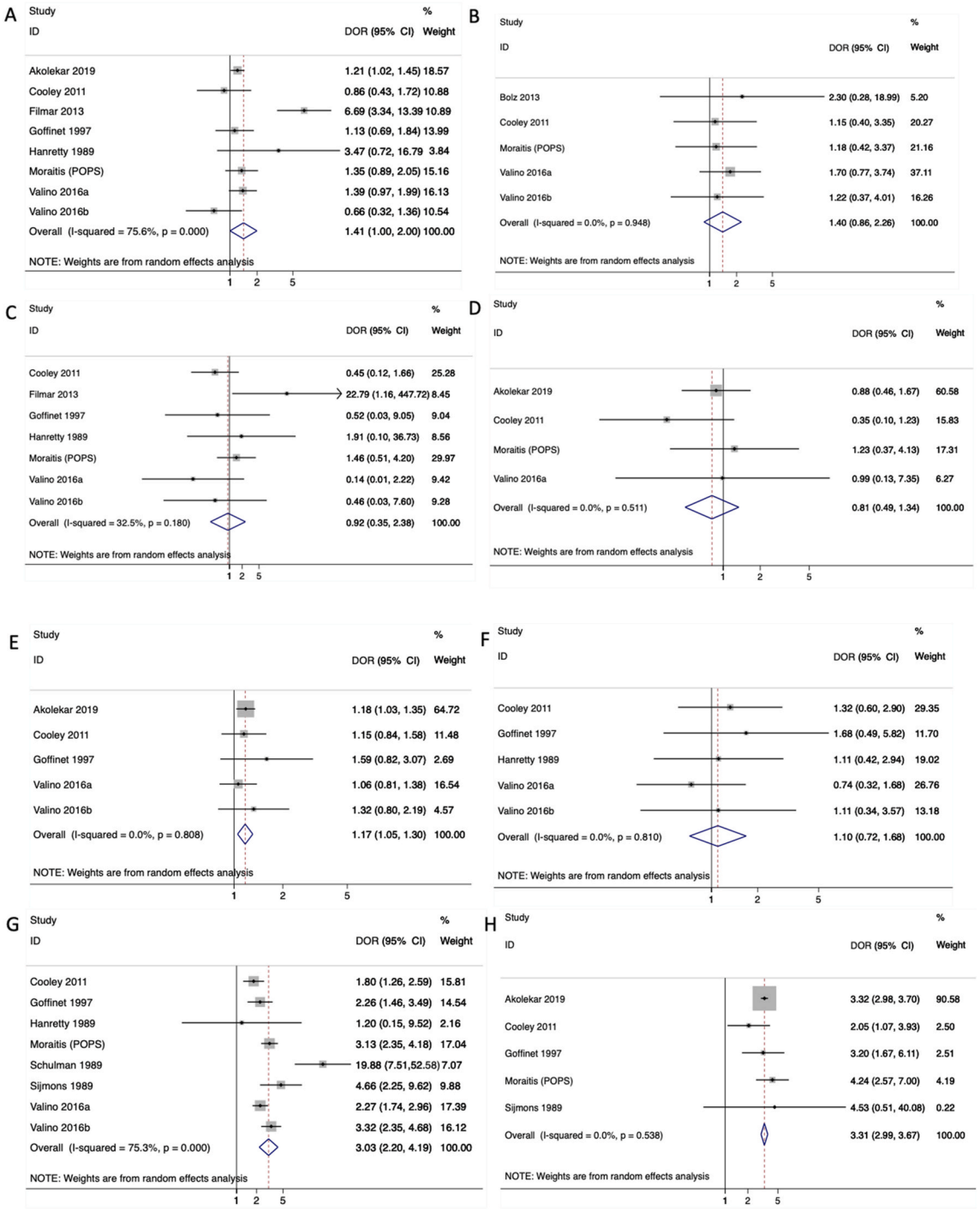


Fig. 4. Meta-analysis of DORs of elevated UA Doppler at predicting: A. NICU admission, B. Neonatal metabolic acidosis, C. 5-minute Apgar score <7, D. Severe adverse perinatal outcome, E. Cesarean section for fetal distress, F. Pre-eclampsia, G. SGA (<10th centile), H. Severe SGA (<3rd centile).

more strongly predictive of adverse neonatal outcome in fetuses which were actually SGA and this has been confirmed in a previous analysis of the POP study [6].

A recent systematic review [28] has also shown that there is considerable heterogeneity in the methodological quality of ultrasound studies on which reference ranges for UA Doppler indices are based. These differences may at least partly explain some of the discrepancies seen in perinatal outcomes following an index above the 90th centile - essentially, what is a normal index on one reference range may be abnormal in another; thus, the review showed clinical cut-offs varied significantly and could lead to important differences in clinical management, demonstrating that up to 40% of fetuses may be classified as having an abnormal result by using one chart rather than another [29].

Given that UA Doppler appears to be moderately predictive of FGR in low risk women it might be regarded as surprising that the RCTs of its use as a screening test failed to demonstrate any benefit. However, a previous analysis of required sample sizes of screening and intervention to prevent stillbirth demonstrated that, even if a test had a positive LR of 5 for perinatal death, and was observed in 5% of women, and even if the test was coupled to an intervention that reduced the risk of perinatal death by 50%, an RCT of screen versus no screen would need to recruit

Table 2
Diagnostic performance of UA PI > 90th centile at predicting adverse pregnancy outcome in the POP study (N = 3615).

Outcome	True Positive/False Positive	True Negative/False Negative	Sensitivity (95% CI)	Specificity (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)	Diagnostic Odds Ratio (95% CI)
SGA <10th centile	72/274	3016/253	22.2% (17.6–26.7%)	91.7% (90.7–92.6%)	2.66 (2.11–3.36)	0.85 (0.80–0.90)	3.13 (2.35–4.18)
SGA <3rd centile	23/323	3215/54	29.9% (19.6–40.1%)	90.9% (89.9–91.8%)	3.27 (2.29–4.68)	0.77 (0.67–0.89)	4.24 (2.57–7.00)
Any neonatal morbidity[a]	32/314	3045/224	12.5% (8.4–16.6%)	90.7% (89.7–91.6%)	1.34 (0.95–1.88)	0.97 (0.95–1.01)	1.39 (0.94–2.04)
NICU admission	27/319	3076/193	12.3% (7.9–16.6%)	90.6% (89.6–91.6%)	1.31 (0.90–1.89)	0.97 (0.92–1.02)	1.35 (0.89–2.05)
5-min Apgar score <7	4/342	3243/26	13.3% (1.2–25.5%)	90.5% (89.5–91.4%)	1.40 (0.56–3.50)	0.96 (0.83–1.10)	1.46 (0.51–4.20)
Metabolic acidosis	4/342	3237/32	11.1% (0.8–21.4%)	90.4% (89.5–91.4%)	1.16 (0.46–2.95)	0.98 (0.88–1.10)	1.18 (0.42–3.37)
Severe neonatal morbidity[b]	3/343	3246/23	11.5% (0.7–23.8%)	90.4% (89.5–91.4%)	1.21 (0.41–3.52)	0.98 (0.85–1.12)	1.23 (0.37–4.13)

SGA, Small for gestational age; LR, Likelihood ratio; CI, Confidence intervals.

[a] Any neonatal morbidity was defined as ≥1 of the following: a 5 min Apgar score less than 7, delivery with metabolic acidosis (defined as a cord blood pH < 7.1 and a base deficit of >10 mmol/L) or admission to the neonatal unit at term (defined as admission <48 h after birth at ≥37 weeks gestational age and discharge ≥48 h after admission).

[b] Severe neonatal morbidity 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 >12 mmol/L.

~300,000 to achieve 90% power see Supplementary Figure 10 in Flenady et al., 2016 [30]. Thus, the Cochrane meta-analysis of low-risk pregnancies is significantly underpowered to identify a reduction in perinatal death [31].

However, despite the lack of evidence, the policy of universal third-trimester ultrasound screening has been applied locally and nationally. This can have unintentional consequences as shown in France, where this policy resulted in a significantly higher risk of provider-initiated preterm birth, neonatal morbidity and obstetric intervention in those pregnancies that were wrongly identified as SGA (false positives) compared to those that were truly SGA but were missed by the screening (false negatives) [32]. These results could partly be explained by the gestational age at screening which was about 32 weeks' gestation. Offering routine ultrasound screening at around 36 weeks' gestation with a policy of induction of labor (IOL) for the screen positives reduces the risk of iatrogenic herm related to prematurity. However, early term deliveries (between 37 + 0 and 38 + 6 weeks' gestation) are still at higher risk of intrapartum asphyxia [33], respiratory distress, prolonged hospitalisation [34] and also higher risk of educational needs later in life [35].

Some recent studies have combined the UA Doppler with the middle-cerebral artery (MCA) Doppler to create the cerebro-placental ratio (CPR) which have been reported to increase the detection of adverse pregnancy outcomes including emergency Cesarean section [36]. A recent systematic review [37] showed moderate accuracy of the CPR for predicting perinatal death and low accuracy for neonatal morbidity in pregnancies with suspected FGR. There is no clear evidence about the use of CPR in low risk populations. In the POP study, the MCA Doppler was not measured so we were unable to investigate this in our population. However, a recent large cohort study of unselected women found that the CPR had poor predicition of adverse neonatal outcome [15].

In conclusion, a high resistance pattern of UA Doppler is moderately predictive of the risk of delivering an SGA infant. The strength of prediction was similar using a blinded 36wkGA scan in unselected nulliparous women in the POP study as it was in a systematic review including the rest of the eligible literature.

Funding

The work was supported by the NIHR Health Technology Assessment (grant 15/105/01). The funding body had no role in any aspect of the conduct, analysis or presentation of this study. The POP study was funded by the NIHR Cambridge Biomedical Research Centre, and supported by the NIHR Cambridge Clinical Research Facility. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Declaration of competing interest

Alexandros Moraitis, Thomas Bainton, Peter Brocklehurst, Jim Thornton, and Stephen Robson have no interest to declare. Ulla Sovio reports grants from NIHR Cambridge Biomedical Research Centre during the conduct of the study. Gordon Smith reports grants and personal fees from GlaxoSmithKline Research and Development Ltd, grants from Sera Prognostics Inc, non-financial support from Illumina Inc, and personal fees from Roche Diagnostics Ltd, outside the submitted work. In addition, Gordon Smith and Ulla Sovio have a patent in preparation for a novel predictive test for fetal size pending. Aris Papageorghiou report personal fees from educational events/lectures, clinical services in the private sector and from Consultancy via Oxford University Innovation, royalties from published works, and editorial work for UOG and BJOG, outside the submitted work.

Acknowledgement

GE Healthcare (Fairfield, CT, USA) donated two Voluson I ultrasound systems for the POP study. This study was also supported by the NIHR Cambridge Clinical Research Facility, where all visits at about 20, 28, and 36 weeks took place. 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.placenta.2021.03.011.

References

[1] R.K. Morris, G. Malin, S.C. Robson, J. Kleijnen, J. Zamora, K.S. Khan, Fetal umbilical artery Doppler to predict compromise of fetal/neonatal wellbeing in a high-risk population: systematic review and bivariate meta-analysis. *Ultrasound in obstetrics & gynecology*, the official journal of the International Society of Ultrasound in Obstetrics and Gynecology 37 (2011) 135–142.

[2] Z. Alfirevic, T. Stampalija, T. Dowswell, Fetal and umbilical Doppler ultrasound in high-risk pregnancies, Cochrane Database Syst. Rev. 6 (2017), Cd007529.

[3] Z. Alfirevic, T. Stampalija, N. Medley, Fetal and umbilical Doppler ultrasound in normal pregnancy, Cochrane Database Syst. Rev. (2015), https://doi.org/10.1002/14651858.CD001450.pub4. Cd001450.

[4] G.C. Smith, Researching new methods of screening for adverse pregnancy outcome: lessons from pre-eclampsia, PLoS Med. 9 (2012), e1001274.

[5] G. Smith, Should we implement universal screening with late pregnancy ultrasound to prevent stillbirth? BJOG An Int. J. Obstet. Gynaecol. 125 (2018) 101–103.

[6] U. Sovio, I.R. White, A. Dacey, D. Pasupathy, G.C.S. Smith, 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) 386 (2015) 2089–2097.

[7] P.F. Whiting, A.W. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, M. M. Leeflang, J.A. Sterne, P.M. Bossuyt, QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, Ann. Intern. Med. 155 (2011) 529–536.

[8] D. Pasupathy, A. Dacey, E. Cook, D.S. Charnock-Jones, I.R. White, G.C. Smith, Study protocol. A prospective cohort study of unselected primiparous women: the pregnancy outcome prediction study, BMC Pregnancy Childbirth 8 (2008) 51.

[9] J.V. Freeman, T.J. Cole, S. Chinn, P.R. Jones, E.M. White, M.A. Preece, Cross sectional stature and weight reference curves for the UK, 1990, Arch. Dis. Child. 73 (1995) 17–24.

[10] D.L. Simel, G.P. Samsa, D.B. Matchar, Likelihood ratios with confidence: sample size estimation for diagnostic test studies, J. Clin. Epidemiol. 44 (1991) 763–770.

[11] C.M. Rutter, C.A. Gatsonis, A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations, Stat. Med. 20 (2001) 2865–2884.

[12] J.B. Reitsma, A.S. Glas, A.W. Rutjes, R.J. Scholten, P.M. Bossuyt, A.H. Zwinderman, Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews, J. Clin. Epidemiol. 58 (2005) 982–990.

[13] J.J. Deeks, Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests, BMJ (Clinical research ed) 323 (2001) 157–162.

[14] J.J. Deeks, P. Macaskill, L. Irwig, The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed, J. Clin. Epidemiol. 58 (2005) 882–893.

[15] R. Akolekar, A. Ciobanu, E. Zingler, A. Syngelaki, K.H. Nicolaides, Routine assessment of cerebroplacental ratio at 35-37 weeks' gestation in the prediction of adverse perinatal outcome, Am. J. Obstet. Gynecol. (2019), https://doi.org/10.1016/j.ajog.2019.03.002.

[16] N. Bolz, K.D. Kalache, H. Proquitte, T. Slowinski, J.P. Hartung, W. Henrich, C. Bamberg, Value of Doppler sonography near term: can umbilical and uterine artery indices in low-risk pregnancies predict perinatal outcome? J. Perinat. Med. 41 (2013) 165–170.

[17] S.M. Cooley, J.C. Donnelly, T. Walsh, C. MacMahon, J. Gillan, M.P. Geary, The impact of umbilical and uterine artery Doppler indices on antenatal course, labor and delivery in a low-risk primigravid population, J. Perinat. Med. 39 (2011) 143–149.

[18] G. Filmar, G. Panagopoulos, V. Minior, Y. Barnhard, M.Y. Divon, Elevated umbilical artery systolic/diastolic ratio in the absence of fetal growth restriction, Arch. Gynecol. Obstet. 288 (2013) 279–285.

[19] R.L. Fischer, K.A. Kuhlman, R. Depp, R.J. Wapner, Doppler evaluation of umbilical and uterine-arcuate arteries in the postdates pregnancy, Obstet. Gynecol. 78 (1991) 363–368.

[20] F. Goffinet, J. Paris, N. Heim, I. Nisand, G. Breart, Predictive value of Doppler umbilical artery velocimetry in a low risk population with normal fetal biometry. A prospective study of 2016 women, Eur. J. Obstet. Gynecol. Reprod. Biol. 71 (1997) 11–19.

[21] K.P. Hanretty, M.H. Primrose, J.P. Neilson, M.J. Whittle, Pregnancy screening by Doppler uteroplacental and umbilical artery waveforms, Br. J. Obstet. Gynaecol. 96 (1989) 1163–1167.

[22] H. Schulman, D. Winter, G. Farmakides, J. Ducey, E. Guzman, A. Coury, B. Penny, Pregnancy surveillance with Doppler velocimetry of uterine and umbilical arteries, Am. J. Obstet. Gynecol. 160 (1989) 192–196.

[23] E.A. Sijmons, P.J. Reuwer, E. van Beek, H.W. Bruinse, The validity of screening for small-for-gestational-age and low-weight-for-length infants by Doppler ultrasound, Br. J. Obstet. Gynaecol. 96 (1989) 557–561.

[24] N. Valino, G. Giunta, D.M. Gallo, R. Akolekar, K.H. Nicolaides, Biophysical and biochemical markers at 30-34 weeks' gestation in the prediction of adverse perinatal outcome. *Ultrasound in obstetrics & gynecology*, the official journal of the International Society of Ultrasound in Obstetrics and Gynecology 47 (2016) 194–202.

[25] N. Valino, G. Giunta, D.M. Gallo, R. Akolekar, K.H. Nicolaides, Biophysical and biochemical markers at 35-37 weeks' gestation in the prediction of adverse perinatal outcome. *Ultrasound in obstetrics & gynecology*, the official journal of the International Society of Ultrasound in Obstetrics and Gynecology 47 (2016) 203–209.

[26] Z. Weiner, A. Reichler, M. Zlozover, A. Mendelson, I. Thaler, The value of Doppler ultrasonography in prolonged pregnancies, Eur. J. Obstet. Gynecol. Reprod. Biol. 48 (1993) 93–97.

[27] A.A. Moraitis, A.M. Wood, M. Fleming, G.C. Smith, Birth weight percentile and the risk of term perinatal death, Obstet. Gynecol. 124 (2014) 274–283.

[28] D. Oros, S. Ruiz-Martinez, E. Staines-Urias, A. Conde-Agudelo, J. Villar, E. Fabre, A. T. Papageorghiou, Reference ranges for Doppler indices of umbilical and fetal middle cerebral arteries and cerebroplacental ratio: systematic review, Ultrasound Obstet. Gynecol. : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology 53 (2019) 454–464.

[29] S. Ruiz-Martinez, A.T. Papageorghiou, E. Staines-Urias, J. Villar, R.G. de Aguero, D. Oros, Clinical impact of Doppler reference charts to manage fetal growth restriction: need for standardization. Ultrasound in obstetrics & gynecology, the official journal of the International Society of Ultrasound in Obstetrics and Gynecology (2019), https://doi.org/10.1002/uog.20380.

[30] V. Flenady, A.M. Wojcieszek, P. Middleton, D. Ellwood, J.J. Erwich, M. Coory, T. Y. Khong, R.M. Silver, G.C. Smith, F.M. Boyle, J.E. Lawn, H. Blencowe, S. H. Leisher, M.M. Gross, D. Horey, L. Farrales, F. Bloomfield, L. McCowan, S. J. Brown, K.S. Joseph, J. Zeitlin, H.E. Reinebrant, C. Ravaldi, A. Vannacci, J. Cassidy, P. Cassidy, C. Farquhar, E. Wallace, D. Siassakos, A.E. Heazell, C. Storey, L. Sadler, S. Petersen, J.F. Froen, R.L. Goldenberg, Stillbirths: recall to action in high-income countries, Lancet (London, England) 387 (2016) 691–702.

[31] L. Bricker, N. Medley, J.J. Pratt, Routine ultrasound in late pregnancy (after 24 weeks' gestation), Cochrane Database Syst. Rev. (2015), https://doi.org/10.1002/14651858.CD001451.pub4. Cd001451.

[32] I. Monier, B. Blondel, A. Ego, M. Kaminiski, F. Goffinet, J. Zeitlin, Poor effectiveness of antenatal detection of fetal growth restriction and consequences for obstetric management and neonatal outcomes: a French national study, BJOG An Int. J. Obstet. Gynaecol. 122 (2015) 518–527.

[33] M. Eskes, S. Ensing, F. Groenendaal, A. Abu-Hanna, A. Ravelli, The risk of intrapartum/neonatal mortality and morbidity following birth at 37 weeks of gestation: a nationwide cohort study, BJOG An Int. J. Obstet. Gynaecol. 126 (2019) 1252–1257.

[34] C.Y. Spong, B.M. Mercer, M. D'Alton, S. Kilpatrick, S. Blackwell, G. Saade, Timing of indicated late-preterm and early-term birth, Obstet. Gynecol. 118 (2011) 323–333.

[35] D.F. MacKay, G.C. Smith, R. Dobbie, J.P. Pell, Gestational age at delivery and special educational need: retrospective cohort study of 407,503 schoolchildren, PLoS Med. 7 (2010), e1000289.

[36] S. Sabdia, R.M. Greer, T. Prior, S. Kumar, Predicting intrapartum fetal compromise using the fetal cerebro-umbilical ratio, Placenta 36 (2015) 594–598.

[37] A. Conde-Agudelo, J. Villar, S.H. Kennedy, A.T. Papageorghiou, Predictive accuracy of cerebroplacental ratio for adverse perinatal and neurodevelopmental outcomes in suspected fetal growth restriction: systematic review and meta-analysis, Ultrasound Obstet. Gynecol. : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology 52 (2018) 430–441.

PLOS MEDICINE

RESEARCH ARTICLE

Screening for breech presentation using universal late-pregnancy ultrasonography: A prospective cohort study and cost effectiveness analysis

David Wastlund[1,2]*, **Alexandros A. Moraitis**[3], **Alison Dacey**[3], **Ulla Sovio**[3], **Edward C. F. Wilson**[1,4], **Gordon C. S. Smith**[3]

1 Cambridge Centre for Health Services Research, Cambridge Institute of Public Health, Cambridge, United Kingdom, **2** The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom, **3** Department of Obstetrics and Gynaecology, University of Cambridge, NIHR Cambridge Comprehensive Biomedical Research Centre, Cambridge, United Kingdom, **4** Health Economics Group, Norwich Medical School, University of East Anglia, Norwich, United Kingdom

* Jdw78@medschl.cam.ac.uk

Citation: Wastlund D, Moraitis AA, Dacey A, Sovio U, Wilson ECF, Smith GCS (2019) Screening for breech presentation using universal late-pregnancy ultrasonography: A prospective cohort study and cost effectiveness analysis. PLoS Med 16(4): e1002778. https://doi.org/10.1371/journal.pmed.1002778

Academic Editor: Jenny E. Myers, University of Manchester, UNITED KINGDOM

Received: August 30, 2018

Accepted: March 11, 2019

Published: April 16, 2019

Data Availability Statement: The terms of the ethical permission for the POP study do not allow publication of individual patient level data. Requests for access to patient level data will usually require a Data Transfer Agreement, and should be made to Mrs Sheree Green-Molloy at the Department of Obstetrics and Gynaecology, Cambridge University, UK (paoandghod@medschl.cam.ac.uk).

Abstract

Background

Despite the relative ease with which breech presentation can be identified through ultrasound screening, the assessment of foetal presentation at term is often based on clinical examination only. Due to limitations in this approach, many women present in labour with an undiagnosed breech presentation, with increased risk of foetal morbidity and mortality. This study sought to determine the cost effectiveness of universal ultrasound scanning for breech presentation near term (36 weeks of gestational age [wkGA]) in nulliparous women.

Methods and findings

The Pregnancy Outcome Prediction (POP) study was a prospective cohort study between January 14, 2008 and July 31, 2012, including 3,879 nulliparous women who attended for a research screening ultrasound examination at 36 wkGA. Foetal presentation was assessed and compared for the groups with and without a clinically indicated ultrasound. Where breech presentation was detected, an external cephalic version (ECV) was routinely offered. If the ECV was unsuccessful or not performed, the women were offered either planned cesarean section at 39 weeks or attempted vaginal breech delivery. To compare the likelihood of different mode of deliveries and associated long-term health outcomes for universal ultrasound to current practice, a probabilistic economic simulation model was constructed. Parameter values were obtained from the POP study, and costs were mainly obtained from the English National Health Service (NHS). One hundred seventy-nine out of 3,879 women (4.6%) were diagnosed with breech presentation at 36 weeks. For most women (96), there had been no prior suspicion of noncephalic presentation. ECV was attempted for 84 (46.9%) women and was successful in 12 (success rate: 14.3%). Overall, 19 of the 179 women delivered vaginally (10.6%), 110 delivered by elective cesarean section (ELCS)

Funding: This study was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme, grant number 15/105/01. EW is part funded by the NIHR Cambridge Biomedical Research Centre. US is funded by the NIHR Cambridge Comprehensive Biomedical Research Centre. The views expressed here are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health (https://www.nihr.ac.uk/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: GS 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. GS is an Academic Editor on PLOS Medicine's editorial board. DW, AM, AD, US, and EW declare no competing interests.

Abbreviations: AFI, amniotic fluid index; AGA, appropriate for gestational age; CrI, credibility interval; ECV, external cephalic version; ELCS, elective cesarean section; EMCS, emergency cesarean section; FTE, full-time education; HCHS, Hospital and Community Health Services; ICER, incremental cost effectiveness ratio; IMD, Index of Multiple Deprivation; LGA, large for gestational age; MOD, mode of delivery; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; POP, Pregnancy Outcome Prediction; PSA, probabilistic sensitivity analysis; QALY, quality-adjusted life-year; SGA, small for gestational age; wkGA, weeks of gestational age; RCOG, Royal College of Obstetricians and Gynaecologists; SRC, spontaneous reversion to cephalic; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

(61.5%) and 50 delivered by emergency cesarean section (EMCS) (27.9%). There were no women with undiagnosed breech presentation in labour in the entire cohort. On average, 40 scans were needed per detection of a previously undiagnosed breech presentation. The economic analysis indicated that, compared to current practice, universal late-pregnancy ultrasound would identify around 14,826 otherwise undiagnosed breech presentations across England annually. It would also reduce EMCS and vaginal breech deliveries by 0.7 and 1.0 percentage points, respectively: around 4,196 and 6,061 deliveries across England annually. Universal ultrasound would also prevent 7.89 neonatal mortalities annually. The strategy would be cost effective if foetal presentation could be assessed for £19.80 or less per woman. Limitations to this study included that foetal presentation was revealed to all women and that the health economic analysis may be altered by parity.

Conclusions

According to our estimates, universal late pregnancy ultrasound in nulliparous women (1) would virtually eliminate undiagnosed breech presentation, (2) would be expected to reduce foetal mortality in breech presentation, and (3) would be cost effective if foetal presentation could be assessed for less than £19.80 per woman.

Author summary

Why was this study done?

- Risks of complications at delivery are higher for babies that are in a breech position, but sometimes breech presentation is not discovered until the time of birth.
- Ultrasound screening could be used to detect breech presentation before birth and lower the risk of complications but would be associated with additional costs.
- It is uncertain if offering ultrasound screening to every pregnancy is cost effective.

What did the researchers do and find?

- This study recorded the birth outcomes of pregnancies that were all screened using ultrasound.
- Economic modelling and simulation was used to compare these outcomes with those if ultrasound screening had not been used.
- Modelling demonstrated that ultrasound screening would lower the risk of breech delivery and, as a result, reduce emergency cesarean sections and the baby's risk of death.

What do these findings mean?

- Offering ultrasound screening to every pregnancy would improve the health of mothers and babies nationwide.

- Whether the health improvements are enough to justify the increased cost of ultrasound screening is still uncertain, mainly because the cost of ultrasound screening for presentation alone is unknown.
- If ultrasound screening could be provided sufficiently inexpensively, for example, by being used during standard midwife appointments, routinely offering ultrasound screening would be worthwhile.

Introduction

Undiagnosed breech presentation in labour increases the risk of perinatal morbidity and mortality and represents a challenge for obstetric management. The incidence of breech presentation at term is around 3%–4% [1–3], and fewer than 10% of foetuses who are breech at term revert spontaneously to a vertex presentation [4]. Although breech presentation is easy to detect through ultrasound screening, many women go into labour with an undetected breech presentation [5]. The majority of these women will deliver through emergency cesarean section (EMCS), which has high costs and increased risk of morbidity and mortality for both mother and child.

In current practice, foetal presentation is routinely assessed by palpation of the maternal abdomen by a midwife, obstetrician, or general practitioner. The sensitivity of abdominal palpation varies between studies (range: 57%–70%) and depends on the skill and experience of the practitioner [6,7]. There is currently no guidance on what is considered an acceptable false negative rate when screening for breech presentation using abdominal palpation. In contrast, ultrasound examination provides a quick and safe method of accurately identifying foetal presentation.

Effective interventions exist for the care of women who have breech presentation diagnosed near term. The Royal College of Obstetricians and Gynaecologists recommends 'that all women with an uncomplicated breech presentation at term should be offered external cephalic version (ECV)' [2]. The rationale for this is to reduce the incidence of breech presentation at term and avoid the risks of vaginal breech birth or cesarean section. The success rate of ECV is considered to be approximately 50% [2,8,9], but it differs greatly between nulliparous and parous women (34% and 66%, respectively) [9]. ECV is overall safe, with less than 1% risk to the foetus and even smaller risk to the mother [10]; despite this, a significant number of women decline ECV for various reasons [11]. Should ECV be declined or fail, generally women are offered delivery by planned (elective) cesarean section, as there is level 1 evidence of reduced risk of perinatal death and severe morbidity compared with attempting vaginal breech birth, and it is also associated with lower costs [3,12,13]. However, some women may still opt for an attempt at vaginal breech birth if they prioritise nonintervention over managing the relatively small absolute risks of a severe adverse event [1,14].

We sought to assess the cost effectiveness of universal late-pregnancy ultrasound presentation scans for nulliparous women. We used data from the Pregnancy Outcome Prediction (POP) study, a prospective cohort study of >4,000 nulliparous women, which included an ultrasound scan at 36 weeks of gestational age (wkGA) [15]. Here, we report the outcomes for pregnant nulliparous women with breech presentation in the study and use these data to perform a cost effectiveness analysis of universal ultrasound as a screening test for breech presentation.

Methods

Study design

The POP study was a prospective cohort study of nulliparous women conducted at the Rosie Hospital, Cambridge (United Kingdom) between January 14, 2008 and July 31, 2012, and the study has been described in detail elsewhere [15–17]. Ethical approval for the study was obtained from the Cambridgeshire 2 Research Ethics Committee (reference 07/H0308/163), and all participants provided informed consent in writing. Participation in the POP study involved serial phlebotomy and ultrasound at approximately 12 wkGA, 20 wkGA, 28 wkGA, and 36 wkGA [16]. The outcome of pregnancy was obtained by individual review of all case records by research midwives and by linkage to the hospital's electronic databases of ultrasonography, biochemical testing, delivery data, and neonatal care data. The research ultrasound at 36 wkGA was performed by sonographers and included presentation, biometry, uteroplacental Doppler, and placental location. The ultrasound findings were blinded except in cases of breech presentation, low lying placenta, or foetal concerns such as newly diagnosed foetal anomaly and an amniotic fluid index (AFI) < 5 cm. This study was not prospectively defined in the POP study protocol paper [16] but required no further data collection.

If the foetus was in a breech presentation at 36 wkGA, women were counselled by a member of the medical team. In line with guidelines from the National Institute for Health and Care Excellence (NICE), ECV was routinely offered unless there was a clinical indication that contraindicated the procedure, e.g., reduced AFI (<5 cm) [18]. ECV was performed by 1 of 5 obstetric consultants in the unit between 36–38 wkGA, patients were scanned before the procedure to confirm presentation, and it was performed with ultrasound assessment; 0.25 mg terbutaline SC was given prior to the procedure at the discretion of the clinician. If women refused ECV or the procedure failed, the options of vaginal breech delivery and elective cesarean section (ELCS) were discussed and documented. The local guideline for management of breech presentation, including selection criteria for vaginal breech delivery, was based upon recommendations from the Royal College of Obstetricians and Gynaecologists (RCOG) [1]. We extracted information about ECV from case records that were individually reviewed by research midwives. Finally, we obtained delivery-related information from our hospital electronic database (Protos; iSoft, Banbury, UK).

Foetal outcomes included mode of delivery (MOD), birth weight, and gestational age at delivery. We used the UK population reference for birthweight, with the 10th and 90th percentile cut-offs for small and large for gestational age, respectively; the centiles were adjusted for sex and gestational age [19]. Maternal age was defined as age at recruitment. Smoking status, racial ancestry, alcohol consumption, and BMI were taken from data recorded at the booking assessment by the community midwife. Socioeconomic status was quantified using the Index of Multiple Deprivation (IMD) 2007, which is based on census data from the area in the mother's postcode [20]. Ethical approval for the study was obtained from the Cambridgeshire 2 Research Ethics Committee (reference 07/H0308/163), and all participants provided informed consent in writing.

This study is reported as per the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline.

Statistical analysis

Data are presented as median (interquartile range) or *n* (%), as appropriate. *P* values are reported for the difference between groups calculated using the two-sample Wilcox rank-sum (Mann–Whitney) test for continuous variables and the Pearson Chi-square test for categorical

variables, with trend tests when appropriate. Comparisons were performed using Stata (version 15.1). Missing values were included in the presentation of patient characteristics and outcomes but were excluded from the economic analysis and estimation of parameters.

Economic model and analysis

To evaluate the cost effectiveness of routinely offering late-pregnancy presentation scans, a decision-tree simulation model was constructed using R (version 3.4.1) [21–24]. The time horizon of the economic analysis was from the ultrasound scan (36 wkGA) to infant lifetime, and costs were from the perspective of the English National Health Service (NHS). Costs for modes of delivery were obtained from NHS reference costs [25]; since these do not list a separate cost for vaginal breech delivery, we assumed that the cost ratio between vaginal breech and ELCS deliveries was the same as in another study (see Supporting information, S1 Text) [12].

The population of interest is unselected nulliparous women. The model compares the outcomes at birth for two strategies: 'universal ultrasound' and 'selective ultrasound' (Fig 1). For universal ultrasound, we assumed that all breech presentations at the time of scanning would be detected (i.e., assumed 100% sensitivity and specificity for the test). For selective ultrasound, the breech presentation was diagnosed either clinically (by abdominal palpation followed by ultrasound for confirmation) or as an incidental finding during a scan for a different indication. These assumptions were based upon current practice and derived from the POP study.

Compared to a standard antenatal ultrasound for which, typically, multiple measurements are made, an ultrasound scan for foetal presentation alone is technically simple. We theorised that such a scan could be provided by an attending midwife in conjunction with a standard antenatal visit in primary care, using basic ultrasound equipment. Since a specific unit cost for a scan for foetal presentation alone is not included in the national schedule of reference costs [25], we estimated the cost of ultrasound to include the midwife's time, the cost of equipment, and room. More details are presented in the Supporting information, S1 Text. The cost of ECV was obtained from James and colleagues [26] and converted to the 2017 price level using the Hospital and Community Health Services (HCHS) index [27]. The probability of ECV uptake and success rate as well as MOD were obtained from the POP study. All model inputs are presented in Table 1 and S1 Table, and the calculation of cost inputs is shown in Supporting information, S1 Text.

The end state of the decision tree was the MOD, which was either vaginal, ELCS, or EMCS. Delivery could be either cephalic or breech. EMCS could be either due to previously undiagnosed breech presentation or for other reasons. All cases of breech could spontaneously revert to cephalic presentation. However, we assumed the probability of this to be lower if ECV had been attempted and failed [28]. If ECV was successful, a reversion back to breech presentation was possible. It is currently unclear whether the probability of MOD varies depending on whether cephalic presentation is the result of successful ECV or spontaneous reversion [2,10,29–31], but we assumed that the probabilities differed.

Long-term health outcomes were modelled based upon the mortality risk associated with each MOD. The risk of neonatal mortality was taken from the RCOG guidelines. For breech presentation, these risks were 0.05% for delivery through ELCS and 0.20% for vaginal delivery. The risk of neonatal mortality for cephalic presentation with vaginal delivery was 0.10% [1]. There were no randomised clinical trials that allowed us to compare the outcomes of ELCS versus vaginal delivery for uncomplicated pregnancies with cephalic presentation; however, most observational studies found no significant difference in neonatal mortality and serious morbidity between the two modes [32–34]. For this reason, we assumed the mortality risk for

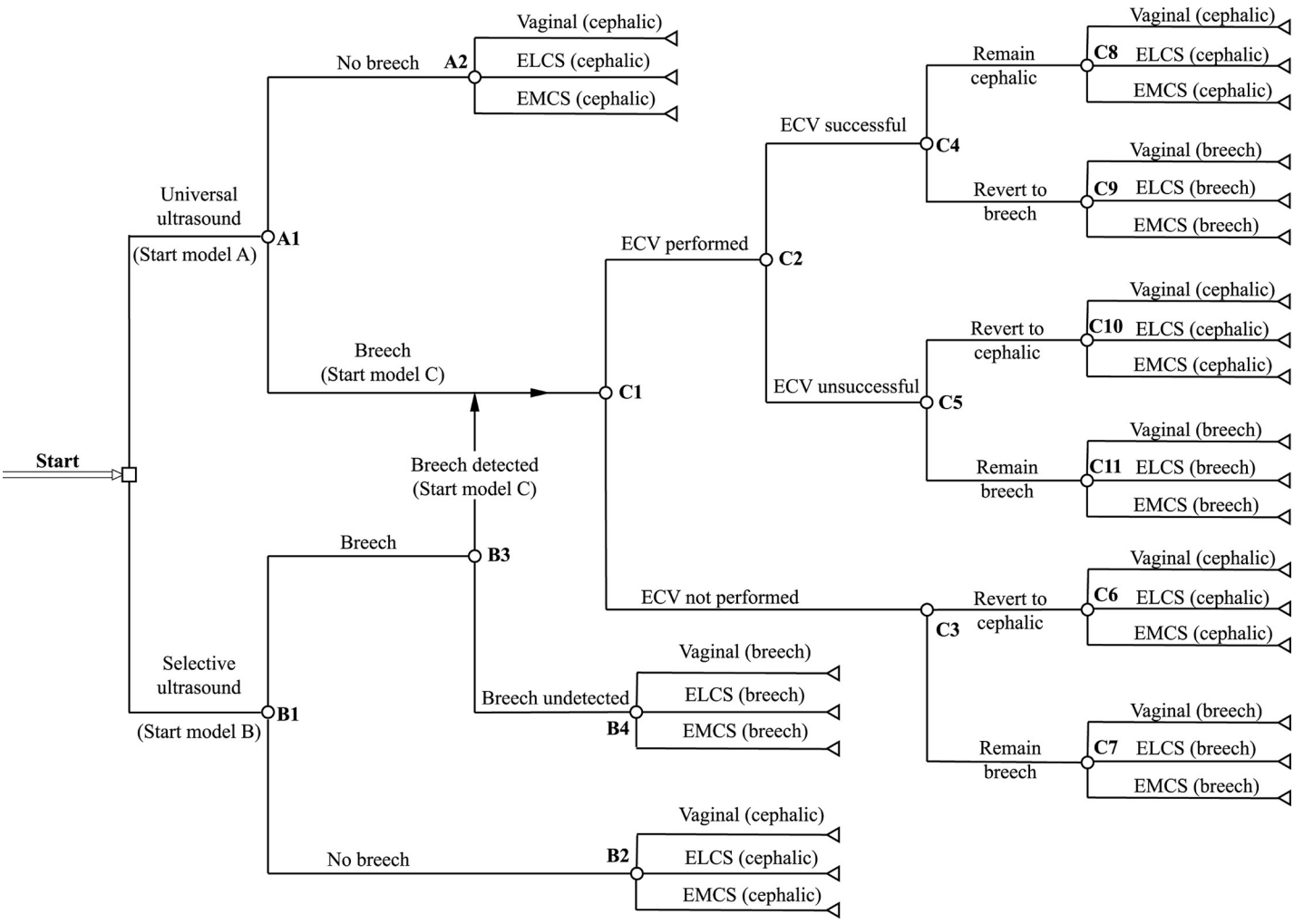


Fig 1. Simulation model structure. Structure of economic simulation model. 'Universal ultrasound' strategy starts in Model A, and patients with breech presentation enter Model C. 'Selective ultrasound', i.e., no routine ultrasound, starts in Model B, and only those with a detected breech presentation enter Model C. The letter–number codes for each node are equivalent to the codes in Table 1. ELCS, elective cesarean section; EMCS, emergency cesarean section.

https://doi.org/10.1371/journal.pmed.1002778.g001

cephalic vaginal and ELCS deliveries to be identical. We also assumed that EMCS would have the same mortality rate as ELCS, both for cephalic and breech deliveries. Studies have found that the MOD for breech presentation affects the risk of serious neonatal morbidity in the short term but not in the long term [1,3,35]. For this reason, we focused the economic analysis on the effect from mortality only. The average lifetime quality-adjusted life-years (QALYs) per member of the UK population was estimated using data on quality of life from Euroqol, weighted by longevity indexes from the Office for National Statistics (ONS) [36,37]. Using the annual discount rate of 3.5%, as recommended by NICE, the net present value for the average lifetime QALYs at birth was 24.3 [38].

The model was probabilistic, capturing how uncertainty in the input parameters affected the outputs by allowing each parameter to vary according to its distribution. Binary and multivariable outcomes were modelled using the beta and the Dirichlet distributions, respectively [39]. Probabilities of events were calculated from the POP study and presented in Table 1. On top of the probabilistic sensitivity analysis (PSA), the sensitivity of individual parameters was also explored through one-way sensitivity analyses modifying probabilities by +/− 1

Table 1. Inputs for costs and probabilities for the economic model.

Costs	Costs	Source				
Ultrasound scanning	20.7	Expert opinion*				
ECV	297.4	James et al. (2001) [26] †				
CV delivery	2,297.3	NHS Reference costs 2015–16 [25] ‡				
Elective cesarean delivery	3,438.1	NHS Reference costs 2015–16 [25] ‡				
Emergency cesarean delivery	4,553.4	NHS Reference costs 2015–16 [25] ‡				
VB delivery	3,999.7	Expert opinion*				
Probabilities	**Alpha**	**Beta**	**Mean**	**Node**	**Source**	
Breech prevalence at approximately 36 wkGA	179	3,700	0.046	A1 and B1	POP study	
ECV attempted	84	93	0.475	C1	POP study	
Detection without ultrasound	79	96	0.451	B3	POP study	
Successful ECV	12	72	0.143	C2	POP study	
SRC (ECV not attempted)	21	72	0.226	C3	POP study	
SRB	1	11	0.083	C4	POP study	
SRC (failed ECV)	3	127	0.023	C5	Ben-Meir and colleagues [28]§	
MOD	**CV**	**ELCS**	**EMCS**	**VB**	**Node**	**Source**
No breech	2,813	141	735	0	A2 and B2	POP study
Cephalic (successful ECV)	8	0	3	0	C8	POP study
Cephalic (spontaneous reversion)	11	1	9	0	C6 and C10	POP study
Breech (ECV not attempted)	0	52	20	0	C7	POP study
Breech (unsuccessful ECV)	0	54	18	0	C11	POP study
Breech (spontaneous reversion)	0	0	15	11	C9	Leung and colleagues [5]
Undetected breech	0	0	15	11	B4	Leung and colleagues [5]

Abbreviations: CV, cephalic vaginal; ELCS, elective cesarean section; EMCS, emergency cesarean section; MOD, mode of delivery; NHS, National Health Service; POP, Pregnancy Outcome Prediction; SRB, spontaneous reversion to breech; SRC, spontaneous reversion to cephalic; VB, vaginal breech.

Costs given per unit/episode. For probabilities, alpha represent case of event and beta case of no event. MOD shows input values for Dirichlet distribution. Node refers to the chance nodes in Fig 1.

*Details on how this value was estimated is provided as Supporting information, S1 Text.

†Cost for ECV (high staff cost), converted to 2017 price level using the HCHS index [27].

‡Weighted average of all complication levels (Total HRGs).

§Due to the small sample size for these parameters in the POP study, the model used inputs for MOD for undetected breech instead.

https://doi.org/10.1371/journal.pmed.1002778.t001

percentage point and costs by +/− £10 to see which parameters had the greatest impact on cost effectiveness estimates.

Total costs depended on the distribution of MOD, the number of expected mortalities, and the cost of ultrasound scanning and ECV. Nationwide costs for each screening strategy were calculated for 585,489 deliveries, i.e., the number of births in England from 2016–2017, assuming 92% occur after 36 wkGA [15,40]. Model parameters were sampled from their respective distributions in a PSA of 100,000 simulations for each strategy. To determine cost effectiveness, we used two different willingness-to-pay thresholds: £20,000 and £30,000 [38]. A copy of the model code is available from the corresponding author (EW) upon request.

Results

Recruitment to the POP study cohort is shown in Fig 2 and has been previously described [17]. Information about presentation at the 36-week scan was available for 3,879 women who delivered at the Rosie Hospital, Cambridge, UK; 179 of these had a breech presentation.

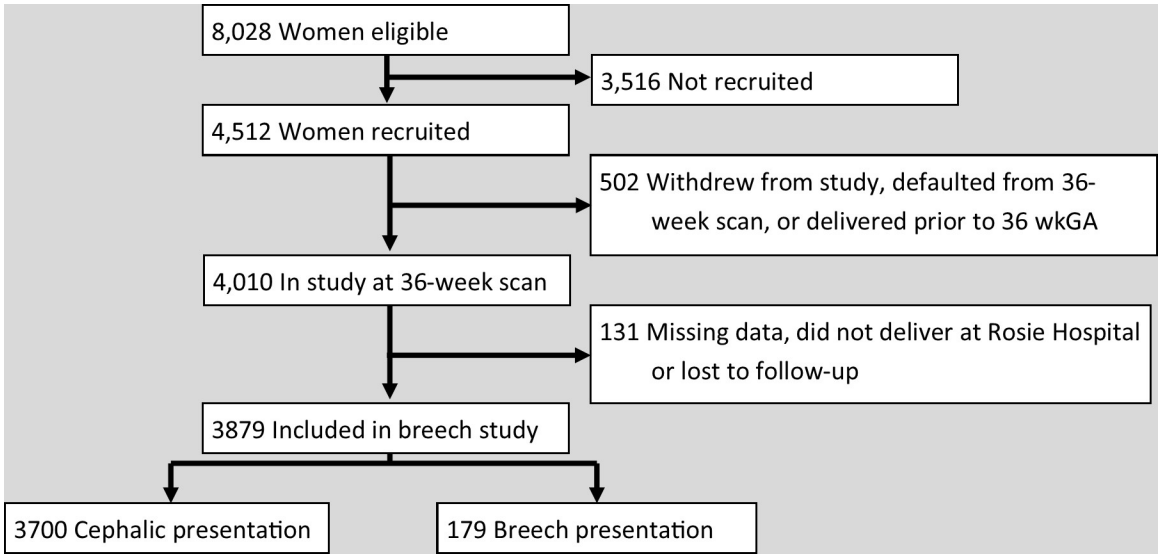


Fig 2. Patient recruitment. Schedule of patient recruitment in the POP study shown by foetal presentation. POP, Pregnancy Outcome Prediction.

https://doi.org/10.1371/journal.pmed.1002778.g002

We compared maternal and foetal characteristics of the 179 women with breech presentation at 36 weeks to the women with a cephalic presentation (Table 2). Women diagnosed with breech presentation were, on average, a year older than women with a cephalic presentation, but other maternal characteristics did not differ. The babies of women diagnosed breech were smaller and born earlier, but their birth weight centile and the proportions of small for gestational age (SGA) or large for gestational age (LGA) were not markedly different. There were no differences in maternal BMI between the groups. As expected, women with breech presentation were more likely to deliver by ELCS or EMCS.

Breech presentation was suspected before the 36-wkGA scan for 79 (44.1%) of the women with breech presentation through abdominal palpation by the midwife or doctor; out of these, 27 had a clinically indicated scan between 32–36 weeks in which the presentation was reported. For 96 women, the breech presentation was unsuspected before the 36-week scan. Information on suspected breech position was missing for 4 women. There were no differences in BMI between the 79 women with suspected breech and the 96 women misdiagnosed as cephalic prior to the scan (median BMI was 24 in both groups, Wilcoxon rank-sum test $P = 0.31$).

MOD by ECV status is shown in Table 3. ECV was performed for 84 women, declined by 45 women, and unsuitable for 23; contraindications included low AFI at screening (18 women), uterine abnormalities (2), and other reasons (3). For 25 women, an ECV was never performed despite consent; 17 babies turned spontaneously, 6 had reduced AFI on the day of the ECV, and 2 went into labour before ECV. When performed, ECV was successful for 12 women; in one case, the baby later reverted to breech presentation before delivery. Information on ECV uptake was missing for 2 women. Foetal presentation and ECV status in the structure of the economic model is shown in Supporting information, S1 Fig.

The results from the economic analysis are presented in Table 4. On average, universal ultrasound resulted in an absolute decrease in breech deliveries by 0.39%. It also led to fewer vaginal breech deliveries (absolute decrease by 1.04%) and overall EMCS deliveries (0.72%) than selective ultrasound but increased overall deliveries through ELCS (1.51%). Resulting from the more favourable distribution of MOD, the average risk of mortality fell by 0.0013%.

Table 2. Characteristics and delivery outcomes in the POP study by presentation at 36 weeks.

Characteristics	Breech (*N* = 179)	Cephalic (*N* = 3,700)	*P* value
Maternal			
Age (years)	31 (28–34)	30 (27–33)	0.002
Age stopped FTE (years)	21 (18–23)	21 (18–23)	0.19
Missing	5 (3%)	105 (3%)	
Racial ancestry			
White European	172 (96%)	3,437 (93%)	0.38
Missing	0 (0%)	66 (2%)	
Alcohol consumption	7 (4%)	172 (5%)	0.65
Missing	0 (0%)	1 (<0.1%)	
Smoker	4 (2%)	179 (5%)	0.11
BMI, kg/m^2	24 (22–27)	24 (22–27)	0.69
Missing	0 (0%)	1 (<0.1%)	
Deprivation quartile			0.08
1 (lowest)	46 (26%)	899 (24%)	
2	53 (30%)	873 (24%)	
3	39 (22%)	886 (24%)	
4 (highest)	33 (18%)	892 (24%)	
Missing	8 (4%)	150 (4%)	
Foetal or neonatal			
Female sex	96 (54%)	1,841 (50%)	0.31
Missing	0 (0%)	1 (<0.1%)	
Birth weight (grams)	3,310 (2,995–3,560)	3,445 (3,145–3,750)	<0.001
Gestational age (weeks)	39.1 (38.7–39.7)	40.4 (39.4–41.3)	<0.001
Birth weight centile	49 (25–70)	44 (24–66)	0.22
Birth weight centile category			0.32
SGA	12 (7%)	332 (9%)	
AGA	158 (88%)	3,199 (86%)	
LGA	9 (5%)	168 (5%)	
Missing	0 (0%)	1 (<0.1%)	
MOD			<0.001
Spontaneous vaginal cephalic	11 (6.1%)	1,885 (50.9%)	
Instrumental vaginal cephalic	8 (4.5%)	928 (25.1%)	
Elective cesarean section	110 (61.5%)	141 (3.8%)	
Emergency cesarean section	50 (27.9%)	735 (19.9%)	
Missing	0 (0%)	11 (0.3%)	

Abbreviations: AGA, appropriate for gestational age; FTE, full-time education; LGA, large for gestational age; MOD, mode of delivery; POP, Pregnancy Outcome Prediction; SGA, small for gestational age.

Statistics are presented as *n* (%) for binary outcomes and median (interquartile range) for continuous variables. The "Missing" category was not included in statistical tests. For variables without a "Missing" category, data were 100% complete. *P* values are reported for the difference between groups using the two-sample Wilcox rank-sum test for continuous variables and the Pearson Chi-square test for categorical variables, with trend test as appropriate (i.e., for deprivation quartile and birth weight centile category).

https://doi.org/10.1371/journal.pmed.1002778.t002

On average, 40 women had to be scanned to identify one previously unsuspected breech presentation (95% Credibility Interval [CrI]: 33 to 49); across England, this would mean that 14,826 (95% CrI: 12,048–17,883) unidentified breech presentations could be avoided annually.

The expected per person cost of universal ultrasound was £2,957 (95% CrI: £2,922–£2,991), compared to £2,949 (95% CrI: £2,915–£2,984) from selective ultrasound, a cost increase of

Table 3. MOD by presentation and response to ECV for POP study participants with breech presentation at 36-week scan (n = 179).

ECV status	Vaginal	ELCS	EMCS	Total
ECV successful	8	1	3	12
ECV unsuccessful	0	54	18	72
ECV not offered*	1	17	5	23
ECV discussed but declined	1	32	12	45
ECV accepted but not performed†	9	5	11	25
Missing	0	1	1	2
Total	19	110	50	179

Abbreviations: ECV, external cephalic version; ELCS, elective cesarean section; EMCS, emergency cesarean section; MOD, mode of delivery.

*Eighteen women were contraindicated due to low AFI at screening, 2 for uterine abnormalities, and 3 for other reasons.

†Seventeen babies turned spontaneously, 6 had reduced AFI on the day of the ECV, and 2 went into labour before ECV.

https://doi.org/10.1371/journal.pmed.1002778.t003

£7.29 (95% CrI: 2.41–11.61). Across England, this means that universal ultrasound would cost £4.27 million more annually than current practice. The increase stems from higher costs of ultrasound scan (£20.3 per person) and ECV (£3.6 per person) but is partly offset by the lower delivery costs (−£16.5 per person). The distribution of differences in costs between the two strategies is shown as Supporting information, S2 Fig. The simulation shows that universal ultrasound would, on average, increase the number of total ELCS deliveries by 8,858 (95% CrI: 7,662–10,068) but decrease the number of EMCS and vaginal breech deliveries by 4,196 (95% CrI: 2,779–5,603) and 6,061 (95% CrI: 6,617–8,670) per year, respectively.

The long-term health outcomes are presented in Table 4. Nationwide, universal ultrasound would be expected to lower mortality by 7.89 cases annually (95% CrI: 3.71, 12.7). After discounting, this means that universal ultrasound would be expected to yield 192 QALYs annually (95% CrI: 90,308). The cost effectiveness of universal ultrasound depends on the value assigned to these QALYs. The incremental cost effectiveness ratio (ICER) was £23,611 (95%

Table 4. Simulated cost and MOD distribution for universal ultrasound and no ultrasound.

	Universal ultrasound	Selective ultrasound	Difference (per patient)	Difference (total population)
Total cost	2,956.59	2,949.30	7.29	4,268,004
Screening cost	20.70	0.43	20.27	11,867,159
ECV cost	6.52	2.94	3.57	2,093,048
Delivery cost	2,927.78	2,944.31	−16.53	−9,679,396
Mortality cost	1.59	1.62	−0.02	−12,806
Vaginal cephalic	0.6850	0.6826	0.0024	1,399
ELCS cephalic	0.0442	0.0441	0.0001	84
EMCS cephalic	0.2321	0.2305	0.0016	918
VB	0.0007	0.0110	−0.0104	−6,061
ELCS breech	0.0273	0.0123	0.0150	8,774
EMCS breech	0.0107	0.0194	−0.0087	−5,115
Total mortality	0.000982	0.000995	−0.000013	−7.89
Total QALY	24.27615	24.27582	0.000327	191.73

Abbreviations: ECV, external cephalic version; ELCS, elective cesarean section; EMCS, emergency cesarean section; MOD, mode of delivery; QALY, quality-adjusted life years; VB, vaginal breech.

Costs (£) are presented per patient, except in column for 'total population' (n = 585,489).

https://doi.org/10.1371/journal.pmed.1002778.t004

CrI: 8,184, 44,851), which is of borderline cost effectiveness (given NICE's willingness to pay of £20,000 to £30,000) [38]. The number needed to scan per prevented mortality was 74,204 (95% CrI: 46,124–157,642).

One-way sensitivity analysis showed that the probability parameter with the greatest impact upon the cost effectiveness of universal ultrasound was the prevalence of breech: increasing this parameter by 1 percentage point was associated with a relative reduction of costs for universal ultrasound by £3.07. The results were less sensitive to the ECV success rate; an increase by 1 percentage point led to a relative reduction in the cost of universal ultrasound by £0.12. The most important cost parameter was the unit cost of ultrasound scan; an increase in this parameter by £10 led to a relative increase for universal ultrasound by £9.79 (see Supporting information, S3 Fig). Keeping all other parameters equal, universal ultrasound would be cost effective if ultrasound scanning could be provided for less than £19.80 or £23.10 per mother, for a willingness-to-pay threshold of £20,000 or £30,000, respectively. For universal ultrasound to be cost saving, scans would need to cost less than £12.90 per mother.

Discussion

In a prospective cohort study of >3,800 women having first pregnancies, a presentation scan at approximately 36 wkGA identified the 4.6% of women who had a foetus presenting by the breech, and for more than half of these, breech presentation had not previously been clinically suspected. The majority of these women were ultimately delivered by planned cesarean section, some experienced labour before their scheduled date and were delivered by EMCS, and a small proportion had a cephalic vaginal delivery following either spontaneous cephalic version or ECV. No woman in the cohort had a vaginal breech delivery or experienced an intrapartum cesarean for undiagnosed breech. The low uptake of vaginal breech birth is likely to reflect the fact that this is a nulliparous population, and it is generally accepted that the risks associated with vaginal breech delivery are lower in women who have had a previous normal birth.

Our economic analysis suggests that a universal late-pregnancy presentation scan would decrease the number of foetal mortalities associated with breech presentation and that this is of borderline cost effectiveness, costing an estimated £23,611 per QALY gained. The key driver of cost effectiveness is the cost of the scan itself. In the absence of a specific national unit cost, we have identified the maximum cost at which it would be cost effective. This is £19.80 per scan to yield an ICER of £20,000 per QALY and £23.10 at £30,000. These unit costs may be possible if assessment of presentation could be performed as part of a routine antenatal visit. Portable ultrasound systems adequate for presentation scans are available at low cost, and a presentation scan is technically quite simple, so the required level of skill could be acquired by a large cadre of midwives. This would result in a small fraction of the costs associated with a trained ultrasonographer performing a scan in a dedicated space using a high-specification machine. If universal ultrasound could be provided for less than £12.90 per scan, the policy would also be cost saving.

Our sensitivity analysis shows that the unit cost of ultrasound scans and the prevalence of breech presentation were by far the biggest determinants of the cost and cost effectiveness of universal ultrasound. The detection rate with abdominal palpation (i.e., for selective ultrasound) is the most important parameter aside from these. By contrast, the costs, attempt, and success rates for ECV have modest impact upon the choice of scanning strategy. It appears that the main short-term cost benefit from late-pregnancy screening lies in the possibility of scheduling ELCSs when breech presentation is detected, rather than turning the baby into a cephalic position.

This analysis may have underestimated the health benefits of universal late-pregnancy ultrasound. In the absence of suitable data on long-term outcomes by MOD and foetal presentation, we made the simplifying assumption that mortality rates were equal for ELCSs and EMCSs. Relaxing this assumption would likely favour universal ultrasound, as this strategy would reduce EMCSs, and these are associated with higher risks of adverse outcomes than ELCSs [41–44]; on top of health benefits, this may also reduce long-term NHS costs. It is also possible that an EMCS for a known breech presentation is less expensive and has better health outcomes than one for which breech is detected intrapartum, although lack of separate data for these two scenarios prevented us from pursuing this analysis further.

Our analysis shows that universal late-pregnancy ultrasound screening would increase total number of cesarean sections. Evidence suggests that cesarean delivery may have long-term consequences on the health of the child (increased risk of asthma and obesity), the mother (reduced risk of pelvic organ prolapse and increased risk of subfertility), and future pregnancies (increased risk of placenta previa and stillbirth) [45,46]. There is no evidence that these are related to the type of the cesarean section (elective versus emergency) [45,46]. Our economic modelling has not been able to capture these complex effects due to the model's endpoints and the focus on the current pregnancy only. However, accounting for these effects, it seems plausible that universal late-pregnancy ultrasound would be more favourable for mothers than children or future pregnancies.

Our results are also driven by vaginal delivery yielding worse long-term health outcomes than ELCS for breech presentation [1]. However, even though the rate of vaginal breech birth declined after the Term Breech Study, in many cases, the outcomes are not inferior to that of ELCS, and the RCOG guidelines state that vaginal breech delivery may be attempted following careful selection and counselling [1,3,47]. It is hard to assess how an increase in vaginal breech delivery would affect the cost effectiveness of universal ultrasound; while decreased mortality risk from vaginal breech delivery would decrease the importance of knowing the foetal presentation, universal screening would facilitate selection for attempted vaginal breech delivery.

One limitation of this study is that foetal presentation was revealed to all women in the POP study. Consequently, this study cannot say what would have happened without routine screening. However, we felt that it was appropriate to reveal the presentation at the time of the 36-wkGA scan, as there is level 1 evidence that planned cesarean delivery reduces the risk of perinatal morbidity and mortality in the context of breech presentation at term [44]. Another weakness was that the study was being undertaken in a single centre only and that the sample size was too small to avoid substantial parameter uncertainty for rare events. Moreover, less than half of all breech presentations in the POP study were detected by abdominal palpation. It is unclear whether the detection rates were affected by midwives knowing that the women were part of the POP study and, hence, would receive an ultrasound scan at 36 wkGA.

The prevalence of breech presentation in this study (4.6%) appears higher than the 3%–4% that is often reported in literature [1]. However, this study is unique in that it reports the prevalence at the time of ultrasound scanning, approximately 36 wkGA. Taking into account the number of spontaneous reversions to cephalic and that some cases of successful ECV may have turned spontaneously without intervention, our finding is consistent with the literature. The ECV success rate in the POP study was considerably lower than reported elsewhere in the literature; it was even lower than the 32% success rate that has been reported as the threshold level for when ECV is preferred over no intervention at all [48]. This might partly reflect the participants in the POP study; they were older and more likely to be obese than in many previous studies, and the cohort consisted of nulliparous women, who have higher rates of ECV failure than parous women [9,49,50]. It is also possible that the real-world ECV success rate is lower than in the literature due to publication bias. However, sensitivity analysis indicates that

the impact from an increased ECV success rate would be modest (an increase in ECV success rate by 10 percentage points lowers the incremental cost of universal ultrasound by £0.91 per patient).

The findings from this study cannot easily be transferred to another health system due to the differences in healthcare costs and antenatal screening routines. Some countries, e.g., France and Germany, already offer a third-trimester routine ultrasound scan. However, these scans are offered prior to 36 wkGA, and as many preterm breech presentations revert spontaneously, it would have limited predictive value for breech at term [51]. Whether screening for breech presentation in lower-income settings is likely to be cost effective largely depends on the coverage of the healthcare system; while screening may be relatively more costly, the benefits from avoiding undiagnosed breech presentation may also be relatively larger.

Whether the findings of this study could be extrapolated beyond nulliparous women is hard to assess. The absence of comparable data on screening sensitivity without universal ultrasound for parous women is an important limitation. The risks associated with breech birth also differ between nulliparous and parous women [52,53]. Compared to nulliparous women, parous women have higher success rates for ECV but also higher risk of spontaneous reversion to breech after 36 wkGA [9,28]. Also, the risks associated with vaginal breech delivery are lower in women who have had a previous vaginal birth [30].

Breech presentation is not the only complication that could be detected through late-pregnancy ultrasound screening. The same ultrasound session could also be used to screen for other indicators of foetal health, such as biometry and signs of growth restriction. Whether also scanning for other complications could increase the benefits from universal ultrasound has been and currently is subject to research [54,55]. Exploring the consequences from such joint screening strategies goes beyond the scope of this paper but has important implications for policy-makers and should therefore be subject to further research.

Conclusion

This study shows that implementation of universal late-pregnancy ultrasound to assess foetal presentation would virtually eliminate undiagnosed intrapartum breech presentation in nulliparous women. If this procedure could be implemented into routine care, for example, by midwives conducting a routine 36-wkGA appointment and using a portable ultrasound system, it is likely to be cost effective. Such a programme would be expected to reduce the consequences to the child of undiagnosed breech presentation, including morbidity and mortality.

Supporting information

S1 STROBE checklist. STROBE, strengthening the reporting of observational studies in epidemiology.
(DOC)

S1 Text. Cost input estimation.
(DOCX)

S1 Table. Input costs and probabilities for the economic model, detailed.
(DOCX)

S1 Fig. Foetal presentation and ECV status in the POP breech study. ECV, external cephalic version; POPs, Pregnancy Outcome Prediction.
(TIF)

S2 Fig. PSA of cost differences between universal ultrasound and selective ultrasound.
PSA, Probabilistic Sensitivity Analysis.
(TIFF)

S3 Fig. One-way sensitivity analysis of the difference in costs between universal ultrasound and selective ultrasound.
(TIFF)

Author Contributions

Conceptualization: Alison Dacey, Gordon C. S. Smith.

Formal analysis: David Wastlund, Alexandros A. Moraitis, Ulla Sovio, Edward C. F. Wilson.

Investigation: Alison Dacey.

Methodology: David Wastlund, Ulla Sovio, Edward C. F. Wilson, Gordon C. S. Smith.

Writing – original draft: David Wastlund, Alison Dacey.

Writing – review & editing: David Wastlund, Alexandros A. Moraitis, Ulla Sovio, Edward C. F. Wilson, Gordon C. S. Smith.

References

1. Impey LWM, Murphy D, Griffiths M, Penna LK on behalf of the Royal College of Obstetricians and Gynaecologists. Management of Breech Presentation. *BJOG*: *an international journal of obstetrics and gynaecology* 2017; 124:e151–e77.
2. Impey LWM, Murphy D, Griffiths M, Penna LK on behalf of the Royal College of Obstetricians and Gynaecologists. External Cephalic Version and Reducing the Incidence of Term Breech Presentation. *BJOG*: *an international journal of obstetrics and gynaecology* 2017; 124:e178–e92.
3. Hannah ME, Hannah WJ, Hewson SA, Hodnett ED, Saigal S, Willan R. Planned caesarean section versus planned vaginal birth for breech presentation at term: a randomised multicentre trial. Term Breech Trial Collaborative Group. *Lancet* 2000; 356(9239):1375–83. PMID: 11052579
4. Hunter LA. Vaginal breech birth: can we move beyond the Term Breech Trial? *Journal of midwifery & women's health* 2014; 59(3):320–27.
5. Leung WC, Pun TC, Wong WM. Undiagnosed breech revisited. *BJOG*: *an international journal of obstetrics and gynaecology* 1999; 106(7):638–41.
6. Watson WJ, Welter S, Day D. Antepartum identification of breech presentation. *The Journal of reproductive medicine* 2004; 49(4):294–6. PMID: 15134156
7. Nassar N, Roberts CL, Cameron CA, Olive EC. Diagnostic accuracy of clinical examination for detection of non-cephalic presentation in late pregnancy: cross sectional analytic study. *BMJ* 2006; 333 (7568):578–80. https://doi.org/10.1136/bmj.38919.681563.4F PMID: 16891327
8. Hofmeyr GJ, Kulier R, West HM. External cephalic version for breech presentation at term. *The Cochrane database of systematic reviews* 2015(4):Cd000083. https://doi.org/10.1002/14651858.CD000083.pub3 PMID: 25828903
9. Beuckens A, Rijnders M, Verburgt-Doeleman GH, Rijninks-van Driel GC, Thorpe J, Hutton EK. An observational study of the success and complications of 2546 external cephalic versions in low-risk pregnant women performed by trained midwives. *BJOG*: *an international journal of obstetrics and gynaecology* 2016; 123(3):415–23.
10. Nassar N, Roberts CL, Barratt A, Bell JC, Olive EC, Peat B. Systematic review of adverse outcomes of external cephalic version and persisting breech presentation at term. *Paediatric and perinatal epidemiology* 2006; 20(2):163–71. https://doi.org/10.1111/j.1365-3016.2006.00702.x PMID: 16466434
11. Rosman AN, Vlemmix F, Fleuren MA, Rijnders ME, Beuckens A, Opmeer BC, et al. Patients' and professionals' barriers and facilitators to external cephalic version for breech presentation at term, a qualitative analysis in the Netherlands. *Midwifery* 2014; 30(3):324–30. https://doi.org/10.1016/j.midw.2013.03.013 PMID: 23664312

12. Palencia R, Gafni A, Hannah ME, Ross S, Willan AR, Hewson S, et al. The costs of planned cesarean versus planned vaginal birth in the Term Breech Trial. *CMAJ* 2006; 174(8):1109–13. https://doi.org/10.1503/cmaj.050796 PMID: 16606959
13. Henderson J, Petrou S. The economic case for planned cesarean section for breech presentation at term. *CMAJ* 2006; 174(8):1118–9. https://doi.org/10.1503/cmaj.060171 PMID: 16606960
14. Tunde-Byass MO, Hannah ME. Breech vaginal delivery at or near term. *Seminars in perinatology* 2003; 27(1):34–45. PMID: 12641301
15. 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(10008):2089–97. https://doi.org/10.1016/S0140-6736(15)00131-2 PMID: 26360240
16. Pasupathy D, Dacey A, Cook E, Charnock-Jones DS, White IR, Smith GCS. Study protocol. A prospective cohort study of unselected primiparous women: the pregnancy outcome prediction study. *BMC pregnancy and childbirth* 2008; 8:51. https://doi.org/10.1186/1471-2393-8-51 PMID: 19019223
17. Gaccioli F, Lager S, Sovio U, Charnock-Jones S, Smith GCS. The pregnancy outcome prediction (POP) study: Investigating the relationship between serial prenatal ultrasonography, biomarkers, placental phenotype and adverse pregnancy outcomes. *Placenta* 2017; 59(Suppl 1):S17–S25.
18. National Institute for Health and Care Excellence. Quality standard: Antenatal care. In: National Institute for Health and Care Excellence, ed., 2012.
19. Freeman JV, Cole TJ, Chinn S, Jones PR, White EM, Preece MA. Cross sectional stature and weight reference curves for the UK, 1990. *Archives of disease in childhood* 1995; 73(1):17–24. PMID: 7639543
20. Noble M, Mclennan D, Wilkinson K, Whiteworth A, Exley S, Barnes H, et al. The English indices of deprivation 2007, 2007.
21. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2017.
22. Martin AD, Quinn KM, Park JH. MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software* 2011; 42(9):22.
23. Fan FY. FinCal: Time Value of Money, Time Series Analysis and Computational Finance. R package version 063, 2016.
24. Baio G, Heath A. BCEA: Bayesian Cost Effectiveness Analysis, 2016.
25. Department of Health. NHS Reference Costs 2015–16, 2016.
26. James M, Kevin H, Robin B, Richard J. A decision analytical cost analysis of offering ECV in a UK district general hospital. *BMC Health Services Research* 2001; 1(6).
27. Curtis L, Burns A. Unit Costs of Health and Social Care 2017: Personal Social Services Research Unit, 2017.
28. Ben-Meir A, Elram T, Tsafrir A, Elchalal U, Ezra Y. The incidence of spontaneous version after failed external cephalic version. *American journal of obstetrics and gynecology* 2007; 196(2):157.e1–3.
29. Rosman AN, Vlemmix F, Ensing S, Opmeer BC, Te Hoven S, Velzel J, et al. Mode of childbirth and neonatal outcome after external cephalic version: A prospective cohort study. *Midwifery* 2016; 39:44–8. https://doi.org/10.1016/j.midw.2016.02.014 PMID: 27321719
30. Krueger S, Simioni J, Griffith LE, Hutton EK. Labour Outcomes After Successful External Cephalic Version Compared With Spontaneous Cephalic Version. *Journal of obstetrics and gynaecology Canada*: *JOGC = Journal d'obstetrique et gynecologie du Canada*: *JOGC* 2018; 40(1):61–67. https://doi.org/10.1016/j.jogc.2017.05.020 PMID: 28780214
31. Hants Y, Kabiri D, Elchalal U, Arbel-Alon S, Drukker L, Ezra Y. Induction of labor at term following external cephalic version in nulliparous women is associated with an increased risk of cesarean delivery. *Archives of gynecology and obstetrics* 2015; 292(2):313–9. https://doi.org/10.1007/s00404-015-3643-z PMID: 25657092
32. Lavender T, Hofmeyr GJ, Neilson JP, Kingdon C, Gyte GM. Caesarean section for non-medical reasons at term. *The Cochrane database of systematic reviews* 2012(3):Cd004660. https://doi.org/10.1002/14651858.CD004660.pub3 PMID: 22419296
33. National Institute of Health. NIH State-of-the-Science Conference Statement on cesarean delivery on maternal request. *NIH consensus and state-of-the-science statements* 2006; 23(1):1–29. PMID: 17308552
34. American College of Obstetricians and Gynecologists. ACOG committee opinion no. 559: Cesarean delivery on maternal request. *Obstetrics and gynecology* 2013; 121(4):904–7. https://doi.org/10.1097/01.AOG.0000428647.67925.d3 PMID: 23635708
35. Whyte H, Hannah ME, Saigal S, Hannah WJ, Hewson S, Amankwah K, et al. Outcomes of children at 2 years after planned cesarean birth versus planned vaginal birth for breech presentation at term: the

International Randomized Term Breech Trial. *American journal of obstetrics and gynecology* 2004; 191 (3):864–71. https://doi.org/10.1016/j.ajog.2004.06.056 PMID: 15467555

36. Szende A, Janssen B, Cabasés J. Self-Reported Population Health: An International Perspective based on EQ-5D: Springer Dordrecht Heidelberg New York London, 2014.
37. Office for National Statistics. National Life Tables, United Kingdom, 1980–82 to 2014–16: Office for National Statistics, 2017.
38. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal 2013, 2013.
39. Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. Oxford: Oxford University Press 2006.
40. Digital NHS. NHS Maternity Statistics 2016–17: Summary report: NHS Digital—Health and Social Care Information Centre, 2017.
41. Petrou S, Kim SW, McParland P, Boyle EM. Mode of Delivery and Long-Term Health-Related Quality-of-Life Outcomes: A Prospective Population-Based Study. *Birth* 2017; 44(2):110–19. https://doi.org/10.1111/birt.12268 PMID: 27862265
42. Mackay DF, Wood R, King A, Clark DN, Cooper SA, Smith GCS, et al. Educational outcomes following breech delivery: a record-linkage study of 456947 children. *International journal of epidemiology* 2015; 44(1):209–17. https://doi.org/10.1093/ije/dyu270 PMID: 25613426
43. Sovio U, Smith GCS. Blinded ultrasound fetal biometry at 36 weeks and risk of emergency Cesarean delivery in a prospective cohort study of low-risk nulliparous women. *Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* 2018; 52(1):78–86. https://doi.org/10.1002/uog.17513 PMID: 28452133
44. Hofmeyr GJ, Hannah M, Lawrie TA. Planned caesarean section for term breech delivery. *The Cochrane database of systematic reviews* 2015(7):Cd000166. https://doi.org/10.1002/14651858.CD000166.pub2 PMID: 26196961
45. Keag OE, Norman JE, Stock SJ. Long-term risks and benefits associated with cesarean delivery for mother, baby, and subsequent pregnancies: Systematic review and meta-analysis. PLoS Med. 2018; 15(1):e1002494. https://doi.org/10.1371/journal.pmed.1002494 PMID: 29360829
46. Moraitis AA, Oliver-Williams C, Wood AM, Fleming M, Pell JP, Smith GCS. Previous caesarean delivery and the risk of unexplained stillbirth: retrospective cohort study and meta-analysis. *BJOG: an international journal of obstetrics and gynaecology* 2015; 122(11):1467–74.
47. Pasupathy D, Wood AM, Pell JP, Fleming M, Smith GCS. Time trend in the risk of delivery-related perinatal and neonatal death associated with breech presentation at term. *International journal of epidemiology* 2009; 38(2):490–8. https://doi.org/10.1093/ije/dyn225 PMID: 18977783
48. Tan JM, Macario A., Carvalho B, Druzin ML, El-Sayed YY. Cost-effectiveness of external cephalic version for term breech position. *BMC pregnancy and childbirth* 2010.
49. Lau TK, Lo KW, Wan D, Rodgers MS. Predictors of successful external cephalic version at term: a prospective study. *BJOG: an international journal of obstetrics and gynaecology* 1997; 104(7):798–802.
50. Cho LY, Lau WL, Lo TK, Tang HHT, Leung WC. Predictors of successful outcomes after external cephalic version in singleton term breech pregnancies: A nine-year historical cohort study. *Hong Kong Medical Journal* 2012; 18(1):11–19. PMID: 22302905
51. Witkop CT, Zhang J, Sun W, Troendje J. Natural history of fetal position during pregnancy and risk of nonvertex delivery. *Obstetrics and gynecology* 2008; 111(4):875–80. https://doi.org/10.1097/AOG.0b013e318168576d PMID: 18378746
52. Macharey G, Gissler M, Ulander VM, Rahkonen L, Vaisanen-Tommiska M, Nuutila M, et al. Risk factors associated with adverse perinatal outcome in planned vaginal breech labors at term: a retrospective population-based case-control study. *BMC pregnancy and childbirth* 2017; 17(1):93. https://doi.org/10.1186/s12884-017-1278-8 PMID: 28320344
53. Parissenti TK, Hebisch G, Sell W, Staedele PE, Viereck V, Fehr MK. Risk factors for emergency caesarean section in planned vaginal breech delivery. *Archives of gynecology and obstetrics* 2017; 295(1):51–58. https://doi.org/10.1007/s00404-016-4190-y PMID: 27631406
54. Bricker L, Medley N, Pratt JJ. Routine ultrasound in late pregnancy (after 24 weeks' gestation). *The Cochrane database of systematic reviews* 2015(6):Cd001451. https://doi.org/10.1002/14651858.CD001451.pub4 PMID: 26121659
55. Henrichs J, Verfaille V, Viester L, Westerneng M, Molevijk B, Franx A, et al. Effectiveness and cost-effectiveness of routine third trimester ultrasound screening for intrauterine growth restriction: study protocol of a nationwide stepped wedge cluster-randomized trial in The Netherlands (The IRIS Study). *BMC pregnancy and childbirth* 2016; 16(1):310. https://doi.org/10.1186/s12884-016-1104-8 PMID: 27737654

BJOG An International Journal of Obstetrics and Gynaecology

DOI: 10.1111/1471-0528.15809
www.bjog.org

General obstetrics

The cost-effectiveness of universal late-pregnancy screening for macrosomia in nulliparous women: a decision analysis

D Wastlund,[a,b] AA Moraitis,[c] JG Thornton,[d] J Sanders,[e] IR White,[f] P Brocklehurst,[g] GCS Smith,[c] ECF Wilson[b,h]

[a] The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK [b] Cambridge Centre for Health Services Research, Cambridge Institute of Public Health, University of Cambridge, Cambridge, UK [c] Department of Obstetrics and Gynaecology, NIHR Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK [d] Division of Child Health, Obstetrics and Gynaecology, School of Medicine, University of Nottingham, Nottingham, UK [e] School of Healthcare Sciences, Cardiff University, Cardiff, UK [f] MRC Clinical Trials Unit, University College London, London, UK [g] Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, UK [h] Health Economics Group, Norwich Medical School, University of East Anglia, Norwich, UK
Correspondence: ECF Wilson, Health Economics Group, Norwich Medical School, University of East Anglia, Norwich NR4 7TJ, UK.
Email: Ed.Wilson@uea.ac.uk

Accepted 16 April 2019. Published Online 6 June 2019.

Objective To identify the most cost-effective policy for detection and management of fetal macrosomia in late-stage pregnancy.

Design Health economic simulation model.

Setting All English NHS antenatal services.

Population Nulliparous women in the third trimester treated within the UK NHS.

Methods A health economic simulation model was used to compare long-term maternal–fetal health and cost outcomes for two detection strategies (universal ultrasound scanning at approximately 36 weeks of gestation versus selective ultrasound scanning), combined with three management strategies (planned caesarean section versus induction of labour versus expectant management) of suspected fetal macrosomia. Probabilities, costs and health outcomes were taken from literature.

Main outcome measures Expected costs to the NHS and quality-adjusted life-years (QALYs) gained from each strategy, calculation of net benefit and hence identification of most cost-effective strategy.

Results Compared with selective ultrasound, universal ultrasound increased QALYs by 0.0038 (95% CI 0.0012–0.0076), but also costs by £123.50 (95% CI 99.6–149.9). Overall, the health gains were too small to justify the cost increase given current UK thresholds cost-effective policy was selective ultrasound coupled with induction of labour where macrosomia was suspected.

Conclusions The most cost-effective policy for detection and management of fetal macrosomia is selective ultrasound scanning coupled with induction of labour for all suspected cases of macrosomia. Universal ultrasound scanning for macrosomia in late-stage pregnancy is not cost-effective.

Keywords Economic modelling, health economics, macrosomia, pregnancy, screening, third-trimester, ultrasound.

Tweetable abstract Universal late-pregnancy ultrasound screening for fetal macrosomia is not warranted.

Linked article This article is commented on by BD Einerson, p. 1251 in this issue. To view this mini commentary visit https://doi.org/10.1111/1471-0528.15851.

Please cite this paper as: Wastlund D, Moraitis AA, Thornton JG, Sanders J, White IR, Brocklehurst P, Smith GCS, Wilson ECF. The cost-effectiveness of universal late-pregnancy screening for macrosomia in nulliparous women: a decision analysis. BJOG 2019;126:1243–1250.

Introduction

The detection and management of macrosomia, i.e. excessive fetal growth, poses a challenge to maternity care. Macrosomia is associated with increased perinatal mortality and morbidity, e.g. shoulder dystocia leading to brachial plexus injury, as well as increased risk of maternal morbidity.[1–3] The definition of macrosomia varies, but is usually defined as a birthweight >4000 or >4500 g. It is differentiated from, but closely related to, the concept of large-for-gestational-age, which is a relative measure: weight greater than the 90th centile for a given gestational age.[1,4] Macrosomia can only be definitively diagnosed by weighing the infant following delivery. However,

ultrasound scans can be used to estimate the fetal weight antenatally, although this approach is known to have low predictive value.[1] There is no general agreement on how to manage macrosomia if it is suspected following ultrasound.[1,4–6] Possible interventions include scheduling an elective Caesarean section (CS), or early induction of labour. However, uncertainty regarding the clinical effectiveness of these interventions persists.[1,5] Furthermore, if given without clinical need, intervention may cause unnecessary harm, e.g. neonatal respiratory morbidity, and the increased maternal risks of CS.[1,4,7,8]

There is currently no national programme that couples screening for macrosomia with a proven, disease-modifying intervention.[4,9] Currently, clinical examination of third-trimester pregnancies does not routinely include ultrasound, but women may be selected for ultrasound scanning following clinical suspicion of macrosomia (selective ultrasound). An alternative approach would be to prospectively scan all women for macrosomia (universal ultrasound) at around 36 weeks of gestation, but whether the benefits of such an approach would justify the increased costs and risk of harmful interventions is unclear. A previous study showed only modest health benefits from universal ultrasound, and the cost for every prevented severe adverse outcome was too high to justify routine scanning.[10] However, this study is now over 20 years old and only considered one management strategy for suspected macrosomia: delivery by planned CS. Following recent research and changes in obstetric care, we sought to re-evaluate the case for universal ultrasound screening for macrosomia.[11]

In this study, we identify the most cost-effective strategy for detection and management of macrosomia in late pregnancy among nulliparous women in the setting of the UK National Health Service (NHS).

Methods

Model structure

The scope of this model was limited to screening for macrosomia rather than any other complication of pregnancy. To compare the cost-effectiveness of different policies for detection and management, we constructed a decision tree simulation model using R (Figure 1).[12–14] Each policy had two components: one for the detection of macrosomia, and one for the management of suspected macrosomia. The detection strategy was either universal ultrasound in the third trimester (around 36 weeks of gestation), or selective ultrasound, i.e. clinical examination through abdominal palpation, where ultrasound would be offered only where macrosomia was suspected. The management strategy for suspected macrosomia was either to schedule an elective CS (Planned CS), induce labour (Induction), or expectant management awaiting spontaneous labour onset. If macrosomia was not suspected, expectant management was used. There are therefore a total of six discrete detection/management policies.

The model structure for detection and management for macrosomia is shown in Figure 1(A). Four different screening statuses were possible: true positives, false negatives, false positives and true negatives. The likelihood of each state was driven by the sensitivity and specificity of the test used for detection, as well as the prevalence of macrosomia. When macrosomia was suspected, the pregnancy was managed according to the management strategy being evaluated: planned CS, induction of labour, or expectant management. If macrosomia was not suspected, it was assumed that vaginal delivery would be attempted, with a risk of emergency CS. To accurately capture the consequences of a false-positive diagnosis of macrosomia, we distinguished between expectant management when macrosomia was suspected or not suspected; suspected macrosomia increased the risk of Caesarean delivery following expectant management.[8]

Five neonatal delivery outcomes were possible: No complications, Respiratory morbidity, Shoulder dystocia, Other acidosis (i.e. acidosis not induced by shoulder dystocia) and perinatal mortality. Their respective likelihoods were affected by both screening and management strategies (see below). The fetal delivery outcomes were then extrapolated into long-term costs and quality-adjusted life-years (QALYs) through the model shown in Figure 1.

Model inputs

Probabilities

For each adverse outcome (respiratory morbidity, shoulder dystocia, other acidosis and mortality), we obtained the baseline risk of that outcome; i.e. the risk if infant was a non-large and non-induced neonate with vaginal delivery. We then multiplied this risk with the relative risk of each present risk factor (macrosomia, induction, delivery through elective CS and delivery through emergency CS). For technical details, see Supplementary material (Appendix S1).

Model input parameters are shown in the Supplementary material (Table S1). Values were identified from literature by AM and DW, prioritising values from systematic reviews and UK data where possible. Ideally, every input should be based upon a systematic review, reflecting current state of knowledge. However, resources only permitted identification of suitable data, rather than performing a meta-analysis. For this reason, sources that provided a distribution for the likely parameter values were prioritised, so that the overall uncertainty associated with this parameter could be assessed through probabilistic sensitivity analysis.[15] Where multiple sources were available the source was chosen by

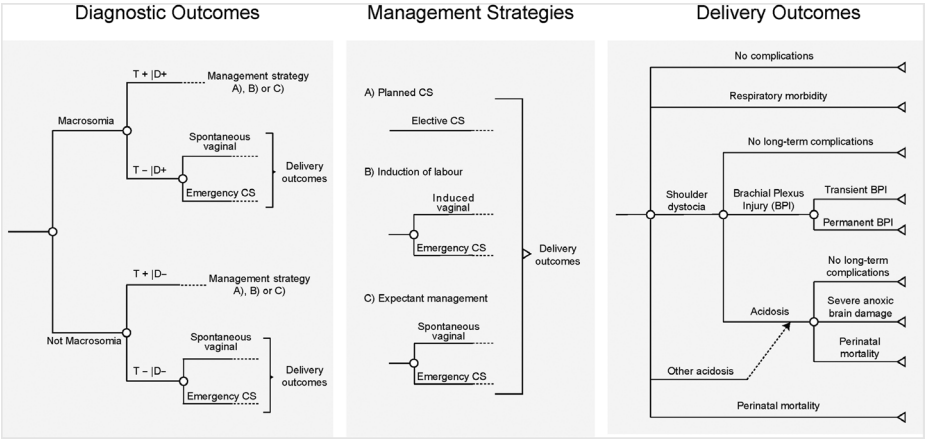


Figure 1. Structure of simulation model. The figure shows the model structure, from screening to long-term health outcomes. Part A (left) shows the pathway from screening to the mode of delivery. When macrosomia is suspected ('T+'), the mode of delivery depends on the management strategy as shown in part B (middle). Part C (right) shows the different delivery outcomes, and their associated long-term outcomes. BPI, brachial plexus injury; D+, disease-positive; D−, disease-negative; T+, test-positive; T−, test-negative.

consensus or through arbitration by GS. Where no credible values for a model parameter could be identified from the literature, AM and GS identified lower and upper limits to the value that the parameter could reasonably assume; the model then sampled input values from this interval using a uniform distribution.

Macrosomia was defined as estimated fetal weight ≥90th centile, i.e. the same as large-for-gestational-age. The sensitivity and specificity for detection of macrosomia, as well as the prevalence of macrosomia, were taken from the POP study, a prospective cohort study of unselected nulliparous women in which all women had fetal biometry at 36 weeks of gestation, where the result of the scan was blinded.[16,17] Using data from this study allowed for a comparison between diagnostic performance of universal and selective ultrasound. Detection with selective ultrasound was based upon clinical suspicion before 36 weeks of gestation following measurement of symphyseal–fundal height, and confirmed with a clinically indicated ultrasound.[17] The baseline risk of each adverse outcome was defined as the risk for a normal-size neonate, where labour was not induced and resulted in a vaginal delivery. We used odds ratios from the literature when directly presented, otherwise we calculated unadjusted odds ratios from prevalence data.[18] Odds ratios were assumed to be log-normally distributed.

Long-term outcomes

Unit costs and health state utilities are shown in the Supplementary material (Table S1). The average costs for induction of labour and respiratory morbidity were calculated from the NHS reference costs (see Supplementary material, Appendix S2).[19] Brachial plexus injury could be either transient or permanent, this was modelled using a β distribution.[20] We assumed that brachial plexus injury would require the same resource usage as reported by Culligan et al., and obtained the costs for these resources from the NHS reference costs (see Supplementary material, Appendix S2).[19,21] We assumed that all cases of nonsevere asphyxia would be treated in the neonatal unit for 1–3 days, but that no additional costs would accrue beyond this. To estimate the long-term outcomes from 'severe anoxic brain damage', we made the simplifying assumption that the costs, consequences and likelihood mirrored those of neonatal encephalopathy. Evidence shows that providing therapeutic hypothermia reduces the likelihood of adverse outcomes from neonatal encephalopathy, and this treatment is routine clinical practice.[22,23] We assumed that all cases of neonatal encephalopathy would receive therapeutic hypothermia, and adjusted costs and consequences from neonatal encephalopathy accordingly; for this reason, we reduced the likelihood of mortality and severe anoxic brain damage following asphyxia by 11.1%.[24] The costs from severe anoxic brain damage included hospital- and community-care costs for all survivors in the cooled group as reported by Regier et al.;[22] the hospital costs were for the first 18 months only, but we assumed that the community-care costs after discharge would accrue annually for the entirety of the model's time horizon. We made the simplifying assumption that the cost of death would be the same regardless of reason.

Quality-adjusted life-years combine the utility of a health-state with its duration, where utility is based upon quality of life (QOL). Quality of life can be expressed as a numeric value, where 1 is equivalent to full health and 0 is equivalent to death.[25,26] Maternal QALYs were based upon the mode of delivery, and QOL weights were obtained from Petrou et al.;[27] these QOL weights were derived using EQ-5D, as recommended by NICE.[28,29] For surviving

infants, we calculated the expected QALYs based upon the assumptions above; per definition, fetal QALYs were zero for death.

Model scope

The expected cost and QALYs gained from six different policies for screening and management of macrosomia were calculated over a 20-year time horizon. Costs and QALYs were discounted by 3.5% annually, as recommended by NICE.[29] Probabilistic sensitivity analysis was used to capture the overall effect of uncertainty in the model parameters. Costs associated with potential litigation claims or potential effects upon subsequent pregnancies were not included. Results were based upon 100 000 simulations and results presented as expected values, incremental cost and QALYs, incremental cost-effectiveness ratios (ICERs) (the ratio of incremental cost to incremental QALYs), and net benefits (defined as QALYs multiplied by the willingness to pay [WTP] for a QALY less the cost). The WTP per QALY threshold was assumed to be £20,000 (the lower of NICE's stated thresholds).[29] Decision uncertainty is illustrated using cost-effectiveness acceptability curves.[29,30] The model's sensitivity towards key parameters was explored through one-way sensitivity analysis (see Supplementary material, Appendix S4). Given the paucity of data relating to maternal quality of life, an additional scenario was conducted including neonatal QALYs alone. Further scenarios explored the impact of assigning zero additional costs for induction of labour, and assuming that induction of labour is cost saving (due to reduced antenatal assessments).[29,30] All costs are from the third-party payer (i.e. NHS) perspective, and the price year is 2016/17. Costs from other years were inflated to the price year of the analysis using the Hospital & Community Health Services index.[31] As this is a secondary analysis/synthesis of existing data, no patients nor the public were involved in the study.

Results

The expected costs and QALYs for each policy are shown in Table 1. The least expensive option is selective ultrasound with expectant management and the most expensive option is universal ultrasound with planned CS. The least effective option (in terms of QALYs gained) is universal ultrasound with planned CS and the most effective option is universal ultrasound with induction of labour. Three strategies (selective US + planned CS, universal ultrasound + expectant management, and universal ultrasound + planned CS) are dominated or extended-dominated by other strategies. Taking into account the balance between costs and outcomes (and with a WTP threshold of £20,000 per QALY), the most cost-effective strategy is selective ultrasound plus induction of labour where macrosomia is suspected. Although universal ultrasound plus induction is expected to yield marginally greater QALYs (+0.002), the added cost (+£113) yields an ICER of £52,719. This is above the threshold and is not, therefore, cost-effective. The expected distribution of mode of delivery and neonatal delivery outcomes is detailed in the Supplementary material (Appendix S3 and Table S2).

We investigated the value of universal ultrasound alone by comparing the results for universal and selective ultrasound when using the same management strategy. When the management strategy was planned CS, universal ultrasound was associated with a cost increase of £123.50 (95% CI £99.60–£149.90), and a QALY increase of 0.0038 (95% CI 0.0012–0.0076). The ICER for this strategy was £35,755 (95% CI £15,962–£98,506). The comparable ICERs for induction of labour and expectant management were even higher, indicating that universal ultrasound screening is unlikely to be cost-effective.

The probability of each policy being the most cost-effective as a function of the WTP threshold is shown by the cost-effectiveness acceptability curves (Figure 2). Selective ultrasound coupled with induction of labour for suspected macrosomia had the greatest chance of being cost-effective for NICE's recommended thresholds of £20,000–£30,000 per QALY.[29] Sensitivity analysis showed that the choice of policy was most sensitive towards the specificity of ultrasound (both universal and selective), maternal QOL for delivery through elective CS, and the prevalence of macrosomia (see Supplementary material, Appendix S4 and Table S3). Although influential, the cost of ultrasound screening alone appears insufficient to determine whether universal screening would be cost-effective; analysis showed that if other parameters remained unchanged, universal ultrasound would only be cost-effective if the cost of ultrasound was £26.56 or lower.

Excluding maternal QALYs from the analysis, selective ultrasound plus planned CS was the preferred management strategy, compared with induction of labour, under the base case (see Supplementary material, Table S4). No other assumptions tested in the alternative scenarios affected the conclusions; selective ultrasound with induction of labour remained the preferred strategy for all other scenarios.

Discussion

Main findings

This study has compared the cost-effectiveness of different policies for detection and management of fetal macrosomia in late-stage pregnancy among nulliparous women. The most cost-effective policy was selective ultrasound coupled with induction of labour for all cases of suspected fetal macrosomia. Although universal ultrasound scanning leads to higher identification of suspected macrosomia, this only

Table 1. Expected costs and QALYs per screening and management strategy

Strategy	Cost (95% CI)	QALY (95% CI)*	ICER	NMB (95% CI)
Selective ultrasound + expectant	2821 (2409–3236)	27.441 (27.262–27.621)	—	546 007 (542 803–549 204)
Selective ultrasound + induction	**2826 (2412–3242)**	**27.446 (27.267–27.626)**	**904**	**546 098 (542 890–549 298)**
Selective ultrasound + planned CS	2833 (2436–3230)	27.417 (27.244–27.588)	Dominated	545 501 (542 424–548 561)
Universal ultrasound + expectant	2933 (2502–3366)	27.441 (27.261–27.621)	Dominated	545 884 (542 695–549 070)
Universal ultrasound + induction	2939 (2506–3374)	27.448 (27.268–27.628)	52 719	546 028 (542 829–549 214)
Universal ultrasound + planned CS	2955 (2549–3360)	27.396 (27.224–27.565)	Dominated	544 956 (541 919–547 978)

NMB, net monetary benefit.
Options ordered from lowest to highest expected cost. ICERs calculated beginning with least expensive option, and comparing with next most expensive, non-dominated option; a policy was dominated/extended-dominated if any other policy or weighted average of two policies was associated with both lower costs and higher QALYs. Net monetary benefit (NMB) was calculated using a WTP threshold of £20,000; higher NMB value means greater cost-effectiveness. Option with the highest expected net monetary benefit highlighted in bold. All costs and NMB are given in pounds sterling (£).
*The maximum QALYs for two people over 20 years, discounted at 3.5%, is 29.42.

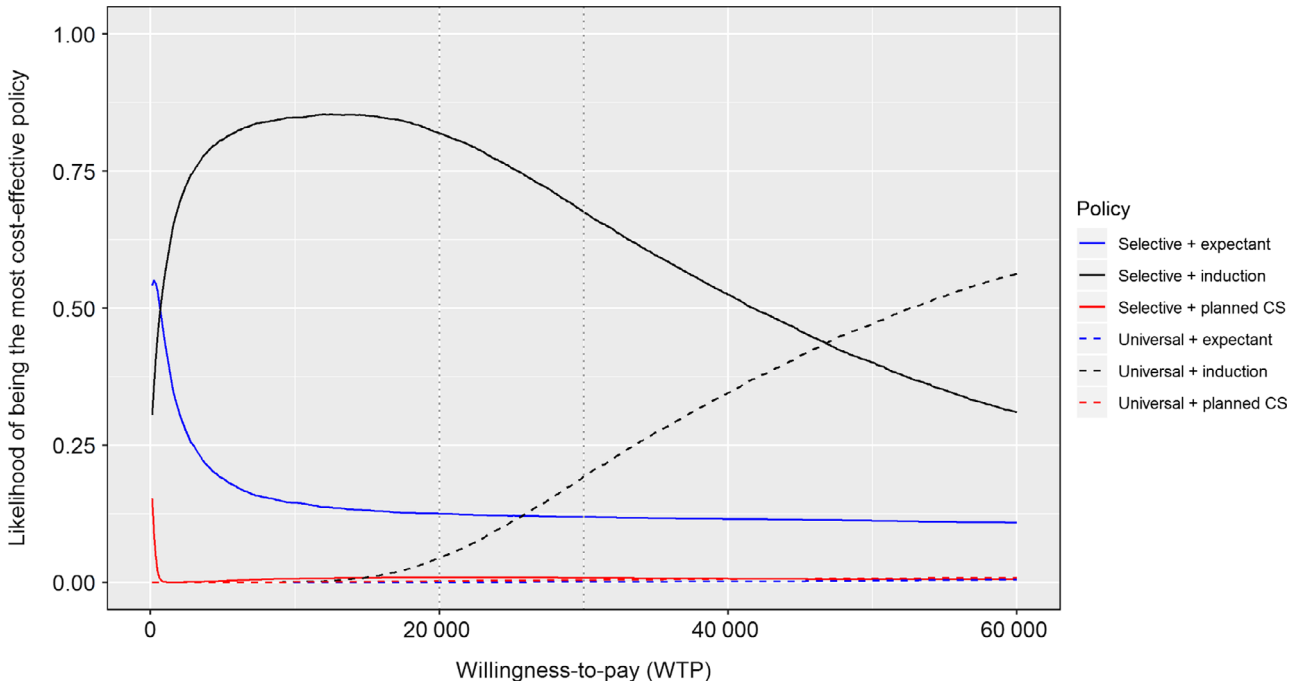


Figure 2. Cost-effectiveness acceptability curve for policies for detection and management of fetal macrosomia. Cost-effectiveness acceptability curve showing the chance of each policy of being the most cost-effective for different levels of WTP. Policies with universal ultrasound are shown as dashed lines and selective ultrasound as solid. Higher values for WTP imply a higher valuation of a QALY. The conventional WTP threshold for cost-effectiveness is £20,000–£30,000 (marked in figure).[29]

translates into modest improvements of overall long-term health outcomes, which are not justified by the added cost of the ultrasound scan. The expected health gain (0.003 QALYs over 20 years) is small because of both the low risk of severe neonatal outcomes resulting from undiagnosed macrosomia and the risk of interventions themselves causing harm.

Where macrosomia is suspected following ultrasound scanning, intervention is generally preferred to awaiting spontaneous labour onset. Although currently subject to further research,[32] this study found that induction of labour is the preferred intervention. However, it is worth noting that from the infant's perspective alone, the best option is an elective CS (see Supplementary material, Table S4, scenario 'Maternal QALYs excluded').

Universal (rather than selective) ultrasound coupled with induction of labour has the potential to be the most cost-effective policy, but only at very high valuations of health

gain: the small added benefit does not currently justify the cost. Sensitivity analysis shows that the relative cost-effectiveness of the policies is sensitive to changes in the cost of ultrasound scanning, as well as the costs of CS and induction of labour, and the sensitivity and specificity of ultrasound scanning. Hence, if the cost of the scan falls substantially in the future, a universal scanning policy could be cost-effective; analysis shows that this would happen at a cost below £26.56 (a cost reduction of 74.4%). Further, macrosomia is not the only fetal complication that can be assessed through ultrasound screening, so when combined with a scan for other anomalies, such as breech presentation, the marginal cost of detecting macrosomia may be sufficiently low to render the overall policy cost-effective. However, further work is needed to explore this.

Strengths and limitations

The strength of this study is that it evaluates strategies for both detection and management of fetal macrosomia jointly. There has been a lack of studies evaluating screening strategies coupled with clear evidence-based interventions. Economic modelling allows us to estimate how neonatal and maternal health outcomes would be affected if ultrasound screening were to be routinely implemented in clinical practice. However, the robustness of the conclusions is only as strong as the data available to inform them. Indeed, many parameters were informed by a single study, and where no data were available we relied on expert opinion. Critically, as a part of this process we elicited a range of plausible values to represent the inherent uncertainty. The probabilistic sensitivity analysis incorporates this uncertainty to determine how much it affects the overall results.

We have limited our analysis to nulliparous women. It is unclear whether our findings could be extended to parous women as well, especially given the absence of data on screening performance for universal and selective ultrasound for this group. The economic modelling also relies upon simplifying assumptions regarding the long-term outcomes from the mode of delivery and fetal delivery outcomes and did not take account of alterations to planned place of birth following ultrasound. The interplay between fetal macrosomia and long-term outcomes may be too complex to capture entirely within our model; macrosomia can lead to more complications than those explored in this analysis. However, in the absence of more detailed data on many of these complications, this model is still based upon the best current understanding of macrosomia and its consequences.

The probability of delivery outcomes in this analysis relied upon the assumption of no interaction between macrosomia and the intervention. In reality, this assumption may not hold perfectly; for example, elective CS may yield a greater relative risk reduction for babies with macrosomia. However, data limitations made the assumption necessary in order to model the relevant outcomes, especially given the many different sources used for parameters. Also, the relative risks associated with both macrosomia and interventions were included in the analysis, even though interactions were not modelled.

Interpretations

Our conclusion that universal ultrasound screening for fetal macrosomia is not cost-effective aligns with previous findings for macrosomia management based upon ultrasound screening.[10] Universal ultrasound screening strategies were less cost-effective than selective ultrasound for all scenarios. Our analysis demonstrated that universal ultrasound is associated with improved health outcomes, but that these gains are too small to justify its added cost.

This analysis is based in a UK NHS setting. The results will be generalisable to other settings with similar management policies and relative costs: current UK practice is to offer a scan at first and second trimesters but to only offer late-pregnancy scans where clinically indicated (our 'selective ultrasound' policy). Many European countries perform a third scan around 32 weeks.[33] Diagnostic effectiveness at 32 weeks for predicting complications related to macrosomia at delivery is likely to be poorer than at the 36–37 weeks assumed in our analysis, given the longer interval between the scan and time of birth.[16] This would suggest that earlier scans are even less likely to be cost-effective.

As stated above, the impact of CS on maternal QOL was a key driver of the results. To the best of our knowledge, the study by Petrou et al.[27] is the only study that reports maternal QOL as a function of the mode of delivery, using an adequate time horizon and a measure for QOL recommended by NICE.[29] However, it reported lower QOL for women who underwent elective CS than their counterparts who delivered through emergency CS, a finding that appears counterintuitive. If maternal QOL had been higher following elective CS than emergency CS, the economic analysis would have been more favourable towards policies with planned CS. Against this should be weighted the research that has shown that CS is associated with increased risk of a range of complications in subsequent pregnancies.[34–36] These risks are not captured in our simulation model because the perspective was for the current pregnancy, but implies that managing suspected macrosomia through planned CS may be more detrimental than suggested in this analysis.

This analysis has compared interventions based upon suspicion of macrosomia alone. However, in clinical practice more factors influence antenatal management than just whether ultrasound screening indicates fetal macrosomia. This analysis offers valuable information for policymaking,

but it does not rule out the use of planned CS or expectant management in individual cases.

Conclusion

Universal ultrasound scanning in the third trimester is not cost-effective at detecting macrosomia in nulliparous women at current UK cost-effectiveness threshold limits. If fetal macrosomia is suspected following ultrasound, induction of labour is likely to be the most cost-effective management option.

The conclusions are based on a single scan for macrosomia alone. A strategy that combines scanning for macrosomia with other conditions, e.g. breech presentation (and growth restriction), might be cost-effective. Future research should focus on whether joint screening for multiple fetal complications would be cost-effective, as well as on the long-term health consequences from delivery outcomes, especially how maternal health is affected by the mode of delivery.

Disclosure of interests

PB reports grants and personal fees from MRC, grants from MRC, NIHR HS&DR, NIHR HTA, Wellcome Trust, and personal fees from AG Biotest, outside the submitted work. GS reports grants and personal fees from GlaxoSmithKline Research and Development Limited, grants from Sera Prognostics Inc., grants and personal fees from Roche Diagnostics Ltd, and non-financial support from Illumina Inc., outside the submitted work. In addition, GS has a patent pending: United Kingdom Patent Application No. 1808489.7 'Novel Biomarkers'. AM, DW, EW, IRW, JS and JGT declare no competing interests. Completed disclosure of interest forms are available to view online as supporting information.

Contribution to authorship

AM, DW, EW and GS contributed to the study concept. Data collection was by AM and DW, and economic analysis by DW and EW. AM, DW, EW, GS, IRW, JS, JGT and PB all contributed to the manuscript preparation.

Details of ethics approval

Not applicable.

Funding

This study was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme, grant number 15/105/01. EW is part funded by the NIHR Cambridge Biomedical Research Centre. IRW was supported by the Medical Research Council Programme MC_UU_12023/21. The views expressed here are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Input values for simulated model.

Table S2. Expected share of mode of delivery and level of fetal complications per screening and management strategy.

Table S3. Sensitivity of net monetary benefit towards input parameters.

Table S4. Expected costs and QALYs per screening and management strategy, alternative scenarios.

Appendix S1. The risk of adverse neonatal delivery outcomes.

Appendix S2. Detailed derivation of model inputs.

Appendix S3. Expected distribution of outcomes.

Appendix S4. Sensitivity of results. ■

References

1 American College of Obstetricians and Gynecologists. Practice Bulletin No. 173: Fetal Macrosomia. *Obstet Gynecol* 2016;128: e195–209.

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

3 Royal College of Obstetrics & Gynaecologists. *Green-top Guideline No. 42: Shoulder Dystocia*. London: Royal College of Obstetrics & Gynaecologists, 2012.

4 Culliney KA, Parry GK, Brown J, Crowther CA. Regimens of fetal surveillance of suspected large-for-gestational-age fetuses for improving health outcomes. *The Cochrane database of systematic reviews*. 2016;4:Cd011739.

5 Campbell S. Fetal macrosomia: a problem in need of a policy. *Ultrasound Obstet Gynecol* 2014;43:3–10.

6 Rossi AC, Mullin P, Prefumo F. Prevention, management, and outcomes of macrosomia: a systematic review of literature and meta-analysis. *Obstet Gynecol Surv* 2013;68:702–9.

7 Morrison JJ, Rennie JM, Milton PJ. Neonatal respiratory morbidity and mode of delivery at term: influence of timing of elective caesarean section. *Br J Obstet Gynaecol* 1995;102:101–6.

8 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;200:340.e1–e3.

9 National Collaborating Centre for Women's and Children's Health. *Antenatal Care - Routine Care for the Healthy Pregnant Women*. London: Royal College of Obstericians and Gynecologists, 2008.

10 Rouse DJ, Owen J, Goldenberg RL, Cliver SP. The effectiveness and costs of elective cesarean delivery for fetal macrosomia diagnosed by ultrasound. *JAMA* 1996;276:1480–6.

11 Boulvain M, Senat MV, Perrotin F, Winer N, Beucher G, Subtil D, et al. Induction of labour versus expectant management for large-for-date fetuses: a randomised controlled trial. *Lancet (London, England)* 2015;385:2600–5.

12 R Core Team. *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2017.

13 Fan FY. FinCal: Time Value of Money, Time Series Analysis and Computational Finance. R package version 0632016.

14 Baio G, Heath A. *BCEA: Bayesian Cost Effectiveness Analysis*. London, 2016. R package version 2.2-6. http://www.statistica.it/gianluca/BCEA

15 Shemilt I, Mugford M, Vale L, Marsh K, Donaldson C. *Evidence-based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice*. Chichester: John Wiley & Sons, 2011.

16 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 (London, England)* 2015;386:2089–97.

17 Sovio U, Moraitis AA, Wong HS, Smith GCS. Universal vs selective ultrasonography to screen for large-for-gestational-age infants and associated morbidity. *Ultrasound Obstet Gynecol* 2018;51:783–91.

18 Welton N, Sutton A, Cooper N, Abrams K. *Evidence Synthesis for Decision Making in Healthcare*. Chichester: John Wiley & Sons, Ltd, 2012.

19 NHS Improvement. *National Schedule of Reference Costs, 2016-17 - NHS Trusts and NHS Foundation Trusts*. London: NHS Improvement, 2017.

20 Briggs A, Claxton K, Sculpher M. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press, 2006.

21 Culligan PJ, Myers JA, Goldberg RP, Blackwell L, Gohmann SF, Abell TD. Elective cesarean section to prevent anal incontinence and brachial plexus injuries associated with macrosomia – a decision analysis. *International urogynecology journal and pelvic floor dysfunction*. 2005;16:19–28; discussion.

22 Regier DA, Petrou S, Henderson J, Eddama O, Patel N, Strohm B, et al. Cost-effectiveness of therapeutic hypothermia to treat neonatal encephalopathy. *Value Health* 2010;13:695–702.

23 Tagin MA, Woolcott CG, Vincer MJ, Whyte RK, Stinson DA. Hypothermia for neonatal hypoxic ischemic encephalopathy: an updated systematic review and meta-analysis. *Arch Pediatr Adolesc Med* 2012;166:558–66.

24 Edwards AD, Brocklehurst P, Gunn AJ, Halliday H, Juszczak E, Levene M, et al. Neurological outcomes at 18 months of age after moderate hypothermia for perinatal hypoxic ischaemic encephalopathy: synthesis and meta-analysis of trial data. *BMJ (Clinical research ed)* 2010;340:c363.

25 Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. *J Health Econ* 1996;15:209–31.

26 York Health Economics Consortium. Quality-Adjusted Life Year (QALY) York: York Health Economics Consortium, 2016 [http://www.yhec.co.uk/glossary/quality-adjusted-life-year-qaly/]. Accessed 16 May 2019.

27 Petrou S, Kim SW, McParland P, Boyle EM. Mode of delivery and long-term health-related quality-of-life outcomes: a prospective population-based study. *Birth (Berkeley, Calif)* 2017;44:110–9.

28 EuroQol Group. EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy (Amsterdam, Netherlands)* 1990;16:199–208.

29 National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal*. London: National Institute for Health and Clinical Excellence, 2013.

30 Baio G, Dawid AP. Probabilistic sensitivity analysis in health economics. *Stat Methods Med Res* 2015;24:615–34.

31 Curtis L, Burns A. *Unit Costs of Health and Social Care 2017: Personal Social Services Research Unit*. Canterbury: University of Kent, 2017.

32 Quenby S, Gornall A, Fisher J, Deshpande S, Dewdney J, Hillyer K, et al. Induction of labour for predicted macrosomia. *'The Big Baby Trial'* 2018 [https://warwick.ac.uk/fac/sci/med/research/ctu/trials/bigbaby/protocol/big_baby_protocol_v5.0_19dec18.pdf]. Accessed 16 May 2019.

33 Institute for Quality and Efficiency in Health Care (IQWiG). Pregnancy and birth: Ultrasound scans in pregnancy Cologne, Germany: Institute for Quality and Efficiency in Health Care (IQWiG), 2012. [https://www.ncbi.nlm.nih.gov/books/NBK343308/]. Accessed 25 March 2015.

34 Smith GC, Pell JP, Dobbie R. Caesarean section and risk of unexplained stillbirth in subsequent pregnancy. *Lancet (London, England)* 2003;362:1779–84.

35 Gray R, Quigley MA, Hockley C, Kurinczuk JJ, Goldacre M, Brocklehurst P. Caesarean delivery and risk of stillbirth in subsequent pregnancy: a retrospective cohort study in an English population. *BJOG* 2007;114:264–70.

36 Keag OE, Norman JE, Stock SJ. Long-term risks and benefits associated with cesarean delivery for mother, baby, and subsequent pregnancies: systematic review and meta-analysis. *PLoS Med* 2018;15:e1002494.

ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Economic Evaluation

Late Pregnancy Ultrasound to Screen for and Manage Potential Birth Complications in Nulliparous Women: A Cost-Effectiveness and Value of Information Analysis

Edward C.F. Wilson, PhD, David Wastlund, MSc, Alexandros A. Moraitis, MSc, Gordon C.S. Smith, DSc

ABSTRACT

Background: Fetal growth restriction is a major risk factor for stillbirth. A routine late-pregnancy ultrasound scan could help detect this, allowing intervention to reduce the risk of stillbirth. Such a scan could also detect fetal presentation and predict macrosomia. A trial powered to detect stillbirth differences would be extremely large and expensive.

Objectives: It is therefore critical to know whether this would be a good investment of public research funds. The aim of this study is to estimate the cost-effectiveness of various late-pregnancy screening and management strategies based on current information and predict the return on investment from further research.

Methods: Synthesis of current evidence structured into a decision model reporting expected costs, quality-adjusted life-years, and net benefit over 20 years and value-of-information analysis reporting predicted return on investment from future clinical trials.

Results: Given a willingness to pay of £20 000 per quality-adjusted life-year gained, the most cost-effective strategy is a routine presentation-only scan for all women. Universal ultrasound screening for fetal size is unlikely to be cost-effective. Research exploring the cost implications of induction of labor has the greatest predicted return on investment. A randomized, controlled trial with an endpoint of stillbirth is extremely unlikely to be a value for money investment.

Conclusion: Given current value-for-money thresholds in the United Kingdom, the most cost-effective strategy is to offer all pregnant women a presentation-only scan in late pregnancy. A randomized, controlled trial of screening and intervention to reduce the risk of stillbirth following universal ultrasound to detect macrosomia or fetal growth restriction is unlikely to represent a value for money investment.

Keywords: economic evaluation, pregnancy, third trimester, ultrasound, value of information analysis.

VALUE HEALTH. 2021; 24(4):513–521

Introduction

Complications of pregnancy, both to mother and baby, are a major determinant of the Global Burden of Disease.[1] Stillbirth, defined as the baby born dead at 24 weeks gestational age or later, is a major contributor to this: there was a total of 2689 stillbirths in England and Wales in 2018, equating to approximately 0.4% of all births.[2] Fetal growth restriction (FGR) occurs when the baby fails to achieve its genetically determined growth potential, and it is a major risk factor for stillbirth.[3] It is possible that offering a routine ultrasound scan to every mother in late pregnancy (around 36 weeks gestational age) could help detect FGR, allowing intervention to reduce the risk of stillbirth. Furthermore, an ultrasound scan has the potential to detect other conditions that place the pregnancy at risk, such as macrosomia (birthweight >4 kg) and fetal presentation (cephalic or breech).

Under current guidelines in England and the rest of the United Kingdom,[4,5] an ultrasound scan after 28 weeks is offered only where clinically indicated (eg, relevant medical history or concerns following clinical examination). An alternative approach is to offer an ultrasound scan to all late-stage pregnancies. This would be expected to identify more pregnancies in need of intervention. However, this could also increase false positive diagnoses, leading to unnecessary, and possibly harmful, interventions. The overall balance of risk to harm to fetal health and whether such a screening program would represent the best use of healthcare resources is unknown, and the need to evaluate this has been highlighted previously.[6–8]

A Cochrane review (searching to August 2014) of routine ultrasound in late-stage pregnancy concluded that there was insufficient evidence to recommend universal screening.[9] However, none of the 13 trials studied screening followed by an intervention; the

different trials applied different definitions of screen positive and performed assessments at different gestational ages, and even the meta-analysis was underpowered for plausible estimates of diagnostic and interventional effectiveness.[10]

The key pieces of information that can be obtained from a scan around 36 weeks are whether the fetus measures small gestational age (SGA) or large for gestational age (LGA; defined as fetal size in the 1st or 10th decile of the distribution, respectively) and whether the fetus is in a cephalic (head down) presentation. An SGA fetus may be suffering FGR and hence may be at increased risk of stillbirth, whereas an LGA fetus may be macrosomic at delivery (defined as birthweight over 4 kg), which increases the risk of complications during delivery. We previously reported analyses of a level 1 study of diagnostic effectiveness[11] (where the results of the ultrasound scan were blinded) in relation to extremes of fetal size,[12,13] and we have also reported that, in the same cohort study, a late pregnancy scan identified about 2.5% of women with a previously undiagnosed breech presentation at 36 weeks.[14] Our previous work has also estimated the cost-effectiveness scanning for each of these individually, concluding that scanning for LGA[15] and SGA (Wastlund, et al [unpublished data]) is unlikely to be worthwhile. However, we predict that a presentation scan could prevent around 8 perinatal deaths per annum and could be cost-neutral to the English National Health Service (NHS) if able to be performed by a midwife as part of a routine antenatal appointment.[14]

In this article we build on this work, comparing all screening and management strategies simultaneously within one decision model framework. Critically, we use our framework to estimate overall decision uncertainty and perform a value of information analysis[16–18] to determine whether there is sufficient evidence to make a policy recommendation or whether investment in further research—for example, a randomized, controlled trial or other data-gathering exercise—would represent value for money for a major public sector funder of research (the National Institute for Health Research, England, United Kingdom). This is of particular importance, given that most existing studies (and systematic review[19]) were underpowered to detect a statistically significant difference in stillbirth rates between routine and selective screening arms. A new and sufficiently powered clinical trial would need to be extremely large, and thus expensive. It is vital, therefore, to consider whether this is the best use of scarce public funds, or whether more health could be generated for the population from investment in other studies or direct patient care.

Methods

Population

The target population is singleton nulliparous pregnancies (ie, babies born to new mothers) in England.

Comparator Strategies

The comparator strategies comprise both a screening option and subsequent management. Screening options are "selective," "universal breech," and "universal." All scans are assumed to take place at between 36 weeks and 36 weeks +6 days gestational age. "Selective" screening means only those mothers who are clinically indicated for a late pregnancy scan receive one, assumed to reflect the status quo.[4,5] The universal breech scanning strategy offers all mothers a simple presentation-only scan, that is, solely to determine the orientation of the fetus. It is assumed performed by a midwife using a point-of-care ultrasound device as part of a routine antenatal contact. Universal screening is defined as all mothers receiving an ultrasound scan incorporating measurements to estimate fetal size. Given the simplicity of establishing fetal presentation, this scan would also identify any babies in the breech position. Findings from a presentation scan can be either cephalic or breech, and fetal size could be either appropriate, small, or large for gestational age (AGA, SGA, and LGA, respectively).

If a breech presentation is identified, all mothers are assumed to be offered external cephalic version (ECV, manual manipulation of the mother's belly to turn the fetus to a cephalic presentation), unless contraindicated. If this is declined or unsuccessful, an elective Caesarean section may be scheduled. If LGA is detected, the mother may be offered either induction of labor or expectant management. If SGA is detected, all mothers are offered induction of labor.

We therefore compare 6 alternative screening and management policies comprising 3 possible screening modes and 2 alternative management plans, numbered 1 through 6 and summarized in Table 1.

Model Structure

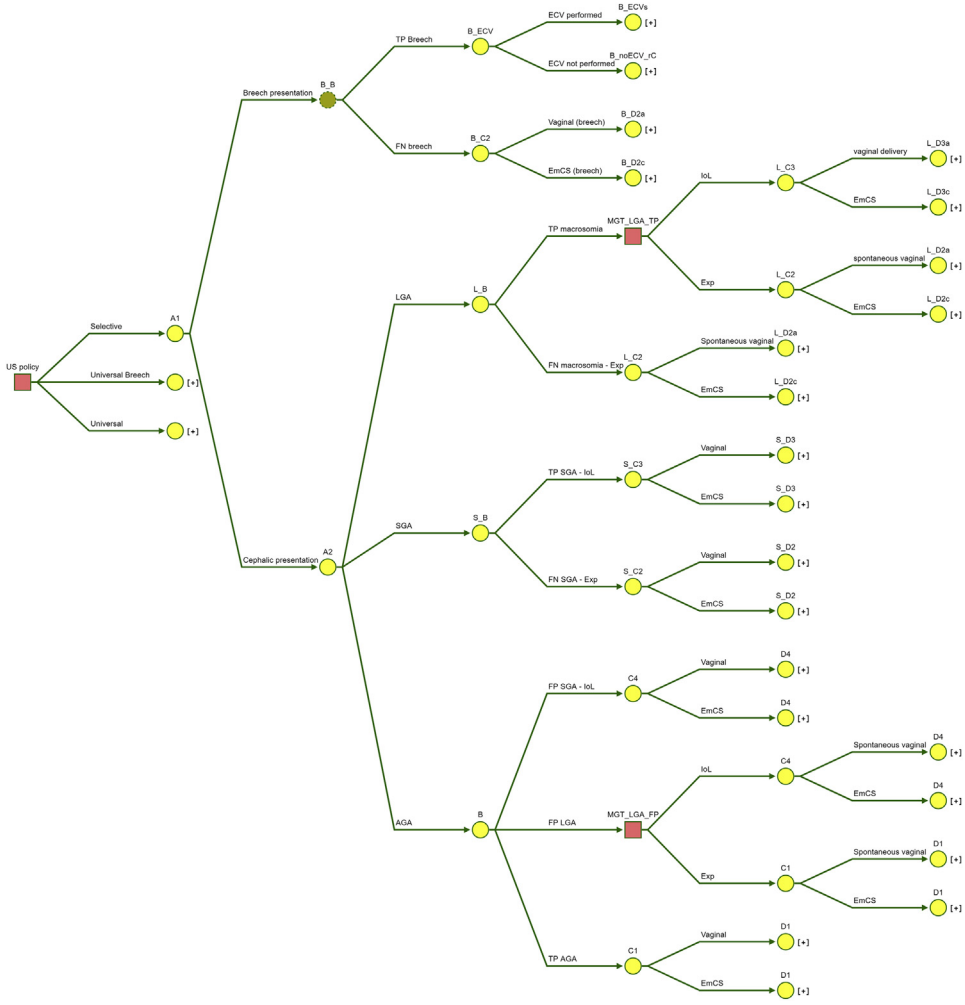
The model structure is a decision tree with 4 sections covering breech, LGA, SGA, and AGA (Fig 1 and Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005). It was established by discussion among the study team, comprising

Table 1. Comparator strategies and policies.

Strategy	Screen	Offered management if diagnosed:		
		Breech+	Macrosomia+	SGA+
1	Selective	ECV	IoL	IoL
2	Selective	ECV	Exp	IoL
3	Universal Breech	ECV	IoL	IoL
4	Universal Breech	ECV	Exp	IoL
5	Universal	ECV	IoL	IoL
6	Universal	ECV	Exp	IoL

ECV indicates external cephalic version; Exp, expectant management; IoL, induction of labor; SGA, small for gestational age.

Figure 1. Model structure overview: Screening-management options and fetal conditions. [+] = subbranches of model collapsed for clarity; see Appendix 1 (in the Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005) for expanded nodes. Nodes with the same letter have identical subsequent structures, while a different number and lowercase letter indicates different probabilities assigned to the next subbranch. The prefix before the underscore indicates a set of probabilities relevant to breech (B_), LGA (L_), or SGA (S_) For example, nodes D1 and D4 have identical substructures, but D1 relates to AGA babies delivered spontaneously, whereas D4 relates to AGA babies wrongly diagnosed as SGA or LGA and undergoing induction of labor unnecessarily.



AGA indicates appropriate gestational age; ECV, external cephalic version; EmCS, emergency Caesarean section; Exp, expectant management; FN, false negative; FP, false positive; IoL, induction of labor; LGA, large gestational age; SGA, small gestational age; TP, true positive; TN, true negative; US, ultrasound.

economists and clinicians. For parsimony, we assume they are all mutually exclusive. This is logically true for LGA, SGA, and AGA, but a baby may be both breech and LGA, for example. The structure is arranged hierarchically, with breech position first because this is most easily and reliably identified.

We assume a presentation-only scan is perfectly predictive of breech (ie, 100% sensitive and specific). However, our model allows for false negatives that are interpreted as undetected breech deliveries under the selective scanning strategy (node B_B, Fig 1). Where breech is detected, ECV is offered, which may be successful or not. If unsuccessful, an elective Caesarean section may be scheduled. In either case, the baby may spontaneously revert to breech or cephalic position. Reversion to breech can lead to a vaginal breech delivery or emergency Caesarean section. Outcomes from delivery comprise none, moderate or severe morbidity, or stillbirth. Surviving infants could subsequently have no long-term complications, special educational needs, severe neurological morbidity, or neonatal/infant mortality. The risk of long-term complications increases with neonatal morbidity severity (see Appendix Figure 1.1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005).

An LGA baby may or may not be diagnosed as such, determined by the sensitivity of the scan (Fig 1, node L_B). A positive scan can be managed either with induction of labor or expectant management, determined by the overall strategy (Table 1, Fig 1, node MGT_LGA_TP). Induction is assumed to reduce the risk of emergency Caesarean section (Fig 1, nodes L_C3 and L_C2, respectively). Delivery of a macrosomic baby leads to either no complications or respiratory morbidity, shoulder dystocia (trapping of the shoulder behind the mother's pubic bone) with attendant risk of transient or permanent brachial plexus injury (damage to the nerves of the arm) and acidosis (lowered blood pH usually due to build-up of carbon dioxide), other acidosis (ie, not related to shoulder dystocia), or neonatal

mortality (Appendix Figure 1.2 in Supplemental Materials found online at https://doi.org/10.1016/j.jval.2020.11.005). Long-term complications are divided into none, special educational needs, severe neurological morbidity, and neonatal/infant mortality (mirroring the structure of the breech arms; see Appendix Figure 1.1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005).

An SGA baby diagnosed as such will undergo induction of labor, with either a vaginal or emergency Caesarean section as the delivery mode. Undetected SGA babies are not induced and undergo either vaginal or emergency Caesarean section, with differing probabilities (Figure 1, nodes S_B, S_C3 and S_C2, respectively). Infants are then at risk of none, moderate or severe morbidity, or stillbirth, with long-term outcomes comprising no complications, special educational needs, severe neurological morbidity, and neonatal/infant mortality (Appendix Figure 1.3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005), mirroring the structure of the breech arms (Appendix Figure 1.1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005).

AGA babies may be falsely diagnosed as SGA or LGA, in which case the management and patient pathways are as per the true positives described above (Figure 1, node B). However, the risks of adverse outcomes vary as described below ("Model Data"). Babies correctly identified as AGA undergo routine deliveries, with a "background" risk of conversion to emergency Caesarean section for reasons other than fetal size or presentation (Figure 1, node C1). The expanded tree for AGA babies is shown in Appendix Figure 1.4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005.

Model Data

Data to populate the model were extracted from multiple sources in the literature[12–14,20–62] (Appendix Table 2.1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005). Good-quality systematic reviews and meta-analyses were prioritized, followed by large, good-quality clinical trials or cohort studies as appropriate. Where possible, probabilities were expressed as a baseline and odds ratio (or relative risk where odds were not calculable). Unit costs pertained to a 2016-2017 price year. Care was taken to appropriately reflect uncertainty in all parameters, as specified in the assigned probability distributions (see Appendix Table 2.1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005). Where no evidence for a parameter existed, we relied on expert opinion either to judge whether a study in a related area provided a sufficient proxy, or to provide a central estimate and credible interval representing beliefs about plausible values for the parameter. Source data for parameters were assigned a subjective quality rating, with high representing a source of directly relevant data and low representing use of indirectly relevant or no data, revised with expert opinion. Model inputs and details of derivation are reported in Appendix 2 (in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005).

Analysis

The model was analyzed via Monte Carlo simulation, with the appropriate number of simulations determined by the trade-off between minimizing Monte Carlo error and computational expense (Appendix 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005). Model outcomes comprised mean, variance, and covariance of costs and quality-adjusted life-years (QALYs), reported as mean and 95% credibility intervals for cost, QALYs, and net benefits calculated at £20 000 per QALY. We also report incremental net benefit relative to strategy 1 (selective scanning and induction of labor for SGA and LGA). Decision uncertainty is illustrated with cost-effectiveness acceptability curves. All costs were from a third-party payer perspective (the English NHS), and the health consequences from a fetal perspective only. All costs and QALYs were discounted by 3.5% annually, as recommended by NICE.[63] The time horizon was 20 years in the base-case scenario. Costs in other currencies were converted to GBP (£) by the exchange rate of the respective year. All prices were updated to the price level of 2016-2017 using the hospital and community health services (HCHS) index.[55]

To complement the probabilistic sensitivity analysis, we also investigated the model's sensitivity to key parameters through 1-way sensitivity analysis. Further, our base-case analysis assumed early labor induction would only affect long-term fetal outcomes via its impact on neonatal outcomes. However, there is evidence suggesting that induction of labor may of itself increase the risk of special educational needs in later life.[39] We therefore explore the impact of an independent effect of induction of labor on the risk of special educational needs.

We report the per-patient (ie, per mother/infant dyad) and population expected value of perfect information (EVPI) at a willingness to pay of £20 000/QALY and the expected value of perfect parameter information (EVPPI) for each parameter individually using the Sheffield Accelerated Value of Information (SAVI) tool.[64] Parameters with a positive EVPPI were grouped into those that could be collected within a single research study and the EVPPI for that group of parameters calculated. The expected value of sample information (EVSI) for any parameters or groups of parameters was then calculated using the method of moment matching with 30 nested samples.[65] EVPPI and EVSI calculations are traditionally extremely computationally expensive. The SAVI and moment matching methods generate statistical approximations, allowing calculation within a feasible timeframe. Briefly, SAVI estimates the EVPPI via a generalized additive model with nonparametric smoothing applied to the sampled input parameter set and resulting net benefits. Our implementation of the moment matching method relies on the conjugate distribution of the respective prior to estimate the preposterior distribution for a given study sample size (see Appendix 3 for code and walk-through). Population values are calculated over a time horizon of 10 years and as a "conservative" estimate, assuming the information is only of value to singleton nulliparous pregnancies resulting in a beneficial population of 1 689 663 and again with a broader estimate that assumes the information is of value to all pregnancies in England (n = 5 477 940; Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005).

The model was coded in R[66] and associated packages.[67–73] Full model code is available from the corresponding author upon request.

Results

Economic evaluation results are presented based on 100 000 simulations of the model. Value of information analysis statistics are based on 10 000 simulations (stability testing results reported in Appendix 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005).

Given current evidence and assuming a willingness to pay of £20 000 per QALY, the strategy associated with the highest net benefit is strategy 3: a presentation-only scan for all women (unless further screening is clinically indicated) with induction of labor where LGA or SGA are suspected. The added benefits from universal

Table 2. Cost effectiveness results (per mother scanned).

Screening + management	Cost (£)	QALYs	NB|£20k	INB|£20k	P_CE|£20k
1. Selective US + IoL for LGA	6090 (4420, 7890)	13.640 (13.441, 13.841)	£266 719 (£262 333, £271 079)	£0 (£0, £0)	0.65%
2. Selective US + Exp for LGA	6091 (4424, 7889)	13.639 (13.439, 13.839)	£266 682 (£262 297, £271 040)	–£37.09 (–£124.7, £35.24)	0.22%
3. **Universal US for breech + IoL for LGA** *	**6101 (4443, 7887)**	**13.645 (13.446, 13.846)**	**£266 806 (£262 426, £271 154)**	**£87.36 (£4.88, £205.68)**	**44.19%**
4. Universal US for breech + Exp for LGA	6102 (4446, 7887)	13.644 (13.444, 13.844)	£266 769 (£262 389, £271 120)	£50.29 (–£68.06, £186.43)	15.63%
6. Universal US + Exp for LGA	6178 (4508, 7972)	13.646 (13.446, 13.846)	£266 734 (£262 351, £271 099)	£14.47 (–£133.98, £173.31)	0.51%
5. Universal US + IoL for LGA	6180 (4498, 7983)	13.648 (13.448, 13.849)	£266 779 (£262 386, £271 147)	£60.24 (–£151.43, £281.7)	38.81%

Note. Strategies are listed in order of increasing cost (1,2,3,4,6,5). Management refers to management strategy when LGA is suspected, all cases of suspected SGA are assumed induced and breech to be offered ECV.
Exp indicates expectant Management; INB, incremental net benefit relative to current practice (strategy 1, selective US + induction of labor); IoL, induction of labor; NB, net benefit ; P_CE|£20k, probability of being the most cost-effective strategy given a willingness to pay of £20 000 per QALY gained.
*Strategy with highest expected net benefit (shown in bold).

ultrasound screening for fetal size are unlikely to justify its added cost (Table 2). However, there is substantial uncertainty associated with this recommendation, with only a 44% probability of this yielding the highest net benefit, and a 39% probability of universal screening being optimal (Table 2, Fig 2). As the willingness-to-pay threshold rises, the probability that universal screening becomes the most cost-effective strategy also rises (Fig 2).

One-way sensitivity analyses suggested that the cost-effectiveness outcomes were only sensitive to a few parameters: presentation-only scanning is the most cost-effective option if the time horizon of the analysis is below 45 years, above which universal screening becomes the most cost-effective option. A presentation-only scan remains the most cost-effective option, provided it costs no more than £90, above which status quo is the most cost-effective, and that the baseline stillbirth rate is below 0.28%, at which point universal scanning is most cost-effective. Finally, we found that the impact of induction of labor on risk of special educational needs would change the conclusions only if the relative risk of special educational needs was lower than 0.95 or above 1.3; observational data suggest that the effect is highly unlikely to be outside this range (Appendix 5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005).[39]

The per patient EVPI is £31.56. Given a population who can benefit from the information of 1 689 663 (see Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005), the population EVPI to England is £53.3 million. If the results of the analysis are assumed generalizable to all pregnancies in England, then the population EVPI is £172.9 million (Table 3). Only 5 input parameters yielded a population EVPPI greater than £100 000, and these logically group into 3 clusters of outcome measures that could be collected in possible future studies or RCTs (labeled studies 1, 2, and 3). The parameter with the greatest EVPPI is the difference in net cost of induced versus noninduced deliveries, accounting for 84% of the EVPI. No other parameters individually account for more than 1% of the total EVPI (Table 3).

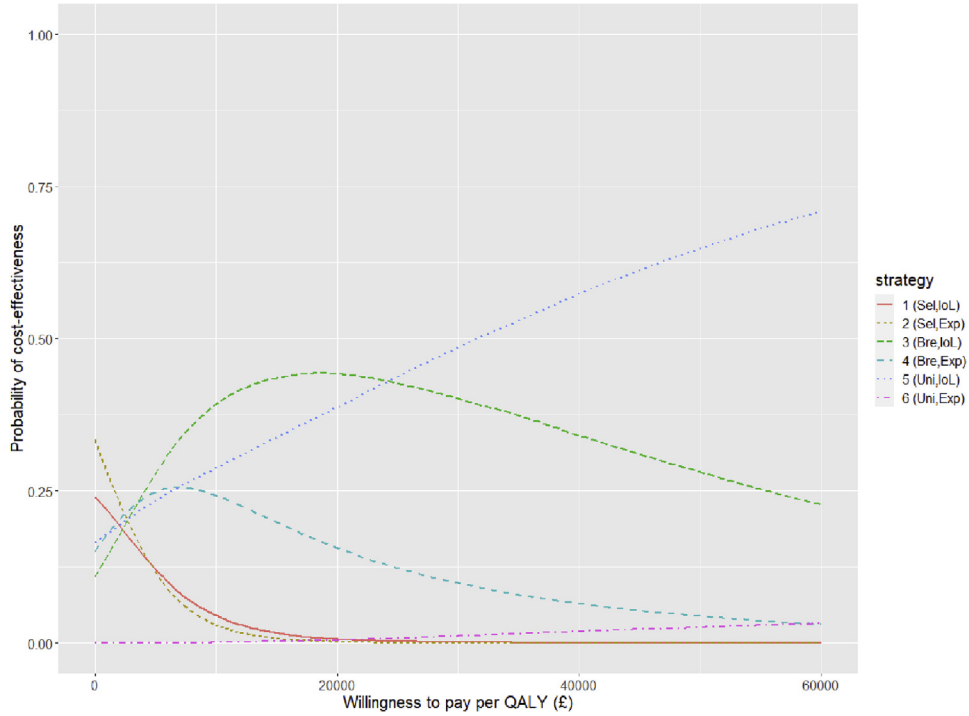
EVSI analysis of study 1, exploring the net cost difference between early labor induction and expectant management suggests scope for it to yield a positive return on investment. For example, a study with 1000 patients (in each arm of a 2-arm study) has an EVSI to England of £11.3 million (or £97.2 million if this information is of value to all pregnancies in England, not just low-risk nulliparous singleton pregnancies; Figure 3). If such a study were to cost £1 million, it would yield a net return on investment of at least £10.3 million. The EVSI algorithm was not able to estimate an EVSI for studies 2 and 3; after investigation, we concluded that for very low EVSIs, the approximation method is not able to return a value. We therefore conclude that the EVSI is very low; thus studies collecting data on the respective parameters are unlikely to be worth more than the cost of collecting them.

Discussion

Given current information, the most cost-effective strategy for late-pregnancy ultrasound scanning is to offer all women a presentation-only scan (those women who are currently indicated to undergo a full third-trimester ultrasound scan to continue to do so), and where SGA or LGA are suspected, the mother should be offered induction of labor, unless otherwise contraindicated. Given current thresholds,[63] universal routine ultrasound screening to assess fetal size is not cost-effective.

There is substantial decision uncertainty around this recommendation. However, the expected value of eliminating all uncertainty is only worth a maximum of £172.9 million or 8644 QALYs to the population of England (assuming £20 000 per QALY). This represents the expected opportunity loss due to the probability that the above recommendation is incorrect (crudely, the probability of being "wrong" multiplied by the consequence of being "wrong"). The majority of the EVPI is concentrated in a single parameter, namely the difference in cost as a result of early induction of labor. This is somewhat surprising, but arises due to the large standard error around the relevant model parameter (Appendix Table 2.1, row "Induction of labor" in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.005). This is because the cost encompasses not only the cost of inducing a pregnancy itself, but the costs of delivery and antenatal visits, which may or may not be avoided, too. Induction also has an uncertain impact on complications and hence long-term cost and outcomes of delivery. On top of this, less than perfect sensitivity and specificity of the scans at detecting LGA and SGA babies magnify the impact of uncertainty in the cost and outcomes of induction of labor. The EVSI of this parameter suggests that a study of "reasonable" size (eg,

Figure 2. Cost-effectiveness acceptability curves showing probability of cost-effectiveness as a function of willingness-to-pay for an additional quality-adjusted life-year.



Bre indicates universal presentation-only scan; Exp, expectant management if LGA suspected; IoL, induction of labor if LGA suspected; Sel, selective scanning; Uni, universal scan of fetal biometry and presentation.

1000 mothers per arm with a cost of £1 million) would likely yield a highly positive return on investment.

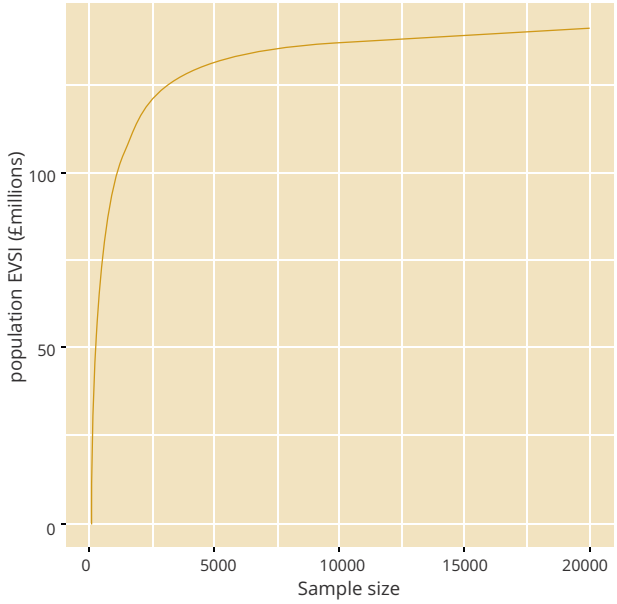
An ideal study design to measure the cost difference would be a study randomizing mothers to induction of labor or not, irrespective of indication. This is likely to raise ethical issues and would require careful consideration of the pros and cons and risks to mothers and their babies, based on the current state of knowledge. A nonrandomized study design (eg, database or cohort analysis) would be feasible but at risk of bias. The mathematics of value of information analysis are blind to whether reducing uncertainty in a parameter is ethical or not, or even possible or not. Instead, as with all economic evaluation, they

Table 3. Expected value of perfect information.

	Per person EVPPI (£expected, SE)	% of EVPI	pEVPPI (£)*	pEVPPI (£)*
Study 1				
Cost difference from early induction of labor	26.51 (0.07)	84	44 790 000	145 200 000
Study 2				
RR for acidosis in macrosomic fetuses if induced early	0.27 (0.04)	1%	456 000	1 478 000
OR for mortality if fetus is macrosomic	0.26 (0.03)	1%	438 900	1 423 000
Group	0.72 (0.07)	2%	1 215 199	3 939 513
Study 3				
RR for emergency CS among SGA fetuses following early labor induction	0.06 (0.01)	0%	99 290	321 900
OR for severe neonatal morbidity if fetus is SGA	0.03 (0.01)	0%	48 740	158 000
Group	0.26 (0.04)	1%	443 104	1 436 484
Expected value of perfect information	31.56 (-)	100%	53 326 764	172 883 786

Note. Standard error around estimates of EVPPI are a result of the SAVI[73] approximation algorithm. The EVPI is calculated directly and thus has no associated standard error. Note sum of EVPPI will not usually equal the EVPI owing to interactions/correlations between input parameters.
CS indicates Caesarean section; EVPI, expected value of information; EVPPI, expected value of partial perfect information; OR, odds ratio; SE, standard error; RR, relative risk.
*First pEVPPI column assumes information is applicable just to the target population (nulliparous singleton pregnancies), second assumes the information is equally applicable to all births in England.

Figure 3. Expected value of sample information of study 1. Expected value of sample information as a function of sample size for a study of the cost-difference between early induction of labor versus expectant management.



provide a guide and input to the decision-making process. An important finding from our analysis is that there is no evidence that a large-scale RCT powered to detect a difference in stillbirth would be a worthwhile investment: the EVPPI from reducing uncertainty in stillbirth rates is worth less than £100 000, a sum for which it is not possible to deliver an RCT.

We believe our analysis represents the most plausible summary of the evidence on the costs and effects of different ultrasound screening and subsequent management strategies in late pregnancy. The decision model translates uncertainty in parameters (crudely, the standard errors around mean estimates of effect, cost, and health state utilities) to decision uncertainty (standard errors around mean estimates of net benefit). The value of information analysis then predicts the likely return on investment from reducing the SEs of the input parameters.

However, the validity of our conclusion rests entirely on the validity of the model. Although we believe we have appropriately captured parameter uncertainty, we have implicitly assumed that the structure of the model itself is "correct." Addressing such structural uncertainty is challenging in decision models. In theory it would require constructing many alternative models and comparing or averaging out the results, which would be prohibitively expensive. However, where possible we did explore structural uncertainty—for example, our base case assumed that all long-term morbidity was mediated through the risk of neonatal morbidity, while there is evidence to suggest an independent effect of induction of labor on risk of special educational needs. We explored this and found our conclusions to be robust to all but implausibly extreme assumptions as to the relative risk. Our analytic perspective was limited to fetal outcomes only, excluding maternal quality of life. This may underestimate the QALY gains from screening and so underestimate cost-effectiveness.

Second, our conclusions regarding the cost-effectiveness of presentation scanning are contingent on midwives being able to undertake the scan as part of a routine antenatal contact. This is currently unknown and requires a feasibility study to test. It should also be noted that the scans will certainly increase the burden on midwives while we predict a reduction in delivery complications. This will require a shift in resources from secondary care to (antenatal) midwifery. The budgetary mechanisms underlying this are not considered in our analysis. It is worth noting that our previous work[14] focusing only on presentation scans (and not including the alternative strategies considered here) concluded that a presentation-only scan was cost-effective so long as it could be provided for £19.80 or less. Our analysis here, which models longer-term costs and outcomes in greater detail, suggests greater scope for cost-effectiveness, with our 1-way sensitivity analysis suggesting the scan remaining cost-effective so long as it can be provided for less than approximately £90 (Appendix Figure 4.2).

To our knowledge, this is the first value of information analysis estimating the return on investment from future research into late pregnancy ultrasound scans. Economic evaluations of obstetric investigations commonly include estimates of the value of perfect information—for example, there may be value in future studies on quality-of-life gains and costs of early detection of gestational diabetes,[74] the effects of interventions to prevent postnatal depression,[75] the cost-effectiveness of financial incentives for smoking cessation during pregnancy,[76] and possibly the effectiveness of a screening program to reduce periconceptional exposure to methylmercury.[77] However, we are not aware of any attempts to calculate the expected value of sample information from specific study designs in obstetrics.

Conclusion

Our results suggest that universal ultrasound for fetal presentation only may be both clinically and economically justified, but that implementation research is needed before it is adopted into routine care. Specifically, this must explore whether a scan can be conducted by a midwife during a routine antenatal visit. Universal ultrasound including estimation of fetal weight is of borderline cost-effectiveness and sensitive to certain assumptions. Our formal value of information analysis suggests that future research should be focused on the net cost of induction of labor compared to expectant management, and that there is unlikely to be value in a large-scale RCT of routine versus selective ultrasound screening powered to detect a difference in stillbirth rates.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2020.11.005.

Article and Author Information

Accepted for Publication: November 3, 2020

Published Online: January 21, 2021

doi: https://doi.org/10.1016/j.jval.2020.11.005

Author Affiliations: Health Economics Group, Norwich Medical School, University of East Anglia, Norwich, UK (Wilson); The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK (Wilson, Wastlund); Parexel Access Consulting, Parexel International, Stockholm, Sweden (Wastlund); Department of Obstetrics and Gynaecology, University of Cambridge, NIHR Cambridge Biomedical Research Centre, Cambridge, UK (Moraitis, Smith).

Correspondence: Edward C.F. Wilson, PhD, Health Economics Group, Norwich Medical School, University of East Anglia, Norwich, NR4 7TJ, United Kingdom. Email: ed.wilson@uea.ac.uk

Author Contributions: *Concept and design*: Wilson, Wastlund, Moraitis, Smith
Acquisition of data: Wilson, Smith
Analysis and interpretation of data: Wilson, Wastlund, Moraitis, Smith
Drafting of the manuscript: Wilson, Wastlund, Smith
Critical revision of the paper for important intellectual content: Wilson, Wastlund, Moraitis, Smith
Statistical analysis: Wilson, Wastlund
Provision of study materials or patients: Smith
Obtaining funding: Wilson, Smith
Supervision: Wilson, Smith

Conflict of Interest Disclosures: Drs Wilson, Moraitis, and Smith reported receiving grants from the National Institute for Health Research Health Technology Assessment Program during the conduct of the study. Dr Smith also reported receiving grants and personal fees from GlaxoSmithKline Research and Development Limited outside the submitted work; grants from Sera Prognostics Inc outside the submitted work; nonfinancial support from Illumina Inc outside the submitted work; and personal fees and nonfinancial support from Roche Diagnostics Ltd outside the submitted work. In addition, Dr Smith has a patent application pending for a novel predictive test for fetal growth restriction. No other disclosures were reported.

Funding/Support: This study was funded by the National Institute for Health Research Health Technology Assessment program under grant number 15/105/01.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The views expressed here are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, or the Department of Health.

Acknowledgment: The authors wish to thank Alex Heazell for comments on the manuscript.

REFERENCES

1. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet (London, England)*. 2017;390(10100):1151–1210.
2. Office for National Statistics. Births in England and Wales: summary tables, release 31st July 2019. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/birthsummarytables. Accessed January 11, 2021.
3. Gardosi J, Madurasinghe V, Williams M, et al. Maternal and fetal risk factors for stillbirth: population based study. *BMJ (Clinical Research Ed)*. 2013;346:f108-f08.
4. National Institute for Health and Care Excellence. *Clinical Guideline: Antenatal Care for Uncomplicated Pregnancies*. 2008.
5. Royal College of Obstetrics and Gynaecologists. *The Investigation and Management of the Small-for-Gestational-Age Fetus (Green-top guideline No. 31)*. 2013.
6. Henrichs J, Verfaille V, Viester L, et al. Effectiveness and cost-effectiveness of routine third trimester ultrasound screening for intrauterine growth restriction: study protocol of a nationwide stepped wedge cluster-randomized trial in The Netherlands (The IRIS Study). *BMC Pregnancy Childbirth*. 2016;16(1):310.
7. Le Ray C, Lacerte M, Iglesias MH, et al. Routine third trimester ultrasound: what is the evidence? *J Obstet Gynaecol Can*. 2008;30(2):118–122.
8. Le Ray C, Morin L. Routine versus indicated third trimester ultrasound: is a randomized trial feasible? *J Obstet Gynaecol Can*. 2009;31(2):113–119.
9. Bricker L, Medley N, Pratt JJ. Routine ultrasound in late pregnancy (after 24 weeks' gestation). *Cochrane Database Syst. Rev.* 2015;(6):CD001451.
10. Smith GCS. Universal screening for foetal growth restriction: best practice and research. *Clin Obstetrics Gynaecol*. 2018;49:16–28.
11. Pasupathy D, Dacey A, Cook E, et al. Study protocol. A prospective cohort study of unselected primiparous women: the pregnancy outcome prediction study. *BMC Pregnancy Childbirth*. 2008;8:51.
12. Sovio U, White IR, Dacey A, et al. 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):2089–2097.
13. Sovio U, Moraitis AA, Wong HS, et al. Universal vs selective ultrasonography to screen for large-for-gestational-age infants and associated morbidity. *Ultrasound Obstet Gynecol*. 2018;51(6):783–791.
14. Wastlund D, Moraitis AA, Dacey A, et al. Screening for breech presentation using universal late-pregnancy ultrasonography: a prospective cohort study and cost effectiveness analysis. *PLoS Med*. 2019;16(4):e1002778.
15. Wastlund D, Moraitis AA, Thornton JG, et al. The cost-effectiveness of universal late-pregnancy screening for macrosomia in nulliparous women: a decision analysis. *BJOG*. 2019;126(10):1243–1250.
16. Pratt J, Raiffa H, Schlaifer R. *Introduction to Statistical Decision Theory*. Cambridge, MA: Massachusetts Institute of Technology; 1995.
17. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ*. 1999;18(3):341–364.
18. Wilson EC. A practical guide to value of information analysis. *Pharmacoeconomics*. 2015;33(2):105–121.
19. Heazell AE, Hayes DJ, Whitworth M, et al. Biochemical tests of placental function versus ultrasound assessment of fetal size for stillbirth and small-for-gestational-age infants. *Cochrane Database Syst Rev*. 2019;5:Cd012245.
20. Monier I, Blondel B, Ego A, et al. Poor effectiveness of antenatal detection of fetal growth restriction and consequences for obstetric management and neonatal outcomes: a French national study. *BJOG*. 2015;122(4):518–527.
21. Grobman WA, Rice MM, Reddy UM, et al. Labor induction versus expectant management in low-risk nulliparous women. *N Engl J Med*. 2018;379(6):513–523.
22. Blackwell SC, Refuerzo J, Chadha R, et al. Overestimation of fetal weight by ultrasound: does it influence the likelihood of cesarean delivery for labor arrest? *Amer J Obstet Gynecol*. 2009;200(3):340.e1–340.e3.
23. Middleton P, Shepherd E, Crowther CA. Induction of labour for improving birth outcomes for women at or beyond term. *Cochrane Database Syst Rev*. 2018;5:Cd004945.
24. Leung WC, Pun TC, Wong WM. Undiagnosed breech revisited. *Br J Obstet Gynaecol*. 1999;106(7):638–641.
25. Ben-Meir A, Elram T, Tsafrir A, et al. The incidence of spontaneous version after failed external cephalic version. *Amer J Obstet Gynecol*. 2007;196(2):157.e1–157.e3.
26. Morrison JJ, Rennie JM, Milton PJ. Neonatal respiratory morbidity and mode of delivery at term: influence of timing of elective caesarean section. *Br J Obstet Gynaecol*. 1995;102(2):101–106.
27. Ouzounian JG, Gherman RB. Shoulder dystocia: are historic risk factors reliable predictors? *Amer J Obstet Gynecol*. 2005;192(6):1933–1935. discussion 35-38.
28. Moraitis AA, Wood AM, Fleming M, et al. Birth weight percentile and the risk of term perinatal death. *Obstet Gynecol*. 2014;124(2 Pt 1):274–283.
29. Rossi AC, Mullin P, Prefumo F. Prevention, management, and outcomes of macrosomia: a systematic review of literature and meta-analysis. *Obstet Gynecol Survey*. 2013;68(10):702–709.
30. Chongsuvivatwong V, Bachtiar H, Chowdhury ME, et al. Maternal and fetal mortality and complications associated with cesarean section deliveries in teaching hospitals in Asia. *J Obstet Gynaecol Res*. 2010;36(1):45–51.
31. Gibson KS, Waters TP, Bailit JL. Maternal and neonatal outcomes in electively induced low-risk term pregnancies. *Amer J Obstet Gynecol*. 2014;211(3):249.e1–249.e16.
32. Boulvain M, Senat MV, Perrotin F, et al. Induction of labour versus expectant management for large-for-date fetuses: a randomised controlled trial. *Lancet (London, England)*. 2015;385(9987):2600–2605.
33. MacKenzie IZ, Shah M, Lean K, et al. Management of shoulder dystocia: trends in incidence and maternal and neonatal morbidity. *Obstet Gynecol*. 2007;110(5):1059–1068.
34. Sandmire HF, DeMott RK. The Green Bay cesarean section study. IV. The physician factor as a determinant of cesarean birth rates for the large fetus. *Amer J Obstet Gynecol*. 1996;174(5):1557–1564.
35. Stock SJ, Ferguson E, Duffy A, et al. Outcomes of elective induction of labour compared with expectant management: population based study. *BMJ (Clinical Research Ed)*. 2012;344:e2838.
36. Thorngren-Jerneck K, Herbst A. Low 5-minute Apgar score: a population-based register study of 1 million term births. *Obstet Gynecol*. 2001;98(1):65–70.
37. Hofmeyr GJ, Hannah M, Lawrie TA. Planned caesarean section for term breech delivery. *Cochrane Database Syst Rev*. 2015;(7):CD000166.
38. Pasupathy D, Wood AM, Pell JP, et al. Time trend in the risk of delivery-related perinatal and neonatal death associated with breech presentation at term. *Int J Epidemiol*. 2009;38(2):490–498.
39. MacKay DF, Smith GC, Dobbie R, et al. Gestational age at delivery and special educational need: retrospective cohort study of 407,503 schoolchildren. *PLoS Med*. 2010;7(6):e1000289.
40. Persson M, Razaz N, Tedroff K, et al. Five and 10 minute Apgar scores and risks of cerebral palsy and epilepsy: population based cohort study in Sweden. *BMJ (Clinical Research Ed)*. 2018;360:k207.
41. Iliodromiti S, Mackay DF, Smith GC, et al. Apgar score and the risk of cause-specific infant mortality: a population-based cohort study. *Lancet (London, England)*. 2014;384(9956):1749–1755.
42. Department of Health and Social Care. NHS reference costs 2016 to 2017. https://improvement.nhs.uk/resources/reference-costs/#rc1718. Accessed January 11, 2021.

43. Vijgen SM, Boers KE, Opmeer BC, et al. Economic analysis comparing induction of labour and expectant management for intrauterine growth restriction at term (DIGITAT trial). *Eur J Obstet Gynecol Reprod Biol*. 2013;170(2):358–363.
44. Palencia R, Gafni A, Hannah ME, et al. The costs of planned cesarean versus planned vaginal birth in the Term Breech Trial. *Canadian Med Assoc J*. 2006;174(8):1109–1113.
45. James M, Hunt K, Burr R, et al. A decision analytical cost analysis of offering ECV in a UK district general hospital. *BMC Health Serv Res*. 2001;1:6.
46. Alfirevic Z, Keeney E, Dowswell T, et al. Which method is best for the induction of labour? A systematic review, network meta-analysis and cost-effectiveness analysis. *Health Technol Assess*. 2016;20(65):1–584.
47. Culligan PJ, Myers JA, Goldberg RP, et al. Elective cesarean section to prevent anal incontinence and brachial plexus injuries associated with macrosomia: a decision analysis. *Int Urogynecol J Pelvic Floor Dysfunct*. 2005;16(1):19–28: discussion 28.
48. Mistry H, Heazell AE, Vincent O, et al. A structured review and exploration of the healthcare costs associated with stillbirth and a subsequent pregnancy in England and Wales. *BMC Pregnancy Childbirth*. 2013;13:236.
49. Barrett B, Mosweu I, Jones CR, et al. Comparing service use and costs among adolescents with autism spectrum disorders, special needs and typical development. *Autism*. 2015;19(5):562–569.
50. Access Economics. *The Economic Impact of Cerebral Palsy in Australia in 2007*. Sydney, NSW; 2008.
51. Young NL, Rochon TG, McCormick A, et al. The health and quality of life outcomes among youth and young adults with cerebral palsy. *Arch Phys Med Rehabil*. 2010;91(1):143–148.
52. Szende A, Janssen B, Cabasés J. *Self-Reported Population Health: An International Perspective Based on EQ-5D*. New York, NY: Springer; 2014.
53. Leigh S, Granby P, Turner M, et al. The incidence and implications of cerebral palsy following potentially avoidable obstetric complications: a preliminary burden of disease study. *BJOG*. 2014;121(13):1720–1728.
54. NHS Digital. *NHS Staff Earnings: Estimates ot September 2017*. Provisional Statistics; 2017.
55. Curtis L, Burns A. *Unit costs of health and social care 2017*. Personal Social Services Research Unit; 2017.
56. Curtis L. Unit Costs of health and Social Care 2008. Cornwallis Building, University of Kent, Canterbury, Kent CT2 7NF: Personal Social Services Research Unit, 2008.
57. NHS Purchasing and Supply Agency. *Cost-effectiveness of ultrasound elastography in the assessment of liver fibrosis*. 2009.
58. Malloy MH, Freeman DH. Respiratory distress syndrome mortality in the United States, 1987 to 1995. *J Perinatol*. 2000;20(7):414–420.
59. Curtis L, Burns A. *Unit costs of health and social care 2016*. Personal Social Services Research Unit; 2016.
60. Office for National Statistics. *National Life Tables, United Kingdom, 1980-82 to 2014-16*. Office for National Statistics; 2017.
61. Palisano R, Rosenbaum P, Walter S, et al. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol*. 1997;39(4):214–223.
62. Health and Social Care Information Centre. *NHS Maternity Statistics 2016-17*. NHS Digital; 2017.
63. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal*. London: NICE; 2013. https://www.nice.org.uk/process/pmg9/chapter/foreword. Accessed January 11, 2021.
64. Strong M, Oakley JE, Brennan A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. *Med Decision Making*. 2014;34(3):311–326.
65. Heath A, Baio G. Calculating the expected value of sample information using efficient nested monte carlo: a tutorial. *Value Health*. 2018;21(11):1299–1304.
66. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2017.
67. Baio G, Berardi A, Heath A. *BCEA: Bayesian Cost Effectiveness Analysis*. 2018.
68. Fan FY. *FinCal: Time Value of Money, Time Series Analysis and Computational Finance*. R package version 63; 2016.
69. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag; 2016.
70. Warnes G, Bolker B, Lumley T. *gtools: Various R Programming Tools*. 2018.
71. Wickham H, Bryan J. *readxl: Read Excel Files*. 2018.
72. Wickham H, Henry L. *tidyr: Easily Tidy Data with "spread()" and "gather()" functions*; 2019.
73. Strong M. *SAVI: SAVI Sheffield Accelerated Value of Information*. 2015.
74. Farrar D, Simmonds M, Griffin S, et al. The identification and treatment of women with hyperglycaemia in pregnancy: an analysis of individual participant data, systematic reviews, meta-analyses and an economic evaluation. *Health Technol Assess*. 2016;20(86):1–348.
75. Morrell CJ, Sutcliffe P, Booth A, et al. A systematic review, evidence synthesis and meta-analysis of quantitative and qualitative studies evaluating the clinical effectiveness, the cost-effectiveness, safety and acceptability of interventions to prevent postnatal depression. *Health Technol Assess*. 2016;20(37):1–414.
76. Boyd KA, Briggs AH, Bauld L, et al. Are financial incentives cost-effective to support smoking cessation during pregnancy? *Addiction*. 2016;111(2):360–370.
77. Gaskin J, Rennie C, Coyle D. Reducing periconceptional methylmercury exposure: cost-utility analysis for a proposed screening program for women planning a pregnancy in Ontario, Canada. *Environ Health Perspect*. 2015;123(12):1337–1344.