A systematic review and cost-effectiveness analysis of the case for screening nulliparous women in late pregnancy using ultrasound

Gordon CS Smith[1*], Alexandros A Moraitis[1], David Wastlund[2], Jim G Thornton[3], Aris Papageorghiou[4], Julia Sanders[5], Alexander EP Heazell[6], Stephen C Robson[7], Ulla Sovio[1], Peter Brocklehurst[8], Edward CF Wilson[2,9]

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

[2] The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom.

[3] Division of Child Health, Obstetrics and Gynaecology, School of Medicine, University of Nottingham, Nottingham, United Kingdom.

[4] Nuffield Department of Obstetrics and Gynaecology, Oxford, United Kingdom.

[5] School of Healthcare Sciences, Cardiff University, Cardiff, United Kingdom.

[6] Medicine and Health, Faculty of Biology, School of Medical Sciences, University of Manchester, Manchester, United Kingdom.

[7] Reproductive and Vascular Biology Group, The Medical School, University of Newcastle, Newcastle, United Kingdom.

[8] Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, United Kingdom.

[9] Health Economics Group, Norwich Medical School, University of East Anglia, Norwich, United Kingdom

* Corresponding author. Professor Gordon Smith: Department of Obstetrics and Gynaecology, Robinson way, Cambridge CB2 0SW, United Kingdom. gcss2@cam.ac.uk

Tel: 01223 763888/763890; Fax: 01223 763889

Declared competing interests of authors: 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. Ulla Sovio reports grants from NIHR Cambridge Biomedical Research Centre during the conduct of the study. Peter Brocklehurst reports grants and personal fees from MRC, grants from MRC, NIHR HS&DR, NIHR HTA, Wellcome Trust, personal fees from AG Biotest, outside the submitted work. Jim Thornton is a member of the NIHR HTA & EME Editorial Board. 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 growth restriction pending.

Key words: Ultrasonography, Pregnancy, Perinatal death, Fetal weight, Breech presentation, Fetal Macrosomia, Cost-Benefit Analysis, Decision Trees, Biometry

Abstract

Background.

Currently, pregnant women are screened using ultrasound at booking and around the middle of pregnancy. Ultrasound scans thereafter are performed for clinical indications only.

Objectives.

We sought to assess the case for offering universal late pregnancy ultrasound to all nulliparous women in the UK. The main questions addressed were to determine the diagnostic effectiveness of universal late pregnancy ultrasound to predict adverse outcome, and the cost effectiveness of either implementing universal ultrasound or conducting further research in this area.

Design

We performed diagnostic test accuracy reviews of five ultrasonic measurements in late pregnancy. We conducted cost effectiveness and value of information (VoI) analysis of screening for fetal presentation, screening for small for gestational age (SGA) fetuses and screening for large for gestational age (LGA) fetuses. We finally conducted a survey and a focus group to determine the willingness of women to participate in a future randomised trial.

Data sources

We searched Medline, EMBASE and the Cochrane library from inception.

Review methods

The protocol for the review was designed a priori and registered. Eligible studies were identified using key words with no restrictions for language or location. The risk of bias in studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 tool. Health economic modelling employed a decision tree analysed via Monte Carlo simulation. Health outcomes were from the fetal perspective and presented as quality-adjusted life years (QALYs). Costs were from the perspective of the public sector defined as the (English) NHS and costs of special educational needs. All costs and QALYs were discounted by 3.5% per annum and the reference case time horizon was 20 years.

Results

Umbilical artery Doppler, cerebro-placental ratio (CPR), severe oligohydramnios, and borderline oligohydramnios were all either non-predictive or weakly predictive of the risk of neonatal morbidity

(summary positive likelihood ratios [LR+] between 1 and 2) and were all weakly predictive of the risk of delivering a SGA infant (summary LR+ between 2 and 4). 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. Very few studies blinded the result of the ultrasound scan and most studies had high risk of bias through treatment paradox, ascertainment bias or iatrogenic harm. Health economic analysis indicated that universal ultrasound for fetal presentation only may be both clinically and economically justified on the basis of existing evidence. Universal ultrasound including fetal biometry was of borderline cost-effectiveness, and sensitive to assumptions. VoI analysis indicated that future research should be focused on the cost difference between IOL and expectant management.

Limitations

The primary literature on the diagnostic effectiveness of ultrasound in late pregnancy is weak. VoI analysis may have underestimated the uncertainty in the literature as it was focused on the internal validity of parameters, which is quantified, whereas the greatest uncertainty may be in the external validity to the research question, which is unquantified.

Conclusions

Universal screening for presentation at term may be justified on the basis of current knowledge. Universal screening for fetal growth disorders cannot currently be justified.

Future work

We describe proof of principle randomised controlled trials which could better inform the case for screening using ultrasound in late pregnancy.

Study registration: This study is registered as PROSPERO CRD42017064093

Funding details: The National Institute for Health Research Health Technology Assessment programme

Word count: 37,835

Contents

List of Tables

List of Figures

List of abbreviations

AFI: Amniotic fluid index

AGA: Appropriate for gestational age

CEAC: Cost-effectiveness acceptability curve

CI: Confidence interval

CP: Cerebral palsy

CPR: Cerebro-placental ratio

CS: Caesarean section

DOR: Diagnostic odds ratio

ECV: External cephalic version

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

INMB: Incremental net monetary benefit

IOL: Induction of labour

LGA: Large for gestational age

LR: Likelihood ratio

NMB: Net monetary benefit

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-pay

Scientific summary

Background

Currently, pregnant women are screening using two-dimensional ultrasound at booking and around the middle of pregnancy. Ultrasound scans thereafter are performed for clinical indications only. Ultrasound has a key role in the management of complicated pregnancies, being used in the assessment of presentation, fetal size, biophysical indicators of fetal well-being and assessment of blood flow using Doppler flow velocimetry. There is evidence that ultrasound might be effective in screening low risk and unselected women. Moreover, induction of labour at term is a reasonable candidate intervention for women who screen high risk. However, the diagnostic accuracy of many ultrasonic features is unknown in low risk populations. Moreover, there is little information on the cost-effectiveness of screening and intervention. Finally, it is uncertain whether further research on screening low risk women is feasible or cost-effective.

Objectives

The objectives of the present study, outlined in the original application, were as follows:

1. To assess the diagnostic effectiveness of late pregnancy ultrasound in nulliparous women based on the existing research literature.
2. Having identified the key ultrasonic findings which identified women as high risk to review the existing literature and current guidelines to identify a management plan for women with high risk characteristics.
3. To conduct a health economic analysis of the likely cost-effectiveness of screening and intervention based on the best available evidence of the costs, diagnostic effectiveness of ultrasound and clinical effectiveness of intervention.
4. To perform a value of information analysis to determine whether there is a strong economic case for funding future research in this area.
5. Conditional on the above, to outline the design a randomised controlled trial which could strengthen the evidence base relating to the issues above.

Methods

We identified the following as key ultrasound measurements 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. We found that there was an

on-going Cochrane Diagnostic Test Accuracy review for SGA, hence we focused on the other five measures. The protocol for the reviews was designed a priori and registered with the PROSPERO register of systematic reviews (CRD42017064093). We searched Medline, EMBASE and the Cochrane library from inception. The studies were identified using a combination of keywords. Selection criteria included cohort or cross-sectional studies with singleton pregnancies which had an ultrasound performed ≥24wkGA. Case-control studies were excluded. 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. The literature search, study selection, and analysis ware performed independently by two researchers using Review Manager 5.3. Any differences were resolved in discussion with the senior author. The risk of bias in each included study was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS 2) tool as described in the Cochrane Handbook of Diagnostic Test Accuracy Studies. 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).

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) respectively. For the data synthesis we used a hierarchal summary receiver operating characteristic curve model. 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. We also pooled the diagnostic odds ratios (DORs) using the method described by Deeks and used the Deeks' funnel plot asymmetry test for publication bias in which $P<0.05$ was defined as significant asymmetry. For the statistical analyses we used the METANDI, METAN and MIDAS packages from STATA version 14 (StataCorp LP, College Station, TX).

We included studies regardless of blinding of the ultrasound to the clinicians but this was reported in the study characteristics. However, revealing the scan result has the potential for multiple biases. We had access to the original data from the Pregnancy Outcome Prediction study (Lancet 2015). This is the larger of only two studies that performed blinded ultrasonic assessment near term in nulliparous

women. The other study (Genesis, Perinatal Ireland) has not yet been widely reported. Given the importance of blinding we performed a number of new analyses of the POP study dataset.

Health economic modelling employed a decision tree analysed via Monte Carlo simulation (repeated sampling from input parameter distributions) and coded in R (an open source statistical software package). Health outcomes were from the fetal perspective and presented as quality-adjusted life years (QALYs). Costs were from the perspective of the public sector, defined as NHS and cost of special educational needs. All costs and QALYs were discounted by 3.5% per annum and the reference case time horizon was 20 years. The health economic analysis evaluated three different strategies for ultrasound screening in late pregnancy, defined as a scan between 36+0 weeks and 36+6 weeks: (i) 'Selective US' (i.e. where ultrasound is only performed following clinical indication of its need), the current standard of care in England, (ii) 'Universal US for presentation only', i.e. scan with the sole purpose of detecting breech presentation, and (iii) 'Universal US for fetal size' i.e. a scan performing ultrasonic assessment of fetal weight plus assessment of presentation.

We assumed that all identified cases of breech presentation would be offered an external cephalic version unless contraindicated, in line with RCOG guidelines. We further assumed that pregnancies identified as SGA (whether correctly diagnosed or not) would be given early induction of labour (IOL) at 37 weeks' gestation. However, for pregnancies diagnosed as LGA, there is uncertainty as to whether intervention (IOL) is beneficial. For this reason, expectant management of suspected LGA pregnancies was also an option. We assumed that selective scanning (i.e. only where clinically indicated) with a policy of offering ECV for suspicion of breech, and IOL for suspicion of SGA or LGA represents an approximation of the status quo from which estimates of incremental net benefit are calculated.

Results

We identified 13 studies of umbilical artery Doppler that met our inclusion criteria including 67,764 patients in total. Umbilical artery 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. We identified 16 studies of CPR that met our inclusion criteria involving 121,607 patients in total. Meta-analysis demonstrated that the 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. Consequently, the 95% confidence intervals for the positive LR are wide and include the point estimates observed for UA Doppler for both SGA and severe SGA. We identified 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, these associations are more difficult to interpret. First, for both of these outcomes, the association was weaker than it was for SGA. Second, in both cases the associations could be a consequence of the scan rather than an outcome predicted by the scan as only two studies containing <5% of the patients included in the meta-analysis blinded the results of the scan. We identified 11 studies of borderline oligohydramnios (including the POP study) that 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. We identified 40 studies of LGA that met our inclusion criteria involving 66,187 patients in total. Ultrasonic suspicion of fetal macrosomia was strongly predictive of the risk of delivering a large baby but was 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.

Based on current information, and assuming a willingness to pay of £20,000 per QALY, offering a universal ultrasound (US) presentation-only scan is on average the most cost-effective strategy. This is associated with an incremental net monetary benefit of £87.36 (95% CI: 4.88, 205.68) per pregnancy compared to current practice. Scaled up to the English population, this equates to a net benefit of £17.1m or 857 QALYs per annual birth cohort. This is the present value of the future flows of expected costs and benefits over a time horizon of 20 years. Due to uncertainties in the evidence base (parameter uncertainty), there is a only a 44.19% probability that this conclusion is correct, i.e. there is a 55.81% probability that this conclusion is incorrect in which case a loss will be incurred. The expected loss associated with this decision uncertainty is £31.56 per pregnancy. Equivalently, this is

the expected gain if uncertainty were to be eliminated (expected value of perfect information, EVPI). Scaled up to the population of England who could benefit from the information from any future studies, this equates to an EVPI of £53.3m. If it is assumed the results of any future study are generalizable to all pregnancies in England, the EVPI is £172.9m.

The parameter with the biggest impact on decision uncertainty was the cost of IOL (specifically, the difference in cost between an induced delivery and expectant management). It should be noted that this does not simply relate to the cost of a procedure to induce delivery; included within this definition is uncertainty as to the timing of IOL, and the impact on for example, antenatal appointments, as well as the cost of the delivery itself. A study of 'reasonable size' to reduce uncertainty in this parameter is likely to yield a positive return on investment. For example, the EVSI of a study of 1000 mothers in each arm is worth in excess of £11m. If this was to be delivered for a cost of £1m, it would yield a greater than 10-fold return on investment. Of note is that studies on the outcomes from SGA or macrosomic deliveries are unlikely to yield a positive return on investment. The results described above relate to a willingness to pay threshold of £20,000 per QALY. At a threshold of £30,600 per QALY (just above the upper end of NICE's stated acceptable range of £20,000 to £30,000, universal scanning becomes the most cost-effective option. Furthermore, our one-way sensitivity analyses suggest there is scope for universal scanning to be cost effective under other assumptions; for example, the most cost-effective option remains a breech-only scan only so long as the time horizon of the analysis is below 45 years.

We then considered the potential for an RCT of screening and intervention using late pregnancy ultrasound in nulliparous women. For the outcomes of perinatal death or severe morbidity, all sample size calculations yielded numbers in excess of 50,000. Hence, trials using these outcomes are unlikely to be realistic. When studying a more general outcome of any perinatal morbidity (with or without maternal preeclampsia), trials which involved randomising women to being screened or not screened generated sample sizes in excess of 10,000 women. Trials screening all women and randomising high risk women to having intervention or the result being masked had sample sizes of <10,000 and this trial design was acceptable to the majority of women assessed by questionnaire and focus group. These trials would also provide data on both screening test performance and the intervention but would not capture the benefits of identifying breech presentation.

Conclusions

Screening for presentation only is likely to be cost-effective. Scanning for fetal biometry and well-being has limited value in predicting neonatal morbidity of low risk women directly but the evidence base is generally weak. Combining ultrasound and intervention appears to have some potential utility but sits at the borderline of acceptable cost effectiveness for the NHS. Better understanding of the cost of IOL compared with expectant management could help inform decision making around the use of ultrasound screening. There is currently no potential for a trial of screening versus no screening with the outcome of perinatal death. However, a range of other options assessing screening and intervention are feasible, each with their own strengths and weaknesses.

Study registration: This study is registered as PROSPERO CRD42017064093

Funding details: The National Institute for Health Research Health Technology Assessment programme

Plain English summary

Ultrasound scans allow doctors to check on the health of the unborn baby. Usually all pregnant women are scanned at about three months and about five months. After that, women are only offered a scan if they have risk factors or a problem develops. Lots of things can go wrong later in pregnancy including problems with baby's growth or which part of baby is coming first. Some of these might have been prevented if a scan had been done, but scans can also get it wrong. When they do, a woman might receive unnecessary treatment, which might even harm her or her baby.

In this study we set out to review previous research about how good ultrasound scanning is at detecting babies with a problem. We focused on detecting if the baby was too big or too small. Unfortunately, many of the studies had not been done to a high standard. Scanning can detect big and small babies pretty well, but it is much less clear whether they can predict complications which might harm the baby during birth. We also studied the costs and outcomes of scanning. We calculated the extra money that would be needed to scan every woman and compared this with the extra benefits from preventing complications. The one thing that came out well was using scan to check whether the baby is presenting head first or bottom first (a 'breech presentation'). Babies presenting by their bottom have high risks of complications. Scanning all women to check whether their baby is breech seems effective and may even pay for itself, although it depends on how much the scan would cost.

Whether it is worthwhile scanning all babies to see if they are too big or too small is less clear. The next step is probably a research study to get some more reliable numbers. We show how such a study should be designed, such that a single study could tell us both how well does scans predict bad outcomes, and does finding out this information actually help?

Chapter 1. Background.

Screening for pregnancy complications

Complications of pregnancy are a major determinant of the Global Burden of Disease, through effects on both the mother and baby.[1] Identifying and managing risk is a key element of antenatal care which aims to reduce the number and severity of adverse outcomes. Current clinical guidelines[2] include multiple methods of identifying high risk women including: (i) identification of maternal risk factors associated with disease (e.g. obesity, age >40 years), (ii) assessment of complications in previous pregnancies, (iii) identification of pre-existing medical conditions (e.g. diabetes mellitus), (iv) clinical presentation with symptoms which are associated with an increased risk of adverse outcome (e.g. antepartum haemorrhage, reduced fetal movements). Additionally, there are multiple tests which are applied to pregnant women to assess risk. Taking the example of screening for Down's syndrome, women's risk is first assessed by maternal age, this background risk is then adjusted for the results of ultrasonic imaging (nuchal translucency) and biomarkers (pregnancy associated plasma protein A and free beta sub-unit of human chorionic gonadotrophin) and the summative risk is used to inform the use of invasive testing (https://www.gov.uk/topic/population-screening-programmes/fetal-anomaly).

Use of ultrasound in pregnancy screening

The first trimester ultrasound scan employed in Down's syndrome screening is an example where all pregnant women are offered a scan as part of their assessment of risk. Routine pregnancy care in the UK also involves a second screening ultrasound scan, performed ≥18 weeks of gestational age (wkGA) and <21wkGA, where the primary purpose of the scan is to identify fetuses with structural abnormalities (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 neural tube defect) 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).

In the UK and USA, universal ultrasound is not recommended after the mid-pregnancy anomaly scan.[2, 3] Rather, it is recommended that ultrasound is offered in a targeted manner and only offered to women where there is a clinical indication. Such indications could include presentation with symptoms

(e.g. antepartum haemorrhage), relevant medical history (e.g. anti-phospholipid antibody syndrome), relevant past medical history (e.g. previous fetal growth restriction [FGR]), or through the results of physical examination (e.g. the uterus is SGA) on clinical examination.

Use of ultrasound in late pregnancy

When scans are performed in late pregnancy, a number of features are commonly reported. Ultrasound allows the estimation of the size (length and circumference) of fetal parts, termed fetal biometry. A variety of methods exist for converting these measurements to an estimated fetal weight (EFW)[4] and a number of reference ranges exist for EFW in relation to the exact gestational age.[5, 6] The interpretation of EFW and the individual biometric measurements generally focuses on two properties: (i) the position of the value on the distribution for the given gestational age, and (ii) 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 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. SGA combined with evidence of reduced fetal growth velocity is regarded as indicating FGR.[7]

Another major category of measurement in ultrasound in late pregnancy is Doppler flow velocimetry (referred to as "Doppler", see Hoffman and Galan for review.[8] In brief, a blood vessel is imaged and electronic callipers on the screen are placed over the vessel. The machine then plots out the velocity of flow on the Y axis, with time on the X axis. The resultant plot is termed a flow velocity waveform. Different blood vessels have different patterns of flow velocity waveform and the pattern is analysed both qualitatively and quantitatively. One of the key blood vessels for study is the umbilical artery. Flow is characterised qualitatively by the direction of flow in end diastole (i.e. immediately prior to the rise in flow that occurs with a heartbeat – systole). The normal state is forward flow, but there can be absent flow or even reversed flow. The waveform can also be analysed mathematically, and a number of indices have been described, such as the pulsatility index (PI) and resistance index (RI). The derivation, calculation and detailed interpretation of these indices is described in detail elsewhere.[8] However, both values correlate positively with the presumed resistance to flow in the vascular bed supplied by the artery. Hence, high values of PI and RI in the umbilical arteries are interpreted as indicating a high 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.[9]

The four most common sites for Doppler are the umbilical arteries, the maternal uterine arteries, the umbilical arteries, the fetal middle cerebral arteries (MCA), and the ductus venosus.[8] In contrast to the other three, it is low resistance in the fetal MCA which is thought to indicate compromise. The interpretation is that a reduced level of oxygen in the fetal blood leads to cerebral vasodilation, hence, reduced measures of resistance in the arteries supplying the brain.

Other features which are examined in late pregnancy include the placenta, the amniotic fluid and fetal presentation. Reporting of the placenta generally focuses on its site in relation to the cervix. Implantation of the placenta over the cervix is called placenta praevia and it can cause massive haemorrhage during labour. Reduced amniotic fluid is called "oligohydramnios" and increased amniotic fluid is called "polyhydramnios". 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, one of the simplest findings on scan is the presentation of the fetus. Near term, >95% of fetuses present by the head. Women are examined close to term to assess presentation but this approach frequently misses babies presenting by the breech.[10] Ultrasound unambiguously establishes the presentation at the time of a scan.

Coupling interventions to scan results

There are a limited number of disease modifying interventions which can be coupled to ultrasound performed in late pregnancy to alter the outcome of pregnancy. Most of the interventions relate to modifications to either the timing of delivery (e.g. IOL) or the mode of delivery (e.g. delivery by pre-labour caesarean section). One exception to this is breech presentation. It has been known for many years that vaginal breech delivery, although safe for the majority of women, could be associated with complications which could have severe consequences for the infant. Breech delivery has a number of specific complications associated with it, such as increased risks of umbilical cord compression and entrapment of the fetal head after delivery of the fetal body. It was demonstrated that vaginal breech birth in the UK was associated with an absolute risk of death during labour or in the first four weeks of life 8.3 per 1,000. Although the absolute risk was low it was much higher than the risk associated with a planned caesarean delivery of 0.3 per 1,000.[11] The awareness of the risks associated with vaginal breech birth (which long predated the epidemiological study confirming the higher risk of

death) were the basis for offering to turn the baby from a breech to a cephalic presentation using manual manipulation of the fetus by a clinician, called “external cephalic version” (ECV). Where this procedure is unsuccessful, generally, delivery by planned caesarean section is recommended.[12] This is based both on the observational data of increased risks associated with vaginal breech birth and on the results of randomised controlled trials of planned caesarean section which have confirmed reduced risks of perinatal death with this procedure, compared with planned vaginal breech birth.[13]

For most of the other diagnoses which might be made by ultrasound, the primary disease modifying intervention in the second half of pregnancy is to deliver the baby, either by IOL or planned caesarean. However, screening may also be used to inform the assessment of fetal well-being to help inform the timing of this intervention. For example, if a baby is found to be SGA and FGR is suspected, there are multiple ways that the well-being of the baby might be assessed. However, these simply reflect another layer of diagnostic and prognostic tests. Ultimately, they are used to target the timing of the disease modifying intervention of delivery. The primary reason for expediting delivery is that IOL removes the subsequent risk of stillbirth (intra-uterine fetal death followed by delivery of a baby showing no signs of life). Most causes of stillbirth are due to complications which can only occur to the fetus in utero (e.g. placental abruption or placental failure), hence, delivery of the fetus removes the risk of stillbirth.[14] This is confirmed by randomised controlled trials which demonstrate that IOL at term is associated with a 67% reduction in the risk of stillbirth.[15]

While early delivery can safely be performed at term, this is not the case preterm. The Cochrane review above described exactly the same reduction in the risk of perinatal death with IOL at term as was observed for stillbirth. Perinatal deaths include both stillbirths and neonatal deaths, hence the favourable effect of IOL on stillbirth was not cancelled out by an unfavourable effect on the risk of neonatal death. However, preterm birth is one of the major determinants of neonatal death, hence, if women are routinely induced preterm, the advantage of reduced risks of stillbirth will be outweighed by the increased risks of intrapartum stillbirth and neonatal death associated with prematurity. The inflection point, i.e. where the risks balance out, has previously been estimated as between 38wkGA and 39wkGA.[16] Hence, although 37wkGA is strictly term, routinely delivering all women at 37 weeks could increase overall perinatal mortality through higher rates of intrapartum stillbirth and neonatal death.[17] It follows, therefore, that screening using a test with a high false positive rate has the potential to cause net harm through increasing iatrogenic prematurity (or early term delivery) in false positives.[18]

Evidence for screening using universal late pregnancy ultrasound

There is strong evidence to support the use of ultrasound in high risk pregnancies. A systematic review of umbilical artery Doppler has shown that it reduces perinatal mortality by about 30% in high risk pregnancies.[19] The mechanism of the effect is likely explained by the fact that its use was also associated with lower rates of IOL and caesarean delivery. Hence, it is likely that the use of Doppler reduced the risk of perinatal death overall by reducing 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.

The fundamental role of ultrasound in the care of high risk women led researchers to explore whether routinely using the same approaches might improve outcomes in low risk women. Disappointingly, a meta-analysis of 13 RCTs including ~35,000 women did not demonstrate any evidence that routine ultrasound improved outcome.[20] It is this finding which has led to the recommendation that ultrasound should not routinely be performed in the second half of pregnancy in the UK and USA. The cautious approach is supported by some evidence arising from countries where universal late pregnancy ultrasound was introduced, despite the lack of strong evidence supporting its clinical effectiveness. A seminal study from France reported rates of adverse perinatal outcome in relation to woman's screening status for SGA.[21] Each woman's screening status was identified (screen positive for SGA or screened normal [AGA = appropriate for gestational age]) and the actual status of the baby at birth was also assessed (SGA or AGA by actual birth weight). The authors subsequently described rates of perinatal morbidity and mortality by true positive and false positive status. As one might have predicted, false positives had higher rates of multiple adverse outcomes compared to AGA babies which were true negatives, and this was explained primarily by higher rates of iatrogenic prematurity in the false positives. Interestingly, the true positive SGA babies also had higher rates of adverse outcome compared with SGA babies which were missed by scan (false negatives). The former observation confirms the potential for iatrogenic harm to false positives. The latter observation questions the rationale for screening for SGA in late pregnancy at all.

Critical analysis of the Cochrane review

While it is generally accepted that a systematic review of RCTs represents the highest level of evidence, there are a number of features of the systematic review of RCTs of universal ultrasound[20] that undermine its main conclusions.

- The 13 studies included in the meta-analysis all used different definitions of screen positive. Moreover, some of the ultrasonic findings were completely divergent. 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. An implicit assumption around combining these studies is that these different ultrasonic tests were all comparable effectiveness, which a subsequent systematic review of diagnostic test accuracy studies has demonstrated is not the case.[22]
- None of the studies were preceded by a high quality assessment of the diagnostic effectiveness of the test in a low risk population. This is problematic for a number of reasons. A key element of study design is a power calculation. It is impossible to perform a power calculation without quantitative information on the diagnostic effectiveness of the test. Moreover, the tests had generally been developed for and evaluated in high risk populations. It is well recognised in screening that test performance differs according to the risk status of the population. One of the key properties of a screening test is the positive predictive value (PPV), i.e. the proportion of women who screen positive who experience the outcome. The positive predictive value of a test is determined by the prior risk of disease multiplied by the positive likelihood ratio (LR+ = the proportional increase in the odds among screen positive women compared with the whole population). Hence, the higher the prior risk of disease, the higher the PPV for a given LR+. Consequently, it is typical that a positive screening test is associated with a much lower PPV in low risk population. As the PPV determines the ratio of true positives to false positives, this will have a major impact on trials of screening.
- None of the 13 RCTs coupled the screening test to an intervention. In all 13 studies the result was revealed to the attending clinicians but there was no specific intervention that was planned. It is self-evident that a screening test could only impact on outcome if it is coupled to an intervention. Moreover, the tests were performed at a wide range of gestational ages. 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. Hence, a positive effect of late pregnancy ultrasound and delivery could have been masked by a negative effect of preterm pregnancy ultrasound scan with higher rates of iatrogenic harm.
- 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.

Parity and the risk of adverse outcome

One of the most important determinants of adverse pregnancy outcome 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,[23] preterm birth[24] stillbirth[25] and FGR[26]. Hence, women experiencing complications in previous pregnancies generally receive enhanced antenatal care. Conversely, the experience of uncomplicated previous pregnancies is strongly predictive of a normal outcome in future pregnancies. Hence, women who have had a previous vaginal delivery of a normally grown live born infant at term following an uncomplicated pregnancy have low absolute risks of complications in future pregnancies.[27] 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.[28] Moreover, first pregnancies, collectively, have high rates of complications than second pregnancies. These qualities have led to the identification of first pregnancies are a priority area for research. Quoting an NIH study description of nulliparous women:

"*This large proportion of women lacks previous pregnancy information to guide risk assessment; as such, adverse outcomes in these first pregnancies are particularly difficult to predict and prevent*."[29]

Summary of the rationale for the focus on nulliparous women in late pregnancy

The characteristics above provide the rationale for the focus of this review. Screening and intervention near term has less of a potential to cause harm than screening and intervention in the preterm period, as the primary intervention – delivery of the baby – is less likely to lead to iatrogenic injury. 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.

The health economics of screening and intervention

A critical consideration in relation to screening and intervention using universal ultrasound is whether it is cost effective. It is possible that, for the individual woman and baby, having a screening ultrasound scan and associated intervention leads to a better outcome but that the impact of the cost of providing the screening test and intervention results in net societal harm as it removes resources from other more cost effective elements of the health care system. The capacity of all health care systems is finite, however, systems differ in their willingness to pay. These questions are addressed quantitatively in health economic analysis by calculating the sum of money required to gain one additional quality adjusted life year (QALY), a subject which is discussed in detail elsewhere.[30] In the English NHS, interventions are considered cost effective if the cost of each QALY is below a given threshold and this is typically between £20,000 and £30,000.

Providing a late pregnancy ultrasound scan will clearly incur direct costs. Managing women who screen high risk will clearly incur further costs. However, these additional costs then have to be set against the reduction in harm, i.e. the QALYs gained by the mother or child because of being screened. Many of the individual elements required for these calculations are associated with uncertainty. Hence, these health economic analyses frequently employ a probabilistic approach running large numbers of simulations where the different parameters for the models are sampled from the presumed plausible range of values from the literature. These methods and their interpretation are discussed in more detail in the relevant chapters.

Value of information (VoI) analysis

The health economic analyses described above relate to the economic case for implementing a given programme of screening and information. VoI analysis addresses the economic case for funding research to try and reduce the uncertainty in the evidence base. Generally speaking, a research question which will be identified as being cost effective from this perspective will have input values which are uncertain, i.e. the confidence intervals for the given parameter in the literature are wide. Moreover, questions which are identified as being cost effective in a VoI will often generate highly variable results in sensitivity analyses where the input value of the parameter is varied within the range of uncertainty. This subject is again dealt with in detail in the relevant chapter.

Designing a randomised controlled trial

Randomised controlled trials (RCTs) of screening have certain differences compared with RCTs of other interventions. Typically, interventions are evaluated in populations with a disease. Hence, the individuals recruited will have high rates of complications as they are experiencing a disease process. Moreover, most of the outcomes in the group are likely to be related to the disease process. In contrast, screening, by design, focuses on individuals before they manifest disease. Hence, the background rate of serious adverse outcomes is likely to be low. Moreover, the experience of adverse outcomes within the population is likely to be due to diverse causes, not simply the disease being screened for. For example, a randomised controlled trial studying mortality in people with cancer is likely to have high rates of death in the different arms of the trial and most of the deaths in both arms are likely to be related to the cancer. In contrast, a randomised controlled trial of screening or not screening a healthy population for the same cancer is likely to have low rates of deaths in both arms and many of the deaths in both arms would be unrelated to the experience of cancer. Both of these properties will tend to increase the sample size in the screening study as there is a low incidence of adverse outcomes and only a subset of the adverse outcomes will be preventable by the given programme of screening and intervention.

We have previously reviewed the approach to screening in pregnancy[31] and highlighted an alternative, namely, that all women in a population are screened and that randomisation is to either revealing the result plus intervention or masking the result with routine care. Using this design, randomisation is being performed in a group which has a higher rate of complications (by virtue of the positive screening test) and a greater proportion of the adverse events will be related to disease process being screened for. This approach has the advantages that the overall number needed to screen for statistical power is substantially reduced and that the screening test can be validated in the same study design through comparing screen negatives with screen positives randomised to have the result masked. These issues are discussed further in Chapter 13 below.

Chapter 2. Objectives.

The objectives of the present study, outlined in the original application, were as follows:

1. To assess the diagnostic effectiveness of late pregnancy ultrasound in nulliparous women based on the existing research literature.
2. Having identified the key ultrasonic findings which identified women as high risk to review the existing literature and current guidelines to identify a management plan for women with high risk characteristics.
3. To conduct a health economic analysis of the likely cost-effectiveness of screening and intervention based on the best available evidence of the costs, diagnostic effectiveness of ultrasound and clinical effectiveness of intervention.
4. To perform a value of information analysis to determine whether there is a strong economic case for funding future research in this area.
5. Conditional on the above, to outline the design a randomised controlled trial which could strengthen the evidence base relating to the issues above.

Chapter 3. Identifying the research questions

We performed a survey of members of a number of professional organisations with the aim of identifying the ultrasonic features which were thought most likely to be informative in a future randomised controlled trial. We also surveyed which outcomes should be prioritised. A web-based questionnaire was designed using the SurveyMonkey platform and was approved by the Ethics Committee of the School of Humanities and Social Sciences at the University of Cambridge. The survey was sent to members of the Royal College of Obstetricians and Gynaecologists, the British Maternal Fetal Medicine Society and the British Association for Perinatal Medicine in May-June 2017. It was also distributed locally at the Rosie Hospital in Cambridge.

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 abnormal fetal biometry or growth velocity (83%), malpresentation (63%), abnormal amniotic fluid volume (63%), high resistance pattern of umbilical artery Doppler flow velocimetry (32%), and 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 hypoxic ischaemic encephalopathy (69%), fetal asphyxia (low umbilical cord blood pH plus a base deficit consistent with metabolic acidosis; 64%), SGA or severe SGA 51%, severe shoulder dystocia (46%), breech presentation diagnosed in labour (41%), admission to neonatal intensive care unit (28%), and low 5-minute Apgar score (21%).

Having completed the survey, we then searched relevant databases (Medline, Embase and Cochrane) to identify any other systematic reviews of Diagnostic Test Accuracy (DTA) which might overlap with our aims. This yielded a protocol for a Cochrane DTA review of ultrasonic diagnosis of SGA (which was subsequently published in 2019).[22] Hence, we did not include this in our own plans. We also identified a previously published systematic review of DTA on severe oligohydramnios which was published in 2014 and included publications up to 2011. We selected the studies in this review which were

performed in low and mixed risk pregnancies and then we performed a literature search for eligible studies that have been published subsequent to the search date the 2014 paper. We then performed a meta-analysis of all the relevant studies.

Based on the priorities gleaned from the review 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. High resistance pattern of umbilical artery Doppler flow velocimetry
2. Low cerebro-placental ratio (CPR)
3. Severe oligohydramnios
4. Borderline oligohydramnios
5. Suspected fetal macrosomia

All five of these were written up as a single study protocol and the analyses were registered on the PROSPERO International prospective register of systematic reviews (CRD42017064093).

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.

High resistance patterns of umbilical artery (UA) Doppler flow velocimetry are thought to reflect placental vascular resistance. This method is currently in widespread clinical use to monitor high risk pregnancies, including those with suspected FGR. 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).[19] 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).[32] 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.[31] 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. The aim of this chapter was to provide Level 1 evidence on 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.[7]

Methods

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.[33] 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,[7] which presented the results on the diagnostic effectiveness of ultrasound as a screening test for SGA.

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 March 2019. The studies were identified using a combination of words related to “ultrasound”, “Doppler”, “umbilical artery”, “pregnancy” and “prenatal diagnosis” (see Appendix 1). 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 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 data extraction

The literature search, study selection, and analysis ware 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 described in the Cochrane Handbook of Diagnostic Test Accuracy Studies.[34] 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).

Statistical and meta-analysis methods

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) respectively. For the data synthesis we used the hierarchal summary receiver operating characteristic curve (HSROC) model of Rutter and Gatsonis.[35] 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.[36] We also pooled the DORs using the method described by Deeks.[37] For the assessment of publication bias we used the Deeks' funnel plot asymmetry test in which $P<0.05$ was defined as significant asymmetry.[38] 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).

Results

The POP study

Initially we analysed the previously unpublished data from the POP study. The analysis included 3615 women that met the inclusion criteria (Appendix 1, Figure 25). All women had a blinded UA ultrasound at 36wkGA and 346 (9.6%) had an UA PI >90th percentile (Appendix 1, Figure 25). The maternal age, socio-economic status, ethnicity, BMI, and rates of alcohol consumption and smoking were similar between the two groups (Appendix 1, Table 18). 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 IOL 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 1. 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.

Table 1. 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 Sovio et al. 2015 for definitions*

Meta-analysis

The literature search PRISMA flowchart is presented in Appendix 1, Figure 26. We identified 13 studies[39-50] that met our inclusion criteria including 67,764 patients in total (these analyses included the previously unpublished POP study results). The study characteristics are presented in Appendix 1, Table 19. Five studies[39, 45, 48, 49] (N=63,436) included unselected pregnancies as part of universal screening, four studies[40, 43, 44, 50] (N=2634) included only low-risk pregnancies and four studies [41, 42, 46, 47](N=1694) included mixed risk pregnancies. Three of the studies[39, 48, 49] that were done in the same hospitals might have had short periods of overlap. Nine studies [40, 41, 43-47, 50](N=8097) were prospective and four[39, 42, 48, 49] (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 Appendix 1, Figure 27. Overall the quality was variable. The main risk of bias was that only six studies[40, 41, 43, 45, 47] (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.

The summary results of the meta-analysis are presented in Table 2. 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 1. 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

Figure 2) and illustrated the variation between studies using forest plots. 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 (Appendix 1, Figure 28). The test showed no evidence of publication bias (P=0.52).

Table 2. 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)
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)

[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.*

Figure 1. 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).

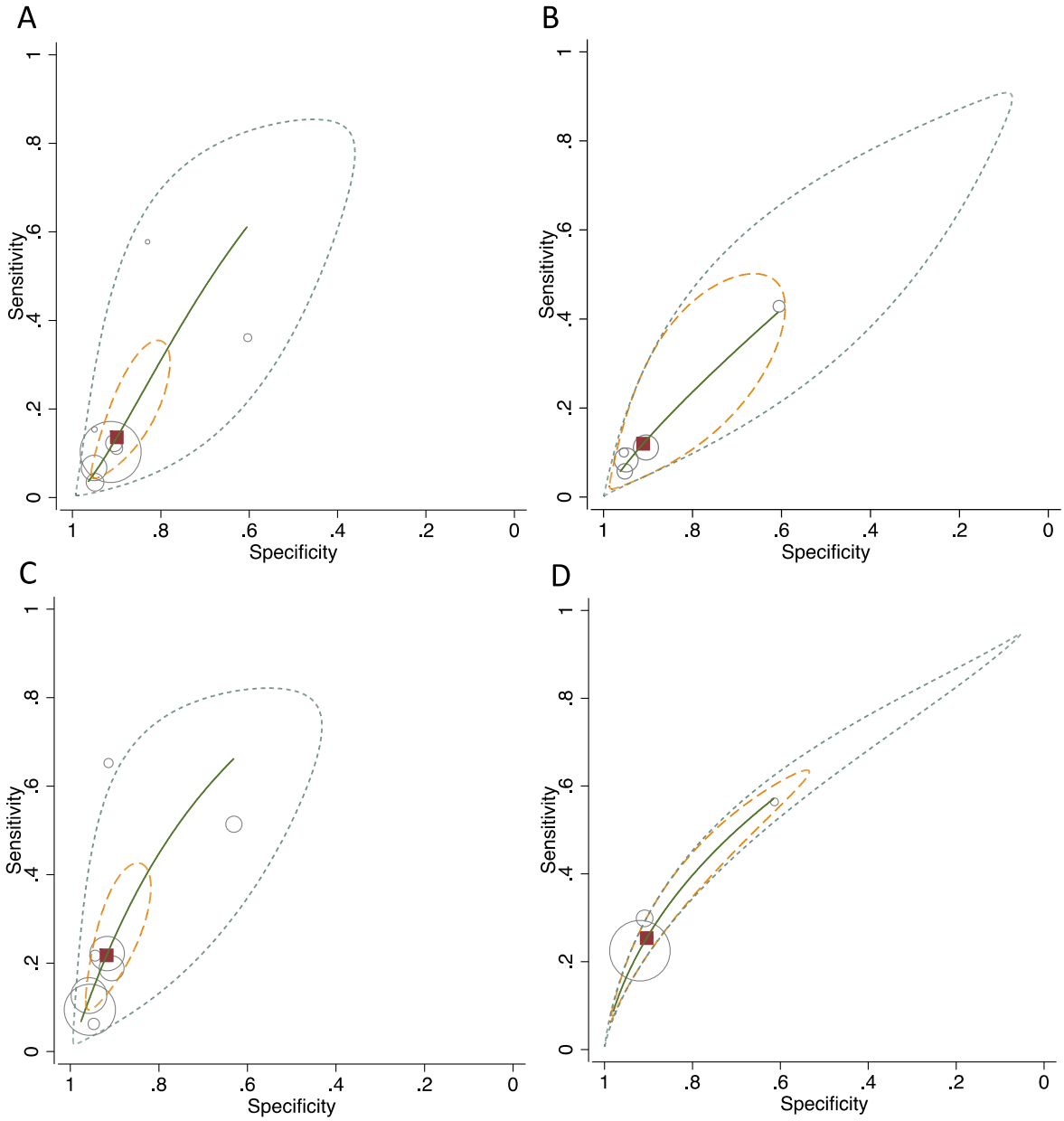
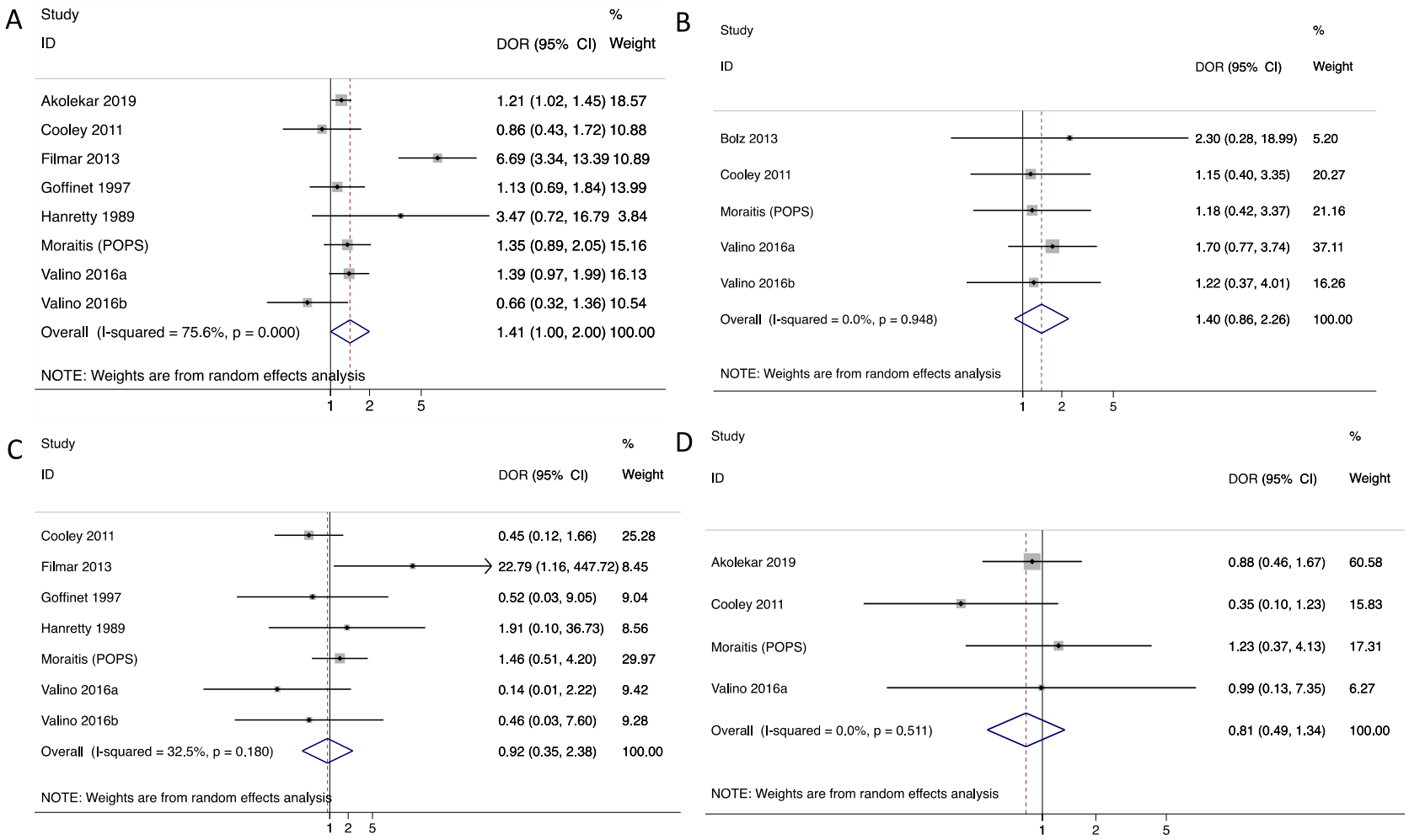
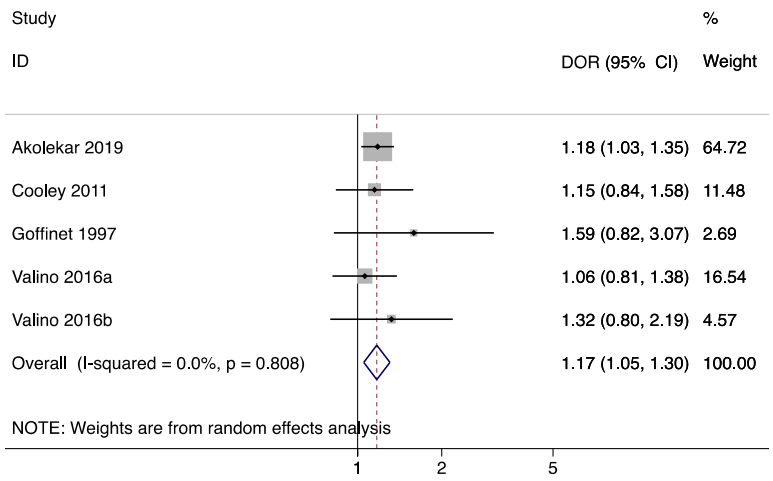


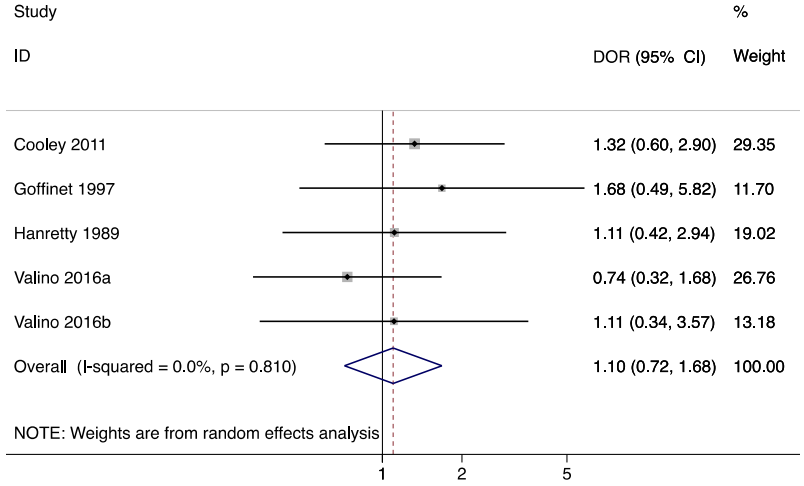
Figure 2. Meta-analysis of DORs of UA Doppler at predicting: A. NICU admission, B. Neonatal metabolic acidosis, C. 5-minute Apgar score <7, D. Severe adverse perinatal outcome, E. Caesarean section for fetal distress, F. Pre-eclampsia, G. SGA (<10th centile), H. Severe SGA (<3rd centile)



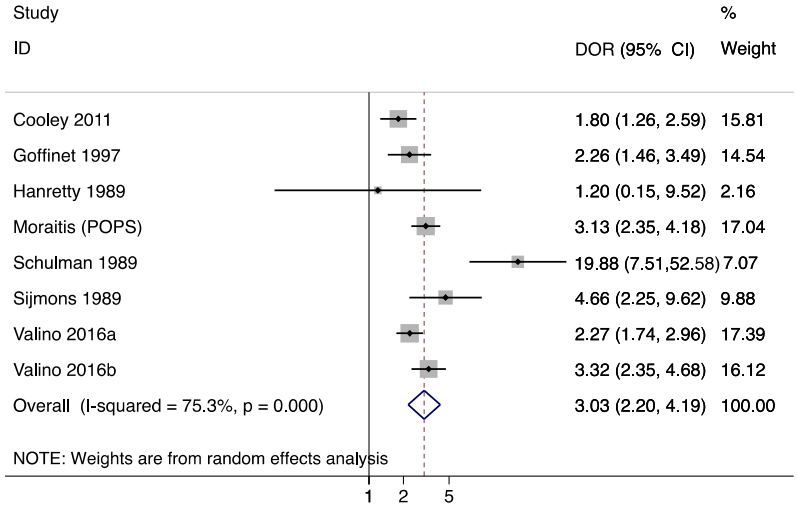
E


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

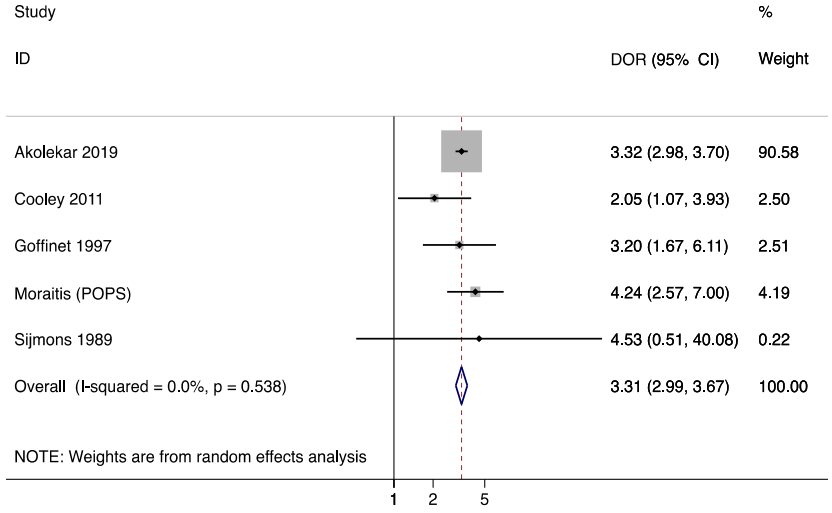
F


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

39. 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.
40. 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**:165-70.
41. 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**:143-9.
42. 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;**288**:279-85.
43. Fischer RL, Kuhlman KA, Depp R, Wapner RJ. Doppler evaluation of umbilical and uterine-arcuate arteries in the postdates pregnancy. *Obstetrics & Gynecology* 1991;**78**:363-8.
44. 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;**71**:11-9.
45. Hanretty KP, Primrose MH, Neilson JP, Whittle MJ. Pregnancy screening by Doppler uteroplacental and umbilical artery waveforms. *British journal of obstetrics and gynaecology* 1989;**96**:1163-7.
46. 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**:192-6.
47. 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**:557-61.
48. 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;**47**:194-202.
49. 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;**47**:203-9.
50. 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**:93-7.

Discussion

The main finding of this study 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. As discussed in the background section, FGR is a theoretical concept with no gold standard. 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 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 of perinatal death at term demonstrated that only 1 in 3 stillbirths at term are associated with abnormal fetal growth.[51] 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.[7]

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 ~300,000 to achieve 90% power see Supplementary Figure 10 in Flenady et al 2016.[52] 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.

The preceding chapter details the fact that a high resistance pattern of flow in the umbilical artery is most strongly associated with severe SGA, which is thought to be most closely reflective of FGR. The abnormal flow in the UA is thought to be related to the pathophysiology of FGR, reflecting impaired perfusion of the placenta due to placental dysfunction. The placenta is the site of gaseous exchange for the fetus. Hence, a consequence of placental dysfunction is that the fetus may have low levels of oxygen in arterial blood. Physiologically, low levels of oxygen are detected by the central and peripheral arterial chemoreceptors (PACs).[53] Activation of these receptors initiates compensatory responses, but these differ comparing fetuses and adults as there is no capacity to reverse the low levels of oxygenation by increased ventilation of the lungs (the chemoreceptors stimulate increased depth and frequency of ventilation in extra-uterine life). In fetal life, one of the key effects of PAC activation is to reduce the resistance to blood flow to the brain. Clinically, this process is assessed using Doppler flow velocimetry and, consistent with the foregoing, hypoxia leads to cerebral vasodilation and a reduced indices of vascular resistance using Doppler of the fetal middle cerebral artery.

One attractive way to develop simple screening tools is to use ratios of values in the presence of opposite associations with an outcome of interest. Hence the cerebro-placental ratio was developed that it would combine measurement of the cause of FGR (placental insufficiency as measured by the UA Doppler) and one of its major consequences (arterial hypoxaemia as measured by MCA Doppler). The aim of the current chapter was to assess the ability of this ratio to predict adverse pregnancy outcome.

Methods

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). The initial search was done in June 2017 and was updated on the 30th of May 2019. No restrictions for language or geographic location were applied. The protocol for the review was designed a priori and registered with the PROSPERO International Prospective Register of Systematic Reviews (Registration number: CRD42017064093). 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. 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. However, if the studies reported different outcomes or performed the ultrasound at different gestational ages we included both in the meta-analysis.

Study quality assessment and data extraction

The literature search, study selection, and analysis ware 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 outlined in the Cochrane Handbook of Diagnostic Test Accuracy Studies. This tool assesses the included studies for potential bias in four domains: patient selection, index test, reference standard and flow and timing. We assessed the risk for flow and timing from the perspective of universal ultrasound screening at 36wkGA. 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 (gestational age at scan, cut-off values used), reference standard (pregnancy outcome, gestational age at delivery, and interval from scan to delivery). We also collected information such as parity and rates of IOL when reported.

Statistical and meta-analysis methods

The statistical methods employed are described in Chapter 4.

Results

The literature search flowchart is presented in Appendix 2, Figure 29. We identified 16 studies[39, 54-68] that met our inclusion criteria involving 121,607 patients in total. The study characteristics are presented in Appendix 2, Table 20. Four studies [39, 54, 55, 65](N= 85,059) included unselected pregnancies, seven studies[56, 57, 59, 60, 63, 64, 67] (N= 12,929) included only low-risk pregnancies and five studies[58, 61, 62, 66, 68] (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,[54] Akolekar 2019,[39] and Bakalis[55] 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,[59] Monaghan[61] and Morales-Rosello[62] studies which reported different outcomes at the same tertiary maternity unit. Moreover, there was also likely population overlap between the Flatley,[58] Sabdia[66] and Twomey[68] 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 Appendix 2, Figure 30. The main risk of bias was for reference standard due to the lack of blinding in the majority of studies. Only five studies [56, 57, 63-65](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 IOL (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.

The summary results for the diagnostic accuracy of CPR at predicting adverse pregnancy outcomes are presented in Table 3. 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 3. Generally, the larger studies reported lower sensitivities and higher specificities for all the outcomes. We also present the pooling of the DORs in **Figure 4**. 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; Appendix 2, Figure 31).

Table 3. 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 3. 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)

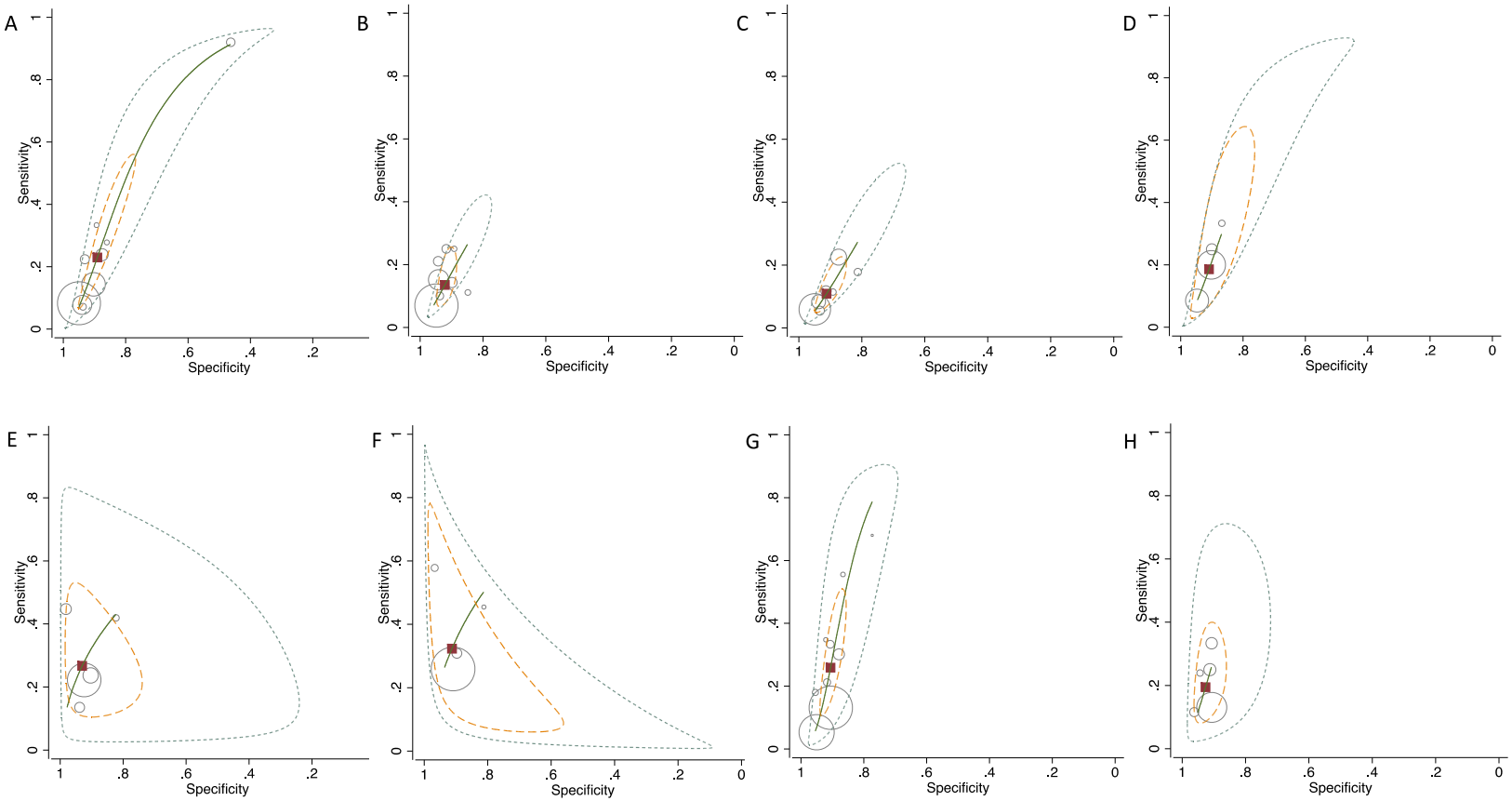
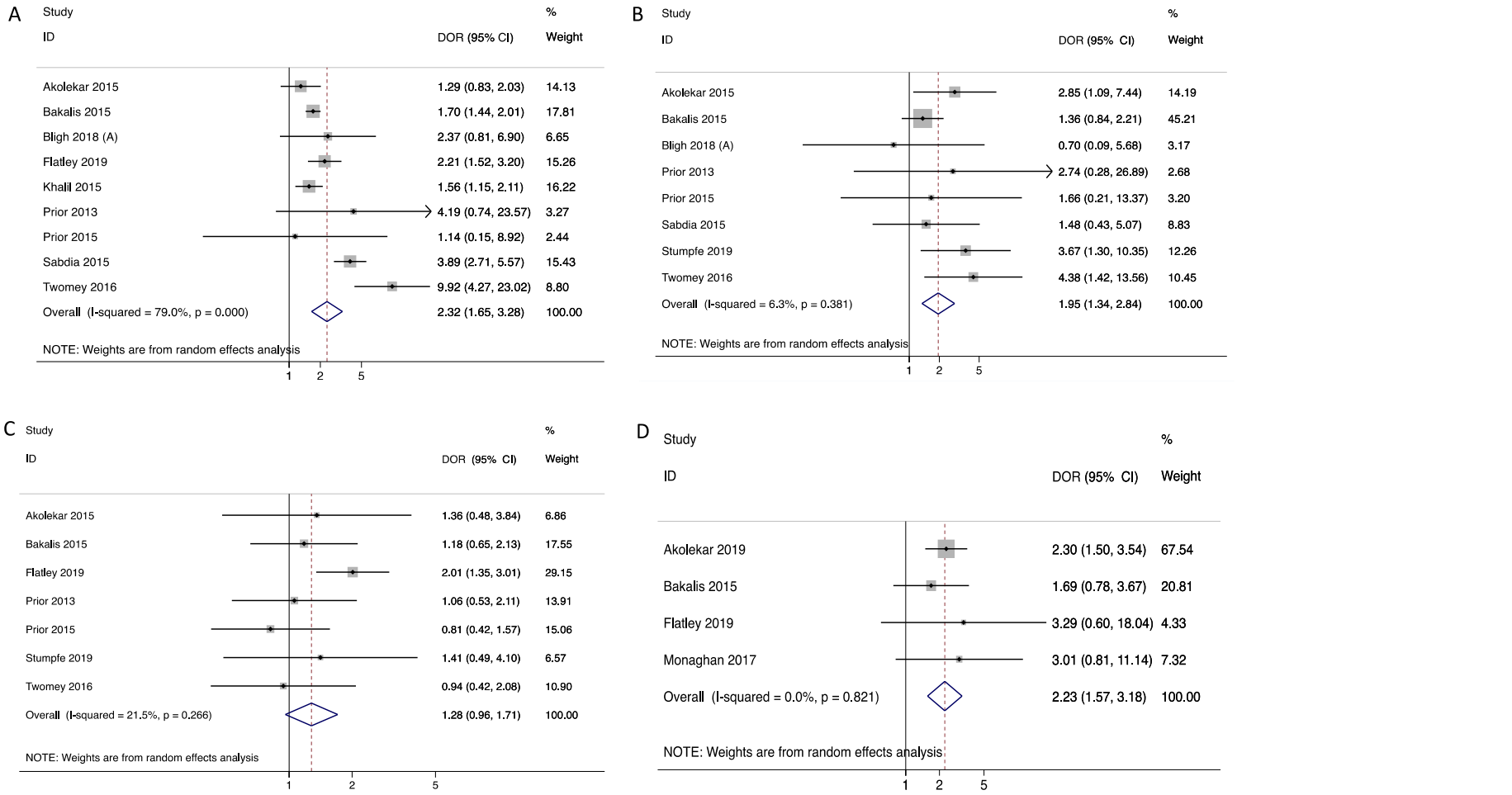
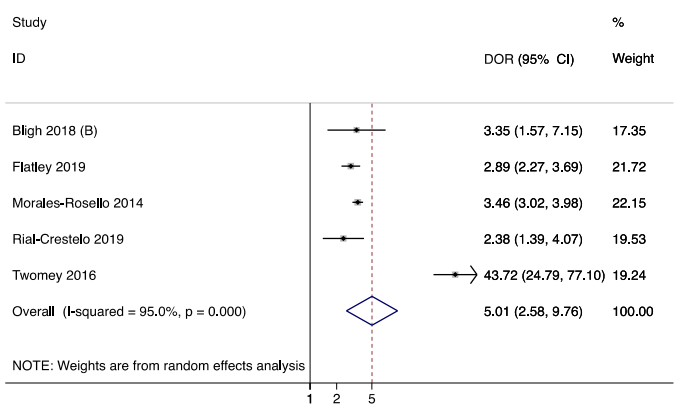


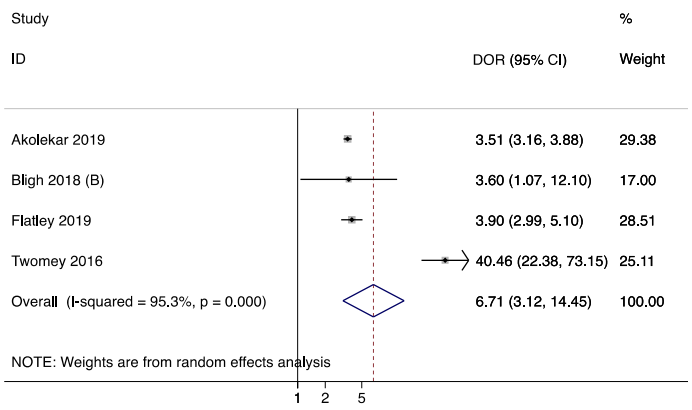
Figure 4. DORs for the diagnostic performance of abnormal cerebroplacental ratio at predicting adverse pregnancy outcomes: A. Neonatal unit admission; B. 5-minute Apgar score <7; C. Neonatal metabolic acidosis; D. Severe adverse perinatal outcome (including stillbirth, neonatal death and hypoxic ischaemic 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)



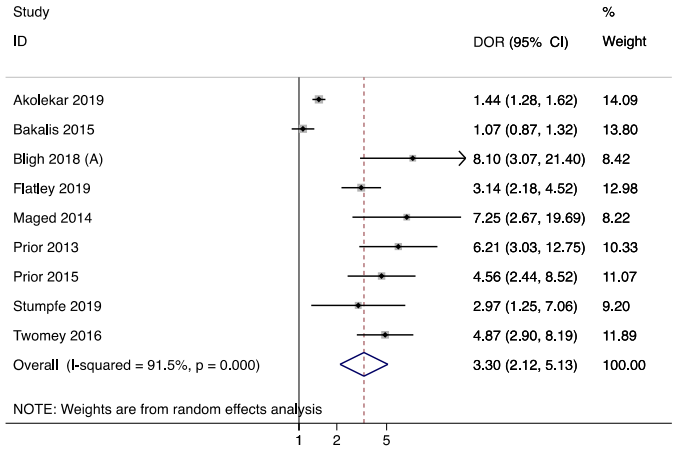
E


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

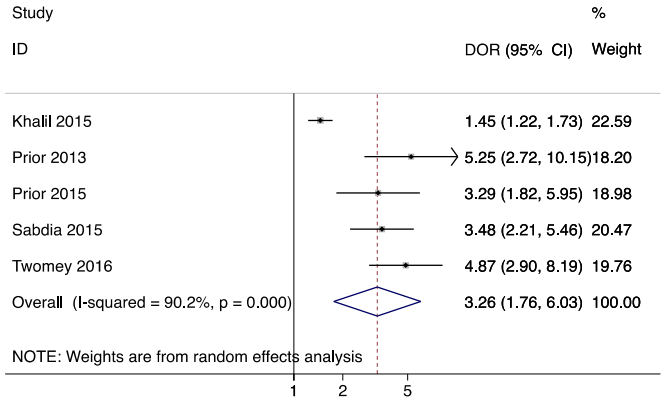
F


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

39. 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.
54. 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;**46**:82-92.
55. 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**:409-20.
56. 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**:340-6.
57. 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**:191-8.
58. 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;**32**:2380-6.
59. 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**:54.e1-10.
60. 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**:1880-5.
61. 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; 10.1002/uog.17500. https://doi.org/10.1002/uog.17500
62. 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**:303-10.
63. 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**:124.e1-6.
64. 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**:460-4.
65. 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;**32**:2554-60.
66. Sabdia S, Greer RM, Prior T, Kumar S. Predicting intrapartum fetal compromise using the fetal cerebro-umbilical ratio. *Placenta* 2015;**36**:594-8.
67. 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;**299**:411-20.

68. 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.

Discussion.

The meta-analysis demonstrated that the 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 compared with 2.5 to 3.0 for UA Doppler. Moreover, unlike UA Doppler, a low level of CPR was associated with an increased risk of neonatal morbidity. However, in this case the strength of prediction was weaker with positive LRs of <2.0. Moreover, in both analyses, there was very significant heterogeneity in relation to both birth weight based outcomes and neonatal morbidity. Consequently, the 95% confidence intervals for the positive LR are wide and include the point estimates observed for UA Doppler for both SGA and severe SGA. Moreover, given that many of the studies were not blinded it is possible that the associations with neonatal morbidity were due to bias. However, the association between CPR and SGA indicates that the ratio is likely to predict FGR. Overall, this analysis indicates that the CPR is indeed predictive of adverse pregnancy outcome. However, it is not clear from the present analysis whether the ratio performs better than simply assessing the UA Doppler, which is used in its calculation anyway. Of the indices assessed in these sections of the report, only the MCA Doppler was not measured in the POP study, hence, unlike the other chapters, we are unable to compare the strength of association in the POP study and the meta-analysis. Our findings contradict the previously published systematic review[69] 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.[69] 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[39, 54], Bakalis[55]) and did not produce any pooled analysis.

There are other issues which should be taken into account when considering the use of MCA Doppler as a screening test in unselected nulliparous women near term. First, the head often engages earlier in nulliparous women and it can be technically difficult to measure MCA Doppler when the head is deeply engaged. Second, the safety of ultrasound has been established in RCTs. However, these studies did not perform MCA Doppler. The main concern around ultrasound is the potential for harm caused by heating tissues. The form of ultrasound that is most strongly associated with heating is pulsed wave Doppler ultrasound. Hence, there is a theoretical safety concern about this use of this method through heating of the baby's brain. In high risk pregnancies, the balance of risks and benefits probably favours gathering additional information. However, screening the entire population using this method may raise some safety concerns. Finally, the method also requires a certain level of

training and implementation of MCA Doppler as a population based screening methods would involve some challenges in relation to implementation.

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

Amniotic fluid evaluation is routinely performed in ultrasonic assessment of fetal wellbeing in the third trimester. Reduced amniotic fluid is called oligohydramnios and increased amniotic fluid is called polyhydramnios. In the second half of pregnancy, the amniotic fluid comes from the fetal urine. Fetuses with no kidneys (renal agenesis) typically have no amniotic fluid at the time of the routine 20wkGA anomaly scan and it remains absent thereafter. However, congenital anomaly is a rare cause of oligohydramnios. One of the common causes of oligohydramnios is rupture of the fetal membranes. In this event, the overall level of fluid is reduced through vaginal loss. Normal fetal production of urine in such cases can be confirmed by filling and emptying of the fetal bladder. However, fetal distress is thought to be a potential cause of oligohydramnios. The mechanism is through reduced fetal urine production. Stress – such as arterial hypoxaemia – results in activation of a number of compensatory responses.[53] These include increased release or arginine vasopressin (aka anti-diuretic hormone) which has a direct effect on the kidney. Fetal hypoxia leads to a chemoreceptor mediated cardiovascular response which increases blood supply to the vital organs (heart and brain) but reduces blood flow to the fetal trunk, including the kidneys. The combination of increased arginine vasopressin and reduced renal blood flow will reduce fetal urine output and lead to oligohydramnios. Hence, assessment of oligohydramnios has been a feature of ultrasonic assessment of fetal well-being for many years.

The most common methods of quantitative assessment 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)[70] and the single deepest pocket (SDP). Severe oligohydramnios is commonly defined as AFI<5cm or SDP<2cm. Given the known association between oligohydramnios and fetal stress, the aim of the present study was to produce level 1 evidence of diagnostic effectiveness of severe oligohydramnios in predicting adverse pregnancy outcomes at or near term and we performed a systematic review and meta-analysis of the literature.

Methods

Sources for meta-analysis

We identified a previous systematic review[71] 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 the latest search date on the 5th of June 2019. 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".

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 that the ultrasound was performed intrapartum. We included studies that reported the following outcomes: stillbirth, neonatal death 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.

Study quality assessment and data extraction

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 risk of bias in each included study was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS 2) tool as outlined in the Cochrane Handbook of Diagnostic Test Accuracy Studies.[34] This tool assesses the included studies for potential bias in four domains: patient selection, index test, reference standard and flow and timing. We assessed the risk for flow and timing from the perspective of universal ultrasound screening at 36wkGA. 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 (gestational age at scan, cut-off values used), reference standard (pregnancy outcome, gestational age at delivery, and interval from scan to delivery). We also collected information such as parity and rates of IOL when reported.

Statistical and meta-analysis methods

The statistical methods employed are described in Chapter 4.

Results

The literature search flowchart is presented in Appendix 3, Figure 32. We identified 14 studies[72-85] that met our inclusion criteria involving 109,679 patients in total. The study characteristics are presented in Appendix 3, Table 21. Two studies[75, 76] (N= 30,555) included unselected pregnancies, ten studies [72-74, 78-83, 85](N= 61,047) included low-risk pregnancies only and two studies[77, 84] (N= 18,077) included mixed risk pregnancies. Six studies [73, 76, 77, 79, 80, 82](N= 5740) were prospective, six [72, 75, 78, 81, 83, 84](N= 97,022) were retrospective, one[74] (N=260) was cross-sectional and one[85] (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 Appendix 3, Figure 33. The main risk of bias was for reference standard due to the lack of blinding in the majority of studies. Only two studies [79, 82](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[73, 83] performed ultrasound prior to IOL or within 4 days from delivery. Two other studies[75, 80] 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[77, 84] as they did not report how they selected their patients and one study[74] 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 4. The most commonly 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 4). 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 5. Generally, the larger studies reported lower sensitivities and higher specificities for all the outcomes. Figure 6 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 (Appendix 3, Figure 34). The test showed no evidence of publication bias (P=0.54).

Table 4. 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 5. 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

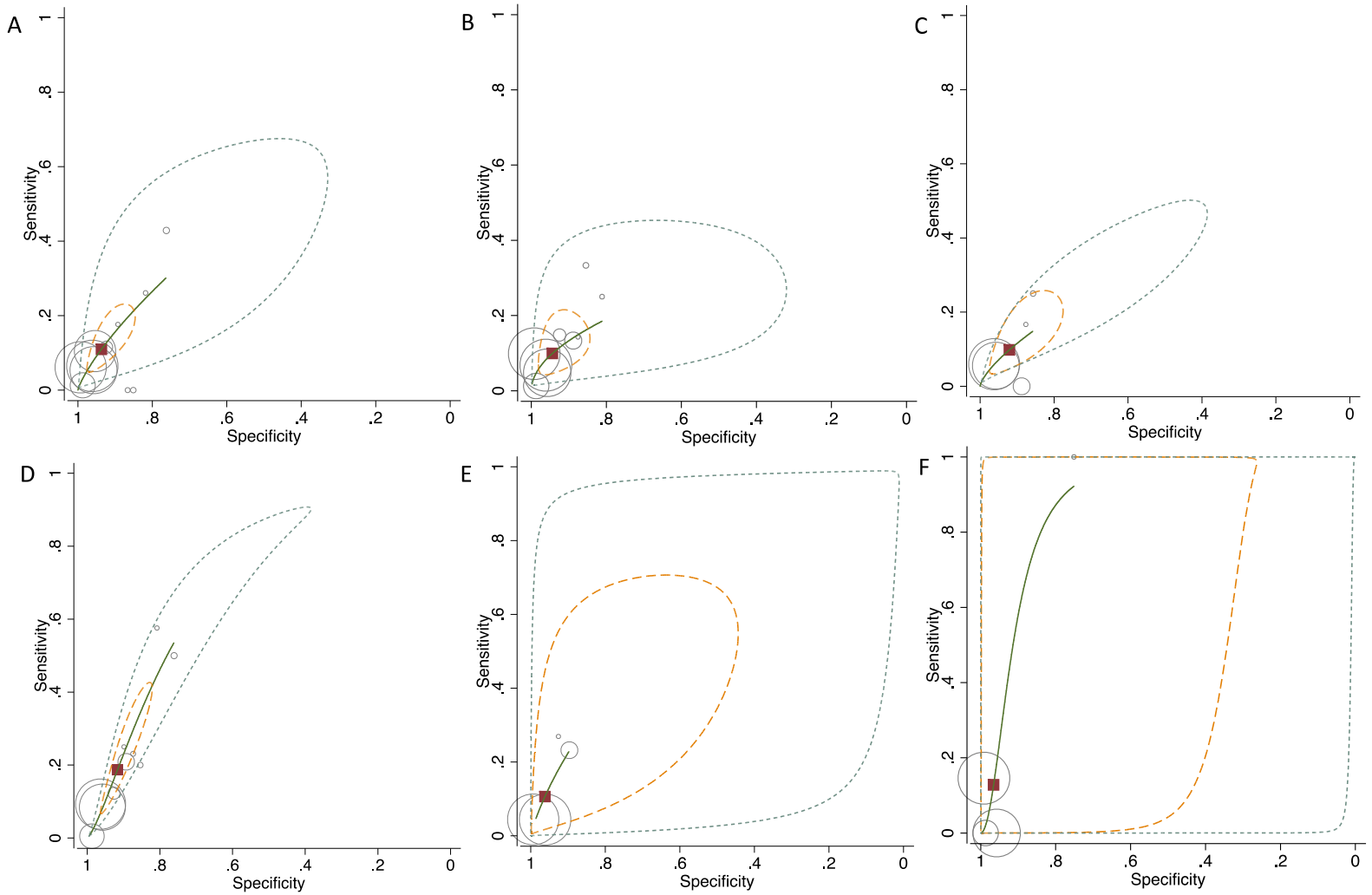
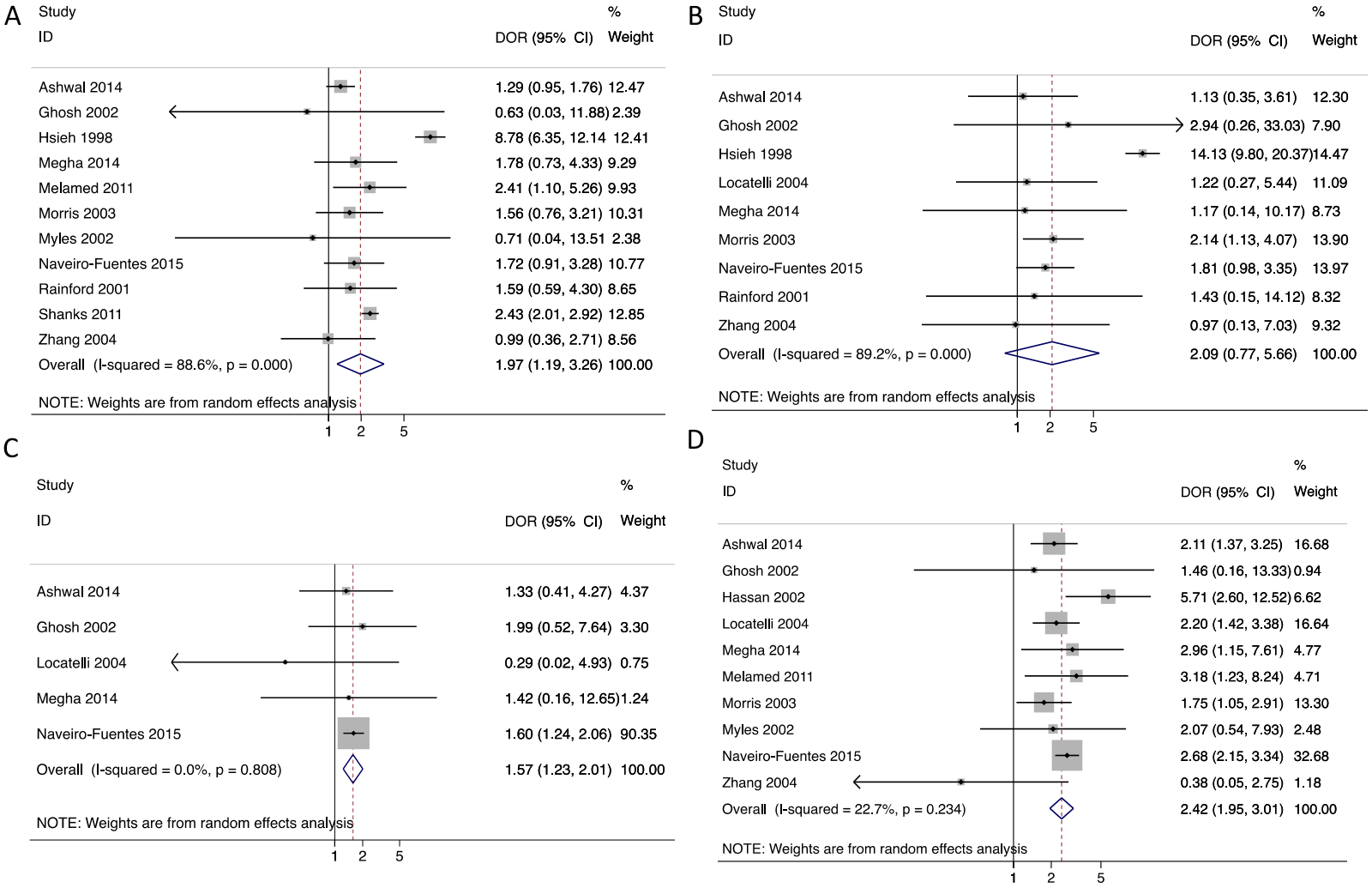
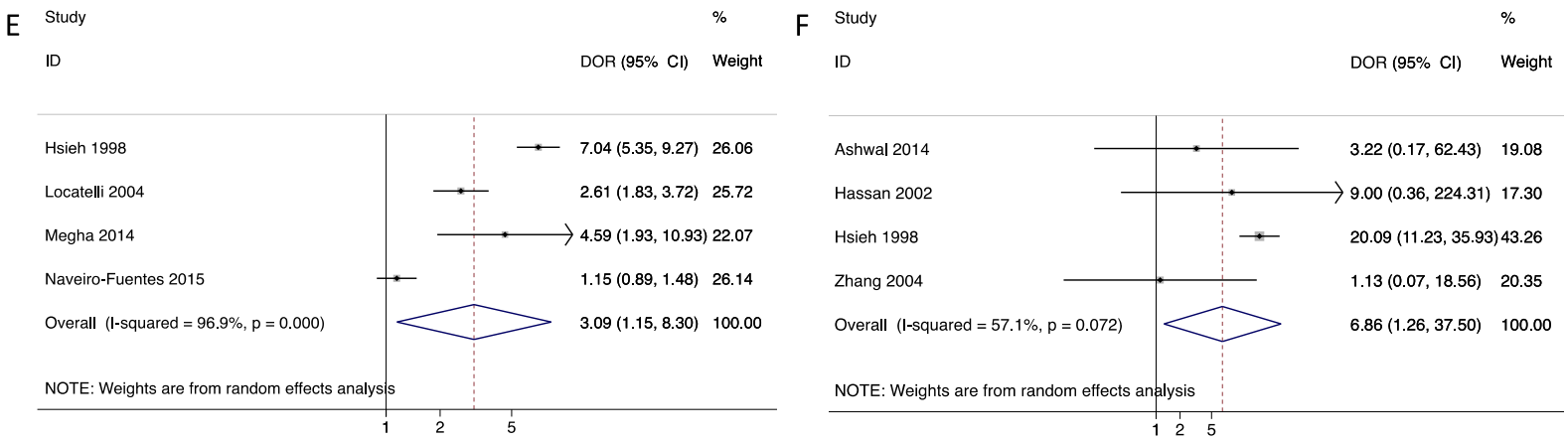


Figure 6. Meta-analysis of DORs for AFI <5cm at predicting adverse pregnancy outcome: A. NICU admission; B. 5-minute Apgar score <7; C. Neonatal metabolic acidosis; D. Caesarean section for fetal distress; E. SGA (<10th centile); F. Neonatal death.





72. 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;**290**:875-81. https://doi.org/10.1007/s00404-014-3292-7

73. 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;**81**:852-5.

74. 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;**15**:85-8. https://doi.org/02.2005/jcpsp.8588

75. 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**:232-6.

76. 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;**269**:130-3. https://doi.org/10.1007/s00404-003-0525-6

77. Megha B, Indu C. Correlation of amniotic fluid index with perinatal outcome. *Journal of Obstetrics and Gynecology of India* 2014;**64**:32-5.

78. 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;**205**:241.e1-6. https://doi.org/10.1016/j.ajog.2011.06.013

79. 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;**110**:989-94.

80. Myles TD, Santolaya-Forgas J. Normal ultrasonic evaluation of amniotic fluid in low-risk patients at term. *The Journal of reproductive medicine* 2002;**47**:621-4.
81. 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.
82. 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**:1319-23.
83. 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;**46**:589-92.
84. 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**:303-7.
85. 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;**111**:220-5.

Discussion

This meta-analysis confirms that a diagnosis of severe oligohydramnios is associated with adverse pregnancy outcome. The key finding was that severe oligohydramnios had a positive LR for SGA of between 2.5 and 3.0. The associations with admission to NICU and emergency Caesarean section for fetal distress are more difficult to interpret. First, for both of these outcomes, the association was weaker than it was for SGA. Second, in both cases the associations could be a consequence of the scan rather than an outcome predicted by the scan. Only two studies containing <5% of the patients included in the meta-analysis blinded the results of the scan. Revealing the results of the scan could explain both associations. In the case of NICU admission, revealing the scan result could lead to a decision to deliver the baby for suspected fetal distress. If this occurs preterm or at early term weeks of gestational age it could lead to an association with NICU admission through iatrogenic prematurity. In the case of caesarean delivery for fetal distress, revealing the result that there is severe oligohydramnios could be used as an indication (in whole or in part) to perform a caesarean section for suspected fetal distress. Alternatively, if a caesarean section was performed for failure to progress it is possible that the operator may include suspected fetal distress in the indication given the presence of the scan finding.

It is, however, also possible that the negative association with adverse neonatal outcome is due to treatment paradox. Given that the diagnosis was known in >95% of cases in the meta-analysis, the attending clinicians may well have put interventions in place that prevented adverse outcome. These could include enhanced levels of fetal monitoring, IOL, or delivery by pre-labour Caesarean section. A further complexity is that the aetiology of severe oligohydramnios may differ between studies as some excluded women with ruptured fetal membranes whereas others did not.

In conclusion, this analysis confirms that severe oligohydramnios is associated with adverse pregnancy outcome. 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.

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, by using borderline oligohydramnios we were able to analyse previously unpublished data which were obtained from the POP study of unselected nulliparous women where a blinded assessment of the presence or absence of borderline oligohydramnios. 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

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 study (POPS), as described in Chapter 4. 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.[7]

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, the Cochrane database of Systematic Reviews (CDSR) and Cochrane Central Register of Controlled Trials (CENTRAL) from inception to June 2019. 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.

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.

Study quality assessment and data extraction

The literature search, study selection, and analysis ware performed independently by two authors (AM and IA) 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 outlined in the Cochrane Handbook of Diagnostic Test Accuracy Studies.[34] 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, cut-off values used), reference standard (pregnancy outcome, gestation at delivery, and interval from scan to delivery).

Statistical and meta-analysis methods

The statistical methods employed are described in Chapter 4.

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 (Appendix 4, Figure 35) and 108 (3.2%) of these women had borderline oligohydramnios (AFI 5-8 cm, Appendix 4). The maternal age, socio-economic deprivation, ethnicity, BMI, and rates of alcohol consumption and smoking were similar between the two groups (Appendix 4, Table 22). 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 5. 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 Appendix 4, Figure 36. We identified 11 studies[86-95] (including the POP study) that met our inclusion criteria involving 37,848 patients in total. The study characteristics are presented in Appendix 4, Table 23. Only the POP study (N=3387) included unselected pregnancies, three studies[88, 94, 95] (N=1890) included only low-risk pregnancies and seven studies[86, 87, 89-93] (N=32,571) included mixed risk pregnancies. Two studies[94] (N=3817) were prospective and nine studies[86-93, 95] (N=34,031) were retrospective. Seven studies[88, 90-94] (N=36,293) defined borderline oligohydramnios as between 5 and 8 cm and four studies[86, 87, 89] (N=1555) as between 5 and 10 cm. The majority of patients in all the studies delivered at term. However, four studies[86, 89, 92, 94] 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 Appendix 4, Figure 37. 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[90] as high risk for selection bias as they used only low risk patients for their comparison group and two studies[86, 87] 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[86, 89, 91, 93, 95]

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 6. The most commonly reported outcomes were SGA <10th centile (9 studies), NICU admission (8 studies), 5 minute Apgar score less than 7 (8 studies), meconium stained amniotic fluid (7 studies) and caesarean section for fetal distress (6 studies). 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 7. Forrest plots of the DORs (Figure 8) demonstrated heterogeneity which was statistically significant for SGA and NICU admission. 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 (Appendix 4, Figure 38). The test showed no evidence of publication bias (P=0.33).

Table 5. 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).

Table 6. 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 7. 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.

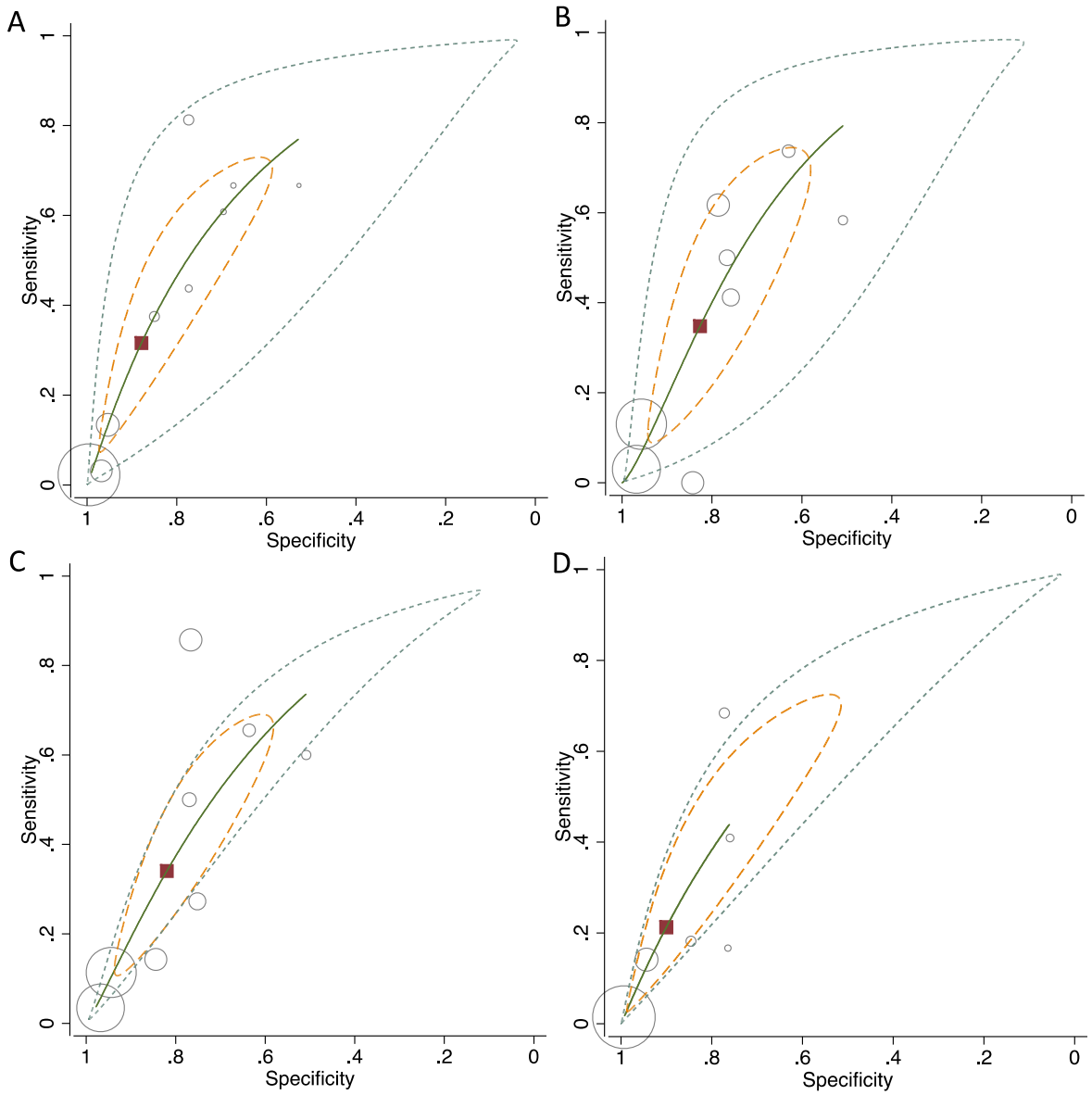
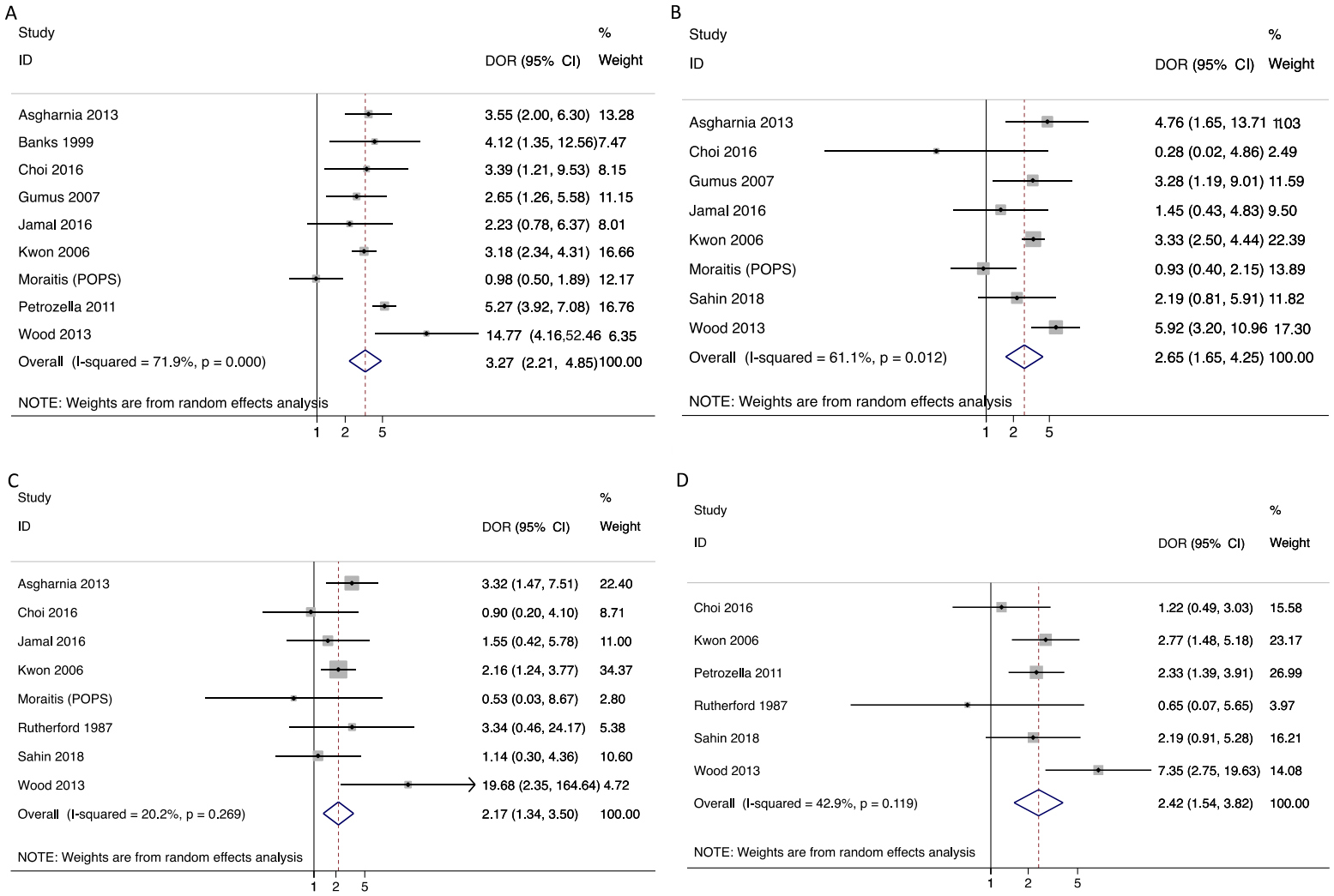


Figure 8. DORs of borderline AFI at predicting A. SGA <10th centile, B. NICU admission, C. 5-minute Apgar score <7, D. Caesarean section for fetal distress.



DOR = Diagnostic odds ratio

86. 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**:705-10.
87. Banks EH, Miller DA. Perinatal risks associated with borderline amniotic fluid index. *American Journal of Obstetrics & Gynecology* 1999;**180**:1461-3.
88. Choi SR. Borderline amniotic fluid index and perinatal outcomes in the uncomplicated term pregnancy. *Journal of Maternal-Fetal & Neonatal Medicine* 2016;**29**:457-60.
89. Gumus, II, Koktener A, Turhan NO. Perinatal outcomes of pregnancies with borderline amniotic fluid index. *Archives of gynecology and obstetrics* 2007;**276**:17-9. https://doi.org/10.1007/s00404-006-0309-x
90. Jamal A, Kazemi M, Marsoosi V, Eslamian L. Adverse perinatal outcomes in borderline amniotic fluid index. *International Journal of Reproductive Biomedicine* 2016;**14**:705-8.
91. 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**:545-9.
92. Petrozella LN, Dashe JS, McIntire DD, Leveno KJ. Clinical significance of borderline amniotic fluid index and oligohydramnios in preterm pregnancy. *Obstetrics & Gynecology* 2011;**117**:338-42.
93. 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**:353-6.
94. 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**:3085-8.
95. 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**:705-11.

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 of pregnancies where ultrasonic fetal biometry indicated an SGA infant demonstrated that early delivery was associated with a significantly decreased the risk of delivering a baby with a birth weight <3rd percentile.[96] 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.[89] 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.[86] 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.[94]

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. Systematic review of the diagnostic effectiveness of universal ultrasonic screening using fetal macrosomia in the prediction of adverse perinatal outcome.

Birth weight is a basic characteristic which defines an individual: the weight and sex of a baby are key themes in discussion following a birth. Similarly, when considering pregnancy outcome and its associations with the subsequent health of the infant, birth weight is critically important. Much of the focus on birth weight is on babies which are SGA due to its association with perinatal mortality. The diagnostic effectiveness of ultrasound in that context was the subject of a Cochrane review of diagnostic effectiveness[22] and this is discussed extensively in the following chapter. However, being born large for gestational age (LGA) is also a predictor of adverse outcome including perinatal mortality and morbidity arising from traumatic delivery and this is the focus of the current chapter.

Ultrasonic estimation of fetal weight (EFW) was first described more than 40 years ago.[97] The most widely employed equation for EFW was published by Hadlock in 1985[4] and a reference range for EFW was published in 1991[5]. A subsequent multi-country study by the World Health Organisation derived very similar EFW percentiles as described by Hadlock in Houston (TX, USA) in the early 1990s. Hence, the diagnostic tools have been available for many years to identify SGA and LGA fetuses. Moreover, an RCT has indicated that routine IOL in the presence of suspected macrosomia may prevent shoulder dystocia, one of the key adverse outcomes associated with an infant being LGA.[98]

Despite the above, it is still not clear whether screening and intervention for suspected fetal macrosomia is clinically effective. The HTA is currently funding an RCT of intervention in women diagnosed with an LGA infant ("Induction of labour for predicted macrosomia: the Big Baby trial"; ISRCTN18229892). However, as universal ultrasound in late pregnancy is not recommended in the UK, these women will have received a clinically indicated scan. Although the trial will determine whether intervention is useful in that group, it will not resolve whether screening and applying the same intervention to screen positive women as it is likely that the diagnostic effectiveness of the test will vary between women is clinically effective scanned routinely and those scanned for a clinical indication. Hence, the aim of the present study was to quantify the diagnostic effectiveness of universal ultrasound in late pregnancy to predict delivery of a large baby and one its major associated complications, namely, shoulder dystocia.

Methods

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). The search was done on then 22nd of October 2018. No restrictions for language or geographic location were applied. The protocol for the review was designed a priori and registered with the PROSPERO International Prospective Register of Systematic Reviews (Registration number: CRD42017064093). 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”.

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 in high risk patients such as patients with pre-existing or gestational diabetes and studies that the ultrasound was performed intrapartum. We included studies regardless of the formula and threshold they used to define macrosomia. We also included studies regardless of blinding of the result to the clinicians. We included studies that reported the following outcomes: LGA (defined as birthweight > 4000g or >90th centile) and severe LGA (birthweight >4500g or above the 97th centile); shoulder dystocia; adverse neonatal outcomes such as neonatal unit admission, 5-minute Apgar score <7, and neonatal metabolic acidosis.

Study quality assessment and data extraction

The literature search, study selection and analysis ware performed independently by two authors (AM and NS) 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 outlined in the Cochrane Handbook of Diagnostic Test Accuracy Studies. This tool assesses the included studies for potential bias in four domains: patient selection, index test, reference standard and flow and timing. We assessed the risk for flow and timing from the perspective of universal ultrasound screening at about 36wkGA. 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 (gestational age at scan, formula and cut-off values used), reference standard (pregnancy outcome, gestational age at delivery, and interval from scan to delivery). We also collected information such as inclusion or exclusion of patients with pre-existing or gestational diabetes.

Statistical and meta-analysis methods

The statistical methods employed are described in Chapter 4.

Results

The literature search flowchart is presented in Appendix 5, Figure 39. We identified 40 studies[99-138] that met our inclusion criteria involving 66,187 patients in total. The study characteristics are presented in Appendix 5, Table 24. Five studies[102, 111, 117, 120, 135] (N=8088) included unselected pregnancies, nine studies[107, 113, 115-117, 119, 126, 128, 136, 137] (N= 6436) included only low-risk pregnancies and 26 studies[99-101, 103-106, 108-110, 112, 114, 118, 121-125, 127, 129-134, 138] (N= 51,663) included mixed risk pregnancies.

The assessment of study quality was performed using the QUADAS-2 tool and is summarized in Appendix 5, Figure 40. The main risk of bias was for reference standard because only two studies, Sovio 2018 (POP study)[135] and Galvin 2017 (GENESIS study)[113] blinded the results to the clinicians. The second more common risk of bias was for flow and timing. This is because six studies had very short interval between ultrasound and delivery (the ultrasound was done either prior to IOL or less than 72 hours from delivery), two studies had long interval (ultrasound performed prior to 33wkGA) and two studies did not specify the gestational age at delivery. Finally, three studies only included prolonged (>41 weeks) pregnancies which were classified as having "high applicability concerns due to patient selection".

The most commonly reported outcomes were birthweight above 4000g (29 studies) followed by birthweight above the 90th centile (7 studies) which we both classified as large for gestational age (LGA). We defined as severe LGA a birthweight above 4500g (6 studies) or above the 95th or 97th centiles (two studies). Shoulder dystocia was reported in 6 studies. Finally neonatal morbidity (any related outcomes) was reported in only two studies, and consequently we could not produce summary results for this outcome. The most commonly used formulas for EFW were those described by Hadlock[4] et al, followed by Shepard. The most common thresholds for suspected LGA on scan were 4000g (21 studies) and 90th centile for the gestational age (9 studies). The abdominal circumference was used in 9 studies with the most common threshold applied being 36 cm (5 studies).

We present the summary diagnostic performance in Table 7. An estimated EFW >4000g or the 90th centile had above 50% sensitivity for predicting LGA at birth and this was similar regardless of the formula used. The positive likelihood ratios for the Hadlock formulas ranged between 7.5 and 12 and for the Shepard formula it was about 5. The AC had similar performance with the EFW. Suspected LGA also had about 70% sensitivity at predicting severe LGA at birth. Finally, an EFW >4000g or 90th centile

had 22% sensitivity at predicting shoulder dystocia with a statistically significant positive likelihood ratio of 2.1.

The summary ROC curves for LGA and shoulder dystocia are presented in Figure 9. We also present the pooling of the DORs (Figure 10). Finally we used the Deeks' funnel plot asymmetry test to assess the risk of publication bias using the outcome of LGA for the analysis (Appendix 5, Figure 41). The test showed potentially significant risk of publication bias (P=0.02).

Table 7. 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)	29	34,198	53.5% (47.3-59.6%)	93.9% (91.8-95.5%)	8.82 (6.83-11.4)	0.49 (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)	6	14,801	57.3% (47.0-67.0)	95.2% (92.3-97.0%)	11.89 (7.81-18.10)	0.45 (0.36-0.56)
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 95th centile)						
	EFW (any) >4000g (or 90th centile)	4	5839	70.2% (42.6-88.2%)	89.2% (74.4-95.9%)	6.49 (2.2-19.1)	0.33 (0.14-0.78)
	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)

Figure 9. 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.

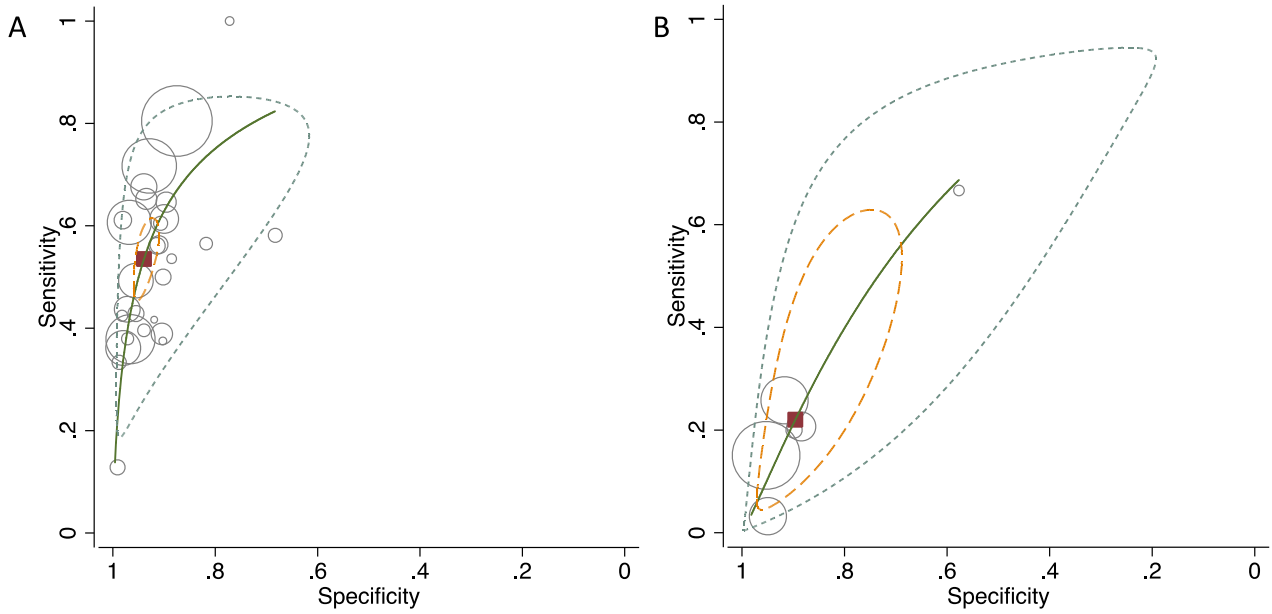
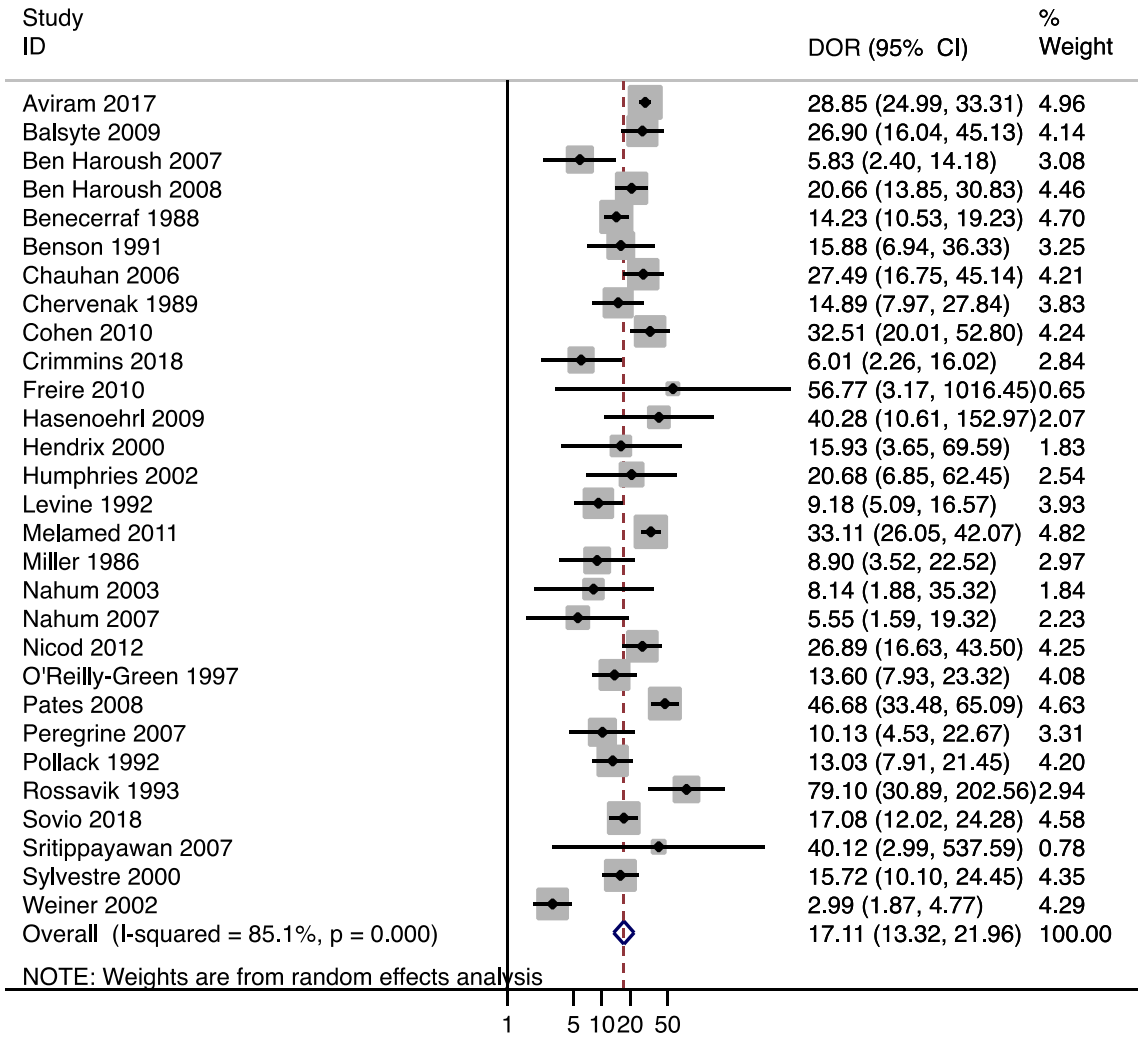


Figure 10. DORs for the diagnostic performance of EFW > 4000g (or 90th centile) at predicting A. LGA at birth (birthweight above 4000g or above the 90th centile) and B. Shoulder dystocia.

A



99. 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**:513-7.

100. 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**:453-8.

102. 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**:187-92.

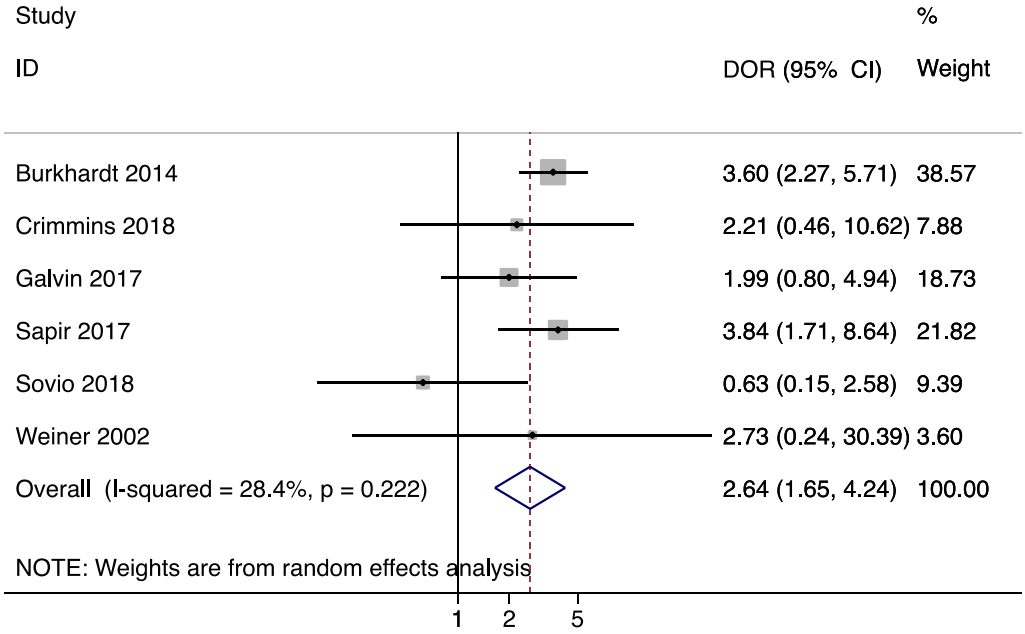
103. Ben-Haroush A, Melamed N, Mashiach R, Meizner I, Yogev Y. Use of the amniotic fluid index combined with estimated fetal weight within 10 days of delivery for prediction of macrosomia at birth. *Journal of Ultrasound in Medicine* 2008;**27**:1029-32.

104. 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**:149-51.

106. 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**:601-6. https://doi.org/10.1016/j.ajog.2006.04.012
107. 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**:753-6.
108. 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**:434-41.
109. 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**:140-5.
112. 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**:4-10.
113. 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**:S68. https://doi.org/https://doi.org/10.1016/j.ajog.2016.11.983
114. 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**:936-8.
115. 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**:585-90.
116. 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**:317-22.
118. 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**:108-12.
119. 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**:566-71. https://doi.org/10.1002/uog.6378
120. 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**:1345-9.
121. 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**:55-8.
122. 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**:74-81.
123. Miller JM, Jr., Korndorffer FA, 3rd, Gabert HA. Fetal weight estimates in late pregnancy with emphasis on macrosomia. *Journal of Clinical Ultrasound* 1986;**14**:437-42.
125. 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**:13-22.
126. 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**:148-56.
127. 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**:2022-7.

128. 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**:403-8. https://doi.org/10.1046/j.1469-0705.1997.09060403.x
129. Pates JA, McIntire DD, Casey BM, Leveno KJ. Predicting macrosomia. *Journal of Ultrasound in Medicine* 2008;**27**:39-43.
130. 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**:304-9.
131. Pollack RN, Hauer-Pollack G, Divon MY. Macrosomia in postdates pregnancies: the accuracy of routine ultrasonographic screening. *American Journal of Obstetrics & Gynecology* 1992;**167**:7-11.
132. Rossavik IK, Joslin GL. Macrosomatia and ultrasonography: what is the problem? *Southern Medical Journal* 1993;**86**:1129-32.
135. 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**:783-91. https://doi.org/10.1002/uog.17491
136. 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**:1058-67.
137. 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**:287-90.
138. 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**:20-4.

B



105. 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**:77-82.

109. 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**:140-5.

113. 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**:S68. https://doi.org/https://doi.org/10.1016/j.ajog.2016.11.983

133. 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.

135. 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**:783-91. https://doi.org/10.1002/uog.17491

138. 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**:20-4.

Discussion

The key findings of the present study 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 and the preceding chapters we have focused analysis on 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[135] 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.

A recurrent theme in all of the chapters 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 were the only two large studies

(>1000) women not to 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.[139]

Finally, the relationship between fetal macrosomia 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.[140-142] 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 9. Conclusions regarding the evidence around universal ultrasound screening of nulliparous women in late pregnancy.

The preceding chapters outline the association between umbilical artery Doppler, CPR, severe oligohydramnios, borderline oligohydramnios, and fetal macrosomia and the risk of adverse pregnancy outcome. The main overall conclusions are as follows:

1. Umbilical artery Doppler, CPR, severe oligohydramnios, borderline oligohydramnios and fetal macrosomia were all either non-predictive or weakly predictive of the risk of neonatal morbidity.
2. Umbilical artery Doppler, CPR, severe oligohydramnios, borderline oligohydramnios were all weakly predictive of the risk of delivering an SGA infant.
3. The vast majority of the studies did not blind the result of the index test. Hence, interpreting the results in relation to prediction of adverse neonatal outcome could be biased against not seeing associations where true associations exist (e.g. through treatment paradox) or biased towards seeing associations where no true associations exist (e.g. through ascertainment bias or iatrogenic harm).
4. Only the POP study has reported the range of ultrasonic findings in late pregnancy in unselected nulliparous women, which is the optimal study design and was conducted in the target population. A second study conducted in Ireland (Genesis) also performed blinded ultrasound scans in late pregnancy in nulliparous women but has not published widely on the results.
5. 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.

During the period of the current project, a systematic review of diagnostic test accuracy in relation to ultrasonic diagnosis of SGA, using EFW was published.[22] In searching the literature, they made the following observation regarding blinded ultrasonic assessment of fetal growth:

"*Sovio 2015 [POP study] blinded clinicians to the results of the universal ultrasonography and Weiner 2016 blinded clinicians to results of all ultrasound methods other than the one they conducted, but in the majority of studies clinicians either were not blinded to test results or this was not reported*".

The Weiner et al. (2016)[143] study was performed on 405 women during active labour and compared clinical assessment of fetal size versus ultrasonic EFW. Hence, the conclusion of the Heazell systematic review is that the only study which performed blinded ultrasonic assessment of SGA relevant for population screening in the antenatal period was the POP study.

We were aware of the Heazell review and did not, therefore, address ultrasonic diagnosis of SGA in the present review.

They reported detection of SGA (birth weight <10th percentile) as follows:

For a specificity of 88%, ultrasonic suspicion of SGA had a sensitivity of 74% (95% confidence intervals 64% to 83%). In the POP study, the sensitivity was 57% for a specificity of 90%. The meta-analysis reported detection of severe SGA (birth weight <3rd percentile) as follows:

For a specificity of 87%, ultrasonic suspicion of SGA had a sensitivity of 66% (95% confidence intervals 56% to 76%). In the POP study, the sensitivity was 77% for a specificity of 87%. The Heazell review is slightly surprising in that one would have predicted better prediction of the more severe outcome. The inconsistency between these two analyses may reflect inclusion of different studies which may have included different populations. However, the review does suggest that the data observed in the POP study were generally comparable to the studies included in the Heazell review.

A further level of complexity in considering these issues is that, generally, an ultrasonic assessment of the fetus typically includes measurement of multiple parameters simultaneously. Hence, a further issue in trying to apply the findings of the Heazell review and our own reviews to health economic analysis and trial design is that none of the reviews completely captures what might be expected to happen clinically. This issue is affected by another layer of complexity, namely, defining the features on a scan that the majority of clinicians would accept as indicating FGR. This last question has been addressed by researchers employing the Delphi consensus method to generating an agreed ultrasonic diagnosis of FGR. The paper arising from this process was published in 2016.[144] These authors described the following criteria for diagnosis of late FGR (32wkGA or beyond):

EFW or AC <3rd percentile

OR

2 *or more of the following:*

(i) EFW/AC <10th (ii) EFW/AC falling 2 quartiles, or (iii) CPR <5th percentile or UA Doppler >95th

In a paper in Lancet CAH in 2018[145], the POP study data were used to compare the Delphi definition of late FGR using the blinded 36wkGA scan with simply an EFW <10th as a predictor of the risk of delivering an SGA infant with complications. The results are reproduced in the Table 8 below:

Table 8. Diagnostic effectiveness of ultrasonic screening at 36wkGA for subsequent delivery of an SGA infant associated with either maternal preeclampsia or perinatal morbidity or mortality.

Screening test	Positive LR (95% CI)	Negative LR (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Ultrasonic EFW <10th	5·1 (4·2-6·3)	0·38 (0·26-0·54)	67·2 (53·8-78·3)	86·9 (85·8-88·0)
Delphi definition of late FGR	5·9 (4·7-7·4)	0·43 (0·31-0·60)	61·4 (47·9-73·4)	89·6 (88·6-90·6)

From Gaccioli et al (2018), Lancet CAH.[145]

In fact, the diagnostic effectiveness appeared to be quite similar comparing the two approaches. It is worth acknowledging that, due to the absence of MCA Doppler, we were unable to include a specific subset of fetuses that would have been defined as FGR by the Delphi method, namely, those where the CPR was <5th but the umbilical artery Doppler was <95th, the EFW>3rd and AC>3rd but the baby filled one of the other two criteria (EFW/AC <10th or EFW/AC falling 2 quartiles). However, given the lack of association between CPR and neonatal morbidity described in Chapter 5, we not believe that it is likely that inclusion of this group would have profoundly altered the results.

Taking the totality of the data, the approach we employed for the health economic analysis was that we defined screen positive as either ultrasonic EFW<10th (suspected SGA) or ultrasonic EFW>90th (suspected LGA). The Heazell et al review demonstrated good diagnostic effectiveness for SGA and the analysis in Chapter 8 demonstrated that ultrasonic suspicion of macrosomia was strongly associated with the risk of delivering a large baby. The attractiveness of this approach was underlined by the fact that there were Cochrane reviews which reported meta-analysis of RCTs of IOL in both situations, and there are extensive epidemiological data on the outcome of SGA and LGA pregnancies. There was one additional further exposure which is detectable by scan and where management is informed by RCT evidence, namely, breech presentation near term. Ultrasound establishes fetal presentation with 100% accuracy at the time of the scan (although the presentation will sometimes change spontaneously after the scan). Hence, we included this in subsequent analyses.

Chapter 10. Evidence based protocol for the care of screen positive women.

The preceding chapter identified three elements in a late pregnancy ultrasound scan where there was evidence that the screening result identified a high risk fetus, namely, breech presentation, an SGA fetus, or an LGA fetus. We next sought to determine the evidence base which existed to inform interventions in women who had these features and employed the search engine of the National Institute for Health and Care Excellence, at https://www.evidence.nhs.uk/

Management plan for breech presentation

This search identified an existing UK based guideline from the RCOG, Management of Breech Presentation (Green-top Guideline No. 20b).[12] In brief, women who do not have a contraindication to external cephalic version are offered this procedure (turning of the fetus by manual manipulation of an un-anaesthetised woman). Where the procedure is contraindicated, declined or unsuccessful women would then have a discussion regarding attempting vaginal breech birth. Where vaginal breech birth was contraindicated or declined, a planned caesarean section would be scheduled at 39 weeks (in the absence of a clinical indication for earlier delivery) with the proviso that the woman would be delivered by emergency caesarean section if she presented in labour before the scheduled date. Women who had a successful ECV would have routine care thereafter, but with midwife checks that the baby had not reverted to breech. In practice, given that the target population is nulliparous, it would be a small minority who would opt for vaginal breech birth and no women took up this option in the POP study.[10] For the purposes of the Markov chain modelling and health economic analysis we used the effect estimates of a Cochrane review which quantified "the effects of planned Caesarean section for singleton breech presentation at term on measures of pregnancy outcome" .[13] Other parameters were obtained from the literature and are detailed in the chapters below.

Management plan for diagnosis of SGA

We next used the NICE evidence search engine to identify existing guidelines around the management of SGA. This search identified an existing UK based guideline from the RCOG, Small-for-Gestational-Age Fetus, Investigation and Management (Green-top Guideline No. 31).[146] Much of this guideline was focused on early pregnancy identification of risk factors and the management of the preterm SGA fetus. However, the following were key points in relation to the detection and management of SGA at term.

- *"In the term SGA fetus with normal umbilical artery Doppler, an abnormal middle cerebral artery Doppler (PI < 5th centile) has moderate predictive value for acidosis at birth and should be used to time delivery.*
- *In the SGA fetus detected after 32 weeks of gestation with normal umbilical artery Doppler, a senior obstetrician should be involved in determining the timing and mode of birth of these pregnancies.*
- *Delivery should be offered at 37 weeks of gestation.*
- *In the SGA fetus with umbilical artery AREDV delivery by caesarean section is recommended.*
- *In the SGA fetus with normal umbilical artery Doppler or with abnormal umbilical artery PI but end–diastolic velocities present, induction of labour can be offered but rates of emergency caesarean section are increased and continuous fetal heart rate monitoring is recommended from the onset of uterine contractions."*

The same search also identified an NHS England Care Bundle which aimed to reduce rates of perinatal death, "*Saving Babies' Lives Version Two: A care bundle for reducing perinatal mortality*". This guideline had a section on management of SGA at term and the following were key relevant sections.

- *"Accepting the proviso that all management decisions should be agreed with the mother in the cases of fetuses <3rd centile and with no other concerning features, initiation of labour and/or delivery should occur at 37+0 weeks and no later than 37+6 weeks gestation.*
- *Fetuses between 3rd – 10th centile will often be constitutionally small and therefore not at increased risk of stillbirth. Care of such fetuses should be individualised and the risk assessment should include Doppler investigations, the presence of any other high risk features for example, recurrent reduced fetal movements, and the mother's wishes. In the absence of any high risk features delivery or the initiation of IOL should be offered at 39+0 weeks."*

However, the context for both the RCOG and NHS England guidelines was the management of women who were identified through the current approach of targeting ultrasound to high risk women. As outlined in the preceding chapter, we have not found evidence that these additional ultrasonic tests are diagnostically effective when used as a screening test. Hence, the management protocol for SGA employed in the health economic analysis is to offer IOL. For the purposes of the health economic analysis we used the effect estimates of a Cochrane review which quantified "the effects of immediate delivery versus expectant management of the term suspected compromised baby on neonatal, maternal and long-term outcomes" [147]. In practice, 90% of the women included in the review came from a trial of IOL for suspected FGR [96]. IOL took place in the intervention group of this trial at an

average of 38 weeks gestational age and we have incorporated this into our management protocol (see below). This does not represent an extreme intervention as a large scale NIH-funded RCT demonstrated no adverse effect of routine IOL at 39 weeks in nulliparous women who lacked risk factors [148]. Other parameters were obtained from the observational literature and are detailed in the chapters below.

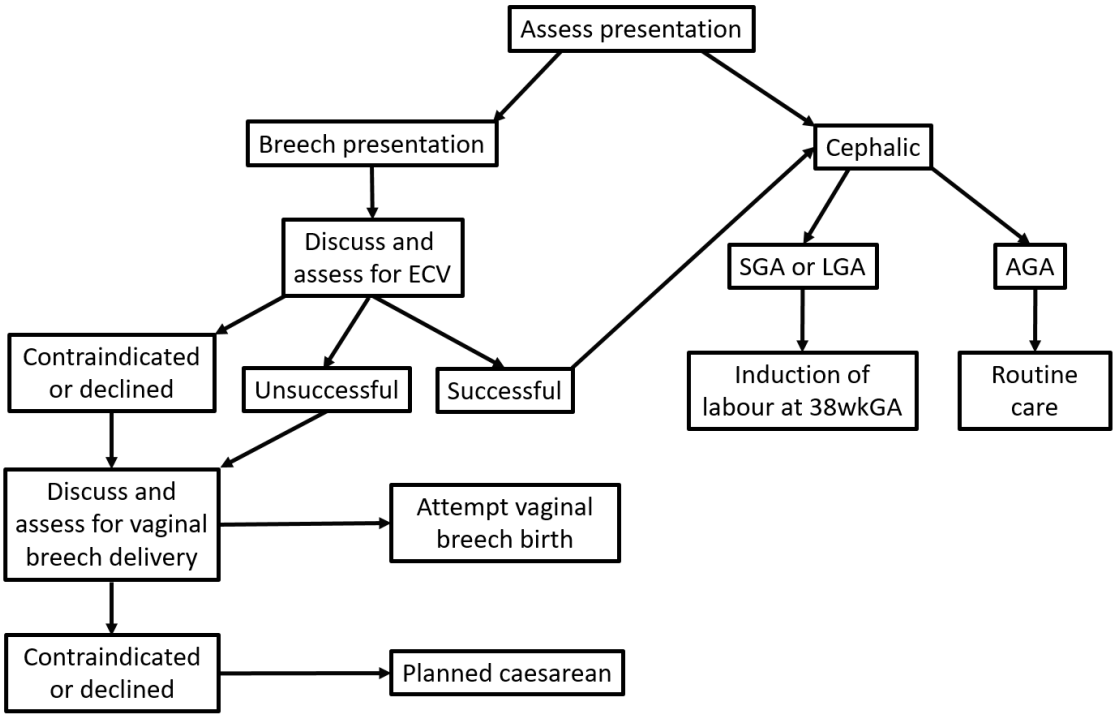
Management plan following diagnosis of LGA.

We next used the NICE evidence search engine to identify existing guidelines around the management of LGA. The only guidelines which we identified using this search related to women with diabetes. These women are routinely scanned during pregnancy and have specific issues and the recommendations relating to this group are not generalisable to the population of interest in the current report. However, the search did identify a number of systematic reviews which addressed IOL and one of these was a Cochrane review.[149] The Cochrane review had the following conclusions:

"induction of labour for suspected fetal macrosomia results in a lower mean birthweight, and fewer birth fractures and shoulder dystocia... [the] trials included in the review suggest that to prevent one fracture it would be necessary to induce labour in 60 women. Since induction of labour does not appear to alter the rate of caesarean delivery or instrumental delivery, it is likely to be popular with many women... the advantages and disadvantages of induction at or near term for fetuses suspected of being macrosomic should be discussed with parents. Although some parents and doctors may feel the evidence already justifies induction, others may justifiably disagree. Further trials of induction shortly before term for suspected fetal macrosomia are needed."

Consistent with this recommendation, the HTA has funded an RCT ["Induction of labour for predicted macrosomia: the Big Baby trial" (ISRCTN18229892)]. Given the uncertainty in the evidence base, it is not possible to develop a robust management plan for management following diagnosis of macrosomia. For the purposes of the Markov chain modelling and health economic analysis, we addressed this uncertainty by comparing multiple strategies, including expectant management, early term IOL and planned caesarean delivery. The effects in relation to IOL were taken from the Cochrane review, as this was assessed as the highest quality evidence available at the time of writing. About 70% of the women came from a single trial[98] where the most common week for IOL was 38wkGA. Other parameters for the modelling and health economic analysis were obtained from the observational literature and are detailed in the chapters below. A summary management plan is outlined in the figure below.

Figure 11. Summary of management plan following 36wkGA scan.



Chapter 11. Economic analysis of universal versus selective ultrasound screening in late stage pregnancy: cost-effectiveness and value of information analyses.

Introduction

This study was commissioned to evaluate the current evidence base as to the costs and effectiveness of performing a routine ultrasound scan in late pregnancy in all nulliparous women combined with appropriate management plans, to identify evidence gaps, and to predict whether future research to plug those gaps is likely to be a cost-effective use of healthcare resources. In this analysis, we use decision modelling to assess the likely outcomes from universal ultrasound screening and determine whether its potential benefits can be clinically and economically justified.

We present a cost utility analysis focusing on three of the main conditions detectable by ultrasound screening that may warrant intervention: breech presentation, SGA and LGA. The cost-effectiveness of universal ultrasound screening for each of these conditions individually has been explored previously.[10, 150] However, here we evaluate the cost-effectiveness of screening for all of these conditions at the same screening session. Furthermore, we use decision uncertainty to predict the expected return on further research. We have applied the simplified management plan outlined above (Figure 11). In essence, women are first assessed for presentation. Where the baby is in a breech presentation ECV is offered. If successful, the woman reverts to expectant management and, if unsuccessful, is delivered by planned caesarean. Where the baby is in a cephalic presentation and the estimated fetal weight is in the normal range, the woman receives expectant management. If the baby is either SGA or LGA, IOL is offered. However, we also compare combined assessment for presentation and fetal biometry with a scan simply for presentation. The rationale for this is that a presentation scan may be readily implemented and relatively inexpensive, and there is much less uncertainty about the utility of knowing the baby's presentation versus estimating the baby's size.

The structure of this chapter is as follows: under 'Methods', we first introduce the general methodology for our economic evaluation. We then summarise the clinical definitions used, as well as the competing strategies evaluated in this study before introducing the structure of the economic simulation model underlying the analysis. Once the model structure and mechanics have been explained, we discuss how we populated the model with the best available data; complete technical details regarding how individual parameters were derived is presented in Appendix 7. Finally, we describe the base case analyses, sensitivity analyses and value of information analysis (VOI) to guide how future research in this area should be prioritized.

Under ‘Results’, we first present the results for the baseline economic evaluation, and sensitivity analsyses. The results from the VOI analysis are then presented: these include the results for the expected value of perfect information (EVPI), the expected value of partial perfect information (EVPPI), and finally the expected value of sample information (EVSI).

Under ‘Discussion’, we summarise key findings, explain the interpretation of our results, and discuss what impact our methodological limitations may have had upon the results.

Methods

To compare long-term health and cost outcomes associated with different strategies of screening in third-trimester pregnancy, we constructed an economic simulation model. We focused the model upon two features for which late-pregnancy ultrasound is amenable to detect: fetal presentation and fetal size. We used a decision tree model, consisting of four sub-trees, one each for breech presentation, LGA, SGA, and AGA. The model structure is largely based upon previous economic analyses of screening for these conditions individually, and the development and key characteristics of these sub-models models have previously been described[10, 150]; a brief summary is provided in Appendix 6. The previous chapters dealt with the diagnostic effectiveness of ultrasound in this setting and outlined how a positive result on scan could influence subsequent care. This chapter focuses on how these sub-models were incorporated into a joint framework, enabling cost-effectiveness analysis of simultaneous screening for all of these conditions.

Scope and population

The analysis relates to nulliparous women in England with singleton pregnancies, excluding those opting for elective CS for any reason except following a diagnosis of breech presentation. The economic analysis uses a public sector perspective defined as NHS and special educational needs costs. Outcomes are from the perspective of the foetus/infant.

Comparators and interventions

This analysis evaluated three different strategies for ultrasound screening in late pregnancy, defined as a scan between 36+0 weeks and 36+6 weeks. 'Selective US' (i.e. where ultrasound is only performed following clinical indication of its need) is the current standard in England.[146] 'Universal US for fetal size' would mean routinely offering a third-trimester ultrasonic assessment of fetal weight in every pregnancy. Given the simplicity of detecting fetal presentation during an ultrasound scan, such a screening strategy would also identify breech presentation. A third option would be to offer 'Universal US for presentation only', i.e. a simpler US scan with the sole purpose of detecting pregnancies with breech presentation. Compared to a standard antenatal ultrasound for which, typically, multiple measurements are made, an ultrasound scan for fetal 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.

We assumed that all identified cases of breech presentation would be offered an ECV unless contraindicated, in line with RCOG guidelines.[151] We further assumed that pregnancies identified as

SGA (whether correctly diagnosed or not) would be given early IOL. However, for pregnancies diagnosed as LGA, there is uncertainty as to whether intervention (IOL) is beneficial. For this reason, expectant management of suspected LGA pregnancies was also an option. We had previously considered also including elective CS for management of macrosomia, however ruled this out because it was inferior to IOL in our cost-effectiveness analysis of ultrasound assessment for macrosomia alone.[150] This conclusion was consistent with a previous decision model analysis.[152] We therefore compare six discrete strategies in the analysis (Table 9).

We assume that selective scanning (i.e. only where clinically indicated) with a policy of offering ECV for suspicion of breech, and IOL for suspicion of SGA or LGA (strategy 2, Table 9) represents an approximation of the status quo from which estimates of incremental net benefit are calculated.

Table 9. Comparator strategies for economic simulation model

Strategy	Screen	Offered management if diagnosed:		
		Breech+	Macrosomia+	SGA+
1	Selective	ECV	IoL	IoL
2	Selective	ECV	Exp	IoL
3	Breech only	ECV	IoL	IoL
4	Breech only	ECV	Exp	IoL
5	Universal	ECV	IoL	IoL
6	Universal	ECV	Exp	IoL

ECV = External cephalic version; Exp = Expectant management; IoL = Induction of labour; LGA = Large for gestational age; SGA = Small for gestational age

As discussed in the preceding chapters there is more uncertainty in relation to the management of LGA than SGA. However, performing fetal biometry will yield a percentile of EFW hence a scan involving fetal biometry can yield three possible outcomes: AGA, SGA or LGA. Consequently, we considered two possible approaches to screening involving fetal biometry. Both included IOL for SGA. However, one also included induction for LGA whereas the other dictated expectant management, given the uncertainty.

Outcomes

In the absence of any trials on third-trimester screening strategies with long enough follow-up, we could not directly estimate long-term health outcomes as a function of screening strategies alone (hence the need for this modelling study). Instead, we simulated outcomes at delivery (survival and different levels of neonatal complications/morbidity), and then simulated long-term health outcomes as a function of these short-term outcomes. Overall health gain was captured as QALYs[153] accrued to the infant. Overall costs for each screening strategy included the cost of the ultrasound scanning, possible intervention, delivery episode, neonatal care and mortality, and long-term care.

Model structure

As stated, the model structure is a decision tree. It was coded in R[154](R Core Team, 2017), using packages: 'BCEA', 'FinCal', 'ggplot2', 'gtools', 'readxl', 'tidyr', and 'SAVI'.[155-161] The code for the model is available from the corresponding author upon request.

Figure 12 shows the structure of the first stages of the decision model. The [+] indicates branches which are collapsed for clarity. Nodes are named to show their relationship to one another: nodes with the same letter have identical structures to the branches of the tree beyond, whilst a different number and/or lower case letter indicates a different set of probabilities. The prefixes B_, L_ and S_ denote nodes with probability sets specific to either breech, large or small for gestational age babies respectively.

At commencement, the scan policy can be set to either selective (i.e. status quo), a universal scan for presentation only, or universal scan for fetal biometry and presentation. The model structure is identical for each case. The difference is the sensitivity and specificity of the scanning policies, and their cost.

A pregnancy will be either in a breech or cephalic presentation (node A1), or be either LGA, SGA, or AGA (node A2). For ease of modelling we assume all four possibilities are mutually exclusive and structure hierarchically, beginning with presentation (breech or cephalic), and then fetal size (LGA, SGA, AGA). The implications of this are considered in the discussion. The probability of being in breech is the prevalence of breech at the time of screening (approximately 4.6%).[10] If the scan policy is universal ultrasound (whether for fetal biometry or presentation only), then, given the ease of interpretation of such a scan, we assume all breeches are detected (i.e. 100% sensitivity and specificity, node B_B). However, under the selective scan policy, approximately 45% of breeches will be undetected[10] due to the mother not having undergone a scan at all (for consistency with the rest

of the model we label these 'false negatives'). Further outcomes relating to breech presentation are described below (section 'outcomes relating to breech').

If the baby has cephalic presentation, it may be LGA, SGA or AGA. The probabilities of each are the prevalence of the conditions (node A2, by definition 10% for each). If a baby is LGA or SGA, the probability of detection is a function of the sensitivity of the scanning policy (nodes L_B and S_B, LGA: 26.55% under selective and presentation only scan policies, 37.85% under universal scan for fetal size.[135] SGA: 19.6% under selective and presentation only scan, 56.53% under universal scan for fetal size[7]). The sensitivity and specificity of ultrasound to detect SGA and LGA were derived from the POP study[7, 135]. The rationale for using the POP study values is that it was conducted in the English NHS, it involved nulliparous women being scanned at 36 weeks, it is the only level 1 study of the diagnostic effectiveness of ultrasound to predict SGA and LGA (i.e. where the test result was blinded) and the values of sensitivity and specificity for SGA were similar to a 2019 Cochrane review of diagnostic test accuracy[22] and the DOR from the POP study for macrosomia was identical to the DOR in the meta-analysis presented in Chapter 8 of this report.

If an LGA baby is correctly diagnosed positive for LGA, the pregnancy is managed according to the defined LGA policy; either IOL or expectant management (node 'MGT_LGA_TP'), in either case leading to either vaginal delivery or emergency CS (nodes L_C3 and L_C2a, odds ratio of EmCS, compared with otherwise healthy baby, 1.79[141]. If an LGA baby is misdiagnosed as AGA (i.e. false negative scan), delivery can be either vaginal or emergency CS. Further outcomes relating to LGA pregnancies are described below (section 'outcomes relating to LGA').

If the baby is SGA and is correctly diagnosed, the pregnancy is induced, leading to either vaginal delivery or emergency CS (node S_C3). False negatives may lead to vaginal delivery or emergency CS (node S_C2). Further outcomes relating to SGA pregnancies are described below (section 'outcomes relating to SGA').

An AGA baby may be misdiagnosed as SGA or LGA (false positive SGA and LGA respectively), or correctly diagnosed as AGA (node B). A false positive SGA baby will be induced unnecessarily, leading to either vaginal delivery or emergency CS (node S_C4). A false positive LGA baby will be managed according to the defined LGA policy; either IOL or expectant management (node 'MGT_LGA_FP'). IOL and expectant management can either lead to spontaneous vaginal or emergency CS deliveries (nodes

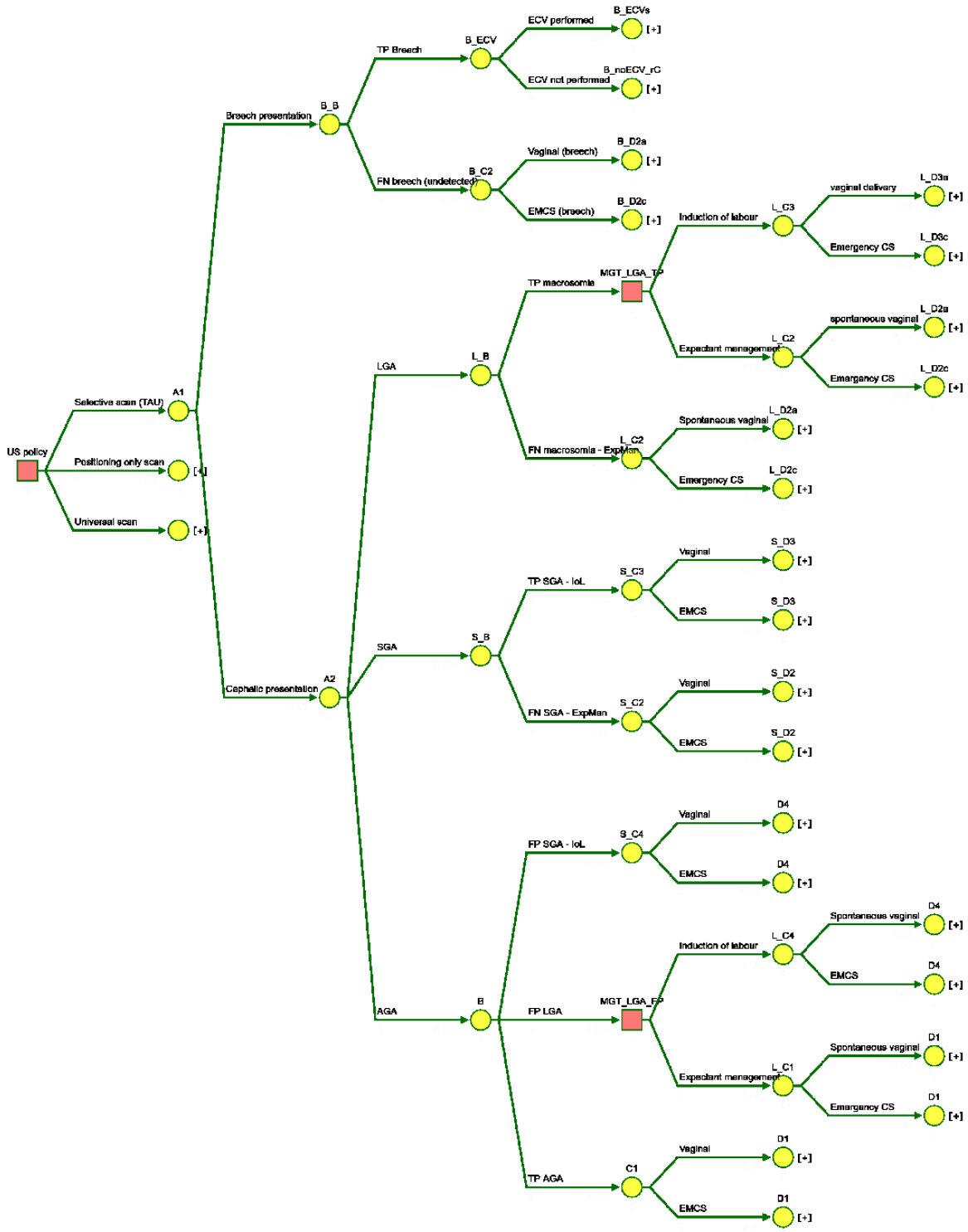
L_C4 and L_C1 respectively). Finally, a correctly diagnosed AGA baby (true negative) can undergo vaginal delivery or emergency CS (node C1).

Short- and long-term outcomes

For all parts of the model, different levels of neonatal morbidity and mortality are possible, although these outcomes are structured slightly differently between the model's sub-trees. For the breech, SGA and AGA models, delivery outcomes include no, moderate, and severe neonatal morbidity, as well as perinatal death. The risks of each level of adverse outcome differ between specific branches, i.e. is affected by the true status of the baby, the mode of delivery, and whether labour was induced early. Long-term outcomes are then modelled as a function of the level of neonatal morbidity at delivery. For the LGA model, delivery and long-term outcomes are modelled differently. This is explained in detail below (section 'outcomes relating to LGA').

Long-term outcomes include 'No long-term complications', 'Special educational needs', 'Severe neurological morbidity', and 'neonatal/infant mortality'. The risk of long-term complications increases with the level of neonatal morbidity (nodes E1, E2, and E3). Unlike delivery outcomes, long-term outcomes are not affected by the actual status of the baby prior to delivery, only by the level of neonatal morbidity at delivery. Importantly, this means that all screening and management options only affect long-term outcomes indirectly, through their impact upon the outcomes at delivery.

Figure 12. Model overview.



[+] = sub-branches of model collapsed for clarity.

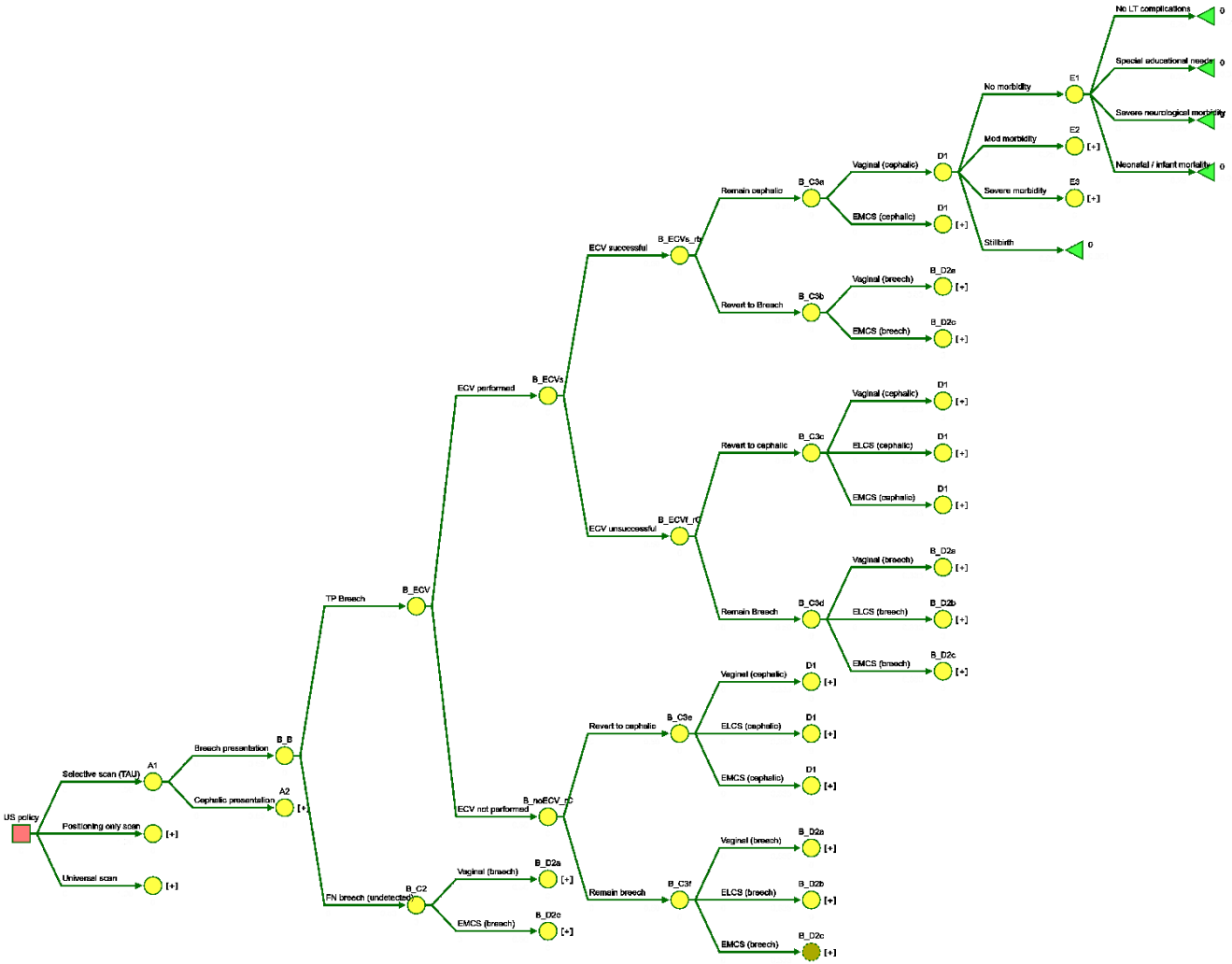
Outcomes relating to breech

Figure 13 shows the decision tree with outcomes relevant to breech expanded and remaining branches collapsed. The prevalence of breech refers to the fetal presentation at the time of screening. We assume that sensitivity and specificity for universal ultrasound is perfect at detecting fetal presentation, whether for size or breech presentation only. For selective ultrasound, the sensitivity is lower since not all women receive US screening, however, we assume that all cases of suspected breech presentation would be either confirmed or rejected through US, so that false positive diagnosis is not an option (i.e. perfect specificity).

On diagnosis of a breech, ECV is offered (node B_ECV). If the ECV is successful (node B_ECVs) and the infant remains cephalic (node B_ECVs_rb), no further intervention will be offered (i.e. expectant management). However, the baby may spontaneously revert back to breech (node B_ECVs_rb). In either case, there is a probability of emergency CS, which is increased if the baby has reverted to breech presentation (nodes B_C3b and B_C3a respectively). If a breech presentation is not diagnosed prior to labour, delivery options include breech vaginal delivery or emergency CS (node B_C2).

Following labour and delivery there is a risk of either no, moderate or severe neonatal complications or perinatal death (node D1), subsequently leading to no long-term complications, special educational needs, severe neurological morbidity or perinatal mortality (node E1). Note we assume no raised risk of neonatal morbidity associated with cephalic emergency CS versus cephalic vaginal delivery per se. We do however allow for a raised risk of complications with an emergency CS following breech compared with a vaginal breech delivery (nodes B_D2a and B_D2c). If ECV is not accepted, or fails, then elective CS may be offered.

Figure 13. Outcomes associated with breech.



[+] indicates collapsed sections of the decision tree

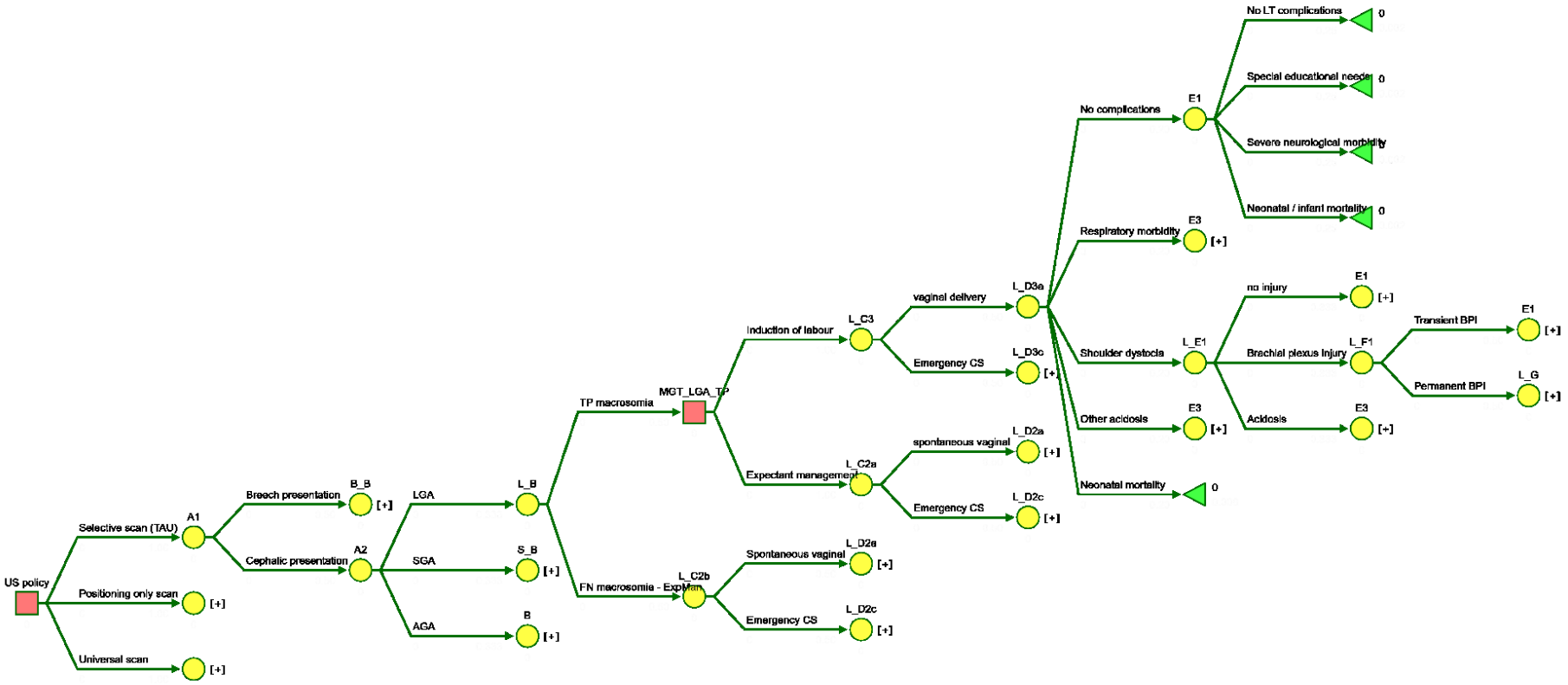
Outcomes relating to LGA

Figure 14. Outcomes associated with LGA. shows the decision tree with outcomes relevant to LGA expanded and remaining branches collapsed. Where LGA is suspected, intervention will be given according to the pre-determined management strategy (IOL or expectant management), both for true positive and false positive LGA diagnosis. The management option will affect the likelihood of both the delivery outcome, as well as the mode of delivery, which can be either vaginal or through emergency CS. Where LGA is not suspected, delivery can be either vaginal or through emergency CS.

Delivery outcomes include 'No complications', 'Respiratory morbidity', 'Shoulder dystocia', 'Other acidosis' (i.e. acidosis not caused by shoulder dystocia), and 'perinatal death'. The risk of each adverse outcome depends on the baseline risk, as well as on the mode of delivery, and whether labour was induced early.

Long-term outcomes depend on the outcome at delivery. For 'No complications', 'Respiratory morbidity', and 'Other acidosis', long-term outcomes included 'No long-term complications', 'Special educational needs' (SEN), 'Severe neurological morbidity', and 'neonatal/infant mortality'. For 'No long-term complications' the risk was equivalent to 'No neonatal morbidity' (node E1), whereas for 'Respiratory morbidity', and 'Other acidosis', the risk of long-term complications were equivalent to 'Severe neonatal morbidity (node E3). Shoulder dystocia (node L_E1) could result in either no complications, brachial plexus injury (BPI, node L_F1), or acidosis. BPI could be either transient or permanent (node L_G), where the latter had the same risk of long-term outcomes as for no neonatal morbidity (node E1), but with a penalty for quality of life. Permanent BPI, special education needs, and severe neurological morbidity were long-term events; any other morbidity was expected to be resolved within the first year of life.

Figure 14. Outcomes associated with LGA.



[+] indicates collapsed sections of the decision tree

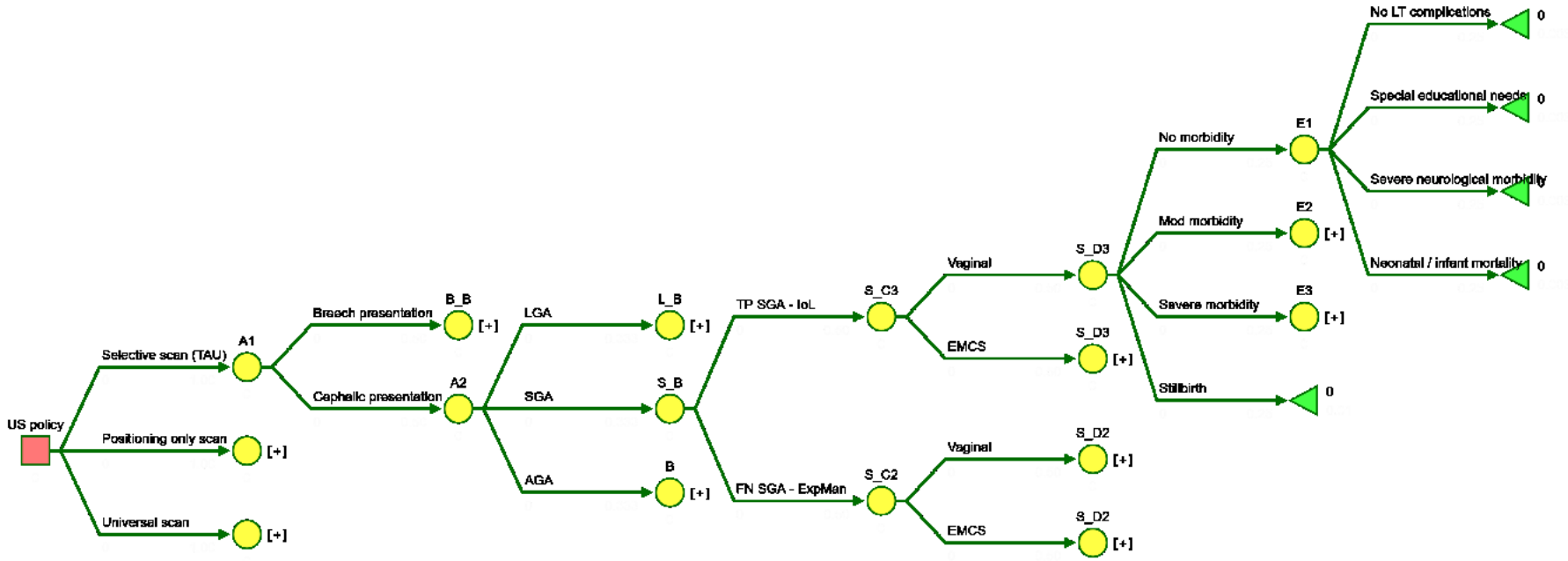
Outcomes relating to SGA

Figure 15 shows the decision tree with outcomes relevant to SGA expanded and remaining branches collapsed. Labour will be induced early for suspected cases of SGA, whether based upon a true or false SGA diagnosis. Deliveries can either be vaginal or through emergency CS. The probability of each mode of delivery is affected by whether or not labour was induced early. However, in order to avoid double counting the health effects of early labour induction, the mode of delivery affects only costs and not health outcomes.

Delivery outcomes include no, moderate, and severe neonatal morbidity, as well as perinatal death. The correctly diagnosed SGA pregnancies (true positives) are offered early IOL, which reduces the risk of morbidity and mortality. Where SGA is unsuspected (false negatives), pregnancies are managed expectantly, with no risk reduction. Note that early labour induction may also increase the risk of morbidity if initiated needlessly, i.e. in an AGA pregnancy falsely suspected of being SGA. However, in a true SGA pregnancy, early labour induction is expected to reduce the risk of morbidity. The scenario with a false positive diagnosis is discussed further below (section 'outcomes relating to AGA').

Long-term outcomes include 'No long-term outcomes', 'Special educational needs' (SEN), 'Severe neurological morbidity' (SNM), and 'neonatal/infant mortality'. Each outcome is possible for all levels of neonatal morbidity. However, the risk of long-term complications increases for moderate and severe neonatal morbidity (nodes E2 and E3).

Figure 15. Outcomes associated with SGA.



[+] indicates collapsed sections of the decision tree

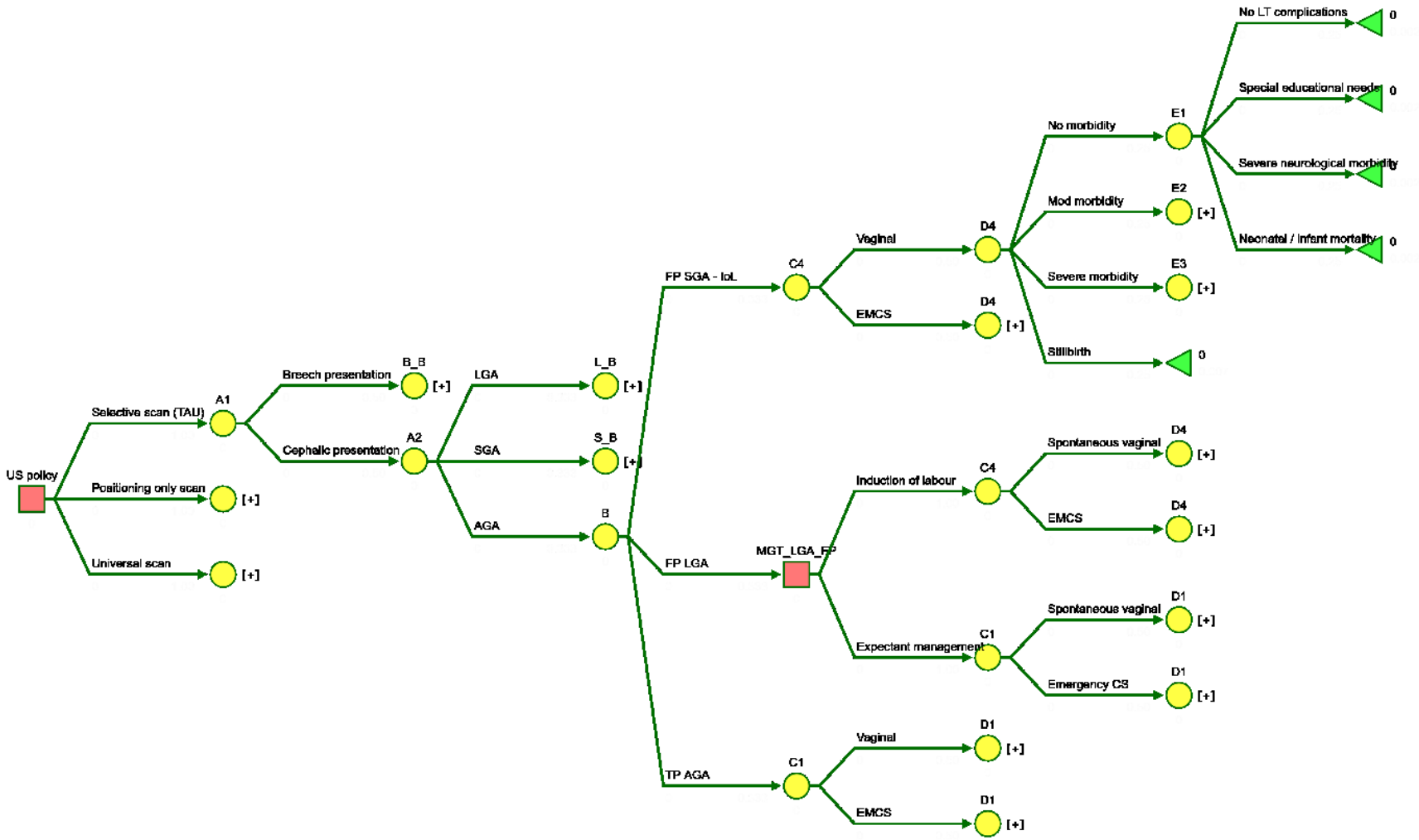
Outcomes relating to AGA

Figure 16 shows the decision tree with outcomes relevant to AGA expanded and remaining branches collapsed. An AGA pregnancy may either be correctly diagnosed, or incorrectly diagnosed as either SGA or LGA (node B). If correctly diagnosed, the mode of delivery can either be vaginal or emergency CS (node C1), after which short-and long-term outcomes will follow as described above (section 'Short- and long-term outcomes').

If an AGA pregnancy is falsely diagnosed as SGA, early IOL is offered. Unlike for true SGA, early labour induction of AGA pregnancies increases the risk of morbidity; however, the risk of perinatal death is still reduced.[162] Short- and long-term outcomes will then follow as described above (section 'Short- and long-term outcomes'). If instead, an AGA pregnancy is misdiagnosed as LGA, the short-and long-term outcomes depend on the management strategy. Compared to expectant management, early IOL decreases the risk of emergency CS and perinatal death, but increases the risk of neonatal morbidity.

Just as for other branches of the model, long-term outcomes include 'No long-term outcomes', 'SEN', 'SNM', and 'Neonatal mortality'. Each outcome is possible for all levels of neonatal morbidity, however, the risk of long-term complications increases for moderate and severe neonatal morbidity (nodes E2 and E3).

Figure 16. Outcomes associated with AGA.



[+] indicates collapsed sections of the decision tree

Data

We populated the model with data from multiple sources from the literature. Where possible, we prioritised inclusion of good quality systematic reviews and meta-analyses, followed by large, good quality clinical trials or cohort studies as appropriate. Where there was no objective evidence for a parameter 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. Data sources were subjectively graded as high, moderate or low, where high represented directly relevant data (i.e. providing the required parameter) from a good quality source (e.g. RCT for relative effects and high quality epidemiological study for baseline risks). A low grade represents situations where evidence on the required parameter was absent from the literature and so are sourced from a related parameter, used as indirect evidence and revised reflecting expert opinion as to the plausible values. Full details of the derivation of model inputs is in Appendix 7 (Tables 25-30), and all parameters are listed in Table 10, Table 11 and Table 12.

Table 10. Model inputs for diagnostic performance.

Parameter	Mean	95%CI	Distribution summary	Node	Source	Quality of evidence
Prevalence of breech	4.60%	3.98%, 5.30%	~B(179, 3700)	A1	Wastlund et al.[10]	H
Prevalence of LGA	10.00%	10%, 10%	N/A	A2	By definition	H
Prevalence of SGA	10.00%	10%, 10%	N/A	A2	By definition	H
Selective US						
Specificity SGA - Selective US	98.10%	97.63%, 98.52%	~B(3556, 69)	B	Sovio et al. (2015)[7]	H
Specificity LGA – Selective US	98.67%	98.28%, 99.02%	~B(3640, 49)	B	Sovio et al. (2018)[135]	H
Sensitivity SGA - Selective US	19.60%	15.63%, 23.90%	~B(69, 283)	S_B	Sovio et al. (2015)[7]	H
Sensitivity LGA - Selective US	26.55%	20.33%, 33.28%	~B(47, 130)	L_B	Sovio et al. (2018)[135]	H
Sensitivity breech – selective US	45.10%	37.85%, 52.54%	~B(79, 96)	B_B	Wastlund et al.[10]	H
Universal US for fetal size and presentation						
Specificity SGA - Universal US	89.99%	88.99%, 90.94%	~B(3262, 363)	B	Sovio et al. (2015)[7]	H
Specificity LGA – Universal US	96.56%	95.95%, 97.12%	~B(3562, 127)	B	Sovio et al. (2018)[135]	H
Sensitivity SGA - Universal US	56.53%	52.33%, 61.67%	~B(199, 153)	S_B	Sovio et al. (2015)[7]	H
Sensitivity LGA - Universal US	37.85%	30.87%, 45.10%	~B(67, 110)	L_B	Sovio et al. (2018)[135]	H
Sensitivity breech – Universal US	100%	100%, 100%	N/A	B_B	Assumption	N/A
Universal US for fetal presentation only						
Specificity SGA - Positioning scan	98.10%	97.63%, 98.52%	~B(3556, 69)	B	Sovio et al. (2015)[7]	H
Specificity LGA – Positioning scan	98.67%	98.28%, 99.02%	~B(3640, 49)	B	Sovio et al. (2018)[135]	H
Sensitivity SGA - Positioning scan	19.60%	15.63%, 23.90%	~B(69, 283)	S_B	Sovio et al. (2015)[7]	H
Sensitivity LGA - Positioning scan	26.55%	20.33%, 33.28%	~B(47, 130)	L_B	Sovio et al. (2018)[135]	H
Sensitivity breech – Positioning scan	100%	100%, 100%	N/A	B_B	Assumption	N/A

[a] *B = Beta distribution*

[b] *Quality assessment: High – good quality directly relevant evidence (e.g. directly relevant population, well conducted RCT for relative effects, or cohort for baseline effects). Med – directly relevant evidence but poorer quality source (e.g. retrospective cohort for relative treatment effect) Low – lack of direct evidence*

/ informed by expert opinion. Direct = source provides required parameter. Indirect = source provides related parameter used as background evidence to inform expert opinion. Note the same source may be used in different contexts, therefore resulting in a different relevance rating to inform different parameters. LGA = Large for gestational age; SGA = Small for gestational age; US = Ultrasound

Probabilities

Where possible probabilities were expressed as a baseline (beta or Dirichlet) for an otherwise healthy baby (i.e. neither breech nor LGA nor SGA), which were then modified by odds ratios or relative risks, depending on the statistic either reported in, or calculable from, the literature. Odds ratios were selected in preference to risk ratios, as these are independent of the baseline risk. Where no relative quantities were identified in the literature, probabilities are reported as independent beta distributions. Sampled values for probabilities were inspected to ensure they were bounded between [0,1]. Where out of range values were sampled, resampling was repeated until within-bounds values were generated.

Where relative effects were expressed as means and 95% CI, standard error of the log of the mean was estimated by dividing the absolute difference between the log mean and log lower or upper 95% confidence interval by 1.96.

Table 11. Model inputs for probabilities.

Parameter	Mean	95%CI	Distribution summary	Node	Source	Quality of evidence
Mode of delivery						
EmCS delivery | AGA and Exp Mgt	20.70%	19.4%, 22.06%	~B(735, 2813)	C1	Wastlund et al.[10]	H
RR EmCS delivery | SGA and Exp Mgt [FN] vs. C1	1.9	1.4, 2.5	~LN(0.642, 0.14)	S_C2	Monier et al.[21]	M
RR EMCS | induced, SGA [TP] vs. C1	2.9	1.8, 4.7	~LN(1.065, 0.246)	S_C3	Monier et al.[21]	L
RR EMCS | induced, AGA, [FP SGA] vs. C1	0.84	0.76, 0.93	~LN(-0.174, 0.052)	C4	Grobman et al.[148]	H
OR of EmCS delivery | LGA and Exp Mgt [FN] vs. C1	1.792	0.718, 4.471	~LN(0.583, 0.466)	L_C2	Blackwell et al.[141]	M
OR of EmCS delivery | LGA and Induce [TP] vs. L_C2	0.92	0.85, 0.99	~LN(-0.083, 0.037)	L_C3	Middleton et al.[15]	L
EmCS delivery | Breech and Exp Mgt [FN]	57.69%	38.67%, 75.62%	~B(15, 11)	B_C2	Leung et al.[163]	M
EmCS delivery | breech, ECV success, remain cephalic	27.27%	6.69%, 55.64%	~B(3, 8)	B_C3a	Wastlund et al.[10]	H
EmCS delivery | breech, ECV success, revert breech	57.69%	38.67%, 75.62%	~B(15, 11)	B_C3b	Leung et al.[163]	M
Vaginal delivery | breech, ECV fail, revert cephalic	52.38%	31.51%, 72.80%	~D(11, 1, 9)	B_C3c	Wastlund et al.[10]	H
ELCS delivery | breech, ECV fail, revert cephalic	4.76%	0.13%, 16.84%	-	B_C3c	Wastlund et al.[10]	
EmCS delivery | breech, ECV fail, revert cephalic	42.86%	23.07%, 63.97%	-	B_C3c	Wastlund et al.[10]	
Vaginal delivery | breech, ECV fail, remain breech	0%	0%, 0%	~D(0, 54, 18)	B_C3d	Wastlund et al.[10]	H
ELCS delivery | breech, ECV fail, remain breech	75%	64.47%, 84.22%	-	B_C3d	Wastlund et al.[10]	
EmCS delivery | breech, ECV fail, remain breech	25%	15.78%, 35.53%	-	B_C3d	Wastlund et al.[10]	
Vaginal delivery | breech, no ECV, revert cephalic	52.38%	31.51%, 72.80%	~D(11, 1, 9)	B_C3e	Wastlund et al.[10]	H
ELCS delivery | breech, no ECV, revert cephalic	4.76%	0.13%, 16.84%	-	B_C3e	Wastlund et al.[10]	
EmCS delivery | breech, no ECV, revert cephalic	42.86%	23.07%, 63.97%	-	B_C3e	Wastlund et al.[10]	
Vaginal delivery | breech, no ECV, remain breech	0%	0%, 0%	~D(0, 52, 20)	B_C3f	Wastlund et al.[10]	H
ELCS delivery | breech, no ECV, remain breech	72.22%	61.38%, 81.88%	-	B_C3f	Wastlund et al.[10]	
EmCS delivery | breech, no ECV, remain breech	27.77%	18.12%, 38.62%	-	B_C3f	Wastlund et al.[10]	

External cephalic version						
ECV attempted	47.46%	40.16%, 54.81%	~B(84, 93)	B_ECV	Wastlund et al.[10]	H
ECV not attempted, spontaneous reversion to cephalic	22.58%	14.72%, 31.56%	~B(21, 72)	B_noECV_rc	Wastlund et al.[10]	H
Probability ECV successful	14.29%	7.70%, 22.48%	~B(12, 72)	B_ECVs	Wastlund et al.[10]	H
Probability of reverting to breech post successful ECV	8.33%	0.23%, 28.49%	~B(1, 11)	B_ECVs_rb	Wastlund et al.[10]	H
Probability of spontaneous revesion to cephalic post ECV failure	2.31%	0.48%, 5.49%	~B(3, 127)	B_ECVf_rc	Ben-Meir et al.[164]	H
Outcomes for LGA model						
Respiratory morbidity, baseline	0.32%	0.20%, 0.46%	~B(22, 6933)	-	Morrison et al.[165]	H
Shoulder dystocia, baseline	0.63%	0.60%, 0.66%	~B(1686, 265542)	-	Ouzounian et al.[166]	M
Other acidosis, baseline	0.68%	0.22%, 1.40%	~B(5, 726)	-	Middleton et al.[15]	H
Perinatal mortality, baseline	0.155%	0.145%, 0.165%	~B(984, 634412)	-	Moraitis et al.[51]	M
RR respiratory morbidity, LGA vs. AGA [FN & ExpMan LGA policy]	0.75	0.5125, 0.9875	~U(0.5, 1)	L_D2a	Expert opinion	L
OR shoulder dystocia, LGA vs. AGA [FN & ExpMan LGA policy]	7.18	2.06, 25.00	~LN(1.971, 0.637)	L_D2a	Rossi et al.[167]	H
OR other acidosis, LGA vs. AGA [FN & ExpMan LGA policy]	2.88	1.34, 6.22	~LN(1.058, 0.393)	L_D2a	Rossi et al.[167]	M
OR perinatal mortality, LGA vs. AGA [FN & ExpMan LGA policy]	1.77	0.30, 10.34	~LN(0.571, 0.901)	L_D2a	Rossi et al.[167]	M
OR respiratory morbidity, LGA vs. AGA, EMCS [FN & ExpMan LGA policy]	5.33	3.50, 7.40	~LN(1.674, 0.167)	L_D2c	Morrison et al.[165]	H
P shoulder dystocia, LGA, EMCS [FN & ExpMan LGA policy]	0	0, 0	N/A	L_D2c	Assumption	H
OR other acidosis, LGA, EMCS [FN & ExpMan LGA policy]	1.867	1.217, 2.865	~LN(0.625, 0.218)	L_D2c	Chongsuvivatwong et al.[168]	M
OR perinatal mortality, LGA, EMCS [FN & ExpMan LGA policy]	1.781	1.266, 2.505	~LN(0.577, 0.174)	L_D2c	Chongsuvivatwong et al.[168]	M
OR respiratory morbidity, LGA, Induction of labour, vaginal delivery [TP]	0.54	0.373, 0.783	~LN(-0.616, 0.19)	L_D3a	Gibson et al.[169]	M
RR shoulder dystocia, LGA, Induction of labour, vaginal delivery [TP]	0.6	0.37, 0.98	~LN(-0.511, 0.25)	L_D3a	Boulvain et al.[98]	M
RR acidosis, LGA, Induction of labour, vaginal delivery [TP]	1.66	0.61, 4.55	~LN(0.507, 0.514)	L_D3a	Middleton et al.[15]	M
RR perinatal mortality, LGA, Induction of labour, vaginal delivery [TP]	0.33	0.14, 0.78	~LN(-1.109, 0.439)	L_D3a	Middleton et al.[15]	M
OR respiratory morbidity, LGA, Induction of labour, EmCS [TP]	0.54	0.373, 0.783	~LN(-0.616, 0.19)	L_D3c	Gibson et al.[169]	M
P shoulder dystocia, LGA, Induction of labour, EmCS [TP]	0	0, 0	N/A	L_D3c	Assumption	H
RR acidosis, LGA, Induction of labour, EmCS [TP]	1.66	0.61, 4.55	~LN(0.507, 0.514)	L_D3c	Middleton et al.[15]	M
RR perinatal mortality, LGA, Induction of labour, EmCS [TP]	0.33	0.14, 0.78	~LN(-1.109, 0.439)	L_D3c	Middleton et al.[15]	M
Risk of acidosis | shoulder dystocia	0.07	0.0630, 0.1112	~B(36, 478)	L_E1	MacKenzie et al.[170]	L

Risk of BPI | shoulder dystocia	0.0856	0.0496, 0.0936	~B(44, 470)	L_E1	MacKenzie et al.[170] [c]	L
Risk of permanent BPI	0.055	0.024, 0.098	~B(8, 137)	L_F1	Sandmire et al.[171] [c]	M
Neonatal morbidity						
Risk of moderate neonatal morbidity (AGA) [FP]	5.62%	0.0488, 0.0641	~B(198, 3325)	D1	The POP study [c]	H
Risk of severe neonatal morbidity (AGA) [FP]	0.62%	0.0039, 0.0091	~B(22, 3501)	D1	The POP study [c]	H
Risk of perinatal death (AGA) [FP]	0.155%	0.145%, 0.165%	~B(984, 634412)	D1	Moraitis et al.[51]	M
OR moderate neonatal morbidity (SGA vs. AGA, ExpMan)	2.48	1.75, 3.51	~LN(0.91, 0.18)	S_D2	The POP Study [c]	H
OR severe neonatal morbidity (SGA vs. AGA, ExpMan)	1.88	0.65, 5.50	~LN(0.63, 0.55)	S_D2	The POP Study [c]	H
OR perinatal death (SGA vs. AGA, ExpMan)	4.39	3.84, 5.03	~LN(1.48, 0.07)	S_D2	Moraitis et al.[51]	H
RR moderate morbidity | induce SGA vs. not inducing SGA [TP]	0.7	0.50, 0.98	~LN(-0.357, 0.172)	S_D3	Middleton et al.[15]	L
RR severe morbidity | induce SGA vs. not inducing SGA [TP]	0.7	0.50, 0.98	~LN(-0.357, 0.172)	S_D3	Middleton et al.[15]	L
RR perinatal death | induce SGA vs. not inducing SGA [TP]	0.33	0.11, 0.96	~LN(-1.109, 0.553)	S_D3	Middleton et al.[15]	L
OR of moderate neonatal morbidity if induce | AGA [FP SGA or LGA]	1.92	1.71, 2.15	~LN(0.652, 0.058)	D4	Stock et al.[162]	H
OR of severe neonatal morbidity if induce | AGA [FP SGA or LGA]	1.92	1.71, 2.15	~LN(0.652, 0.058)	D4	Stock et al.[162]	H
OR of perinatal death if induce | AGA [FP SGA or LGA]	0.15	0.03, 0.68	~LN(-1.897, 0.771)	D4	Stock et al.[162]	H
OR of moderate neonatal morbidity | vaginal breech vs. vaginal cephalic delivery	6.70	5.9, 7.6	~LN(1.902, 0.064)	B_D2a	Thorngren-Jerneck et al.[172]	H
OR of severe neonatal morbidity | vaginal breech vs. vaginal cephalic delivery	6.70	5.9, 7.6	~LN(1.902, 0.064)	B_D2a	Thorngren-Jerneck et al.[172]	H
OR of perinatal death | vaginal breech vs. vaginal cephalic delivery	6.68	2.75, 16.22	~LN(1.899, 0.453)	B_D2a	Moraitis et al.[51]	H
RR of moderate morbidity | ELCS vs. vaginal breech delivery	0.43	0.12, 1.47	~LN(-0.844, 0.627)	B_D2b	Hofmeyr et al.[13]	H
RR of severe morbidity | ELCS vs. vaginal breech delivery	0.11	0.01, 0.87	~LN(-2.207, 1.055)	B_D2b	Hofmeyr et al.[13]	H
RR of perinatal death | ELCS vs. vaginal breech delivery	0.29	0.1, 0.86	~LN(-1.238, 0.555)	B_D2b	Hofmeyr et al.[13]	H
OR of moderate morbidity | EmCS vs. vaginal breech delivery	0.533	0.192, 1.482	~LN(-0.629, 0.522)	B_D2c	Pasupathy et al.[173] [c]	M
OR of severe morbidity | EmCS vs. vaginal breech delivery	0.533	0.192, 1.482	~LN(-0.629, 0.522)	B_D2c	Pasupathy et al.[173] [c]	M
OR of perinatal death | EmCS vs. vaginal breech delivery	0.533	0.192, 1.482	~LN(-0.629, 0.522)	B_D2c	Pasupathy et al.[173] [c]	M
Risk of long-term outcomes from neonatal morbidity						
Risk of SEN | no neonatal morbidity	0.0474	0.0467, 0.0480	~B(18736, 376891)	E1	MacKay et al.[174]	H
Risk of neurological morbidity | no neonatal morbidity	0.0008	0.0007, 0.0008	~B(906, 1193647)	E1	Persson et al.[175]	H

Risk of neonatal/infant mortality | no neonatal morbidity	0.002	0.0020, 0.0021	~B(2074, 1011289)	E1	Iliodromiti et al.[176]	H
OR of SEN | moderate neonatal morbidity	1.55	1.43, 1.67	~LN(0.438, 0.038)	E2	MacKay et al.[174]	H
RR of neurological morbidity | moderate neonatal morbidity	10.4	7.8, 13.9	~LN(2.34, 0.149)	E2	Persson et al.[175]	H
RR of neonatal/infant mortality | moderate morbidity	12.82	9.33, 17.61	~LN(2.551, 0.162)	E2	Iliodromiti et al.[176]	H
OR of SEN | severe neonatal morbidity	1.66	1.46, 1.88	~LN(0.507, 0.063)	E3	MacKay et al.[174]	H
RR of neurological morbidity | severe morbidity	145.5	104.0, 204.1	~LN(4.98, 0.173)	E3	Persson et al.[175]	H
RR of neonatal/infant mortality | severe morbidity	60.61	48.17, 76.26	~LN(4.104, 0.117)	E3	Iliodromiti et al.[176]	H

[a] *Distributions: B = Beta, D = Dirichlet; LN = Log-normal, U = Uniform*

[b] *Quality assessment: High – good quality directly relevant evidence (e.g. directly relevant population, well conducted RCT for relative effects, or cohort for baseline effects). Med – directly relevant evidence but poorer quality source (e.g. retrospective cohort for relative treatment effect) Low – lack of direct evidence / informed by expert opinion. Direct = source provides required parameter. Indirect = source provides related parameter used as background evidence to inform expert opinion. Note the same source may be used in different contexts, therefore resulting in a different relevance rating to inform different parameters.*

[c] *Parameter estimated based upon data from the source, rather than directly from the source. Details are provided in Appendix 7 (Tables 25-30). BPI = Brachial plexus injury, ECV = External cephalic version, ELCS = Elective Caesarean section, EmCS = Emergency Caesarean section, ExpMan = Expectant management, FN = False negative, FP = False positive, LGA = Large for gestational age, OR = odds ratio, RR = relative risk, SEN = Special educational needs, SGA = Small for gestational age, TP = True positive.*

Costs

The price year of the analysis is 2016/17. The majority of costs were sourced from the English national schedule of reference costs[177]. The national schedule of reference costs report different costs depending on how the service was delivered (e.g. elective inpatient, non-elective inpatient, outpatient procedures etc.). We used costs from total HRG's (i.e. weighted by each category by the number of yearly activities), except for cases where only one or a few categories made logical sense. All categories in the schedule reports costs as mean and inter-quartile range. To obtain parameter estimates of costs, we fitted a gamma distribution using these data points. Where multiple cost categories were used, we first calculated a weighted average of the mean and inter-quartile range by the number of yearly activities in each category before fitting the gamma distribution.

Where no directly applicable cost could be identified in the reference schedule, we first attempted to obtain resource usage from literature, and assign costs to these using the reference costs. Where insufficient data on resource usage were available, we adopted the costs directly from literature. Costs reported in currencies other than GBP or 2016/17 costs were converted to GBP at the exchange rate of the year that the source was published and inflated to 2016/17 prices using the hospital & community health services (HCHS) index.[178] Where no credible estimates could be identified from literature, we estimated the costs ourselves assigning a wide credibility interval to represent the uncertainty. Full details on the derivation of all cost parameters is presented in Appendix 7.

Table 12. Model inputs for costs and related probabilities.

Parameter	Mean	95%CI	Distribution summary	Node	Source	Quality of evidence
Ultrasound scan	£107.06	£70.98, 134.92	~G(4.9604, 22.8062)	A	NHS reference costs 2016-17[177 c]	H
Positioning scan only	£48.71	£8.96, 88.46	~U(6.87, 90.55)	A	Expert opinion	N/A
Proportion scanned with US (selective screening)	0.3499	0.3349, 0.3650	~B(1351, 2510)	A	Sovio et al.[7]	H
Induction of labour (difference vs. normal delivery)	£125	-£1343, 1594	~N(125.3, 749.2)	B1, B2	Vijgen et al.[179]	M
Cost of vaginal (cephalic) delivery	£1,834	£1750, 2236	~G(7.2606, 252.5824)	C1 – C4	NHS reference costs 2016-17[177 c]	H
Relative cost difference (vaginal breech vs. cephalic delivery)	1.1633	1.0982, 1.2284	~N(1.1633, 0.0332)	B_C3b, B_C3d, B_C3f, B_C2,	Palencia et al.[180]	M
Cost of ECV	£292.30	£287.5, 297.1	~U(287.22, 297.38)	B_ECV	James et al.[181 c]	M
Cost of emergency Caesarean section	£4,688	£3816, 5443	~G(14.7329, 318.1354)	C1 – C4	NHS reference costs 2016-17[177 c]	H
Cost of elective Caesarean section	£3,412	£2680, 4038	~G(11.1212, 307.0169)	C1 - C4	NHS reference costs 2016-17[177 c]	H
Cost of Special Care Baby Unit admission	£1,064	£487, 1862	~G(9.0371, 117.7307)	D1 - D4	NHS reference costs 2016-17[177 c]	H
Cost of Neonatal High Dependency Unit admission	£1,346	£807, 2020	~G(18.7696, 71.7047)	D1 - D4	NHS reference costs 2016-17[177 c]	H
Cost of Neonatal Intensive Care Unit admission	£2,590	£1280, 4352	~G(10.7403, 241.0768)	D1 - D4	NHS reference costs 2016-17[177 c]	H
Proportion of neonates admitted to SCBU	74%	65%, 82%	~D(74, 7, 19)	D1 - D4	Alfirevic et al.[182]	M
Proportion of neonates admitted to NHDU	7%	3%, 13%	-	D1 - D4	Alfirevic et al.[182]	
Proportion of neonates admitted to NICU	19%	12%, 27%	-	D1 - D4	Alfirevic et al.[182]	
Probability of admission to care | no neonatal morbidity	0.074	0.066, 0.082	~B(292, 3659)	D1 - D4	Sovio et al.[7]	H
Odds ratio of admission to care | Moderate neonatal morbidity	11.29	5.90, 21.60	~LN(2.424, 0.331)	D1 - D4	Sovio et al.[7]	H
Probability of admission to care | Severe neonatal morbidity	1	1, 1	N/A	D1 - D4	Assumption	N/A
Short-term cost of acidosis / anoxia	£3,240	£806, 7328	~G(3.6143, 895.6169)	L_E1, L_D2a	Own estimation [c]	L
Short-term cost of respiratory morbidity	£2,011	£993, 3381	~G(10.7125, 187.6316)	L_D2a, L_D3a	Own estimation [c]	L
Cost of transient BPI	£2,066	£1033, 4132	~LN(7.6334, 0.3536)	L_F1	Culligan et al.[183]	M

Cost of permanent BPI	£14,134	£7068, 28264	~LN(9.5563, 0.03536)	L_F1	Culligan et a.[183] [c]	M
Cost of perinatal or infant mortality	£1,664	£1372, 1956	~U(1357, 1971)	D1 & E1 – 3	Mistry et al.[184]	M
Special educational needs (per annum)	£7,428	£4467, 10389	~N(7428.1, 1511)	E1 – E3	Barrett et al.[185]	M
Severe neurological morbidity (per annum)	£2,930	£1465, 5859	~LN(7.9826, 0.3536)	E1 – E3	Access economics[186] [c]	M

[a] *Distributions: B = Beta, D = Dirichlet; G = Gamma, LN = Log-normal, N = Normal, U = Uniform*

[b] *Quality assessment: High – good quality directly relevant evidence (e.g. directly relevant population, well conducted RCT for relative effects, or cohort for baseline effects). Med – directly relevant evidence but poorer quality source (e.g. retrospective cohort for relative treatment effect) Low – lack of direct evidence or informed by expert opinion. Direct = source provides required parameter. Indirect = source provides related parameter used as background evidence to inform expert opinion. Note the same source may be used in different contexts, therefore resulting in a different relevance rating to inform different parameters.*

[c] *Parameter estimated based upon data from the source, rather than directly from the source. Details are provided in 7.*

BPI = Brachial plexus injury, ECV = External cephalic version, NHDU = Neonatal high-dependency unit, NICU = Neonatal intensive care unit, SCBU = Special care baby unit, US = Ultrasound

All costs presented in pound sterling (£) and updated to the cost-year of 2016-17 using the HCHS Index[178].

Quality of life

We estimated age-specific quality of life for healthy neonates using EuroQol data for a general UK population.[187] Age specific health state utilities were multiplied by age specific survival[188], the discounted sum over the time horizon of the model yielding the expected QALYs gained for an otherwise healthy neonate. Per definition, the quality of life following mortality is zero, and we made the simplifying assumption that all deaths during a particular year of life occurred on the first day of the year. In the absence on suitable evidence of how SEN affect quality of life, we assumed for our base-case scenario that SEN would affect costs only. In the case of severe neurological morbidity, we adjusted the baseline quality of life with a relative decrease following the methodology of Leigh et al.[189], using CP as a proxy for severe neurological morbidity. Full details on the derivation of quality of life parameters is presented in Appendix 7.

Analysis

The model was analysed via Monte Carlo simulation, capturing the overall uncertainty in cost-effectiveness as a function of the uncertainty of the input parameters. Health outcomes were from the fetal perspective only and ultimately presented as QALYs. Cost-effectiveness was explored through ICERs and net monetary benefits (NMB), using a willingness-to-pay (WTP) threshold of £20,000 per QALY. All costs and QALYs were discounted by 3.5% per annum.[190] All costs were from a third-party (payer) perspective, i.e. the English NHS plus special educational needs costs, and the reference case time horizon was 20 years (varied in sensitivity analysis).

Stability testing was conducted to quantify (and thence minimise) Monte Carlo error as a function of the number of simulations. The model was run 30 times with a given number of simulations. The coefficient of variation of the estimates of mean and standard error of mean cost and QALYs for each comparator were calculated. The mean of all of these was used as a summary measure of the Monte Carlo error. We used an arbitrary 2% cut-off to declare the results stable.

Cost-effectiveness: reference case

For each of the six discrete strategies, we present mean and 95% credibility intervals for cost and QALYs gained, net benefit at a willingness to pay of £20,000 per QALY, and INMB relative to the assumed status quo (selective scanning with IOL for macrosomia or SGA, offer of ECV for breech). The option with the highest expected net benefit was identified as the most cost-effective. Decision uncertainty was expressed as the probability of each decision being cost-effective at the reference

case threshold (£20,000/QALY). The cost-effectiveness acceptability curve plots decision uncertainty as a function of willingness to pay per QALY.

Cost-effectiveness: sensitivity and scenario analyses

In addition to the primary analysis, we report a number of scenario analyses and one-way sensitivity analyses to explore specific uncertainties in more detail. Specifically:

- Time horizon
    - The base case analysis assumes a 20 year time horizon. We vary this from 1 to 100 years.
- Cost of scan to assess fetal presentation only
    - The cost of a presentation-only scan is dependent on whether it is feasible to incorporate it into a routine antenatal visit with the scan conducted by a midwife using a hand-held unit, or whether it can only be done at a dedicated visit by an ultrasonographer in a secondary care setting.
- The baseline risks of perinatal death, moderate and severe neonatal morbidity.
    - Baseline risks of each of these were estimated from different sources, yet they are mutually exclusive events. Ideally, these should be modelled as a Dirichlet distribution, but as the data were from different sources we modelled them as independent betas. We thus explore these further in one-way sensitivity analysis.

In addition, due to concerns over the validity of input data, we also explore the difference in risk of acidosis and respiratory morbidity associated with vaginal delivery of an LGA infant (versus AGA), the odds ratio of perinatal death from an emergency caesarean section from a breech baby (versus vaginal delivery), the relative risk of an emergency CS from IOL for an SGA infant (versus expectant management of an AGA infant), the relative risk of SEN as a result of inducing labour (versus expectant management), and its impact on health related quality of life, and the sensitivity of ultrasound scanning at detecting SGA.

Value of information analysis

Uncertainty in cost-effectiveness results (i.e. decision uncertainty) was used to conduct a value of information analysis. Decision uncertainty arises from parameter uncertainty. The expected value of perfect information (EVPI) is the expected value of eliminating all decision uncertainty, which by definition implies eliminating all parameter uncertainty. This therefore provides an upper bound for the value of all research into the decision question. The expected value of perfect parameter information (EVPPI) is the expected value of eliminating uncertainty in a single parameter or group of

parameters. The expected value of sample information (EVSI) is the expected value of a study of sample size n. The EVSI of a study of size n less the cost of conducting it provides a measure of the expected return on investment in that research project (expected net gain of sampling, ENGS).[191-193] An EVPPI above the plausible cost of a research project is a necessary condition for future research to be economically viable. A positive ENGS is the sufficient condition. The efficient sample size of a study is that which maximises the ENGS.

We estimated that there are approximately 196,297 singleton births at ≥37 weeks' gestation to nulliparous women not delivered by elective CS each year. Assuming a time horizon for which the decision question remains valid of 10 years yields a (discounted) beneficial population of 1,689,663. If it is reasonable to assume our analyses are generalisable to all births in England, the beneficial population is 5,477,940.

We report the per-patient (i.e. per mother/infant dyad) and population expected value of perfect information at a willingness to pay of £20,000/QALY. We then report the per patient and population EVPPI for each parameter individually, calculated using the Sheffield Accelerated Value of Information (SAVI) tool.[161] Parameters with a positive EVPPI were grouped into those which could logically be collected within one research study and the EVPPI for that group of parameters calculated (also with the SAVI tool[161]). The expected value of sample information (EVSI) for any parameters or groups of parameters is then calculated using the method of Heath et al.[194] Population values are presented as a 'conservative' estimate, assuming the information is only of value to singleton nulliparous pregnancies (i.e. using the 1,689,663 beneficial population), and a broader estimate which assumes the information is of value to all pregnancies in England (5,477,940 population).

Results

Stability testing

Our analyses showed that we were able to achieve extremely stable results (coefficient of variation of <0.01%) with 100,000 simulations, at a 'reasonable' run time of around 30 seconds (Table 13). We therefore ran our cost-effectiveness analyses with 100,000 simulations. However, due to the need for repeated loops, the EVSI calculations are based on 10,000 simulations.

Table 13. Results from stability testing.

Simulations	Computation time (seconds)	Mean Coefficient of Variation (%)
10	0.10	24.68
100	0.09	7.73
1000	0.33	2.53
10000	2.75	0.56
100000	29.56	<0.01

Cost-effectiveness results

Table 14 shows the overall costs, QALYs, net benefit and incremental net benefit for each of the six screening-management strategies. Net benefit is calculated assuming a willingness to pay of £20,000 per QALY gained. Incremental net benefit is shown relative to the status quo (assumed selective ultrasound scanning and IOL for both suspected SGA and LGA"). Strategies are ordered in terms of increasing cost.

Given current evidence and assuming a willingness to pay of £20,000 per QALY, the strategy associated with the highest net benefit is a presentation only scan for all women (where women with relevant indications also get a full scan"). Where LGA is suspected, the recommended management is IOL: on average IOL is associated with a small improvement in QALYs compared with expectant management (SGA is assumed managed with IOL). Universal ultrasound screening for fetal size is not recommended as its added benefits do not justify its added cost. Decision uncertainty suggests there is a 44.19% probability that this is the most cost-effective strategy (Table 14 and Figure 17).

Table 14. Cost effectiveness results (per mother scanned, mean and 95% Credibility Interval).

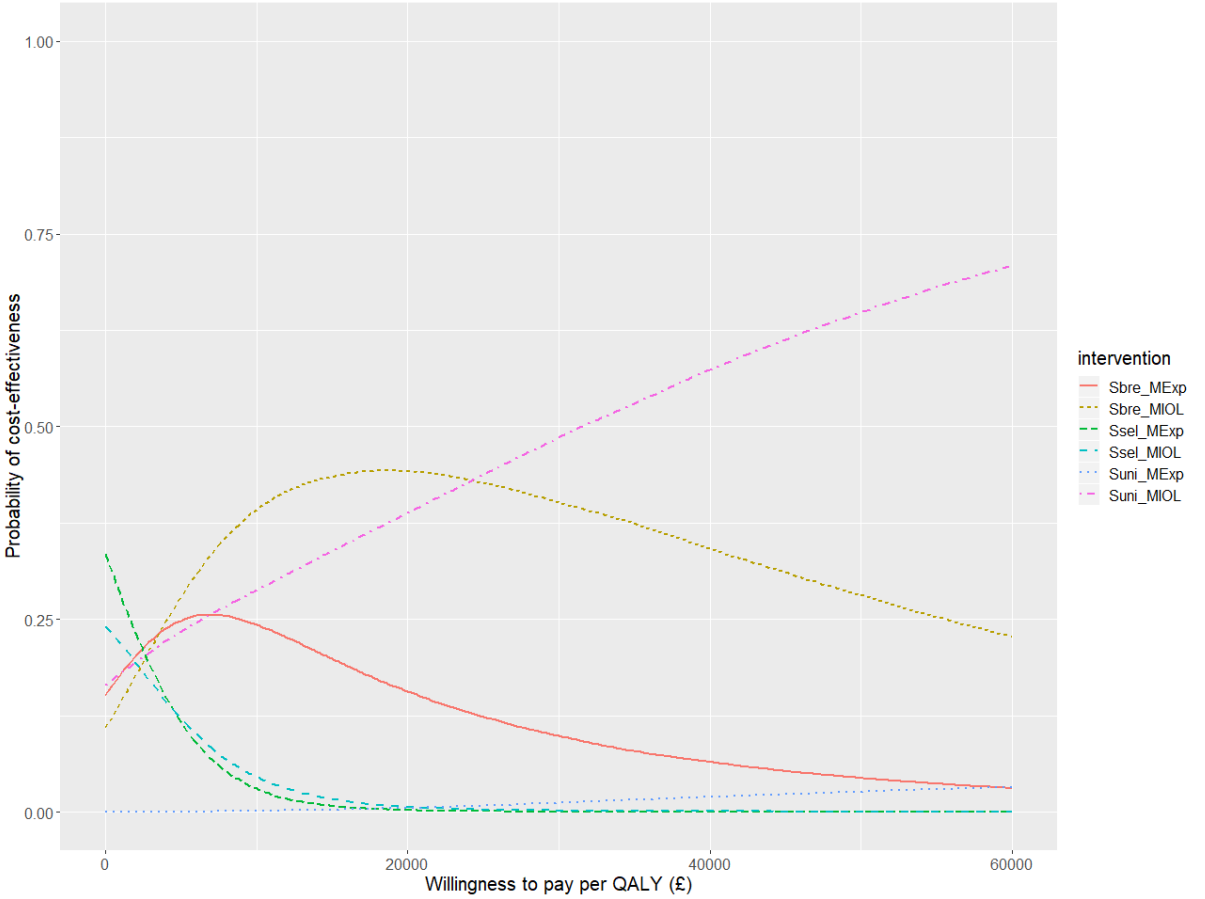
Screening + management	Cost	QALYs	NB|£20k	INB|£20k	P_CE|£20k
Selective US + Induction	6090	13.640	266719	0	0.65%
	(4420, 7890)	(13.441, 13.841)	(262333, 271079)	(0, 0)	
Selective US + Expectant	6091	13.639	266682	-37.09	0.22%
	(4424, 7889)	(13.439, 13.839)	(262297, 271040)	(-124.7, 35.24)	
Universal US for presentation + induction *	6101.	13.645	266806	87.36	44.19%
	(4443, 7887)	(13.446, 13.846)	(262426, 271154)	(4.88, 205.68)	
Universal US for size + Expectant	6102	13.644	266769	50.29	15.63%
	(4446, 7887)	(13.444, 13.844)	(262389, 271120)	(-68.06, 186.43)	
Universal US for size + Expectant	6178	13.646	266734	14.47	0.51%
	(4508, 7972)	(13.446, 13.846)	(262351, 271099)	(-133.98, 173.31)	
Universal US + Induction	6180	13.648	266779	60.24	38.81%
	(4498, 7983)	(13.448, 13.849)	(262386, 271147)	(-151.43, 281.7)	

Management refer to management strategy when LGA is suspected, all cases of suspected SGA are assumed induced.

** Strategy with highest expected net benefit.*

INB = Incremental net benefit relative to current practice (selective US + induction of labour); NB = Net benefit ; P_CE = Probability of being the most cost-effective strategy

Figure 17. Cost-effectiveness acceptability curve for the chance of each strategy being the most cost-effective as a function of willingness-to-pay for an additional quality-adjusted life year.



MExp = Expectant management, MIOL = Induction of labour, Sbre = Universal ultrasound for fetal presentation only, Ssel = Selective ultrasound, Suni = Universal ultrasound for fetal biometry plus presentation.

One-way and scenario analyses

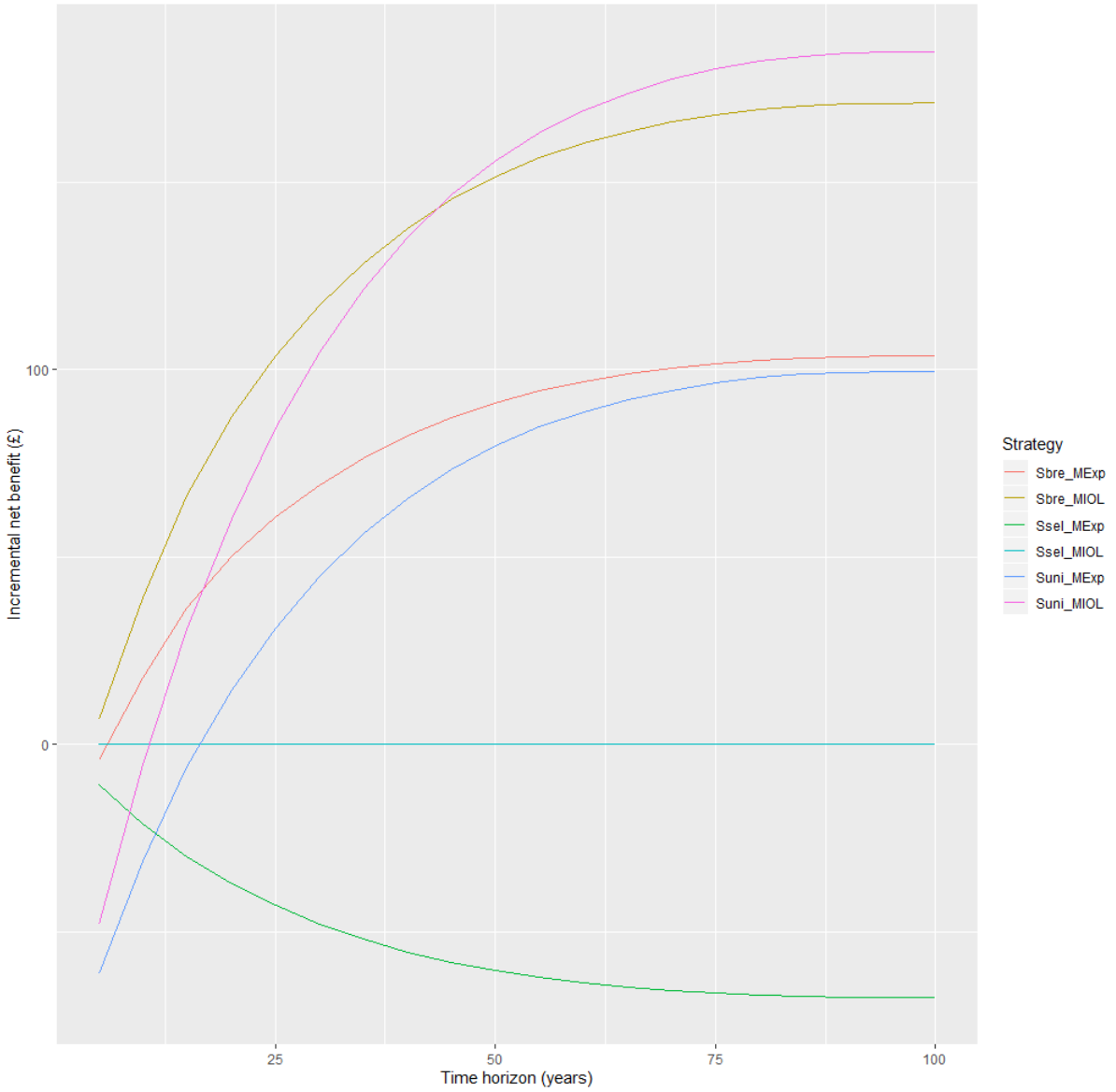
Cost-effectiveness conclusions were sensitive only to the time horizon, the cost of an ultrasound scan for fetal presentation only, the background risk of stillbirth, moderate and severe perinatal complications, and the risk of special educational needs associated with IOL.

With respect to the time horizon, universal ultrasound for fetal presentation only is the most cost-effective option as long as the time horizon of the analysis is below 45 years (Figure 18 below). Above this time horizon, universal ultrasound for size and presentation becomes the most cost-effective option. With respect to the cost of a presentation scan, a presentation-only scan remains the most cost-effective option provided it costs no more than £90. Above this cost, status quo is the most cost-effective (Figure 19 below).

As the background risks of perinatal mortality, moderate and severe perinatal complications rise, the net benefit of a detailed universal scan rises (Figure 20 below). This is because the risks of complications from SGA and LGA infants are modelled relative to the baseline risks: as the baseline risk rises, the risks for SGA and LGA infants rises more than proportionately, thus the benefit from detecting and intervening rises. A breech-only scan remains the most cost-effective option so long as the baseline risk of perinatal death remains below 0.28%, and for moderate and severe complications below 4.8% and 1.12% respectively. Above these values, universal screening becomes the cost-effective option.

Our base case analysis assumed a linear progression through the model whereby long term outcomes were dependent on perinatal outcomes, which were dependent on mode of delivery alone (vaginal versus (emergency or elective) CS). However, there is evidence to suggest that IOL may of itself increase the risk of special educational needs in later life.[174] We therefore explored the impact on the results via a one-way sensitivity analysis. We find that our results remain the same as long as the relative risk of SEN as a result of IOL is between approximately 0.95 and 1.3; and the estimated risk at 38 wkGA lies within this range[174]. Below this risk, the most cost-effective strategy is to perform universal screening for both presentation and estimated fetal weight, and to induce labour where SGA or LGA is suspected. Above this risk, then whilst the recommended scan remains a presentation-only scan, the most cost-effective intervention for suspected SGA or LGA is expectant management (i.e. IOL ceases to be the appropriate intervention, Figure 21). Given this, whilst not captured in our formal value of information analysis (due to structural assumptions), it may be worthwhile exploring the impact of inducing labour on long term risk of special educational needs in future research.

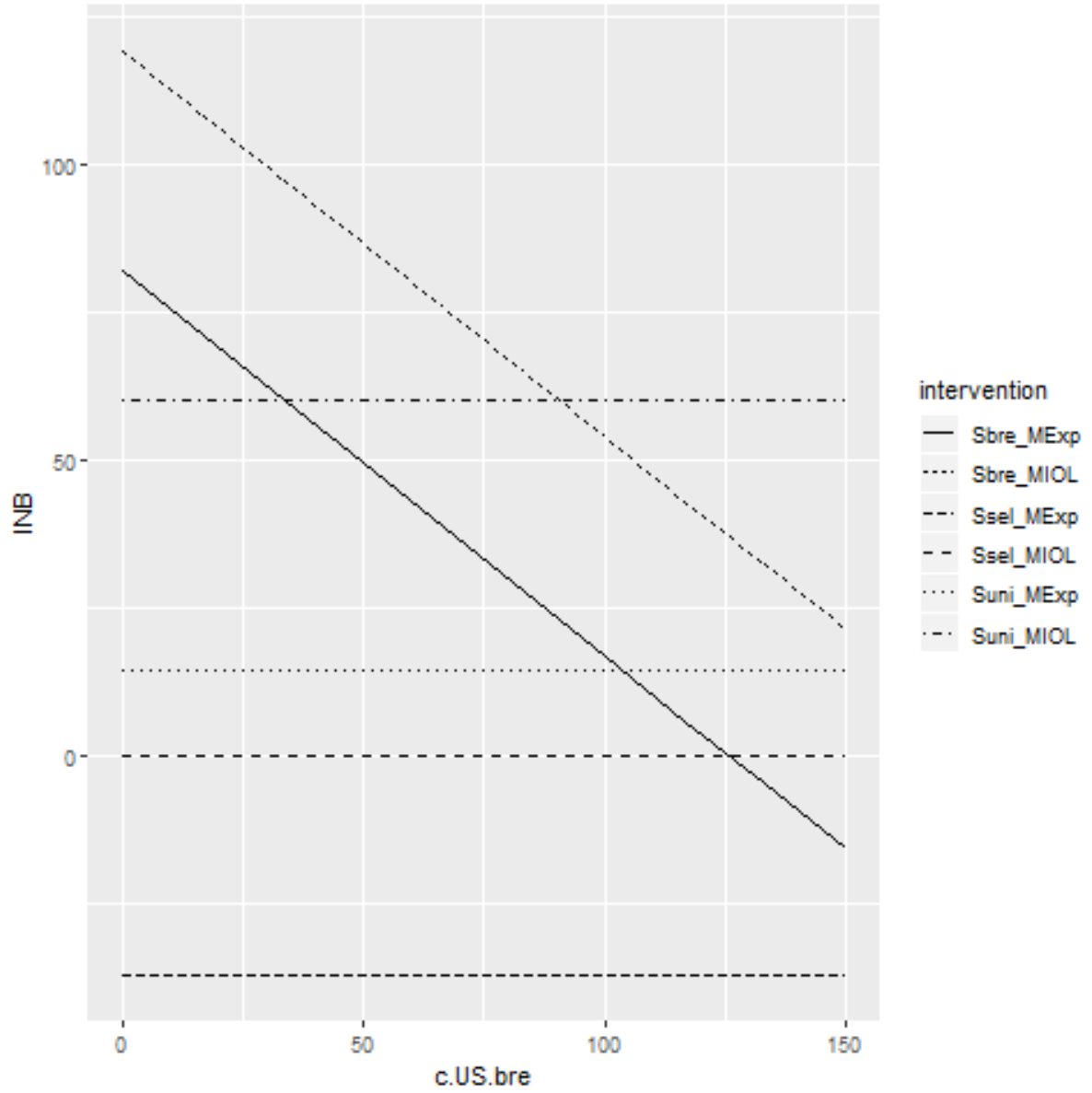
Figure 18. One-way sensitivity analysis on model time horizon.



MExp = Expectant management, MIOL = Induction of labour, Sbre = Universal ultrasound for fetal presentation only, Ssel = Selective ultrasound, Suni = Universal ultrasound for fetal biometry plus presentation.

The figure shows the expected incremental net monetary benefit for different strategies compared to current practice (selective ultrasound with induction of labour for suspected LGA) as a function of the model's time horizon (years). Calculations are based upon a willingness-to-pay (i.e. valuation of one additional quality-adjusted life year) of £20,000.

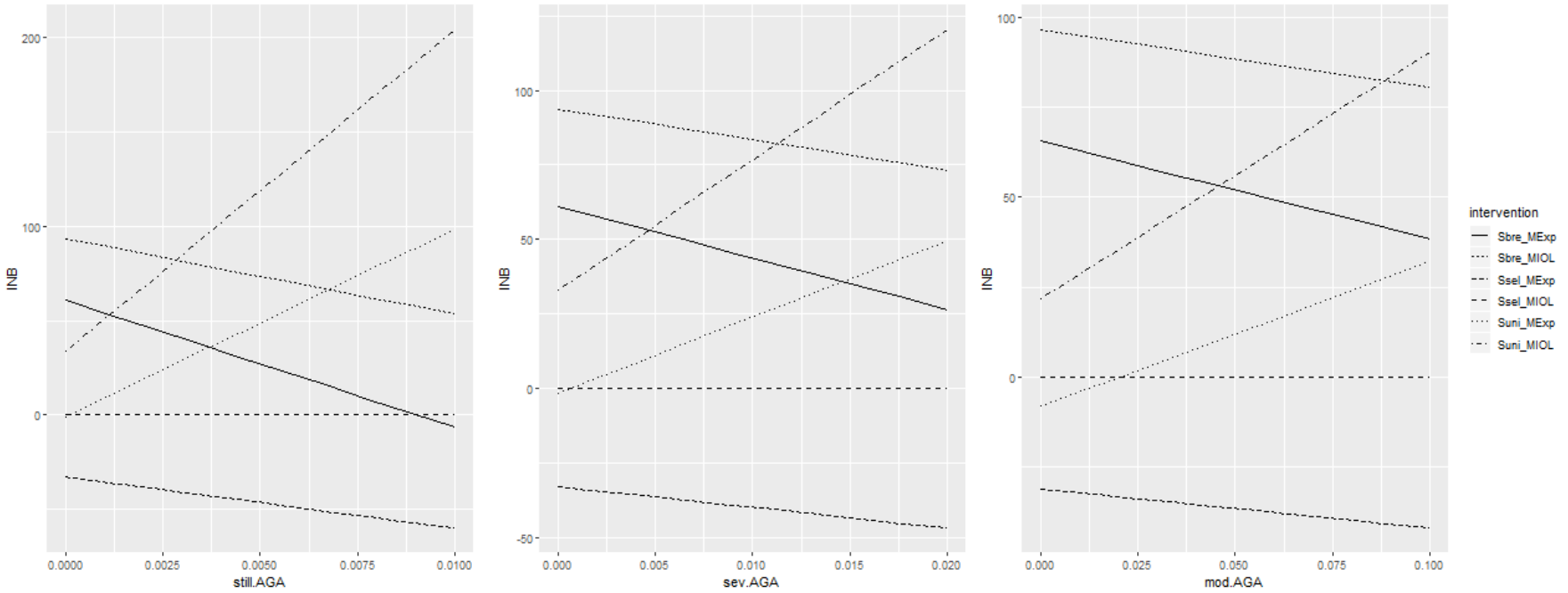
Figure 19. One-way sensitivity analysis on the cost of a scan for fetal presentation only.



MExp = Expectant management, MIOL = Induction of labour, Sbre = Universal ultrasound for fetal presentation only, Ssel = Selective ultrasound, Suni = Universal ultrasound for fetal biometry plus presentation.

The figure shows the expected incremental net monetary benefit for different strategies compared to current practice (selective ultrasound with induction of labour for suspected LGA) as a function of the cost of an ultrasound for fetal presentation only. Calculations are based upon a willingness-to-pay (i.e. valuation of one additional quality-adjusted life year) of £20,000.

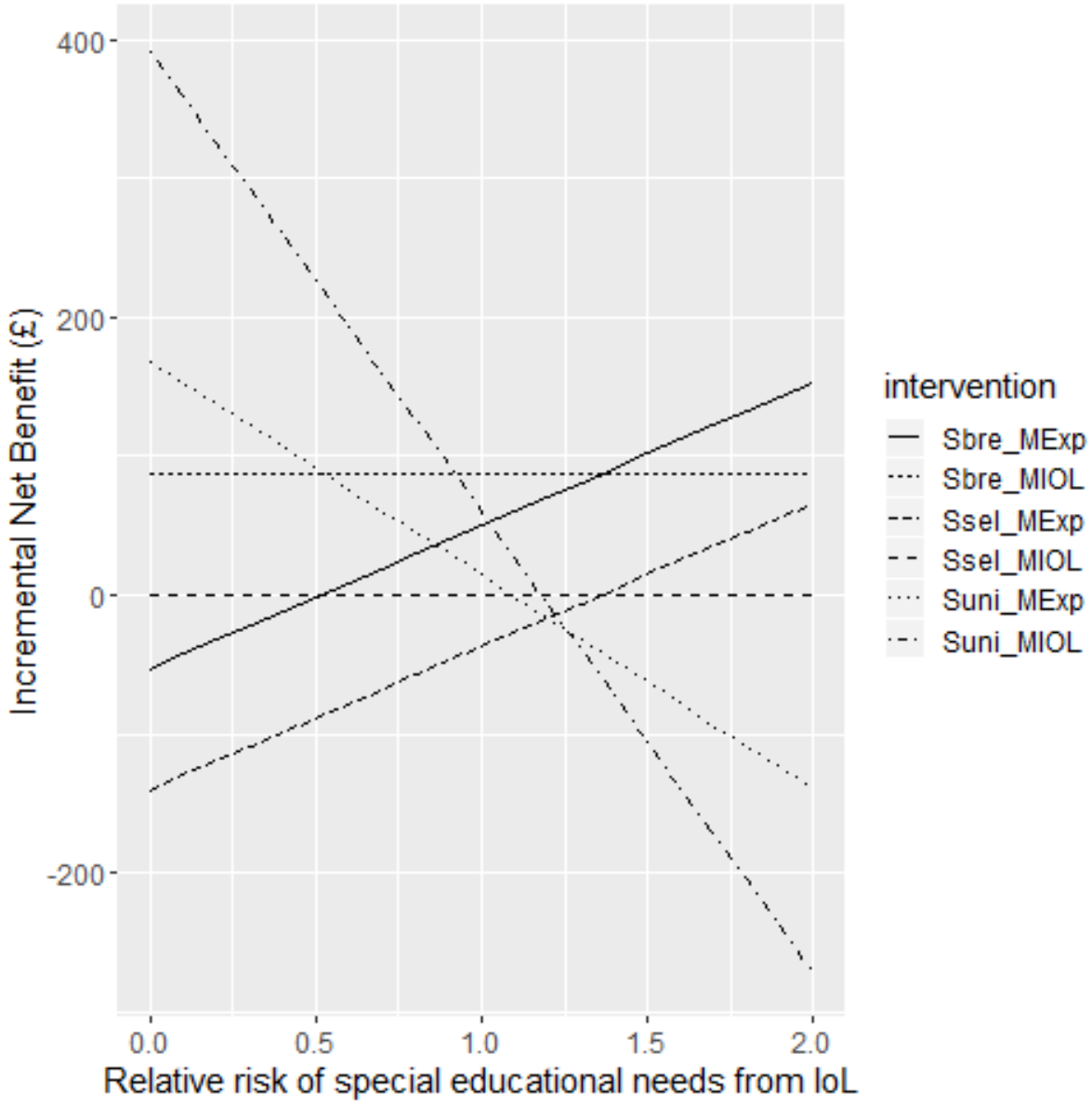
Figure 20. One-way sensitivity analysis on baseline risk of perinatal mortality, moderate and severe morbidity respectively.



MExp = Expectant management, MIOL = Induction of labour, Sbre = Universal ultrasound for fetal presentation only, Ssel = Selective ultrasound, Suni = Universal ultrasound for fetal biometry plus presentation.

The figure shows the expected incremental net monetary benefit for different strategies compared to current practice (selective ultrasound with induction of labour for suspected LGA) as a function of the baseline risk of perinatal mortality (left), moderate neonatal morbidity (middle), and severe neonatal morbidity (right). Calculations are based upon a willingness-to-pay (i.e. valuation of one additional quality-adjusted life year) of £20,000.

Figure 21. One-way sensitivity analysis on relative risk of special educational needs from induction of labour.



MExp = Expectant management, MIOL = Induction of labour, Sbre = Universal ultrasound for fetal presentation only, Ssel = Selective ultrasound, Suni = Universal ultrasound for fetal biometry plus presentation.

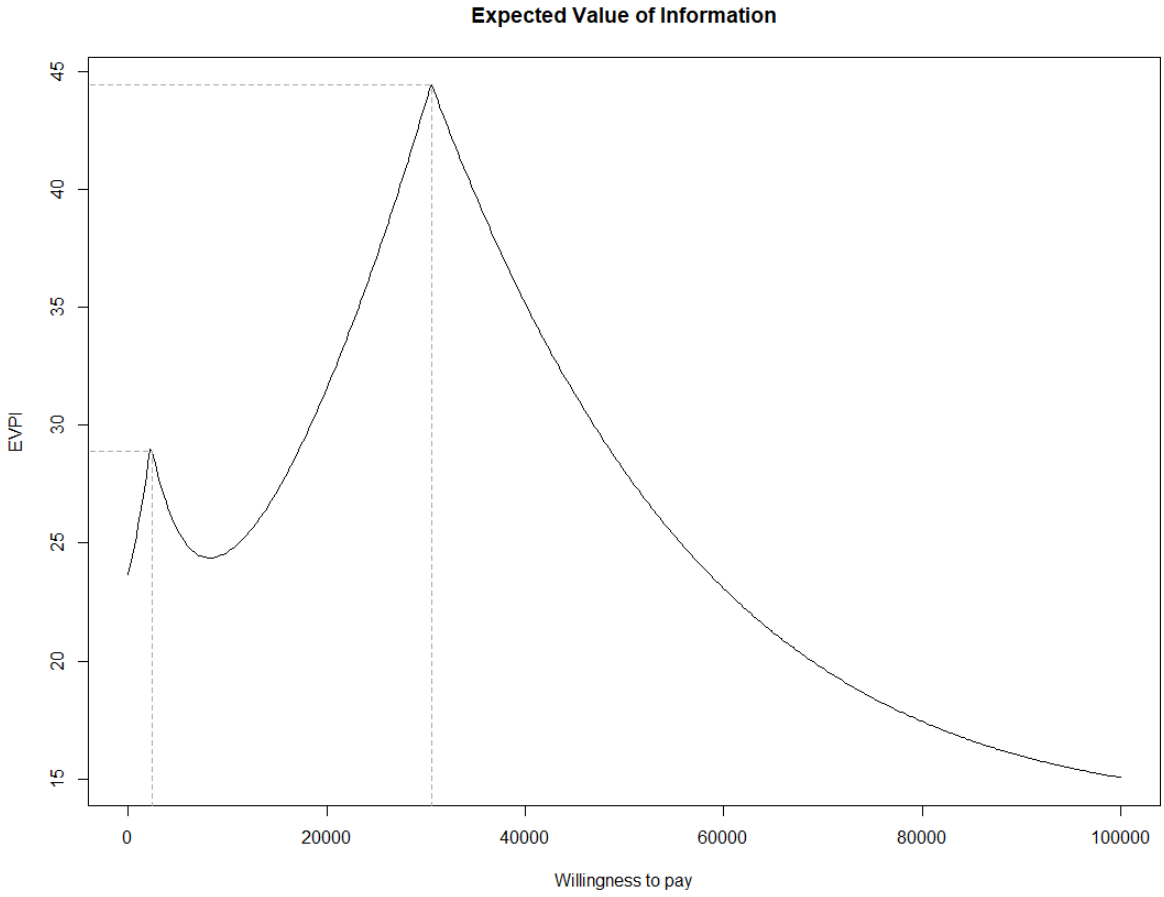
The figure shows the expected incremental net monetary benefit for different strategies compared to current practice (selective ultrasound with induction of labour for suspected LGA) as a function of the relative risk of special education needs if labour is induced early (compared to expectant management). Calculations are based upon a willingness-to-pay (i.e. valuation of one additional quality-adjusted life year) of £20,000.

Value of information analysis

Expected Value of Perfect Information

At a willingness to pay of £20,000 / QALY, the per patient EVPI is £31.56. Given a beneficial population of 1,689,663, the population EVPI to England is £53.3m. If the results of the analysis are assumed generalizable to all pregnancies in England, then the population EVPI is £172.9m. Figure 22 shows the per patient EVPI as a function of the willingness to pay threshold. The two local peaks are where the decision (i.e. which screening strategy is preferred) changes and thus the impact of decision uncertainty is greatest around these thresholds.

Figure 22. Per patient expected value of perfect information (EVPI) as a function of the willingness-to-pay for an additional quality-adjusted life year (QALY).



EVPI = Expected value of information

EVPI presented per person. Willingness to pay refers to monetary valuation of an additional quality-adjusted life year (£).

Expected Value of Perfect Parameter Information and Expected Value of Sample Information

Table 15 shows parameters with an EVPPI exceeding £100,000 under the broader assumption that any future study will be of value to all births in England, not just low risk singleton pregnancies. The most valuable parameter is difference in cost of delivery from induction of labour (c.IOL), accounting for 84% of the EVPI. Except for this cost, no other parameters individually account for more than 1% of the total EVPI. The other parameters with the greatest contribution to EVSI are: the relative risk (LGA versus AGA) of acidosis from a vaginal delivery following IOL ; the odds ratio of perinatal death (LGA versus AGA) from a baby being delivered vaginally without IOL; the relative risk (SGA versus AGA) of Emergency CS following IOL; ´the odds ratio (SGA versus AGA) of severe neonatal morbidity under expectant management.

These five parameters could naturally be collected from three separate studies, namely:

1. a costing study of the difference in cost of delivery associated with IOL versus expectant management
2. an RCT of delivery outcomes relating to LGA
3. an RCT of delivery outcomes relating to SGA babies

The EVPPI of the costing study is either £44.8m or £145.2m, depending on whether the results are considered applicable to only singleton nulliparous pregnancies, or all pregnant mothers respectively. The two RCTs have EVPPIs of up to £3.9m and £1.4m under the broader applicability criteria.

The EVSI of the costing study suggests scope for it to yield a positive return on investment. For example, a study with 1000 patients (in each arm of a two-arm study) has an EVSI to England of £11.3m (or £97.2m if this information is of value to all pregnancies in England, not just low risk nulliparous singleton pregnancies). If such a study was to cost £1m, then it would yield a net return on investment of at least £10.3m (

Figure 23).

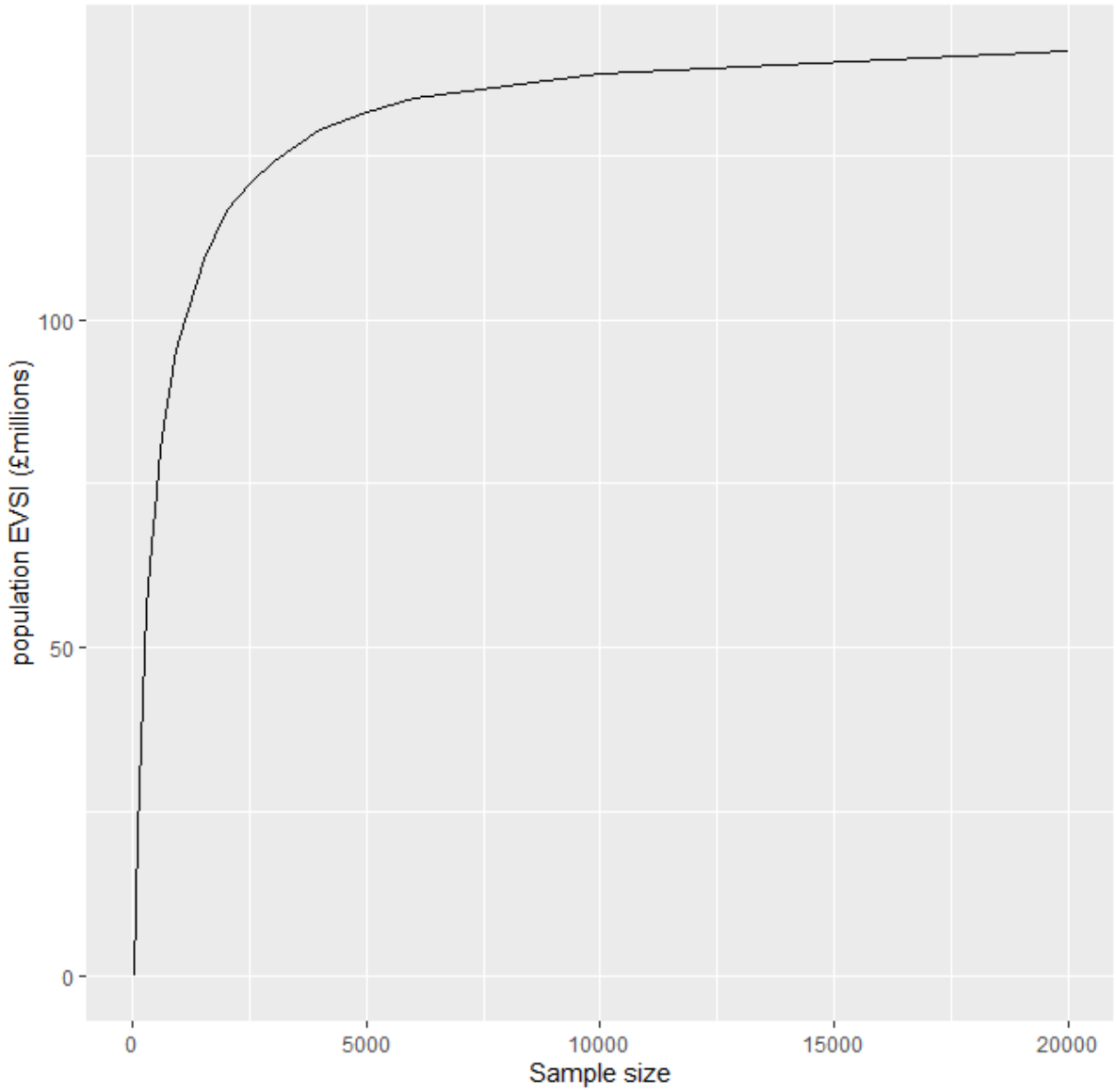
We were not able to calculate non-zero EVSI estimates for studies on macrosomia or SGA outcomes as the per-patient EVPPI is too low.

Table 15. EVPPI for individual parameters and groups of parameters.

	Per Person EVPPI (£)	SE	% of EVPI	pEVPPI (£)	pEVPPI (£)*
c.IOL	26.51	0.07	84	44,790,000	145,200,000
RR.acidosis.macro.IoL.vag	*0.27*	*0.04*	*1%*	*456,000*	*1,478,000*
OR.mort.macro.vag	*0.26*	*0.03*	*1%*	*438,900*	*1,423,000*
Group	0.72	0.07	2%	1,215,199	3,939,513
RR.EMCS.SGA.IoL	*0.06*	*0.01*	*0%*	*99,290*	*321,900*
OR.sev.SGA	*0.03*	*0.01*	*0%*	*48,740*	*158,000*
Group	0.26	0.04	1%	443,104	1,436,484

** Assuming study results are applicable to all births in England.*

Figure 23. Population Expected Value of Sample Information for a study on the cost of induction of labour.



EVSI = Estimated value of perfect information

Expected Value of Perfect Parameter Information under alternative scenarios

EVPPI provides the value of obtaining perfect information for a parameter based upon the magnitude for which perfect information would affect the decision outcome. This mean that even parameters with great impact upon overall cost and QALYs, and for which the value is highly uncertain, may have low EVPPI if perfect information would not change the decision; i.e. which screening strategy is most cost-effective. However, whether the exact value of a parameter affect the decision outcome is highly dependent on context. Through simulating alternative scenarios, we analysed how the EVPPI of key parameters were affected by model assumptions.

Given the uncertainty over which setting an ultrasound scan for fetal presentation only could be provided in, there were some concerns that the cost was not correctly specified in the base-case scenario. We therefore simulated three alternative scenarios where we varied the assumptions underlying the cost calculations: 1) where fetal presentation could be assessed through directly accessed diagnostic services (£52 [95% CI: £24, 91]), 2) where an antenatal standard routine ultrasound scan was required (£108 [95% CI: £97, 118]), and 3) where costs could range between either of these scenarios (£24-118). Results showed that EVPPI was highest where the cost was highest. For this scenario, the EVPPI was £6.07 per person. Depending on the beneficial population, the overall EVPPI was £10.3m (nulliparous women only), or £33.3m (all mothers). It is worth noting that the model's assessment of the value of further studies is in this case at odds with cost-effectiveness. A greater cost for scanning means a lower chance of ultrasound for fetal presentation being cost-effective, but the value of researching this parameter further increases.

The cost of IOL (specifically, the net difference in total cost between pregnancies that were induced early versus expectant management) had the greatest EVPPI in our base-case scenario, and hence the greatest expected benefit from future research. In the base-case scenario, the cost was £125 (95% CI: -£1343, £1594); more details are presented in appendix 7. To test how sensitive the EVPPI was to the exact input values used, we simulated two alternative scenarios: 1) where the standard error of the mean was reduced by 50%, and 2) where costs were instead obtained from the 35/39 trial, where the cost difference was -£236 (95% CI: -£646, £174)[195]; see appendix 7 for details. When the standard error was reduced by 50%, the EVPPI fell by ~80%. When costs were obtained from the 35/39 trial, the EVPPI was £6.3m for the beneficial population (nulliparous women).

Discussion

Main findings

This study has evaluated the cost-effectiveness of alternative screening strategies for ultrasound in the third-trimester in a population of low-risk nulliparous women. Based on current information, and assuming a willingness to pay of £20,000 per QALY, offering a universal ultrasound (US) presentation-only scan is on average the most cost-effective strategy. This is associated with an incremental net monetary benefit of £87.36 (95% CI: 4.88, 205.68) per pregnancy compared to current practice. Scaled up to the English population, this equates to an added net benefit of £17.1m or 857 QALYs per annual birth cohort. This is the present value of the future flows of expected costs and benefits over a time horizon of 20 years.

Third-trimester scans for fetal size should take place only where clinically indicated. We estimate that the added benefits of including estimation of fetal weight in the scan are too small to justify the added cost: more health would be lost elsewhere than would be gained from the added knowledge and subsequent management from these scans. Where LGA is suspected following ultrasound, early IOL is the preferred management, irrespective of whether screening was offered routinely or following clinical indication.

It should be noted that the presentation-only scan policy implies an increased burden on those performing the scan, but that this is partially offset by reductions in the cost of complications from delivery. Implementation would therefore require a reallocation of resources away from delivery and towards antenatal care or ultrasonography.

Due to uncertainties in the evidence base (parameter uncertainty), there is a only a 44% probability that this screening strategy really is the most cost-effective, i.e. there is a 56% probability that this conclusion is incorrect in which case a loss will be incurred. The expected loss associated with this decision uncertainty is £31.56 per pregnancy. Equivalently, this is the expected gain if uncertainty were to be eliminated (expected value of perfect information, EVPI). Scaled up to the population of England who could benefit from the information from any future studies, this equates to an EVPI of £53.3m. If it is assumed the results of any future study are generalizable to all pregnancies in England, the EVPI is £172.9m.

The net difference in cost between an induced delivery and expectant management was the parameter with the biggest impact on decision uncertainty in the base-case scenario, and hence is the

parameter that should be prioritised for future research. It should be noted that this does not simply relate to the cost of a procedure to induce delivery; included within this definition is uncertainty as to the timing of induction, and the impact on for example, antenatal appointments, as well as the cost of the delivery itself. A study of 'reasonable size' to reduce uncertainty in this parameter is likely to yield a positive return on investment. For example, the EVSI of a study of 1000 mothers in each arm is worth in excess of £11m. If this was to be delivered for a cost of £1m, it would yield a greater than 10-fold return on investment. Alternative scenarios found that the value of future research may be less than for the base-case scenario. Nonetheless, although the exact value of future research is hard to determine, the net cost of labour induction appears influential for which screening strategy is the most cost-effective. Of note is that studies on the outcomes from SGA or LGA fetuses are unlikely to yield a positive return on investment based upon the model.

Our base-case scenario showed very limited value in further researching the cost for which an ultrasound scan for fetal presentation only could be provided. However, this was because the model deemed a policy of universal ultrasound for fetal presentation so cost-effective that the cost of the scan was unlikely to change which policy is preferred; one-way sensitivity analysis showed that, all else equal, the cost of a presentation scan would need to exceed £90 before another screening strategy was likely to be more cost-effective. In practice, the cost for which universal ultrasound for fetal presentation only could be provided is uncertain, mainly because it is unclear which type of clinical setting would be required for the scan. Therefore, prior to any roll-out, it is essential to establish whether, for example, midwifes can be trained to perform the presentation scans and find it feasible to incorporate them into routine antenatal visits, or whether this can only be done in a secondary care setting.

The results described above relate to a willingness to pay threshold of £20,000 per QALY. At a threshold of £30,600 per QALY (just above the upper end of NICE's stated acceptable range of £20,000 to £30,000[190], universal scanning becomes the most cost-effective option. Furthermore, our one-way sensitivity analyses suggest there is scope for universal scanning to be cost effective under other assumptions. For example, the most cost-effective option remains a breech-only scan only so long as the time horizon of the analysis is below 45 years. The ideal time horizon for an economic evaluation should be sufficient to capture all relevant differences in cost and outcomes.[190] In many cases this implies a life-time horizon.[196] However, our base case analysis was limited to 20 years. This represents a compromise between the desire for a long time horizon and the inherent uncertainties in extrapolating relatively short-term data into long term outcomes. We therefore acknowledge the

possibility that universal ultrasound scanning may be cost-effective in the long run, but would urge caution in any recommendation of such.

Finally, all else being equal, a presentation-only scan is the most cost-effective option provided it can be accomplished for below £90 a scan. This is higher than we estimated in our previous work, which estimated a maximum cost-effective price of a presentation scan of approximately £20[10](Wastlund et al., 2019a). This difference is due to the more detailed modelling in this analysis; where the previous analysis based QALY gains upon mortalities averted and a set life-expectancy, this analysis included the impact of morbidity on costs and quality of life, and incorporation of explicit survival functions.

Strength and weaknesses

Through incorporating several conditions detectable by US screening into one decision model, this study was able to assess the overall effect that the introduction of universal US may have upon a population of nulliparous women. It also enables assessing the impact that introducing such a programme would have upon NHS budget and whether it is likely to represent good value for money. Further, by incorporating a value of information analysis, this study has the potential to assess, not only where the current gaps are in the evidence base for evaluating the use of universal US screening, but also for which of these gaps future research would have the greatest potential of resulting in meaningful findings.

A key limitation of this study is that only foetal outcomes were considered, excluding those of the mother. Maternal outcomes may be as significant as fetal outcomes, or more so. Further, the well-being of mother and child are sometimes at odds with each other, and clinical decisions frequently trade-off between the two. Incorporating maternal outcomes into the analysis therefore could have an impact upon both the cost-effectiveness of the different strategies (in either direction), and our value of information analyses guiding where future research should be prioritized. However, as per our original protocol, maternal health consequences were not incorporated in this study. The primary justification for this is the lack of sufficiently reliable evidence of how screening outcomes may affect maternal quality of life. We have previously emphasised the need for further such research in this area, particularly surrounding long-term maternal consequences from mode of delivery,[10, 150] and repeat that call here.

Throughout the development of the simulation model, we have attempted to capture clinical probabilities and their respective uncertainties as accurately as possible. However, uncertainty

persists for many parameters, not only over their exact value, but also about how well suited these are for the new decision context. Essentially, this creates two separate types of uncertainties. The internal validity is well captured in the model through the incorporation of parameter uncertainty as quantified by the authors of the respective source. However, there is also the question of external validity, i.e. to which extent that parameter is suitable for our model, which is uncaptured by the model. This means that the true uncertainty of our results are likely to be greater than what is expressed in the confidence intervals of the outputs. While this does not invalidate the model as a tool for decision-making, it means that thoughtful interpretation of the results is needed, and that such interpretation should always acknowledge the inherent uncertainty of combining data from different sources.

Through its focus on breech presentation, SGA and LGA only, this analysis may have underestimated the merits of universal US. Such a screening program would also raise the chance of detecting otherwise unknown complications, e.g. previously undetected congenital anomalies or placenta praevia. Although less prevalent than the conditions included in this analysis, the potential of detecting such complications could be an added benefit of introducing a universal US programme. However, it is important that subsequent management of other such complications follow protocols which have taken the diagnostic performance of US into account. If the risk of false positive diagnoses are high, and the consequences severe, introduction of universal US risks putting patients in a worse position than they would have been in without screening.

The outcomes of economic modelling and especially value of information (VOI) analysis are highly sensitive to the structural assumptions that underlie the simulation model. Throughout this analysis, we have attempted to model the potential outcomes of screening using parameters for which credible data are available. Where parameter uncertainty has been wider, the expected value of future research is generally greater. However, this approach has required us to be able to incorporate a parameter into the model structure. The problem has been capturing effects that we suspect exist, but for which no evidence has been available.

In this analysis, we modelled the risk of long-term outcomes such as SEN as a function of neonatal morbidity. This means that clinical interventions that can alleviate neonatal morbidity are also expected to alleviate the risk of SEN. Similarly, interventions that do not affect neonatal morbidity will have no impact upon the risk of SEN. However, this may not accurately capture how interventions

affect SEN. This model structure has been adopted due to data limitations, and to avoid overestimating the effect of intervention.

There is some evidence that the risk of SEN increases with early IOL, and the perceived risk of this is often influential in the clinical decision of whether or not to induce labour early. Our model structure captures long-term effects on SEN from early labour induction if it is mediated through neonatal morbidity. However, if there is a direct link between gestational age at delivery and the risk of SEN that is not mediated through neonatal morbidity, this is uncaptured in the model. One-way sensitivity analyses exploring this suggest that our results hold so long as the risk of SEN associated with IOL (versus expectant management) is below approximately 1.34. Above this, the recommendation for a presentation-only scan holds, but inducing labour for LGA is no longer recommended. If it is plausible that the increased risk of SEN associated with IOL exceeds 34%, then it may be worthwhile exploring this in future research. However, observational data indicate that delivery at 38 wkGA is associated with less than 34% increase in risk[174].

Whilst macrosomia and SGA are mutually exclusive by definition, we assumed breech was also mutually exclusive with SGA and LGA. This simplification was used because data constraints would not allow a credible estimation of risk adjustments for fetuses that were both breech and SGA/LGA, and for structural simplicity of an already complex model. It was also considered likely that breech presentation would be a stronger determinant of possible clinical interventions than fetal size. Relaxing this assumption would in practice have the same effect in the model as a slight increase of the prevalence of SGA and LGA, however, the effect of this would be limited given the low prevalence of fetuses with a breech presentation and SGA/LGA.

The conclusions for our economic analysis, and especially for the VOI analysis, depend heavily on the exact data used to capture parameter uncertainty in the economic model. However, accurately capturing the uncertainty of a parameter in line of all current evidence is far from straightforward. For many parameters, alternative sources were available, and the combined parameter uncertainty for multiple studies is theoretically smaller than for just the one. Ideally, every input parameter in the model should be subject to a meta-analysis. However, due to the high number of parameters in the model, this was not feasible. Further, in many cases, we suspected that the difference in parameter values between studies were the result of different clinical definitions rather than reflective of the true parameter uncertainty. To address this issue, we conducted extensive one-way sensitivity analyses.

We modelled acidosis risk as that secondary to shoulder dystocia as well as 'other acidosis'. No sources disaggregated that attributable to shoulder dystocia and other causes. We may therefore have overestimated the risk of acidosis via double counting. However, our sensitivity analyses suggested the base case results were insensitive to this parameter.

Comparison with other studies

A previous review of studies of universal ultrasound assessment during the late pregnancy found no clear benefit of universal ultrasound.[20] In this study, we have found that universal ultrasound may be associated with better clinical outcomes. Whether universal screening is cost-effective, however, depends on the features included in such a scan. Our analysis show that universal ultrasound for fetal size is unlikely to be cost-effective, unless the valuation of additional health is higher than those recommended by current UK guidelines.[190] By contrast, universal ultrasound for fetal presentation alone is likely to be cost-effective, although uncertainty persist over whether fetal presentation could be assessed sufficiently cheaply using ultrasound to make such a screening policy feasible.

Further, the findings also align with our cost-effectiveness analyses of universal ultrasound for individual complications only. When exploring the cost-effectiveness of universal ultrasound for breech presentation only, we found that whether such a screening program could be cost-effective largely depended on the price for which fetal presentation could be detected.[10] It seemed unlikely that screening for SGA or LGA only would be cost-effective, however we highlighted that the effectiveness of labour induction was uncertain and may warrant further research. This joint analysis confirms these findings, and has allowed us to point more specifically towards those parameters for which further research may have a meaningful impact upon the decision problem.

Implementation considerations

The purpose of this stream has been to make recommendations on screening policy based upon our current understanding of the evidence base, to identify the current gaps in the evidence, and provide recommendations about which of these gaps should be addressed to allow future policy-making about late-pregnancy ultrasound in the relevant population. We speculate that late-pregnancy ultrasound screening for fetal presentation only could be provided by midwives as part of a routine antenatal assessment. Such a screening setting has obvious benefits for the patient, as an extra appointment (typically in a secondary care setting) could be avoided, saving time and travel costs for mothers and possibly partners as well. However, an US scan in this context would not also assess fetal biometry. It

is important that the introduction of such a screening programme into NHS routine care would not expand the scope of this scan beyond assessing fetal presentation as this may lead to unnecessary intervention. Another potential problem for the NHS would be the implied relocation of budget between units. Though universal US in a primary care setting may be cost-effective for NHS as a whole, in practice this would put extra financial strain on primary care, while the benefits would mostly arise from the avoidance of complications following delivery. To be successful, the implementation of such a screening policy would need to be accompanied by a suitable reallocation of budget from the benefitting units into primary care.

The consequences of future research likely go beyond the perspective employed in this analysis. First, our analysis focused upon nulliparous women with singleton pregnancies, but for many parameters, reducing uncertainty would be helpful to women regardless of parity. To address this, we provided two population values of information: one based on nulliparous singleton pregnancies, and the other on all pregnancies. Second, the scope of our study was limited to England, but many findings are likely to be just as applicable to the rest of the UK, and indeed to other high-income countries as well. If the value of information analyses are considered applicable to the entire UK, the EVPI, EVPPI and EVSI figures should be multiplied by approximately 25% to reflect this (England accounts for approximately 80% of the UK population). Third, the economic perspective of this study was the English NHS and education services only, but many consequences would go beyond this. For instance, it has been estimated that the majority of the costs associated with stillbirth and cerebral palsy are indirect, e.g. from decreased productivity, extra monitoring for subsequent pregnancies, mourning etc.[184, 186, 197] When considering such perspectives, both the attractiveness of universal ultrasound and the value of future research is likely to increase.

Conclusions

The remit of this work was to advise NIHR on the current body of evidence regarding the cost-effectiveness of late pregnancy ultrasound screening, and specifically whether (a) there was value in commissioning further research in the area, and (b) if so what.

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 IOL compared to expectant management.

Chapter 12. The views of recently delivered and currently pregnant women on universal ultrasound screening in late pregnancy.

Aims

The aims of this section were the following:

1. To assess the knowledge of pregnant women on the current antenatal care pathway for low risk pregnancies.
2. To assess their understanding of the potential benefits and drawbacks of third trimester screening.
3. 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.

Methods

In order to evaluate both the quantitative and qualitative aspects of the 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). Amanda Stranks, the head of the PPI department of CUHFT, had active role in the writing and testing of the survey as well as the design, recruitment and running the focus groups as explained below.

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. We excluded any high risk pregnancies with either maternal or fetal pathology. 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. We have attached the final version of the questionnaire in Appendix 8. 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.

The second part this section was to run focus groups 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 including Facebook, Twitter and WhatsApp to address groups of mothers in the broader area of Cambridge. The focus group discussion was designed by AM, GS and Amanda Stranks.

Results

Survey

We collected 100 replies from pregnant women attending for their routine dating or anomaly scan 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.

Table 16. Results of the survey of low risk pregnant women (n=100).

Question	Answer	Number of responses
1)Were you aware that women whose pregnancies are straight-forward are NOT routinely scanned after 20 weeks?	Yes	85
	No	15
2) "I would like to have the option of a scan at around 36 weeks as part of my routine NHS care".	Agree/Strongly agree	84
	Neither agree nor disagree	13
	Disagree/ Strongly disagree	3
3) I would be likely to agree to take part in study A	Agree/Strongly agree	76
	Neither agree nor disagree	17
	Disagree/ Strongly disagree	7
4) I would be likely to agree to take part in study B	Agree/Strongly agree	66
	Neither agree nor disagree	18

	Disagree/ Strongly disagree	16
5) If you are happy to participate in one of the above research projects which one would you prefer?	Study A	10
	Study B	23
	Both	32
	N/A- Missing	35
Maternal age	<30	38
	≥30	60
	Missing	2
Ethnicity	White British	40
	Other British	20
	Other European	17
	Asian/African	8
	Missing	15
Age stopped education	<22	53
	≥22	39
	Missing	8

Focus group

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 focus group was run by Alexandros Moraitis and Amanda Stranks (PPI Lead in CUHFT). The participant characteristics are as below:

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 IOL 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 IOL 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.

Discussion and Conclusions

We were able to collect both quantitative and qualitative data the opinions of women on third trimester ultrasound screening. We saw a clear interest in having an additional ultrasound scan in the third trimester which was also confirmed in the focus group by all but one participant. This also confirms the previously published finding by the Stillbirth Priority Setting Partnership[198], which included responses by over 300 parents and 700 professionals, and concluded that the question whether third trimester ultrasound can reduce stillbirth was one of the most important research priorities. We also found that the majority of women would be happy to participate in a future randomised trial and we would expect a recruitment rate of at least 2 out of 3 women, which is similar

to the recruitment rate of the POP study where ultrasound result was blinded to the women and the clinicians. 66% of women were that replied to our questionnaire and all the participants in the focus group would be happy with the blinding of the ultrasound if there was no severe problem which we will have to define clearly.

Reflections/Clinical perspective

We managed to acquire a large number of replies (as planned in advance) through a questionnaire which gave us an overall view of women's opinion and willingness to participate in a future trial. However, we found it difficult to recruit women for the focus groups. Prior to recruitment, after discussion with the collaborators and the PPI office in CUHFT, we made the decision not to include pregnant women in the focus groups as the discussion could create anxiety with their care. However, it was also difficult to recruit new mothers and they could not easily find the time to participate in a focus group. We managed to recruit four women by arranging for childcare and transport (in one case). The input from the focus group was valuable because we had the opportunity to listen to women that were keen to have an additional scan and a woman that was sceptical about the need of those additional scans. We also gained valuable information on what to include in a future consent form and the timing that this should be done. Overall, we believe that all the above information would affect the design and conduct of a future clinical trial.

Chapter 13. Designing a randomised controlled trial of screening and intervention.

Implications of the health economic analysis

The economic analysis demonstrated that whilst on average, the most cost-effective approach was to screen all nulliparous pregnant women with a presentation-only scan, there was only a 44% probability that this was true and a scan included fetal biometry had a ~39% chance of being the most cost-effective. Moreover, if the timescale was increased, it became likely that such a scan in late pregnancy would be the most cost-effective approach. These observations indicate that implementing such a scan should be seriously considered. However, one of the major obstacles to implementing such a policy is that there is no direct evidence from a randomised controlled trial to show that such screening and intervention is clinically effective. The Cochrane review of universal late pregnancy ultrasound has failed to show any benefit to the mother or baby. However, as discussed in the introduction, this review has a number of methodological issues and it is more accurate to state that it does not provide any information to answer the question of whether universal late pregnancy ultrasound reduces the risk of perinatal death or not.

Interestingly, the VoI analysis highlighted reducing uncertainty about the costs of IOL. Given the above, this may be regarded as somewhat counterintuitive. However, the parameters used in the VoI analysis in relation to the screening performance of ultrasound and the effect of intervention were known with a degree of precision that meant that reducing their uncertainty was not identified as the most cost-effective research question. For example, the ability of ultrasound to predict SGA, the relationship between SGA birth weight and the risk of stillbirth and the ability of IOL to reduce the risk of stillbirth are all known quite precisely and are based on high quality data. Consequently, even though there is no direct evidence to indicate that universal late pregnancy ultrasound would reduce the risk of stillbirth, the model estimates quite a high chance that it is the most cost effective approach and does not highlight reducing the uncertainty in these parameters in the VoI analysis. In contrast, previous health economic analyses of IOL have generated quite wide confidence intervals,[179, 195] hence the model has identified reducing this uncertainty as the key question.

Case for considering a randomised controlled trial of screening and intervention

In this chapter we consider the practicalities of designing an RCT of screening and intervention using fetal biometry in nulliparous women at 36wkGA. We have done this because, even though the parameters in the modelling were reasonably certain, these parameters were calculated from a range of different study designs. i.e. we did not perform the VoI analysis based on the uncertainty of

parameters calculated from a large RCT of late pregnancy screening and intervention in nulliparous women. Rather, we performed the analysis using parameters from a range of observational studies and a range of studies of intervention in women who were deemed high risk for other reasons. The concern in this case is with external validity. The parameters may be reasonably certain in relation to the setting where they were derived but there is an unquantifiable uncertainty in relation to how well they inform our research question. The obvious way to address this would be to perform a study in the setting of interest. Such a study could be the definitive study or it could be a pilot or proof of principle study. The former might be a trial of screening versus not screening with perinatal death as the primary outcome. The latter might exploit alterative study designs and use of proxies. Hence, there are a number of important considerations to take into account when designing an RCT of screening and intervention using universal ultrasound and we will consider each of these in turn.

Candidate primary outcomes

In relation to primary outcome, we believe that the strongest case can be made for perinatal death. First, to lose a baby at term is clearly a devastating outcome for a family. In the absence of a lethal anomaly, preventing the death would lead to an entire life gained which, from a healthcare and health economic perspective is a gain of unique magnitude. Second, the main intervention available is earlier delivery. There is strong evidence that IOL is effective in reducing the risk of perinatal death. Over two thirds of perinatal deaths at term are antepartum stillbirths,[51] i.e. intra-uterine fetal death prior to the onset of labour. Self-evidently, antepartum stillbirth cannot occur after a baby has been delivered.[16] Delivery at or after 38-39 weeks of gestational age carries the same risk of intrapartum stillbirth and neonatal death as delivery at later weeks of gestational age.[16, 199] These epidemiological observations underlie the 67% reduction in the risk of perinatal death associated with IOL at term.[15]

Proxies

The main problem with a primary outcome of perinatal death is that the outcome is uncommon and this will result in major issues of statistical power. Indicators of perinatal morbidity would be an alternative to perinatal death. First, as the same factors might be involved in death and morbidity, the latter could be used as proxies of the former. Second, perinatal morbidity is of importance in its own right. For example, birth asphyxia is one of the major determinants of the burden of litigation in the health service through devastating effects on the later health of the child, such as cerebral palsy. There is evidence that supports the use of a single indicator in both roles. A Apgar score of <4 at five minutes was associated with a relative risk of early neonatal death of ~360[176] and a relative risk of cerebral palsy of >400.[175] Hence, a primary outcome based on perinatal morbidity, such as Apgar <4, could be

clinically important, both as a proxy of death and as a determinant of long term outcome. Morbidity could be a more pragmatic outcome as rates of severe morbidity are much greater than the risks of death, hence it may be easier to design a trial.

Sub-groups

A further refinement to the primary outcome is to study sub-groups of the given event that were actually associated with the baby being born SGA or LGA. It is self-evident that screening for SGA or LGA will primarily impact on outcomes that are related to fetal growth disorder. Many adverse perinatal outcomes, both lethal and non-lethal, are unrelated to fetal growth abnormalities. Consequently, if a screening study of fetal biometry has a primary outcome which includes babies in the full range of birth weight, most of the primary outcomes in both arms of the trial will be unrelated to fetal growth disorder hence not preventable by screening for fetal growth disorder and intervention. This means that the potential for screening to impact on the rate of death is limited and extremely large sample sizes would be required. For example, about one third of perinatal deaths at term are related to being SGA or LGA.[51] The background rate of perinatal death at term is ~2 per 1,000. Even if a screening test was perfect (i.e. detected all cases of growth disorder) and even if the intervention was perfect (i.e. prevented all such deaths), a power calculation still indicates that >100,000 women would have to be recruited to the trial. However, if the primary outcome was perinatal death of an SGA or LGA infant, the sample size would be ~22,000 (note: this is used to illustrate the point, it is not a practical proposition as the screening and intervention characteristics were perfect). An analogy might be in a trial of breast cancer screening. Screening reduces deaths related to breast cancer but does not reduce all-cause mortality.[200] This is likely explained by the fact that no study could be sufficiently powered to detect an effect of screening for breast cancer on all-cause mortality because most deaths are due to other causes. Consequently, one approach to addressing the problems of statistical power in trials of screening using fetal biometry would be to define primary outcomes which were related to fetal growth abnormalities. Insistence on evidence that shows reduction in all cause perinatal death would simply remove the possibility of screening and intervention being implemented, which could lead to avoidable harm which could have been prevented in a cost-effective way.

Early delivery and iatrogenic harm

Routine induction at term had less dramatic effects on the risk of neonatal morbidity, with a 12% reduction in the risk of NICU admission and a 30% reduction in the risk of a low Apgar score. Moreover,

these effects may be lost or even reversed in the context of early term IOL. Most of the trials in the Cochrane review of term induction were at 41wkGA and beyond.[15] As post-term pregnancy is associated with increased risks of neonatal morbidity, preventing this outcome should improve immediate neonatal outcome as well as prevent stillbirth. In the context of IOL <39 weeks, epidemiological data indicate that the intervention may actually increase neonatal morbidity.[162] The potential for earlier intervention to cause harm is increasingly recognised. The AFFIRM study reported a stepped wedge RCT of a programme to inform women about reduced fetal movements and to standardise intervention. Although it did not show a significant reduction in stillbirth, the intervention was associated with increased risks of neonatal morbidity.[201] This trial has some parallels with the current question. Despite the fact that women were selected on the basis of having a risk factor (reduced fetal movements, which is associated with stillbirth), it still failed to demonstrate reduction in stillbirth rates and intervention was associated with increased rates of intervention and adverse outcomes. The result of the trial underlines two key issues (i) the need for better predictors of adverse outcome, (ii) the potential for intervention to cause harm.

Current status of screening tests

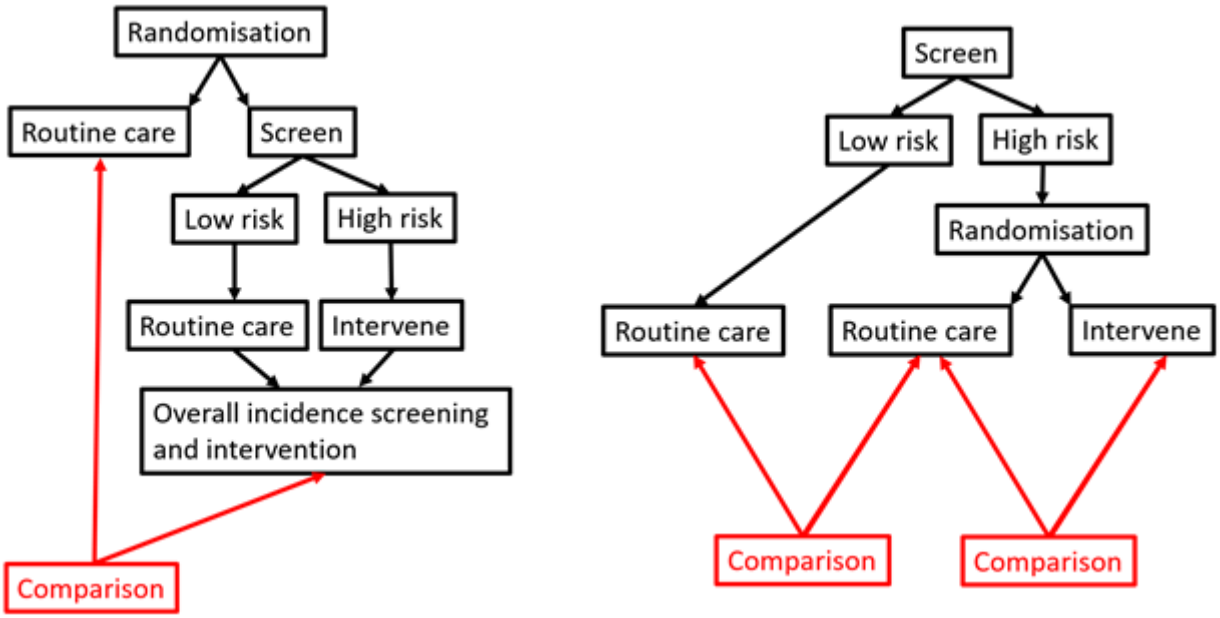
Unfortunately, the results of our systematic reviews of diagnostic effectiveness and a Cochrane DTA review[22] failed to identify any ultrasonic marker that was clearly predictive of the risk of stillbirth in the context of scanning women in late pregnancy. Moreover, if we regard neonatal morbidity as a proxy of stillbirth, again, tests performed very poorly. Finally, actual birth weight <3rd percentile was associated with a 0.9% to 1% risk of perinatal death at term compared with a background risk of just over 0.2%.[51] Hence, even knowing that the actual birth weight was <3rd percentile would be associated with an LR+ of between 4 and 5. In the POP study, of 562 women with a scan indicating that the baby was SGA, only 12% delivered a baby with a birth weight <3rd percentile, a further 23% delivered a baby ≥3rd and <10th percentile but about two thirds of the women delivered a baby ≥10th percentile. Hence, on the basis of the association between estimated fetal weight and actual birth weight, and the relationship between actual birth weight and the risk of stillbirth, it is highly unlikely that detecting an SGA infant is strongly predictive of the risk of stillbirth. Given the lack of information, we model outcomes with variable incidence and assess different screening test values to establish what characteristics would be required of a test to make a trial of screening and intervention feasible.

Possible trial designs

Broadly speaking, there are two main approaches to trial design (Figure 24).[31] First, (hereinafter referred to as screen versus no screen) women might be randomised (a) to be screened, with the offer

of intervention if they screen positive, or (b) to receive routine care, which is currently only to be scanned if there is a conventional clinical indication. The result of this trial design is a simple comparison between the two groups. In the event of a statistically non-significant result, it is impossible to determine whether the result was because the screening test did not work or because the intervention did not mitigate the higher risks in screen positive women. The second approach is that 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), hereinafter referred to as "screen all". 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 the figure below, "screen versus no screen" (left) and "screen all" (right).

Figure 24. Flowcharts of possible trial designs.



Acceptability of the "screen all" approach

When discussing the possibility of randomising women with a high risk screening test some of the co-applicants expressed concerns. Interestingly, however, when we surveyed pregnant women, they actually preferred a study design where all participants were scanned. In the focus group, women tended to be more concerned about being offered interventions. The observations underline the different perspectives of pregnant women and professionals. We envisage that women who were recruited to a "screen all" approach would have some information revealed irrespective of their randomisation status. For example, we do not feel that it would be practical or ethical not to reveal

the presentation of the baby as cephalic or non-cephalic. Hence, this would likely be revealed in a screen all trial design. In the POP study, although scans were blinded, breech presentation was revealed. Subsequent interviews with participants were highly positive about this element of the study where the baby was breech [Dacey 2015; https://www.repository.cam.ac.uk/handle/1810/280595]. A drawback of this approach is, however, that a screen all with reveal breech presentation design would not capture the health benefits of detecting breech presentation. Other features that should be considered for revealing are the presence of previously undiagnosed major congenital anomalies and placenta praevia. In the POP study, there was no case of placenta praevia but two patients had major anomalies diagnosed where revealing the result optimised care and, in one case (unilateral hydrothorax with severe mediastinal shift), likely prevented intra-uterine fetal demise.

Power calculations

In order to determine the feasibility of a randomised controlled trial we have performed power calculations using the two different study designs represented above. The sample size calculations are presented in Table 17. All power calculations are performed for $P<0.05$ (two-sided) with 90% power to detect the effect. We have selected a range of possible primary outcomes, perinatal death, severe neonatal morbidity, any neonatal morbidity, and delivery of an SGA infants with complications. In relation to perinatal death we found no adequately powered studies of the diagnostic effectiveness of ultrasound to predict this outcome and the Cochrane DTA review of SGA also found no data in relation to this question. We have, therefore, modelled a series of possible screening performances, varying the screen positive rate and LR+. In relation to morbidity, we used two studies reporting data from the POP study, from the Lancet 2015[7] and Lancet CAH 2018.[145] As described above, the POP study was one of only two studies (Perinatal Ireland Genesis study being the other) which performed blinded ultrasound in late gestation in nulliparous women. Unfortunately, the Genesis study has not reported the association between SGA and morbidity and the only publication in relation to LGA is in abstract form only and addresses shoulder dystocia. The two POP study publications address the relationship between SGA, SGA combined with reduced growth velocity (which was the best performing predictor of morbidity from a range of candidate predictors of FGR) and the Delphi consensus definition of late FGR.

In all of these calculations we assumed that intervention would reduce the risk of the given event by 50%. Given the lack of data, a range of figures could be considered. We employed this figure as we felt that it was conservative in relation to perinatal death. It could be argued, based on the discussion above, that it is optimistic in relation to neonatal morbidity. However, by concentrating the outcome

for morbidity on infants that are actually SGA, it is plausible that the combined effect of making the diagnosis and intervening could cause a substantial reduction in the rate of adverse events. It should be borne in mind that in the relevant RCT, DIGITAT, randomisation occurred after ultrasonic SGA was suspected. Hence, the group randomised to expectant management would still have received enhanced monitoring and high risk care during labour as the baby was known to be SGA. In contrast, routine care in a trial of screening means that neither antenatal nor intrapartum care are tailored for the suspected SGA status of the fetus.

Table 17. 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).

Implications of sample size calculations

We present the data on sample size calculations but we are not recommending a specific trial design. It is also possible that a trial may be considered where the combination of screening parameters, intervention effect and outcome are not listed in the Table above. The exact design of the trial would depend on the resources available and the research question. We do, however, discuss some of the issues which might motivate a choice.

We believe that the calculations above rule out a trial based either on perinatal death or severe neonatal morbidity as the sample size required is so great that the trial may not be feasible but would inevitably be extremely expensive. Whether the screening test is simple SGA or whether one of the FGR indicators is used will depend on the trade-off between labelling much larger numbers of women screen positive versus sample size. In all calculations, the screen positive rate was higher with SGA but the sample size was lower.

Whether a screen versus no screen or a screen all approach is used will depend on the information required and on the screening test evaluated. A problem with the screen all approach is that is does not recapitulate the real world of comparing not doing something versus doing it. It would also not capture health benefits related to diagnosing non-cephalic presentation at 36wkGA. However, it would provide more information about the evidence base as it would allow quantification of performance of the screening test and intervention separately. Finally, the complicated SGA outcome is delivery of a small baby where either the mother experiences preeclampsia or the baby experiences morbidity. This outcome has the attraction of focusing on the cases which are most likely to reflect true FGR and it is perhaps in this group where the intervention is most likely to yield a positive result. However, a primary outcome that includes morbidity to all infants may be preferred if the priority is to determine the overall effect of screening and intervention. It is also worth noting in the "complicated SGA" outcome that the screen all study design would actually involve performing more scans than the screen versus no screen design if the screening test was simple SGA or the Delphi consensus definition of FGR.

Chapter 14. Overall conclusions and assessment of evidence required for a national screening programme.

Overall conclusions

- Late pregnancy ultrasound is only weakly predictive of neonatal morbidity
- Late pregnancy ultrasound is strongly predictive of SGA and LGA birth weight
- There is a strong health economic case for implementing a scan in late pregnancy to assess fetal presentation
- There is a chance that screening for fetal size in late pregnancy may be cost-effective under the current NHS recommendations, however:
    - The balance of probabilities favours a presentation only scan
    - The case for including assessment of fetal size is sensitive to the assumptions of the model
    - There is no direct evidence from an RCT or meta-analysis that screening and intervention is clinically effective
- The main uncertainty in relation to the health economic case for universal ultrasound (including both presentation and an estimate of fetal size) is uncertainty about the net costs of IOL versus expectant management
- Randomised controlled trials of late pregnancy screening aimed at directly demonstrating a protective effect on the risk of perinatal death or severe morbidity are unlikely to be feasible due to the required sample size
- Randomised controlled trials of late pregnancy screening aimed at directly demonstrating a protective effect on the risk of proxies or sub-groups of outcomes could be feasible due to sample size, but would depend on the exact study design

Consultation with the National Screening Committee

We sent the scientific summary of the project and the chapter on trial design to the UK National Screening Committee (NSC) Evidence Lead who has worked for the UK NSC for >15 years. The UK NSC would be happy to contribute to any further HTA discussions where this is useful. Following preliminary discussion, the applicants plan to submit a proposal to the UK NSC to suggest that they recommend a screening programme for breech near term. Their evidence review process is outlined on their website (https://www.gov.uk/government/publications/uk-nsc-evidence-review-process).

We then discussed the case for a trial of including assessment of fetal size in the same scan. The key questions were as follows:

1. If the uncertainty around the costs of IOL were reduced, how likely is it that the NSC would recommend screening for fetal size near term based on a model that lacked direct evidence from an RCT that involved screening? For example, if the currently funded HTA trial around IOL for suspected fetal macrosomia confirms improved outcomes with intervention, would the combination of the diagnostic effectiveness of ultrasound as a screening test for LGA and the clinical effectiveness of IOL as an intervention in LGA be regarded as acceptable evidence for screening? The issue of interpretation is that screened women are likely to have lower prior odds of complications than women identified as having an LGA fetus through a clinically indicated scan. Hence, extrapolation of the results of the trial may involve an assumption that is untrue.
2. If direct evidence of a beneficial effect of screening from an RCT was required, would this have to come from a screen versus no screen trial or would evidence from a screen all trial suffice?
3. What outcomes would be acceptable? Specifically:
    (i) would screening be recommended on the basis of an effect on proxies?
    (ii) would screening be considered on the basis of an effect on a sub-group, for example, sub-groups of neonatal morbidity or mortality confined to infants which were actually small or large at birth?
    (iii) would screening be considered on the basis of an effect on a composite outcome?

Following discussion, the overview was that the NSC does not have specific “hard stops” but, as one would expect, the stronger the evidence across the 20 criteria for assessing the viability of a screening programme, the more likely it is that a programme would be recommended. For example, because the committee bases recommendations on an assessment of these criteria, it would not necessarily reject a screening programme because the main trial supporting the programme reported a composite outcome in one criterion. But, all other things being equal, a programme would be less likely to be recommended if the study was based on a composite. Hence, none of the questions above were answered by a simple yes/no. But the following were key points:

1. RCTs based on intervention from screen positive women would provide much stronger support for a programme than evidence derived from RCTs of high-risk women (i.e. those not identified through screening the general population).
2. Data from a screen versus no screen study would be preferred to a screen all design. However, one approach if there were absolute methodological obstacles to screen versus no screen

would be to show proof of principle with a screen all study, consider other studies to address any shortfall arising from this design and other criteria, and then to perform a stepped wedged randomised controlled trial when implementing the new test.

Although evidence from trials reporting proxies, sub-groups and composite outcomes would be considered, a strong case for screening would have a simple substantive outcome that reflected the totality of the effect of screening (i.e. benefit to true positives and harm to false positives).

Acknowledgements

Steering group members: David Cromwell, Gianluca Baio, Kathryn Cook, Elizabeth Duff, Neil Marlow, Tracey Mills, Dharmintra Pasupathy

We would like to thank the members of our steering group and our co-applicants Ian White, David Fields and Charlotte Bevan for valuable advice at different stages of the project. We are also grateful towards Amanda Stranks for her strategic help with patient and public involvement and engagement, and Alison Dacey for analysing records on breech presentation in the POP study.

For their roles as second reviewers on one of the systematic reviews, we would also like to thank Tom Bainton (Umbilical artery), Norman Shreeve (Large for gestational age), Illianna Armata (Borderline amniotic fluid index), and Dexter Hayes (Cerebroplacental ratio).

Contribution of authors

Gordon Smith (Professor, Head of Department of Obstetrics and Gynaecology) conceived the project and contributed to protocol development, management of the project, planning of the systematic reviews, conceptualization of the economic models, clinical interpretation of findings, and writing of the report.

Alexandros Moraitis (Research Associate, Obstetrics & Gynaecology) performed the systematic reviews of clinical effectiveness, drafted and edited the final report, designed questionnaires for determining which conditions the systematic reviews should focus on, commissioned the focus group, and contributed to the identification of data and conceptualization of the economic models.

David Wastlund (Research Assistant, Health Economics) conceptualized and programmed the economic models, identified and estimated data for the economic analyses, performed the cost-effectiveness and value of information analyses, performed the systematic review on amniotic fluid index, and drafted and edited the final report.

Jim Thornton (Professor, Obstetrics and Gynaecology) provided input on which systematic reviews to undertake and the design of future research, helped design the questionnaire for identifying topics for systematic reviews, and commented on drafts of the systematic reviews, the economic analyses and the final report.

Aris Papageorghiou (Professor, Obstetrics and Gynaecology) contributed to designing the methods for systematic reviews, provided input on the design of future research, and commented on drafts of the report.

Julia Sanders (Professor, Clinical Nursing & Midwifery) provided input on which systematic reviews to undertake, and edited drafts of the economic analyses and final report.

Alexander Heazell (Professor, Obstetrics) helped design the methods for the systematic reviews, provided input on which clinical areas the systematic reviews should target, and edited the final report.

Stephen Robson (Professor, Fetal Medicine) reviewed chapters on systematic reviews, contributed to the design of patient and public involvement (PPI), and edited the final report.

Ulla Sovio (Senior Research Associate, Applied Medical Statistics) contributed to the statistical analysis of the systematic reviews, reviewed and commented on drafts of the systematic reviews and the economic analyses, contributed to data analysis for the economic models, and edited the final report.

Peter Brocklehurst (Professor of Women's Health) provided input on which systematic reviews to undertake and the design of future research, helped design the questionnaire for identifying topics for systematic reviews, and commented on drafts of the systematic reviews, the economic analyses and the final report.

Edward Wilson (Senior Lecturer, Health Economics) designed and programmed the models for the economic and value of information analysis, designed methods for the quantification of data for the economic analysis, and drafted and edited the final report.

Publications

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

Wastlund D, Moraitis AA, Thornton JG, Sanders J, White IR, Brocklehurst P, Smith GC, Wilson EC. The cost-effectiveness of universal late-pregnancy screening for macrosomia in nulliparous women: a decision-analysis. BJOG: An International Journal of Obstetrics & Gynaecology. 2019

Data sharing statement

"All available data are contained within the report. Requests for patient-level data should be made to Mrs Sheree Green-Molloy at the Department of Obstetrics and Gynaecology, Cambridge University, UK (paoandghod@medschl. cam.ac.uk). All other requests should be made to the corresponding author."

Funding

This study was funded by the National Institute for Health Research Health Technology Assessment programme

References

1. Collaborators GCoD. 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**:1151-210. https://doi.org/10.1016/s0140-6736(17)32152-9
2. NICE NCCfWsaCs, Health. *NICE guidline: Antenatal Care*. London; 2008.
3. ACOG Practice Bulletin No. 175: Ultrasound in Pregnancy. *Obstetrics and gynecology* 2016;**128**:e241-e56. https://doi.org/10.1097/aog.0000000000001815
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**:333-7.
5. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology* 1991;**181**:129-33. https://doi.org/10.1148/radiology.181.1.1887021
6. 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;**14**:e1002220. https://doi.org/10.1371/journal.pmed.1002220
7. 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. https://doi.org/10.1016/s0140-6736(15)00131-2
8. Hoffman C, Galan HL. Assessing the 'at-risk' fetus: Doppler ultrasound. *Curr Opin Obstet Gynecol* 2009;**21**:161-6. https://doi.org/10.1097/GCO.0b013e3283292468
9. 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;**175**:1534-42. https://doi.org/10.1016/s0002-9378(96)70103-5
10. 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;**16**:e1002778. https://doi.org/10.1371/journal.pmed.1002778
11. Thorpe-Beeston JG, Banfield PJ, Saunders NJ. Outcome of breech delivery at term. *BMJ (Clinical research ed)* 1992;**305**:746-7. https://doi.org/10.1136/bmj.305.6856.746
12. 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**:e178-e92. https://doi.org/10.1111/1471-0528.14466
13. Hofmeyr GJ, Hannah M, Lawrie TA. Planned caesarean section for term breech delivery. *The Cochrane database of systematic reviews* 2015; 10.1002/14651858.CD000166.pub2:Cd000166. https://doi.org/10.1002/14651858.CD000166.pub2
14. Smith GC, Fretts RC. Stillbirth. *Lancet (London, England)* 2007;**370**:1715-25. https://doi.org/10.1016/s0140-6736(07)61723-1
15. 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;**5**:Cd004945. https://doi.org/10.1002/14651858.CD004945.pub4
16. Smith GC. Life-table analysis of the risk of perinatal death at term and post term in singleton pregnancies. *American journal of obstetrics and gynecology* 2001;**184**:489-96. https://doi.org/10.1067/mob.2001.109735
17. Smith G. The risk of perinatal death at term. *BJOG : an international journal of obstetrics and gynaecology* 2019; 10.1111/1471-0528.15827. https://doi.org/10.1111/1471-0528.15827

18. Smith G. Should we implement universal screening with late pregnancy ultrasound to prevent stillbirth? *BJOG : an international journal of obstetrics and gynaecology* 2018;**125**:101-3. https://doi.org/10.1111/1471-0528.14782
19. Alfirevic Z, Stampalija T, Dowswell T. Fetal and umbilical Doppler ultrasound in high-risk pregnancies. *The Cochrane database of systematic reviews* 2017;**6**:Cd007529. https://doi.org/10.1002/14651858.CD007529.pub4
20. Bricker L, Medley N, Pratt JJ. Routine ultrasound in late pregnancy (after 24 weeks' gestation). *The Cochrane database of systematic reviews* 2015; 10.1002/14651858.CD001451.pub4:Cd001451. https://doi.org/10.1002/14651858.CD001451.pub4
21. 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;**122**:518-27. https://doi.org/10.1111/1471-0528.13148
22. 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;**5**:Cd012245. https://doi.org/10.1002/14651858.CD012245.pub2
23. 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;**125**:1642-54. https://doi.org/10.1111/1471-0528.15394
24. Phillips C, Velji Z, Hanly C, Metcalfe A. Risk of recurrent spontaneous preterm birth: a systematic review and meta-analysis. *BMJ Open* 2017;**7**:e015402. https://doi.org/10.1136/bmjopen-2016-015402
25. Lamont K, Scott NW, Jones GT, Bhattacharya S. Risk of recurrent stillbirth: systematic review and meta-analysis. *BMJ (Clinical research ed)* 2015;**350**:h3080. https://doi.org/10.1136/bmj.h3080
26. Kinzler WL, Kaminsky L. Fetal growth restriction and subsequent pregnancy risks. *Semin Perinatol* 2007;**31**:126-34. https://doi.org/10.1053/j.semperi.2007.03.004
27. 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;**343**:d7400. https://doi.org/10.1136/bmj.d7400
28. Group SCRNW. Association between stillbirth and risk factors known at pregnancy confirmation. *Jama* 2011;**306**:2469-79. https://doi.org/10.1001/jama.2011.1798
29. Haas DM, Parker CB, Wing DA, Parry S, Grobman WA, Mercer BM*, et al.* A description of the methods of the Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b). *American journal of obstetrics and gynecology* 2015;**212**:539.e1-.e24. https://doi.org/10.1016/j.ajog.2015.01.019
30. Revicki DA, Lenderking WR. Methods and issues associated with the use of quality-adjusted life-years. *Expert Rev Pharmacoecon Outcomes Res* 2012;**12**:105-14. https://doi.org/10.1586/erp.11.100
31. Smith GC. Researching new methods of screening for adverse pregnancy outcome: lessons from pre-eclampsia. *PLoS medicine* 2012;**9**:e1001274. https://doi.org/10.1371/journal.pmed.1001274
32. Alfirevic Z, Stampalija T, Medley N. Fetal and umbilical Doppler ultrasound in normal pregnancy. *The Cochrane database of systematic reviews* 2015; 10.1002/14651858.CD001450.pub4:Cd001450. https://doi.org/10.1002/14651858.CD001450.pub4
33. Pasupathy D, Dacey A, Cook E, Charnock-Jones DS, White IR, Smith GC. Study protocol. A prospective cohort study of unselected primiparous women: the pregnancy outcome prediction study. *BMC Pregnancy Childbirth* 2008;**8**:51. https://doi.org/10.1186/1471-2393-8-51

34. 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**:529-36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009
35. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in medicine* 2001;**20**:2865-84.
36. 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;**58**:982-90. https://doi.org/10.1016/j.jclinepi.2005.02.022
37. Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *BMJ (Clinical research ed)* 2001;**323**:157-62.
38. 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;**58**:882-93. https://doi.org/10.1016/j.jclinepi.2005.01.016
39. 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.
40. 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**:165-70.
41. 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**:143-9.
42. 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;**288**:279-85.
43. Fischer RL, Kuhlman KA, Depp R, Wapner RJ. Doppler evaluation of umbilical and uterine-arcuate arteries in the postdates pregnancy. *Obstetrics & Gynecology* 1991;**78**:363-8.
44. 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;**71**:11-9.
45. Hanretty KP, Primrose MH, Neilson JP, Whittle MJ. Pregnancy screening by Doppler uteroplacental and umbilical artery waveforms. *British journal of obstetrics and gynaecology* 1989;**96**:1163-7.
46. 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**:192-6.
47. 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**:557-61.
48. 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;**47**:194-202.
49. 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;**47**:203-9.
50. 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**:93-7.

51. Moraitis AA, Wood AM, Fleming M, Smith GC. Birth weight percentile and the risk of term perinatal death. *Obstetrics and gynecology* 2014;**124**:274-83. https://doi.org/10.1097/aog.0000000000000388
52. 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;**387**:691-702. https://doi.org/10.1016/s0140-6736(15)01020-x
53. Giussani DA. The fetal brain sparing response to hypoxia: physiological mechanisms. *J Physiol* 2016;**594**:1215-30. https://doi.org/10.1113/jp271099
54. 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;**46**:82-92.
55. 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**:409-20.
56. 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**:340-6.
57. 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**:191-8.
58. 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;**32**:2380-6.
59. 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**:54.e1-10.
60. 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**:1880-5.
61. 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; 10.1002/uog.17500. https://doi.org/10.1002/uog.17500
62. 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**:303-10.
63. 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**:124.e1-6.
64. 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**:460-4.
65. 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;**32**:2554-60.
66. Sabdia S, Greer RM, Prior T, Kumar S. Predicting intrapartum fetal compromise using the fetal cerebro-umbilical ratio. *Placenta* 2015;**36**:594-8.

67. 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;**299**:411-20.
68. 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.
69. 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;**54**:68-75. https://doi.org/10.1016/j.placenta.2017.02.006
70. Phelan JP, Ahn MO, Smith CV, Rutherford SE, Anderson E. Amniotic fluid index measurements during pregnancy. *The Journal of reproductive medicine* 1987;**32**:601-4.
71. 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;**121**:686-99. https://doi.org/10.1111/1471-0528.12589
72. 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;**290**:875-81. https://doi.org/10.1007/s00404-014-3292-7
73. 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;**81**:852-5.
74. 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;**15**:85-8. https://doi.org/02.2005/jcpsp.8588
75. 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**:232-6.
76. 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;**269**:130-3. https://doi.org/10.1007/s00404-003-0525-6
77. Megha B, Indu C. Correlation of amniotic fluid index with perinatal outcome. *Journal of Obstetrics and Gynecology of India* 2014;**64**:32-5.
78. 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;**205**:241.e1-6. https://doi.org/10.1016/j.ajog.2011.06.013
79. 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;**110**:989-94.
80. Myles TD, Santolaya-Forgas J. Normal ultrasonic evaluation of amniotic fluid in low-risk patients at term. *The Journal of reproductive medicine* 2002;**47**:621-4.
81. 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.
82. 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**:1319-23.
83. 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;**46**:589-92.
84. 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**:303-7.

85. 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;**111**:220-5.
86. 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**:705-10.
87. Banks EH, Miller DA. Perinatal risks associated with borderline amniotic fluid index. *American Journal of Obstetrics & Gynecology* 1999;**180**:1461-3.
88. Choi SR. Borderline amniotic fluid index and perinatal outcomes in the uncomplicated term pregnancy. *Journal of Maternal-Fetal & Neonatal Medicine* 2016;**29**:457-60.
89. Gumus, II, Koktener A, Turhan NO. Perinatal outcomes of pregnancies with borderline amniotic fluid index. *Archives of gynecology and obstetrics* 2007;**276**:17-9. https://doi.org/10.1007/s00404-006-0309-x
90. Jamal A, Kazemi M, Marsoosi V, Eslamian L. Adverse perinatal outcomes in borderline amniotic fluid index. *International Journal of Reproductive Biomedicine* 2016;**14**:705-8.
91. 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**:545-9.
92. Petrozella LN, Dashe JS, McIntire DD, Leveno KJ. Clinical significance of borderline amniotic fluid index and oligohydramnios in preterm pregnancy. *Obstetrics & Gynecology* 2011;**117**:338-42.
93. 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**:353-6.
94. 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**:3085-8.
95. 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**:705-11.
96. 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;**341**:c7087. https://doi.org/10.1136/bmj.c7087
97. Campbell S, Wilkin D. Ultrasonic measurement of fetal abdomen circumference in the estimation of fetal weight. *British journal of obstetrics and gynaecology* 1975;**82**:689-97. https://doi.org/10.1111/j.1471-0528.1975.tb00708.x
98. 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. https://doi.org/10.1016/s0140-6736(14)61904-8
99. 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**:513-7.
100. 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**:453-8.
101. Benacerraf BR, Gelman R, Frigoletto Jr FD. Sonographically estimated fetal weight: Accuracy and limitation. *American journal of obstetrics and gynecology* 1988;**159**:1118-21.
102. 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**:187-92.
103. 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**:1029-32.

104. 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**:149-51.
105. 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**:77-82.
106. 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**:601-6. https://doi.org/10.1016/j.ajog.2006.04.012
107. 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**:753-6.
108. 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**:434-41.
109. 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**:140-5.
110. 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**:861-6.
111. 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**:324-9.
112. 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**:4-10.
113. 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**:S68. https://doi.org/https://doi.org/10.1016/j.ajog.2016.11.983
114. 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**:936-8.
115. 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**:585-90.
116. 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**:317-22.
117. 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**:339-42.
118. 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**:108-12.
119. 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**:566-71. https://doi.org/10.1002/uog.6378
120. 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**:1345-9.

121. 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**:55-8.
122. 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**:74-81.
123. Miller JM, Jr., Korndorffer FA, 3rd, Gabert HA. Fetal weight estimates in late pregnancy with emphasis on macrosomia. *Journal of Clinical Ultrasound* 1986;**14**:437-42.
124. 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**:1110-4.
125. 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**:13-22.
126. 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**:148-56.
127. 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**:2022-7.
128. 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**:403-8. https://doi.org/10.1046/j.1469-0705.1997.09060403.x
129. Pates JA, McIntire DD, Casey BM, Leveno KJ. Predicting macrosomia. *Journal of Ultrasound in Medicine* 2008;**27**:39-43.
130. 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**:304-9.
131. Pollack RN, Hauer-Pollack G, Divon MY. Macrosomia in postdates pregnancies: the accuracy of routine ultrasonographic screening. *American Journal of Obstetrics & Gynecology* 1992;**167**:7-11.
132. Rossavik IK, Joslin GL. Macrosomatia and ultrasonography: what is the problem? *Southern Medical Journal* 1993;**86**:1129-32.
133. 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.
134. 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**:186-90.
135. 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**:783-91. https://doi.org/10.1002/uog.17491
136. 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**:1058-67.
137. 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**:287-90.
138. 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**:20-4.
139. Ouzounian JG. Shoulder Dystocia: Incidence and Risk Factors. *Clin Obstet Gynecol* 2016;**59**:791-4. https://doi.org/10.1097/grf.0000000000000227

140. 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**:309.e1-6. https://doi.org/10.1016/j.ajog.2012.06.065
141. 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**:340.e1-3. https://doi.org/10.1016/j.ajog.2008.12.043
142. 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;**45**:17-22.
143. Weiner E, Fainstein N, Mizrachi Y, Elyashiv O, Mevorach-Zussman N, Bar J*, et al.* 410: Comparison between three methods for the detection of macrosomia and growth restriction in patients presenting in active labor--a prospective study. *American Journal of Obstetrics & Gynecology* 2016;**214**:S225-S6. https://doi.org/10.1016/j.ajog.2015.10.451
144. Gordijn SJ, Beune IM, Thilaganathan B, Papageorghiou A, Baschat AA, Baker PN*, et al.* Consensus definition of fetal growth restriction: a Delphi procedure. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* 2016;**48**:333-9. https://doi.org/10.1002/uog.15884
145. 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;**2**:569-81. https://doi.org/10.1016/s2352-4642(18)30129-9
146. . *The Investigation and Management of the Small–for–Gestational–Age Fetus (Green-top guideline No. 31)*; 2013.
147. Bond DM, Gordon A, Hyett J, de Vries B, Carberry AE, Morris J. Planned early delivery versus expectant management of the term suspected compromised baby for improving outcomes. *The Cochrane database of systematic reviews* 2015; 10.1002/14651858.CD009433.pub2:Cd009433. https://doi.org/10.1002/14651858.CD009433.pub2
148. Grobman WA, Rice MM, Reddy UM, Tita ATN, Silver RM, Mallett G*, et al.* Labor Induction versus Expectant Management in Low-Risk Nulliparous Women. *N Engl J Med* 2018;**379**:513-23. https://doi.org/10.1056/NEJMoa1800566
149. 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; 10.1002/14651858.CD000938.pub2:Cd000938. https://doi.org/10.1002/14651858.CD000938.pub2
150. 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* 2019; 10.1111/1471-0528.15809. https://doi.org/10.1111/1471-0528.15809
151. Impey L, Murphy D, Griffiths M, Penna L, on behalf of the Royal College of Obstetricians and Gynaecologists. Management of Breech Presentation. *BJOG: An International Journal of Obstetrics & Gynaecology* 2017;**124**:e151-e77.
152. 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.
153. York Health Economics Consortium. *Quality-Adjusted Life Year (QALY)*. York: York Health Economics Consortium,; 2016. URL: http://www.yhec.co.uk/glossary/quality-adjusted-life-year-qaly/ (accessed 31-07-2018).
154. R Core Team. R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing; 2017.
155. Baio G, Berardi A, Heath A. BCEA: Bayesian Cost Effectiveness Analysis. In; 2018.
156. Fan FY. FinCal: Time Value of Money, Time Series Analysis and Computational Finance. In: R package version 0.6.3.; 2016.
157. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*: Springer-Verlag New York; 2016.
158. Warnes G, Bolker B, Lumley T. gtools: Various R Programming Tools. In; 2018.
159. Wickham H, Bryan J. readxl: Read Excel Files. In; 2018.

160. Wickham H, Henry L. tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions. In; 2019.
161. Strong M. SAVI: SAVI Sheffield Accelerated Value of Information. In; 2015.
162. Stock SJ, Ferguson E, Duffy A, Ford I, Chalmers J, Norman JE. Outcomes of elective induction of labour compared with expectant management: population based study. *BMJ (Clinical research ed)* 2012;**344**:e2838. https://doi.org/10.1136/bmj.e2838
163. Leung WC, Pun TC, Wong WM. Undiagnosed breech revisited. *British journal of obstetrics and gynaecology* 1999;**106**:638-41. https://doi.org/10.1111/j.1471-0528.1999.tb08360.x
164. 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**:157.e1-3. https://doi.org/10.1016/j.ajog.2006.10.889
165. 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;**102**:101-6. https://doi.org/10.1111/j.1471-0528.1995.tb09060.x
166. Ouzounian JG, Gherman RB. Shoulder dystocia: are historic risk factors reliable predictors? *American journal of obstetrics and gynecology* 2005;**192**:1933-5; discussion 5-8. https://doi.org/10.1016/j.ajog.2005.02.054
167. 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. https://doi.org/10.1097/01.ogx.0000435370.74455.a8
168. Chongsuvivatwong V, Bachtiar H, Chowdhury ME, Fernando S, Suwanrath C, Kor-Anantakul O*, et al.* Maternal and fetal mortality and complications associated with cesarean section deliveries in teaching hospitals in Asia. *J Obstet Gynaecol Res* 2010;**36**:45-51. https://doi.org/10.1111/j.1447-0756.2009.01100.x
169. Gibson KS, Waters TP, Bailit JL. Maternal and neonatal outcomes in electively induced low-risk term pregnancies. *American journal of obstetrics and gynecology* 2014;**211**:249.e1-.e16. https://doi.org/10.1016/j.ajog.2014.03.016
170. MacKenzie IZ, Shah M, Lean K, Dutton S, Newdick H, Tucker DE. Management of shoulder dystocia: trends in incidence and maternal and neonatal morbidity. *Obstetrics and gynecology* 2007;**110**:1059-68. https://doi.org/10.1097/01.AOG.0000287615.35425.5c
171. 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. *American journal of obstetrics and gynecology* 1996;**174**:1557-64. https://doi.org/10.1016/s0002-9378(96)70606-3
172. Thorngren-Jerneck K, Herbst A. Low 5-minute Apgar score: a population-based register study of 1 million term births. *Obstetrics and gynecology* 2001;**98**:65-70.
173. Pasupathy D, Wood AM, Pell JP, Fleming M, Smith GC. Time trend in the risk of delivery-related perinatal and neonatal death associated with breech presentation at term. *Int J Epidemiol* 2009;**38**:490-8. https://doi.org/10.1093/ije/dyn225
174. 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;**7**:e1000289. https://doi.org/10.1371/journal.pmed.1000289
175. Persson M, Razaz N, Tedroff K, Joseph KS, Cnattingius S. 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. https://doi.org/10.1136/bmj.k207
176. Iliodromiti S, Mackay DF, Smith GC, Pell JP, Nelson SM. Apgar score and the risk of cause-specific infant mortality: a population-based cohort study. *Lancet (London, England)* 2014;**384**:1749-55. https://doi.org/10.1016/s0140-6736(14)61135-1
177. NHS Improvement. National Schedule of Reference Costs, 2016-17 - NHS trusts and NHS foundation trusts. In; 2017.
178. Curtis L, Burns A. *Unit Costs of Health and Social Care 2017* Personal Social Services Research Unit; 2017.

179. Vijgen SM, Boers KE, Opmeer BC, Bijlenga D, Bekedam DJ, Bloemenkamp KW*, 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**:358-63. https://doi.org/10.1016/j.ejogrb.2013.07.017
180. 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**:1109-13. https://doi.org/10.1503/cmaj.050796
181. James M, Hunt K, Burr R, Johanson R. A decision analytical cost analysis of offering ECV in a UK district general hospital. *BMC Health Serv Res* 2001;**1**:6.
182. Alfirevic Z, Keeney E, Dowswell T, Welton NJ, Medley N, Dias S*, 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**:1-584. https://doi.org/10.3310/hta20650
183. 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. *Int Urogynecol J Pelvic Floor Dysfunct* 2005;**16**:19-28; discussion https://doi.org/10.1007/s00192-004-1203-3
184. Mistry H, Heazell AE, Vincent O, Roberts T. 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. https://doi.org/10.1186/1471-2393-13-236
185. Barrett B, Mosweu I, Jones CR, Charman T, Baird G, Simonoff E*, et al.* Comparing service use and costs among adolescents with autism spectrum disorders, special needs and typical development. *Autism* 2015;**19**:562-9. https://doi.org/10.1177/1362361314536626
186. Access Economics. *The Economic Impact of Cerebral Palsy in Australia in 2007*. Sydney, NSW; 2008.
187. Szende A, Janssen B, Cabasés J. *Self-Reported Population Health: An International Perspective based on EQ-5D* Springer Dordrecht Heidelberg New York London; 2014.
188. Office for National Statistics. National Life Tables, United Kingdom, 1980-82 to 2014-16. In: Office for National Statistics; 2017.
189. Leigh S, Granby P, Turner M, Wieteska S, Haycox A, Collins B. The incidence and implications of cerebral palsy following potentially avoidable obstetric complications: a preliminary burden of disease study. *BJOG : an international journal of obstetrics and gynaecology* 2014;**121**:1720-8. https://doi.org/10.1111/1471-0528.12897
190. . *Guide to the methods of technology appraisal 2013*; 2013.
191. Wilson EC. A practical guide to value of information analysis. *Pharmacoeconomics* 2015;**33**:105-21. https://doi.org/10.1007/s40273-014-0219-x
192. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999;**18**:341-64.
193. Pratt J, Raiffa H, Schlaifer R. *Introduction to Statistical Decision Theory*. Cambridge, MA: Massachusetts Institute of Technology; 1995.
194. Heath A, Baio G. Calculating the Expected Value of Sample Information Using Efficient Nested Monte Carlo: A Tutorial. *Value Health* 2018;**21**:1299-304. https://doi.org/10.1016/j.jval.2018.05.004
195. Walker KF, Dritsaki M, Bugg G, Macpherson M, McCormick C, Grace N*, et al.* Labour induction near term for women aged 35 or over: an economic evaluation. *BJOG : an international journal of obstetrics and gynaecology* 2017;**124**:929-34. https://doi.org/10.1111/1471-0528.14557
196. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--2. *Value Health* 2012;**15**:804-11. https://doi.org/10.1016/j.jval.2012.06.016
197. Heazell AE, Siassakos D, Blencowe H, Burden C, Bhutta ZA, Cacciatore J*, et al.* Stillbirths: economic and psychosocial consequences. *Lancet (London, England)* 2016;**387**:604-16. https://doi.org/10.1016/s0140-6736(15)00836-3

198. Heazell AE, Whitworth MK, Whitcombe J, Glover SW, Bevan C, Brewin J*, et al.* Research priorities for stillbirth: process overview and results from UK Stillbirth Priority Setting Partnership. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* 2015;**46**:641-7. https://doi.org/10.1002/uog.15738
199. Eskes M, Ensing S, Groenendaal F, Abu-Hanna A, Ravelli A. The risk of intrapartum/neonatal mortality and morbidity following birth at 37 weeks of gestation: a nationwide cohort study. *BJOG : an international journal of obstetrics and gynaecology* 2019; 10.1111/1471-0528.15748. https://doi.org/10.1111/1471-0528.15748
200. Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of Breast Cancer Screening: Systematic Review and Meta-analysis to Update the 2009 U.S. Preventive Services Task Force Recommendation. *Annals of internal medicine* 2016;**164**:244-55. https://doi.org/10.7326/m15-0969
201. Norman JE, Heazell AEP, Rodriguez A, Weir CJ, Stock SJE, Calderwood CJ*, et al.* Awareness of fetal movements and care package to reduce fetal mortality (AFFIRM): a stepped wedge, cluster-randomised trial. *Lancet (London, England)* 2018;**392**:1629-38. https://doi.org/10.1016/s0140-6736(18)31543-5
202. Health and Social Care Information Centre. NHS Maternity Statistics 2016-17. In: NHS Digital; 2017.
203. ACOG Practice Bulletin No. 173: Fetal Macrosomia. *Obstetrics and gynecology* 2016;**128**:e195-e209. https://doi.org/10.1097/aog.0000000000001767
204. 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. https://doi.org/10.1002/14651858.CD011739.pub2
205. Benner JS, Morrison MR, Karnes EK, Kocot SL, McClellan M. An evaluation of recent federal spending on comparative effectiveness research: priorities, gaps, and next steps. *Health Aff (Millwood)* 2010;**29**:1768-76. https://doi.org/10.1377/hlthaff.2010.0687
206. . *Green-top Guideline No. 42: Shoulder Dystocia*; 2012.
207. 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;**356**:1375-83. https://doi.org/10.1016/s0140-6736(00)02840-3
208. Gherman RB, Ouzounian JG, Miller DA, Kwok L, Goodwin TM. Spontaneous vaginal delivery: a risk factor for Erb's palsy? *American journal of obstetrics and gynecology* 1998;**178**:423-7. https://doi.org/10.1016/s0002-9378(98)70413-2
209. NHS Digital. NHS Staff Earnings - Estimates ot September 2017, Provisional Statistics. In; 2017.
210. NHS Purchasing and Supply Agency. Cost-effectiveness of ultrasound elastography in the assessment of liver fibrosis. In; 2009.
211. Curtis L. *Unit Costs of health and Social Care 2008*. Cornwallis Building, The University of Kent, Canterbury, Kent CT2 7NF: Personal Social Services Research Unit; 2008.
212. Curtis L, Burns A. Unit costs of health and social care 2016. In: Personal Social Services Research Unit; 2016.
213. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol* 1997;**39**:214-23.
214. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199-208.
215. Young NL, Rochon TG, McCormick A, Law M, Wedge JH, Fehlings D. The health and quality of life outcomes among youth and young adults with cerebral palsy. *Arch Phys Med Rehabil* 2010;**91**:143-8. https://doi.org/10.1016/j.apmr.2009.08.152

Appendix 1. 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

Figure 25. POP study inclusion flowchart.

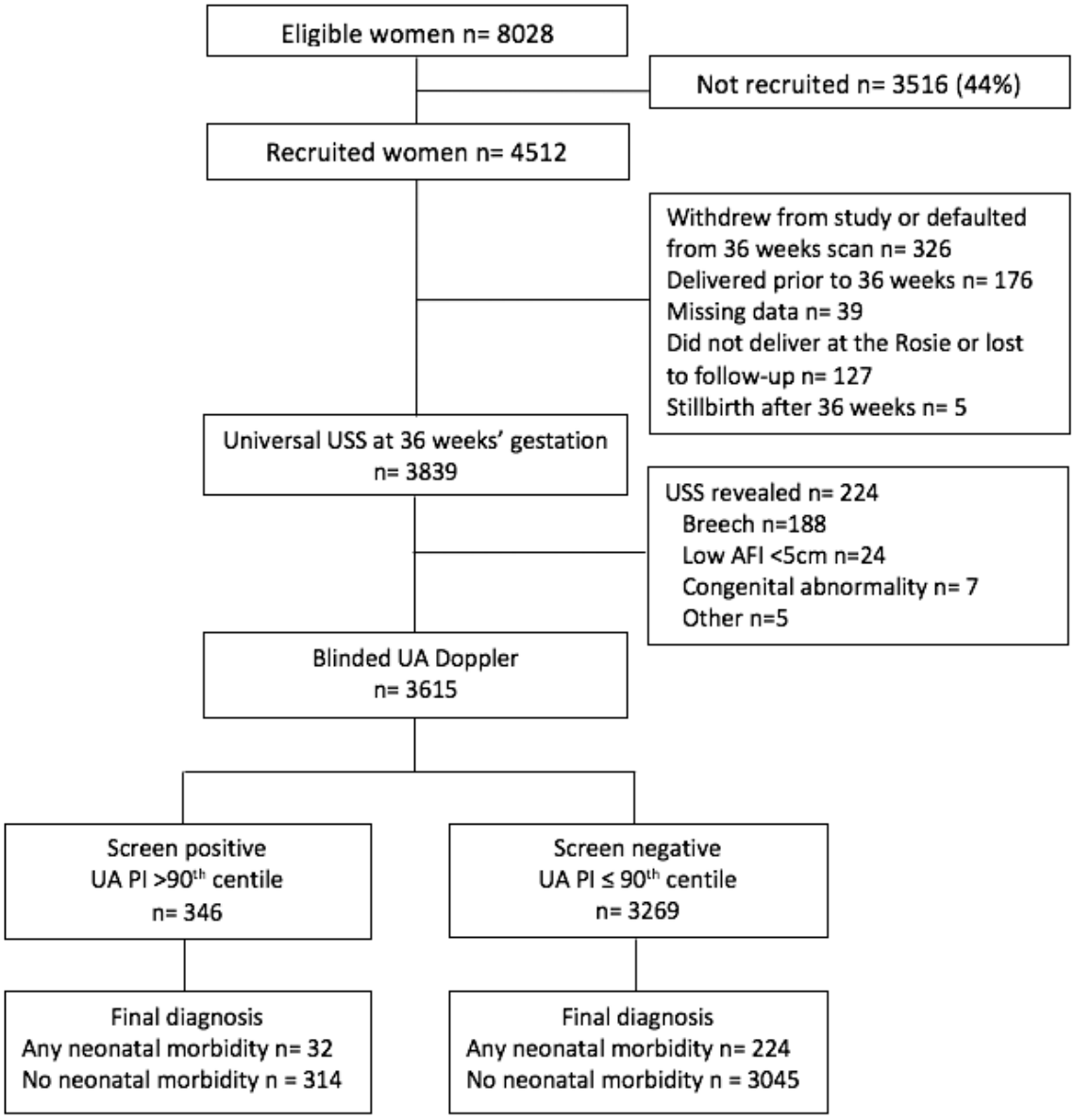


Table 18. Maternal characteristics and birth outcomes of POP study.

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)

Figure 26. Literature search PRISMA flow diagram for the systematic review on umbilical artery Doppler.

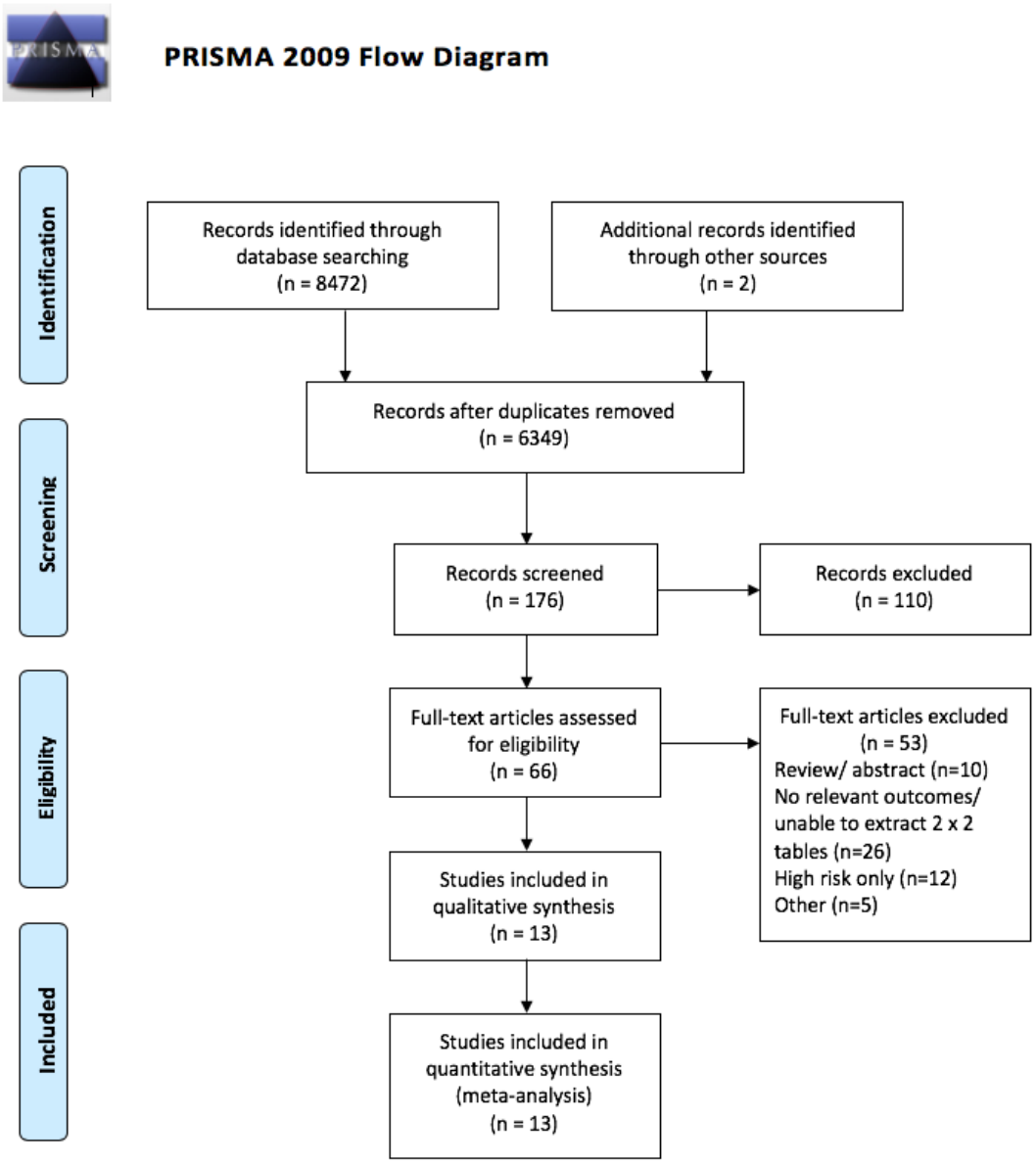
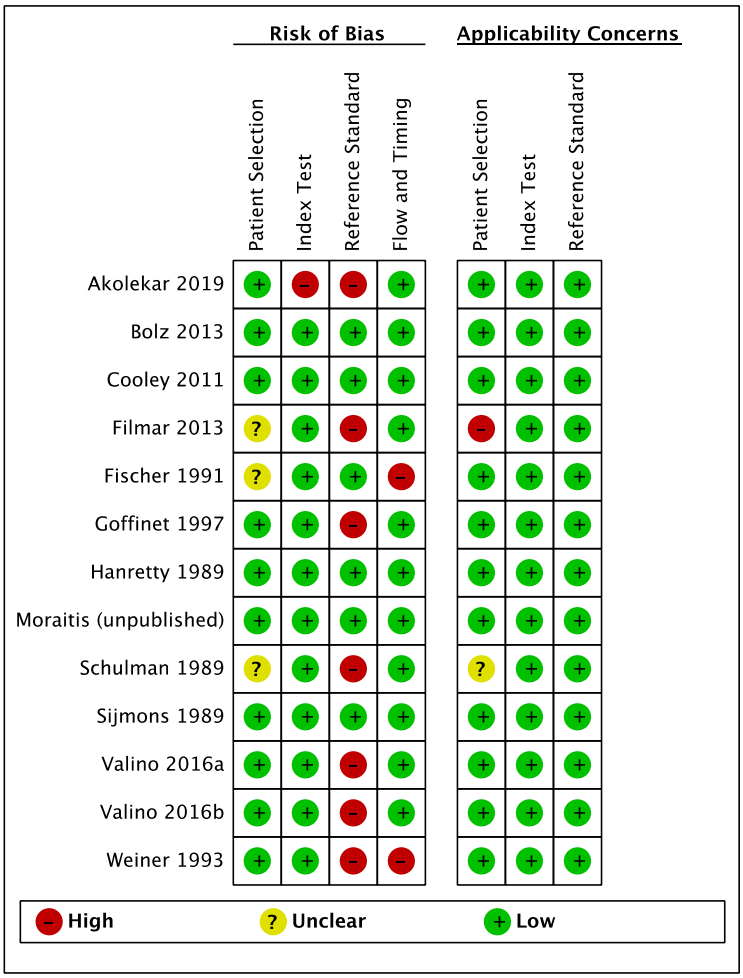


Figure 27. Risk of bias and applicability concerns using the QUADAS-2 tool for the studies included in the meta-analysis of umbilical artery Doppler.



39. 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.

40. 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;41:165-70.

41. 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;39:143-9.

42. 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;288:279-85.

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

44. 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;71:11-9.

45. Hanretty KP, Primrose MH, Neilson JP, Whittle MJ. Pregnancy screening by Doppler uteroplacental and umbilical artery waveforms. British journal of obstetrics and gynaecology 1989;96:1163-7.
46. 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;160:192-6.
47. 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;96:557-61.
48. 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;47:194-202.
49. 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;47:203-9.
50. 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;48:93-7.

Table 19. Characteristics of studies included in the meta-analysis.

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
Akolekar 2019[39]	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.	Adverse perinatal outcome (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[40]	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.

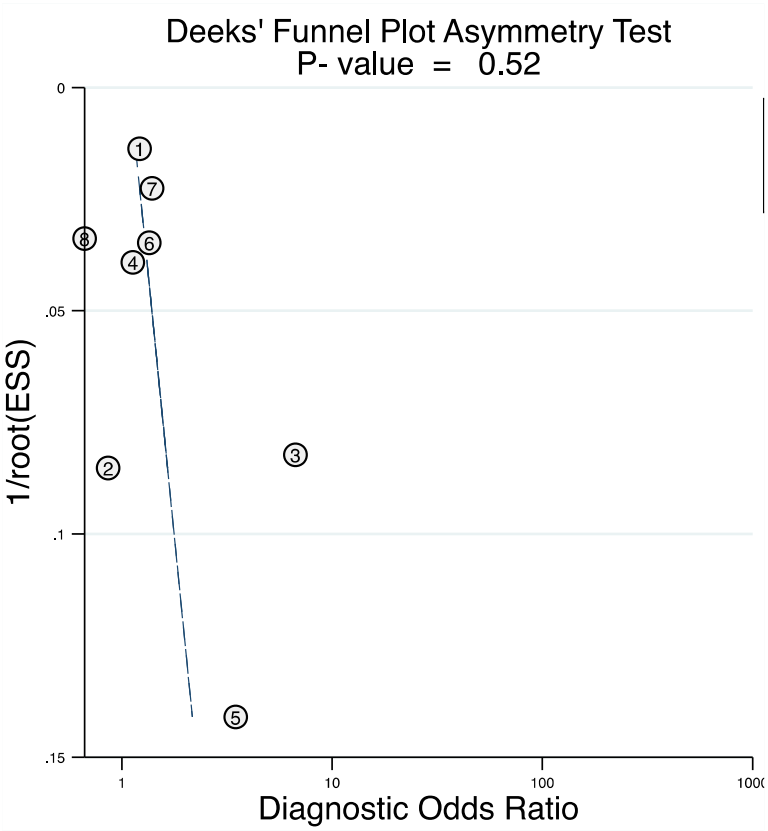
Cooley 2011[41]	Prospective cohort, Single Hospital, Ireland	N=810 Mixed risk, nulliparous only. Only included Caucasian aged 18-40 years.	PI>95th centile UA blinded but EFW not blinded.	Around 36 weeks (not specified)	Emergency CS, PIH, PET, preterm delivery (<37 weeks), SGA <10th centile, SGA <3rd centile, 5-minute Apgar score <7, Cord arterial PH <7.10, NICU admission, Stillbirth	Not reported	Nulliparity: All IOL: 22.4%.
Filmar 2013[42]	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 ga 34.4 for control group.	NICU admission, 5-minute Apgar score <7	Median ga 37 weeks for abnormal UA group, 39 weeks for control group.	Nulliparity: Not reported IOL: Not reported.
Fischer 1991[43]	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)	Mean ga at delivery 292.2 days	Nulliparity: 57% IOL: Not reported

					birthweight <10th centile.		
Goffinet 1996[44]	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[45]	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 (POPS)	Prospective cohort, Single Hospital, Cambridge, UK	N=3615 Universal, nulliparous only, >36 weeks	PI >90th centile Blinded.	Mean 36 weeks	NICU admission, metabolic acidosis, 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[46]	Prospective cohort,	N=255 Mixed	S/D ratio >3 Not blinded.	Around 30 weeks	SGA <15th centile	Not reported	Nulliparity: Not reported

	Single Hospital, NY, USA						IOL: Not reported.
Sijmons 1989[47]	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[48]	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[48]	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[50]	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	Mean 41.8 weeks	Nulliparous: n=43 IOL: Not reported.

					fetal distress, SGA <5th centile		

Figure 28. Deeks' funnel plot for publication bias for umbilical artery doppler for the prediction of neonatal unit admission.



Appendix 2. 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

Figure 29. Literature search PRISMA flow diagram for the systematic review on cerebro-placental ratio.

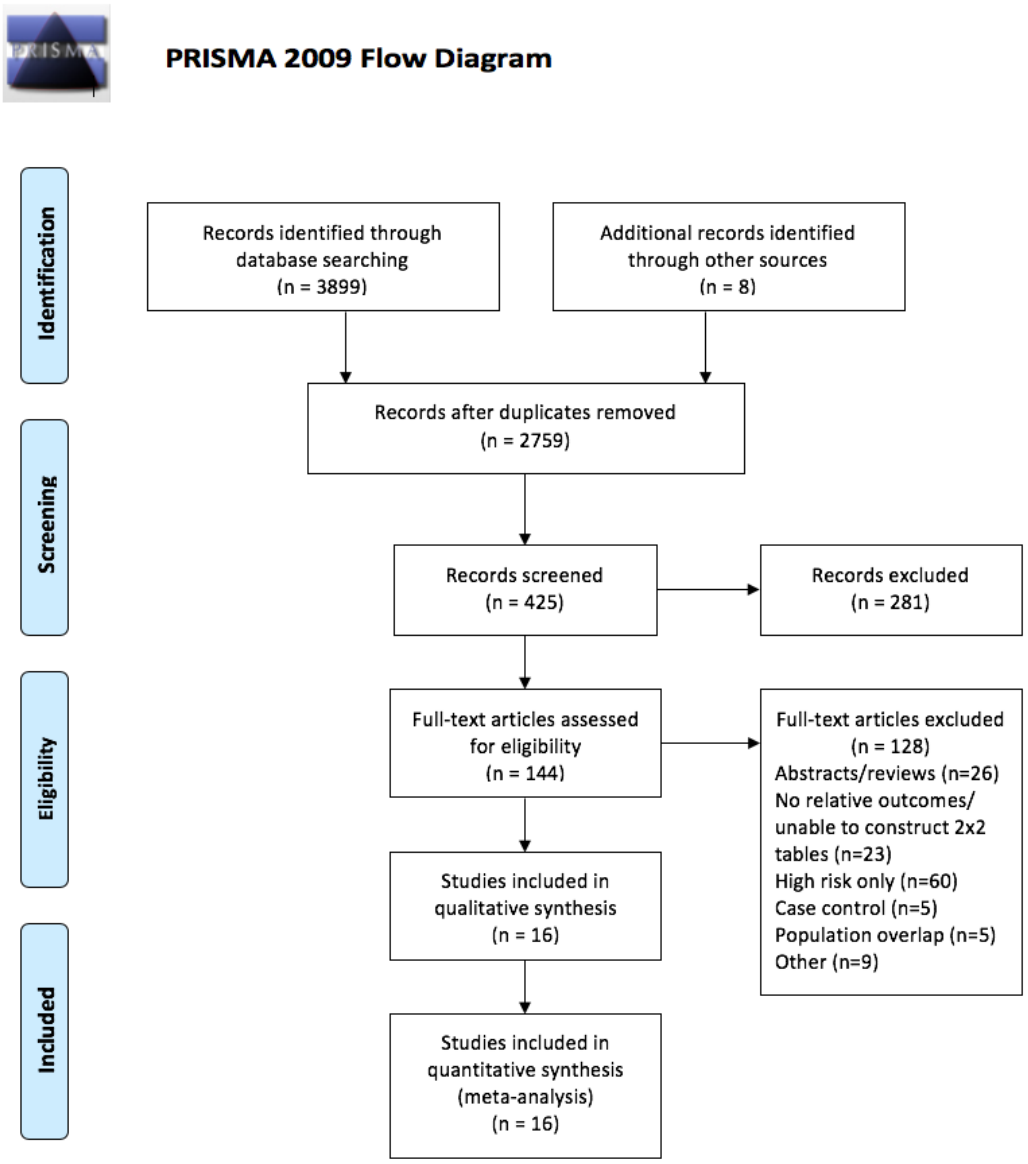
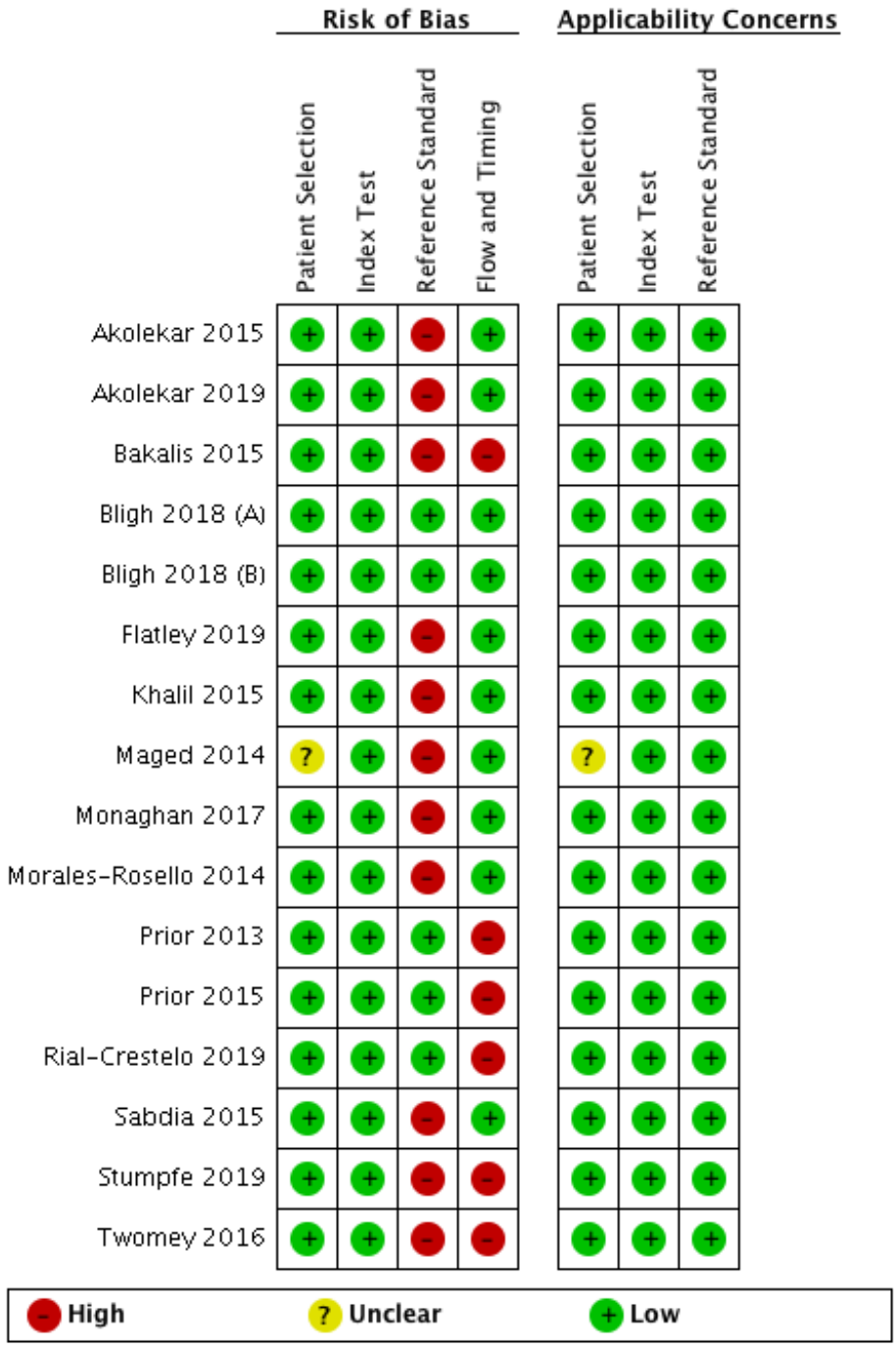


Figure 30. Risk of bias and applicability concerns using the QUADAS-2 tool for the studies included in the meta-analysis of cerebro-placental ratio.



54. 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;**46**:82-92.

39. 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.

55. 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**:409-20.
56. 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**:340-6.
57. 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**:191-8.
58. 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;**32**:2380-6.
59. 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**:54.e1-10.
60. 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**:1880-5.
61. 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; 10.1002/uog.17500. https://doi.org/10.1002/uog.17500
62. 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**:303-10.
63. 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**:124.e1-6.
64. 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**:460-4.
65. 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;**32**:2554-60.
66. Sabdia S, Greer RM, Prior T, Kumar S. Predicting intrapartum fetal compromise using the fetal cerebro-umbilical ratio. *Placenta* 2015;**36**:594-8.
67. 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;**299**:411-20.
68. 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.

Table 20. 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[54]	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[39]	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[55]	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)[56]	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.

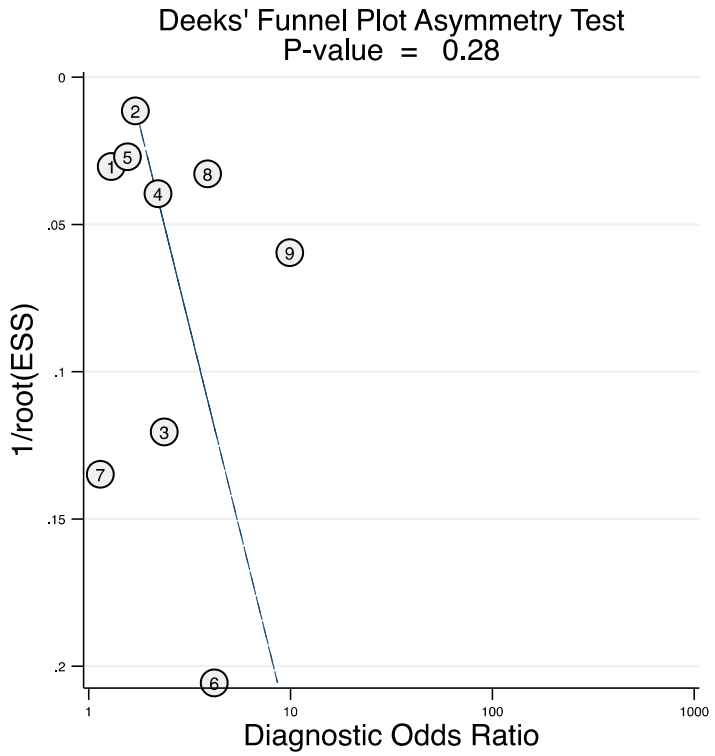
Bligh 2018 (B)[57]	Prospective cohort, 1 hospital, Brisbane, Australia (May 2014 – August 2016)	N= 437 Low risk Uncomplicated, term only.	CPR <10th centile CPR <5th centile Blinded.	From 36 Within 2 weeks of delivery	SGA <10th centile SGA <5th centile	Median 40 (IQR 39.3-40.9)	Nulliparous: 87.4% IOL: Not reported.
Flatley 2019[58]	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 2015[59]	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[60]	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[61]	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[62]	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[63]	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[64]	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[65]	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[66]	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[67]	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[68]	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 31. Deeks' funnel plot for publication bias for cerebroplacental ratio for the prediction of neonatal unit admission.



Appendix 3. 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 05/06/2019)

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

Figure 32. PRISMA flow diagram for the systematic review of severe oligohydramnios.

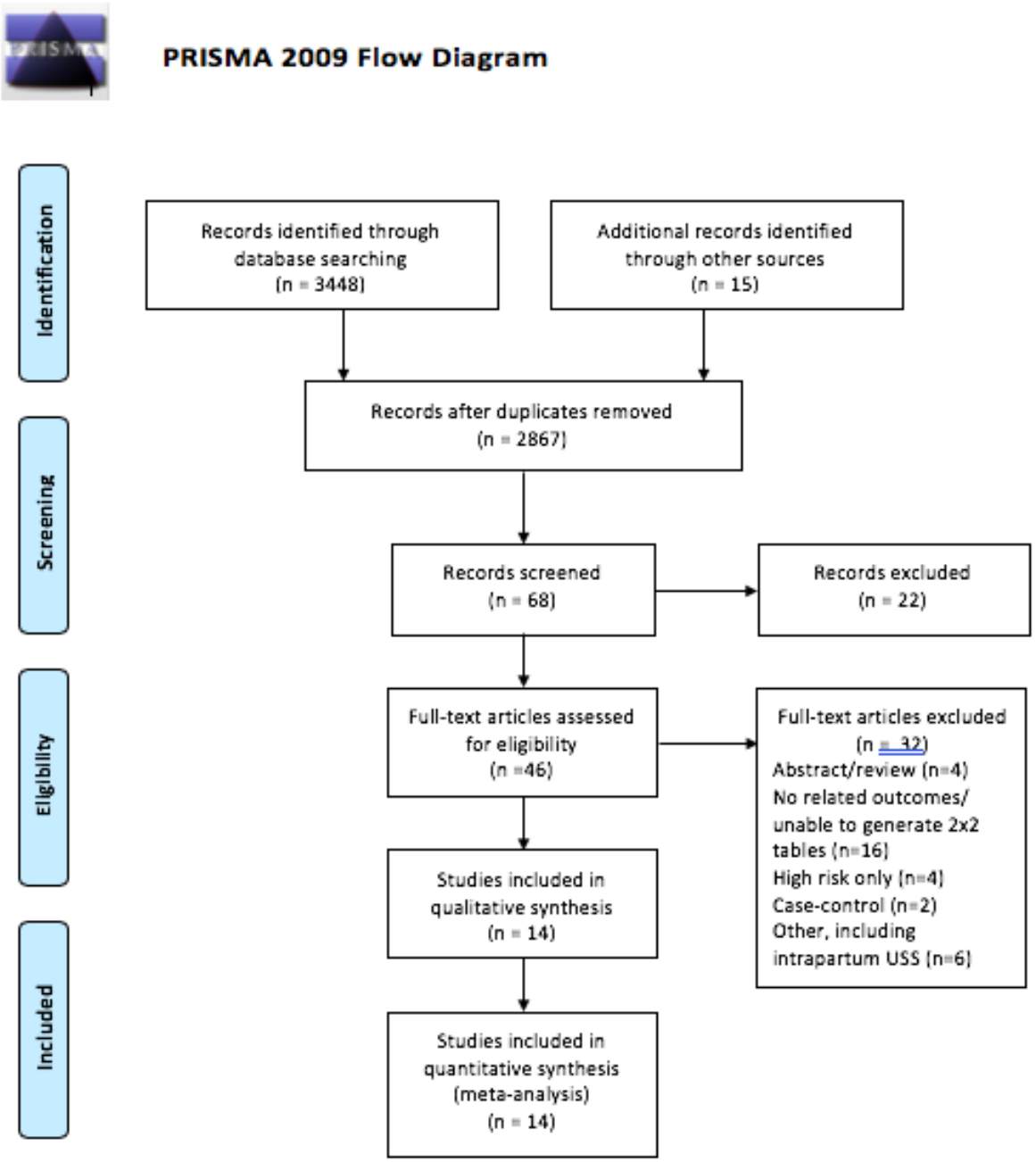


Figure 33. Risk of bias graph of included studies for systematic review of severe oligohydramnios using the QUADAS-2 tool.

72. 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;290:875-81. https://doi.org/10.1007/s00404-014-3292-7

73. 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;81:852-5.

74. 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;15:85-8. https://doi.org/02.2005/jcpsp.8588

75. 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;45:232-6.

76. 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;269:130-3. https://doi.org/10.1007/s00404-003-0525-6

77. Megha B, Indu C. Correlation of amniotic fluid index with perinatal outcome. Journal of Obstetrics and Gynecology of India 2014;64:32-5.
78. 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;205:241.e1-6. https://doi.org/10.1016/j.ajog.2011.06.013
79. 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;110:989-94.
80. Myles TD, Santolaya-Forgas J. Normal ultrasonic evaluation of amniotic fluid in low-risk patients at term. The Journal of reproductive medicine 2002;47:621-4.
81. 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.
82. 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;25:1319-23.
83. 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;46:589-92.
84. 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;30:303-7.
85. 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;111:220-5.

Table 21. Characteristics of studies included in the meta-analysis of 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[72]	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[73]	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[74]	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.

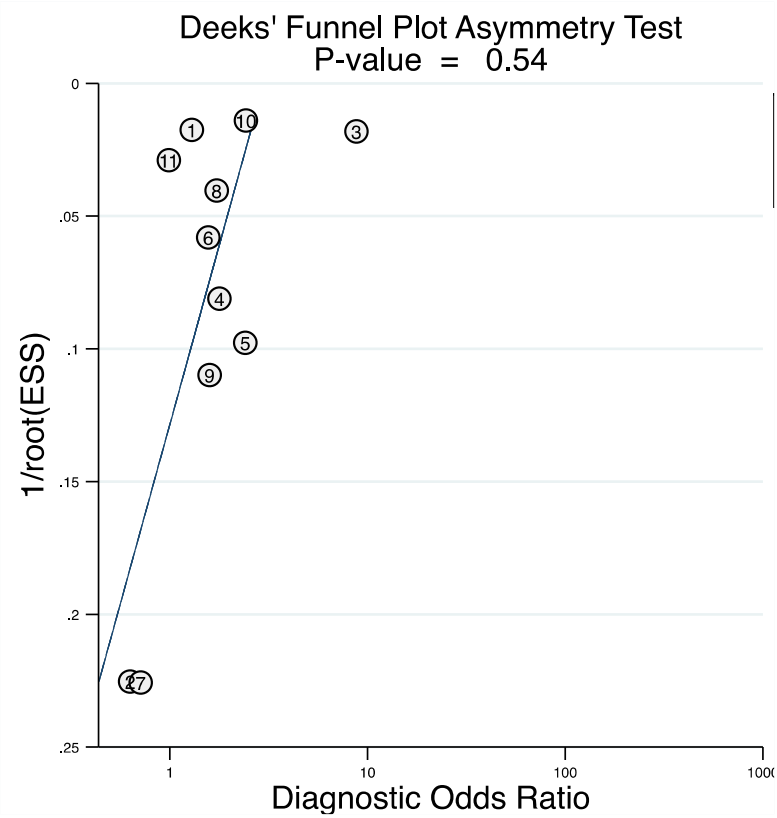
	Single hospital, Pakistan						
Hsieh 1998[75]	Retrospective cohort, Single hospital Taiwan	N=27,506 Universal Excluded those with AFI>24cm, PPROM.	AFI <5cm Not blinded	Not specified	Stillbirth, SGA <10th centile, 5-min Apgar <7, NICU admission, Neonatal death.	Not specified	Nulliparous: Not specified IOL: Not specified.
Locatelli 2004[76]	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[77]	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[78]	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[79]	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[80]	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[81]	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[82]	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[83]	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[84]	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[85]	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 34. Deeks' funnel plot for publication bias for severe oligohydramnios for the prediction of neonatal unit admission.



Appendix 4. 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 18/06/2019)

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

Figure 35. POPS study inclusion flowchart.

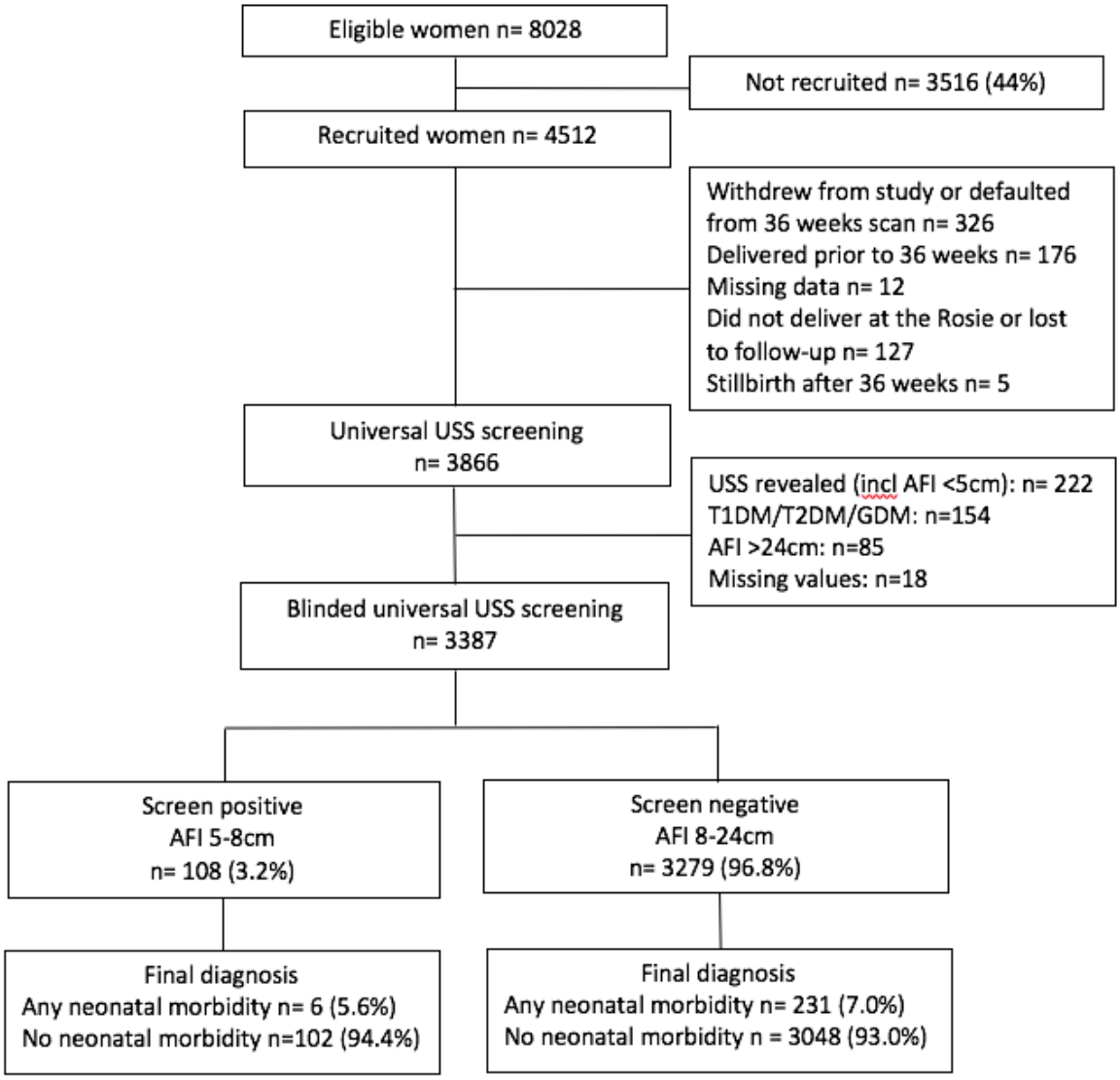


Table 22. Patient characteristics and birth outcomes of POP study.

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 cesarean	13 (12.0)	596 (18.2)		609 (18.0)
	Pre-labor cesarean	6 (5.6)	157 (4.8)		163 (4.8)
	Missing	0 (0.0)	9 (0.3)		9 (0.3)

Figure 36. PRISMA flow diagram for the systematic review of borderline oligohydramnios.

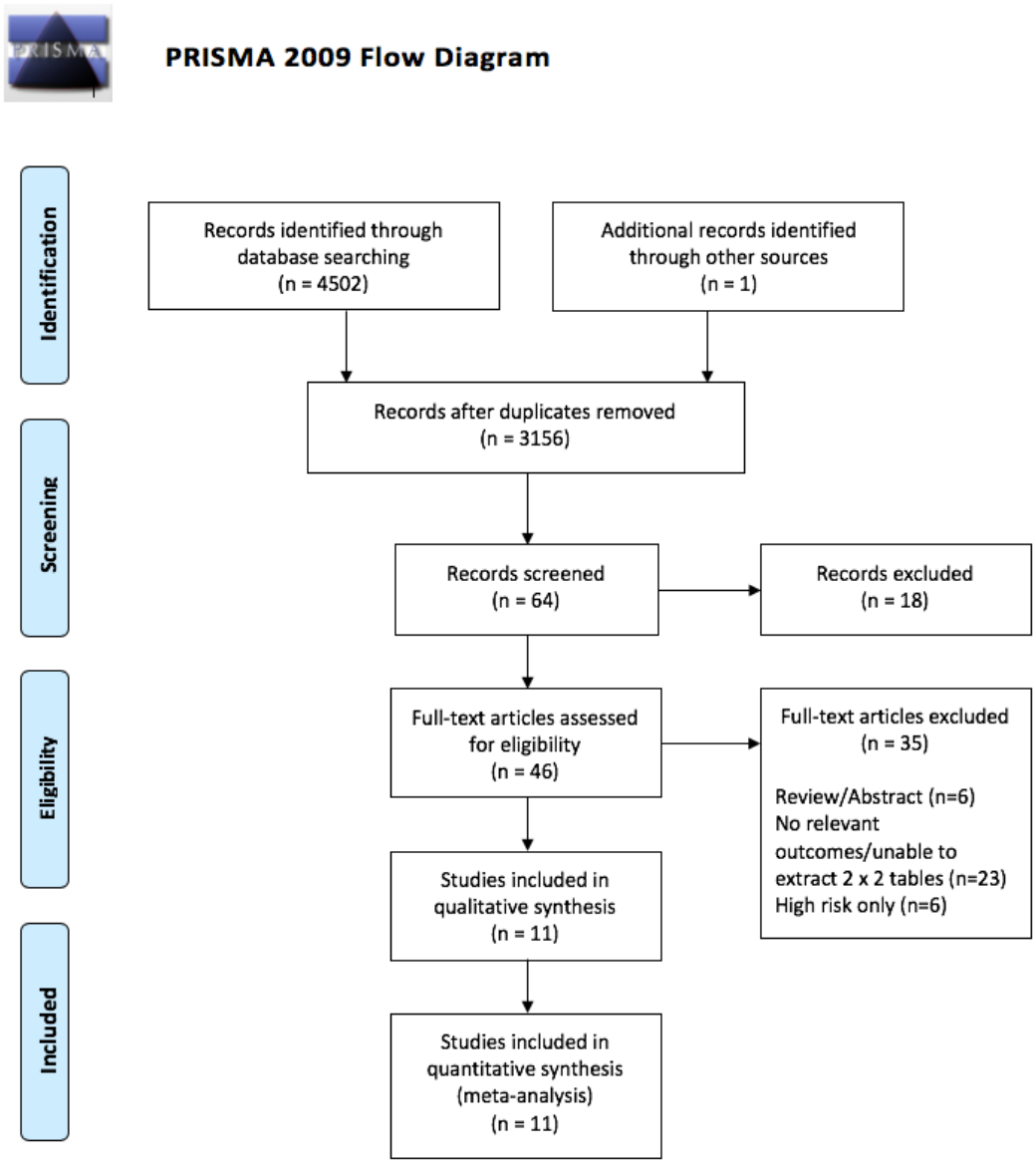
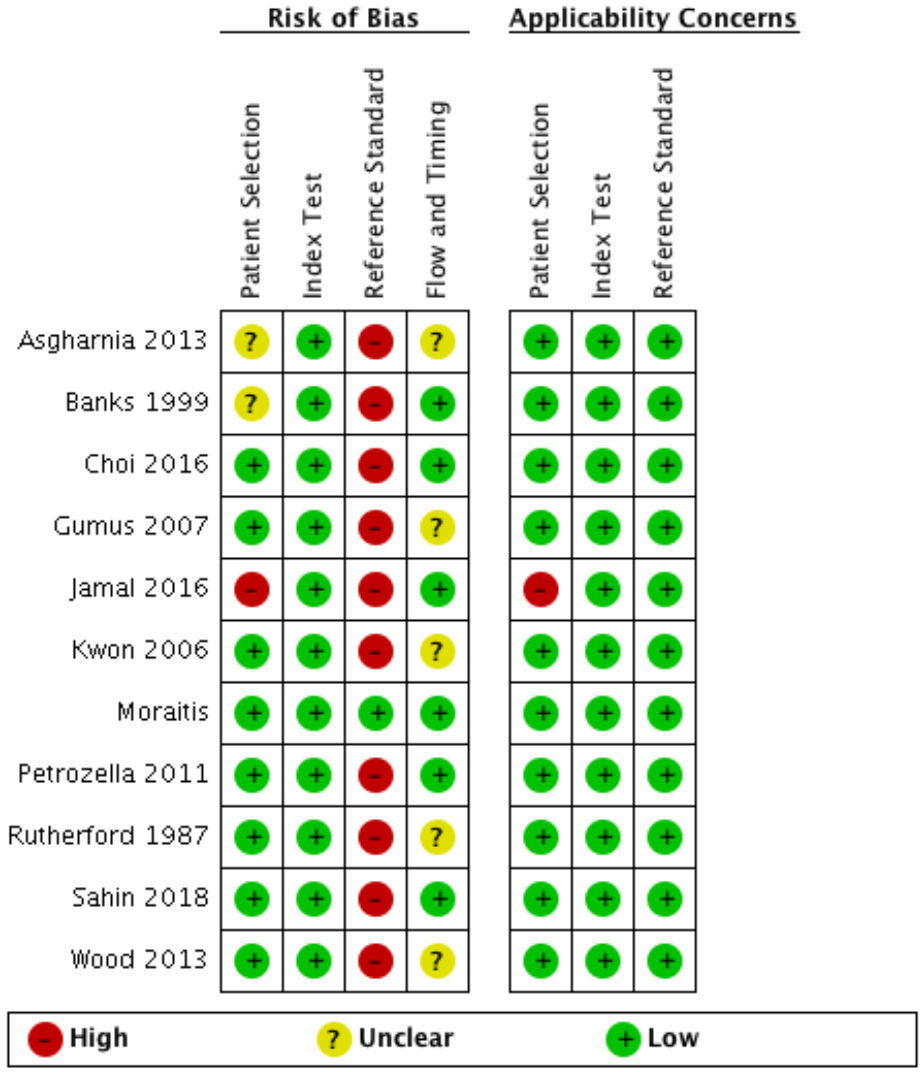


Figure 37. Risk of bias graph of included studies for systematic review of borderline oligohydramnios using the QUADAS-2 tool.



86. 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;11:705-10.
87. Banks EH, Miller DA. Perinatal risks associated with borderline amniotic fluid index. American Journal of Obstetrics & Gynecology 1999;180:1461-3.
88. Choi SR. Borderline amniotic fluid index and perinatal outcomes in the uncomplicated term pregnancy. Journal of Maternal-Fetal & Neonatal Medicine 2016;29:457-60.
89. Gumus, II, Koktener A, Turhan NO. Perinatal outcomes of pregnancies with borderline amniotic fluid index. Archives of gynecology and obstetrics 2007;276:17-9. https://doi.org/10.1007/s00404-006-0309-x
90. Jamal A, Kazemi M, Marsoosi V, Eslamian L. Adverse perinatal outcomes in borderline amniotic fluid index. International Journal of Reproductive Biomedicine 2016;14:705-8.
91. 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;32:545-9.

92. Petrozella LN, Dashe JS, McIntire DD, Leveno KJ. Clinical significance of borderline amniotic fluid index and oligohydramnios in preterm pregnancy. Obstetrics & Gynecology 2011;117:338-42.
93. 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;70:353-6.
94. 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;31:3085-8.
95. 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;33:705-11.

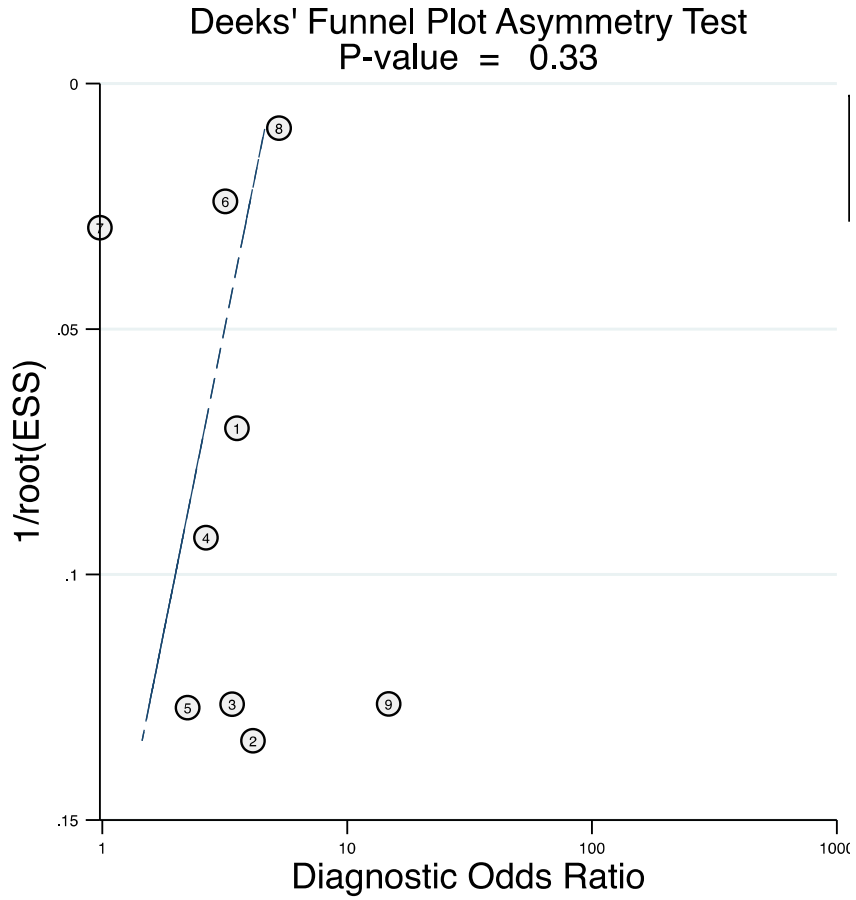
Table 23. Characteristics of the studies included in the meta-analysis of borderline oligohydramnios.

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[86]	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[87]	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[88]	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[89]	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[90]	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[91]	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[92]	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[93]	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[94]	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[95]	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 38. Deeks' funnel plot for publication bias for borderline oligohydramnios for the prediction of SGA <10th centile.



Appendix 5. 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 the 22/10/2018)

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

Figure 39. PRISMA flow diagram for the systematic review of macrosomia.

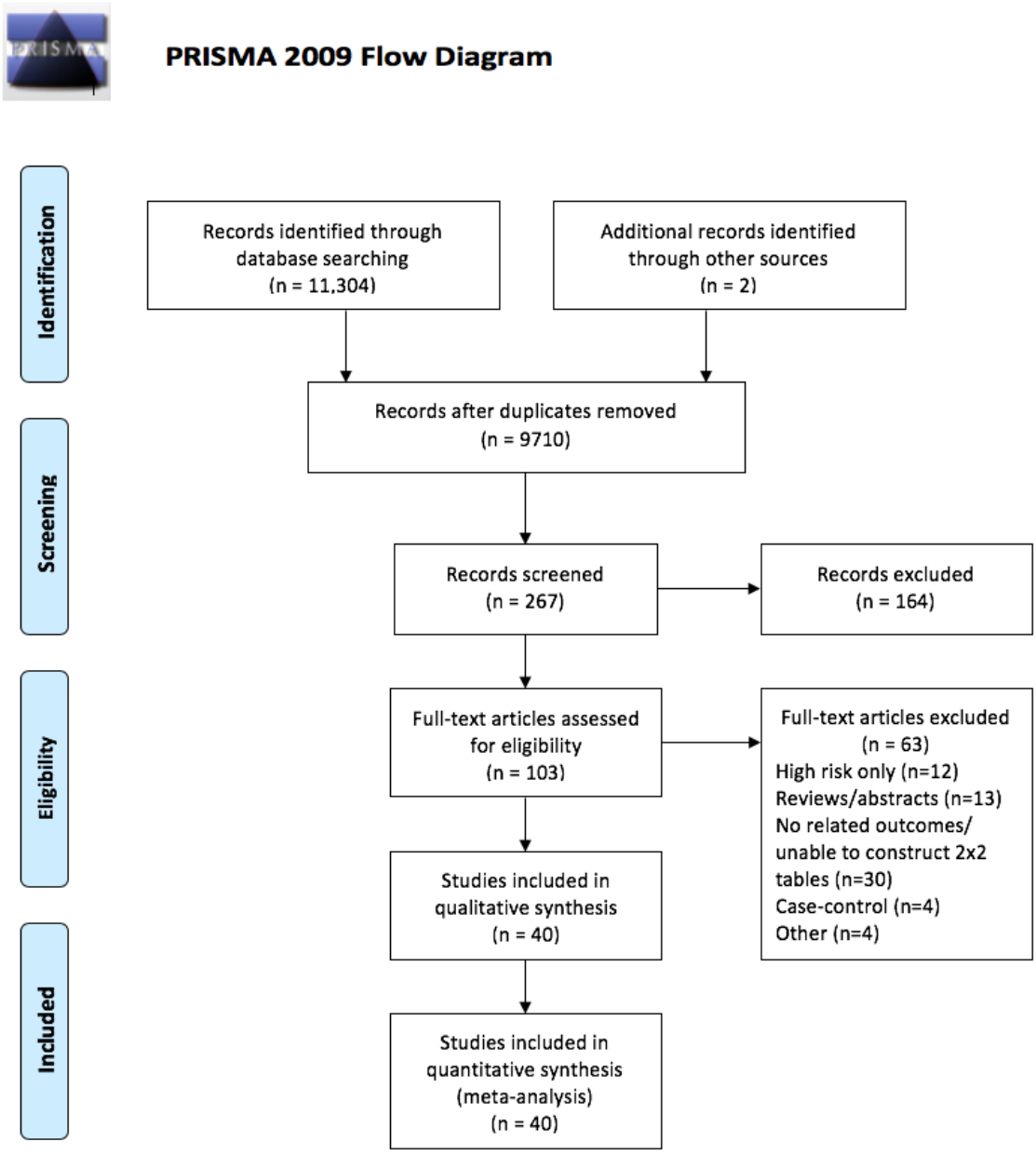
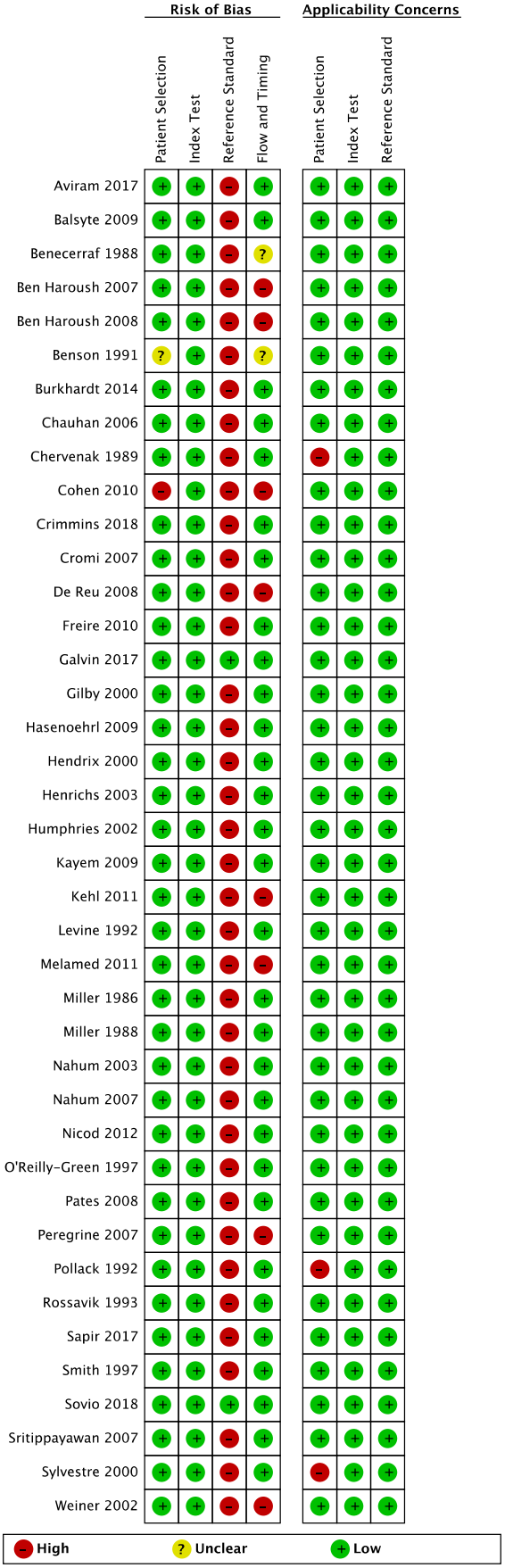


Figure 40. Risk of bias graph of included studies for systematic review of macrosomia.



99. 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;37:513-7.
100. 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;33:453-8.
101. Benacerraf BR, Gelman R, Frigoletto Jr FD. Sonographically estimated fetal weight: Accuracy and limitation. American journal of obstetrics and gynecology 1988;159:1118-21.
102. 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;130:187-92.
103. 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;27:1029-32.
104. Benson CB, Coughlin BF, Doubilet PM. Amniotic fluid volume in large-for-gestational-age fetuses of nondiabetic mothers. Journal of Ultrasound in Medicine 1991;10:149-51.
105. 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;43:77-82.
106. 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;195:601-6. https://doi.org/10.1016/j.ajog.2006.04.012
107. 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;161:753-6.
108. 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;35:434-41.
109. 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;35:140-5.
110. 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;30:861-6.
111. 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;36:324-9.
112. Freire DMC, Cecatti JG, Paiva CSM. Correlation between estimated fetal weight by ultrasound and neonatal weight. [Portuguese]. Revista Brasileira de Ginecologia e Obstetricia 2010;32:4-10.
113. 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:S68. https://doi.org/https://doi.org/10.1016/j.ajog.2016.11.983
114. 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;45:936-8.
115. 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;30:585-90.

116. 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;45:317-22.
117. Henrichs C, Magann EF, Brantley KL, Crews JH, Sanderson M, Chauhan SP. Detecting fetal macrosomia with abdominal circumference alone. Journal of Reproductive Medicine 2003;48:339-42.
118. 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;11:108-12.
119. 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;34:566-71. https://doi.org/10.1002/uog.6378
120. 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;284:1345-9.
121. 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;79:55-8.
122. 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;38:74-81.
123. Miller JM, Jr., Korndorffer FA, 3rd, Gabert HA. Fetal weight estimates in late pregnancy with emphasis on macrosomia. Journal of Clinical Ultrasound 1986;14:437-42.
124. Miller Jr JM, Brown HL, Khawli OF, Pastorek IJG, Gabert HA. Ultrasonographic identification of the macrosomic fetus. American journal of obstetrics and gynecology 1988;159:1110-4.
125. 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;48:13-22.
126. 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;133:148-56.
127. 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;8:2022-7.
128. 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;9:403-8. https://doi.org/10.1046/j.1469-0705.1997.09060403.x
129. Pates JA, McIntire DD, Casey BM, Leveno KJ. Predicting macrosomia. Journal of Ultrasound in Medicine 2008;27:39-43.
130. 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;29:304-9.
131. Pollack RN, Hauer-Pollack G, Divon MY. Macrosomia in postdates pregnancies: the accuracy of routine ultrasonographic screening. American Journal of Obstetrics & Gynecology 1992;167:7-11.
132. Rossavik IK, Joslin GL. Macrosomatia and ultrasonography: what is the problem? Southern Medical Journal 1993;86:1129-32.
133. 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.
134. 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;104:186-90.
135. 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;51:783-91. https://doi.org/10.1002/uog.17491
136. 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;90:1058-67.
137. 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;9:287-90.
138. 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;105:20-4.

Table 24. Characteristics of the studies included in the meta-analysis of macrosomia.

First Author (Year)	Type of Study, Setting	Number of total fetuses (LGA fetuses), 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[99]	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[100]	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[101]	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[102]	Prospective cohort, Single Hospital,	N= 259 (23) Risk: Universal	EFW Hadlock (AC/FL/BPD) Threshold: >90th centile	Mean 32 weeks	BW >4000g	Mean 39 weeks.	DM/GDM: Excluded

	Israel	Selection: Routine scan. Included SGA. Excluded hypertensives and diabetics.	Blinded: No				
Ben-Haroush 2008[103]	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[104]	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[105]	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[106]	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[107]	Prospective cohort	N= 317 (81) Risk: Low	EFW	>41 weeks	BW >4000g	Mean 42 +/- 0.6 weeks	DM/GDM: Excluded

	Single Hospital, New Jersey, USA	Selection: Uncomplicated pregnancies after 41 weeks' gestation.	Hadlock AC/BPD or AC/FL if BPD not available Threshold: >4000g Blinded: Not clear				
Cohen 2010[108]	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[109]	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[110]	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[111]	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[112] (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[113] (GENESIS study) (Abstract)	Prospective cohort Large multi-centre study, Ireland	N= 2336 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[114]	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[115]	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[116]	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[117]	Prospective cohort,	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

	South Carolina, USA						
Humphries 2002[118]	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[119]	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[120]	Prospective cohort, Single Hospiotal, Germany	N= 258 (30) Risk: Universal Selection: Term only with vertex presentation and USS within 3 days of delivery.	AC Threshold: >36cm Blinded: No	Within 3 days of delivery	BW >4000g	40+5 weeks for AC>36cm 39+6 weeks for AC <36cm	DM/GDM: Not reported
Levine 1992[121]	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[122]	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)	Within 3 days of delivery	BW >4000g	Mean 38.1	DM/GDM: Excluded

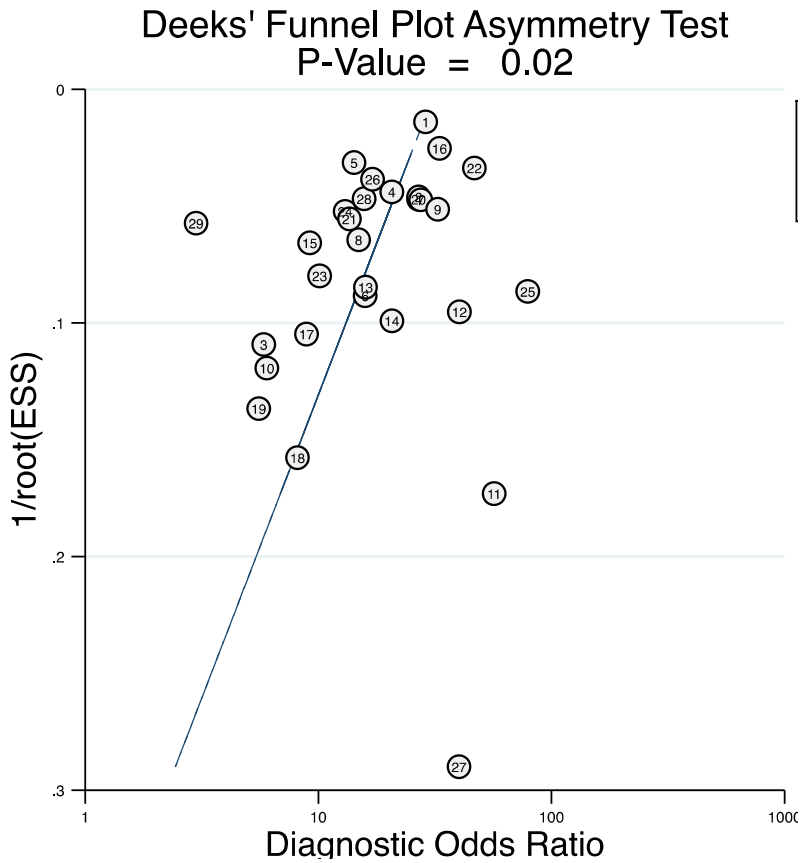
			Shepard (AC/BPD) Threshold: >4000g,>36cm Blinded: No				
Miller 1986[123]	Retrospective cohort, Single Hospital, Luisiana, USA	N= 150 (28) Risk: Mixed Selection: Term only, included diabetes, PET, prior CS. Excluded SGA	EFW Hadlock (AC/FL) Shepard (AC/BPD) Threshold: >4000g Blinded: No	Within 7 days of delivery	BW >4000g	Term (Mean ga not reported)	DM/GDM: Included
Miller 1988[124]	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[125]	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[126]	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[127] (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[128]	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[129]	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[130]	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[131]	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[132]	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[133] (Abstract)	Retrospective cohort Single Hospital, Israel	N=6214 Risk: Mixed Selection: term only, no GDM with scan within 7 days of delivery	EFW, AC Threshold: >4000g, >4500g, AC>39cm Blinded: No	Wiothin 1 week of delivery	Shoulder dystocia	Term (not specified)	DM/GDM: Excluded
Smith 1997[134]	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[135]	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, Neonatal morbidity (composite of metabolic acidosis, 5-min Apgar <7, NICU admission),	Median 40.4 weeks.	DM/GDM: Included (4.3%)

					Severe neonatal morbidity		
Sritippayawan[136] 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[137]	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[138]	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 41. Deeks' funnel plot for publication bias for the prediction of LGA (birthweight >4000g or >90th centile).



Appendix 6. Brief summary of economic analyses of universal screening for breech presentation, large for gestational age, and small for gestational age

Ultrasound screening can be used to detect several different antenatal conditions. Ultrasound assessment could be used to target these conditions individually, or to scan for multiple conditions during the same appointment. However, a screening policy that makes sense for one condition may not be the most cost-effective for a combination of different conditions. In light of this, determining the overall cost-effectiveness of ultrasound screening is a complex task. For this reason, we decided to first target individual conditions and construct economic simulation models capable of evaluating the merits of universal ultrasound for each of these. Once the cost-effectiveness of universal ultrasound for each particular condition had been assessed, we merged these simulation models into a framework that enabled a joint analysis of screening for different combinations of conditions.

In this appendix, we present a brief summary of the economic analyses of universal ultrasound screening for individual antenatal complications. Though neither of these analyses are integral to the final delivery of the study (i.e. the economic analysis of joint screening for different combinations of conditions), they serve as a good introduction to the construction of the joint economic model and the assumptions underlying it. Further, the cost-effectiveness of universal ultrasound for individual conditions may still be relevant for future research and other healthcare systems.

Below we present the economic analysis of universal ultrasound for three conditions: breech presentation, large for gestational age (LGA), and small for gestational age (SGA). The economic analyses of screening for breech presentation and LGA have been published. It should be noted that the term macrosomia was used in the publication of the LGA analysis. Though macrosomia is differentiated from LGA, the two are closely related, and the definition used for macrosomia in this particular analysis was the same as for LGA.

Breech presentation

Background

Despite the relative ease with which breech presentation can be identified through ultrasound screening, the assessment of fetal 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 fetal 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

To estimate the effects of universal ultrasound screening for breech presentation we analysed the outcomes for women with a breech presentation in the Pregnancy Outcome Prediction (POP) study. The POP study was a prospective cohort study between January 14, 2008 and July 31, 2012, where nulliparous women in addition to current clinical practice also attended a research screening ultrasound examination at 36 wkGA. All cases of breech presentation was revealed to both the woman and attending clinician. By analysing the patients' journals, we noted whether breech presentation had been suspected prior to the research scan.

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 CS at 39 wkGA or attempted vaginal breech delivery. We noted whether an ECV had been offered, accepted, performed, and successful; where it was not performed we noted the reason. We also analysed the mode of delivery as a function of the ECV status.

We then used the data to attempt to estimate the consequences of implementing universal ultrasound screening across England. For this purpose, we constructed an economic simulation model capable of comparing outcomes for universal screening with those for current clinical practice. Outcomes included the mode of delivery, which was then extrapolated into long-term fetal health outcomes; due to limited data on long-term morbidity for different modes of delivery, we focused exclusively on mortality risks. The model was probabilistic, capturing overall uncertainty in the outcomes as a function of uncertainty in its input parameters.

Results

Breech presentation was detected in 179 out of 3,879 women (4.6%). For most women (96), there had been no prior suspicion of noncephalic presentation, indicating that up to 54.9% (95% CI: 47.5, 62.1) of all breech presentations may have been undetected in the absence of universal ultrasound. 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 CS (61.5%) and 50 delivered by emergency CS (27.9%). There were no women with undiagnosed breech presentation in labour in the cohort.

On average, 40 scans were needed per detection of a previously undiagnosed breech presentation (95% CI: 33, 49). 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.

We found that a key determinant of the cost-effectiveness of universal ultrasound was the cost of the ultrasound scan itself. We also noted that there was a high degree of uncertainty surrounding this cost, since no NHS cost data was available for an ultrasound scan for fetal presentation only. We therefore estimated the cost thresholds for which universal ultrasound may be cost-effective. We found that universal ultrasound would be cost-effective if fetal presentation could be assessed for £19.80 or less, assuming a willingness-to-pay (WTP) per quality-adjusted life year (QALY) of £20,000; for a WTP-threshold of £30,000, the threshold for cost-effectiveness was £23.10. If the fetal presentation could be assessed for less than £12.90 per mother, universal ultrasound would be cost saving.

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 fetal mortality in breech presentation, and (3) would be cost effective if fetal presentation could be assessed for less than £19.80 per woman.

Large for gestational age (LGA)

Background

Large for gestational age (LGA) pregnancies, i.e. those with an estimated fetal weight in the highest decile, are at increased risk of complications at delivery. This may manifest in increased neonatal morbidity and mortality, as well as maternal morbidity. Ultrasound screening can be used to diagnose LGA antenatally, but this approach is known to have low predictive value. Further, there is no general agreement on how best to manage suspected LGA. Possible interventions include scheduling an elective CS, or early induction of labour. However, uncertainty regarding the clinical effectiveness of these interventions persists, and intervention may cause unnecessary harm if given without clinical need.

There is currently no national programme that couples screening for macrosomia with a proven, disease-modifying intervention. Currently, clinical examination of third-trimester pregnancies does not routinely include ultrasound, but women may be selected for ultrasound scanning following clinical suspicion of LGA (selective ultrasound). An alternative approach would be to prospectively scan all women for LGA (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.

Methods

We constructed a health economic simulation model to compare long-term maternal-fetal health and cost outcomes for different screening programmes for LGA in third trimester pregnancy. The analysis was from a payer perspective and included all nulliparous women within the English NHS. Screening options included universal ultrasound at approximately 36-week's gestation versus selective ultrasound (i.e. current clinical practice). For suspected LGA, possible interventions included elective CS, early induction of labour, or expectant management, i.e. letting the pregnancy take its natural cause.

We simulated outcomes at delivery, using sources of data on probabilities, costs and health outcomes obtained from literature. Outcomes included mode of delivery, as well as respiratory morbidity, shoulder dystocia, acidosis and mortality of the neonate. Long-term neonatal outcomes were then modelled based upon the outcomes at delivery; these included permanent brachial plexus injury (BPI), severe anoxic brain damage, and neonatal mortality. Maternal health outcomes were based upon the

mode of delivery. Probabilistic sensitivity analysis was used to capture overall uncertainty in the outcomes as a function of uncertainty in its input parameters. Overall outcomes included expected costs to the English NHS and quality adjusted life-years (QALYs) gained from each strategy. To identify the most cost-effective screening policy we calculated expected net benefit of each screening-management strategy and compared these using incremental cost-effectiveness ratios (ICER) and cost-effectiveness acceptability curves (CEAC).

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. The most cost-effective policy was selective ultrasound coupled with IOL where macrosomia was suspected.

For suspected LGA, early IOL was always the preferred management strategy from a joint maternal-fetal perspective. However, this was largely explained by the suspected decrease in long-term maternal health associated with elective CS. From a fetal only perspective, elective CS was the preferred management option.

Results were especially sensitive towards changes in maternal health following elective and emergency CS. Our sensitivity analysis also showed that the cost of ultrasound scans and early labour induction were important determinants for which policy was preferred.

Conclusions

The most cost-effective policy for detection and management of fetal macrosomia is selective ultrasound scanning coupled with IOL for all suspected cases of macrosomia. Universal ultrasound scanning for macrosomia in late-stage pregnancy is not cost-effective.

Weaknesses of the analysis includes that LGA was the only criteria evaluated for intervention. In clinical practice, the choice between interventions are typically based upon other factors as well, and not all pregnancies suspected of LGA would be managed in the same way. However, by comparing the outcomes for different interventions, our analysis estimates the value of universal ultrasound screening for LGA. Another weakness was the weak evidence-base for long-term maternal outcomes following different modes of deliveries; this is something that should be subject of future research.

Small for gestational age (SGA)

Background

Small-for-gestational-age (SGA) pregnancies are at higher risk of morbidity and mortality. Ultrasound screening can be used to detect SGA pregnancies, but current clinical guidelines recommend that ultrasound screening be only offered following clinical indications of a problem. Consequently, many SGA pregnancies are not detected. This study sought to evaluate the cost-effectiveness of universal ultrasound screening for SGA in late pregnancy (approx. 36 weeks).

Methods

We constructed a decision model to simulate long-term fetal cost and health outcomes for different screening strategies in the English NHS. Screening strategies were universal ultrasound at 36 weeks' gestation versus ultrasound following clinical indication only. Where the estimated fetal weight (EFW) was <10th percentile, early labour induction was initiated. Cost-effectiveness was assessed using quality-adjusted life years (QALYs), and probabilities, costs, and quality of life weights (QOL) were obtained from literature. Probabilistic sensitivity analysis was used to capturing overall uncertainty in the outcomes as a function of uncertainty in its input parameters. Overall outcomes included expected costs to the English NHS and quality adjusted life-years (QALYs) gained from each strategy.

We focused our analysis on fetal health only, due to the absence of long-term data on maternal quality of life following screening versus no screening. Outcomes at delivery included mode of delivery, level or neonatal morbidity (none, moderate or severe), and survival beyond the first week of life. Long-term outcomes included No long-term complications, special educational needs, severe neurological morbidity, and neonatal mortality. Each long-term outcome was possible for every level of neonatal morbidity, however, the risk of severe outcomes increased with increasing neonatal morbidity.

Results

Universal ultrasound was expected to have minor impact upon long-term neonatal neurological and educational outcomes, but decreased overall fetal mortality slightly (RD: -0.02%; 95% CI: -0.01, -0.03). Compared to selective ultrasound, universal screening was expected to improve overall health by 0.0004 QALYs (95% CI: -0.0001, 0.0002). However, expected costs also increased by £90 (95% CI: -£77, £257), yielding an incremental cost-effectiveness ratio (ICER) of £256,735.

The results rely upon both data and structural assumptions that are uncertain. Probabilistic sensitivity analysis showed that even though the expected ICER was well above the current threshold for cost-effectiveness (£20,000), universal ultrasound still had a 17% chance to be cost-effective due to parameter uncertainty. Further, the assumption that the effect of ultrasound screening upon long-term outcomes is mediated through neonatal morbidity was crucial for the analysis. When this assumption was relaxed, and a direct link between screening and long-term outcomes included in the model, the chance that universal ultrasound would be cost-effective increased greatly.

Conclusions

Universal ultrasound screening in late-stage pregnancy does not appear cost-effective. However, there is great uncertainty surrounding the data informing the model. Future research may be warranted, especially regarding the long-term health consequences of early labour induction.

Appendix 7. Derivation of input parameters for economic simulation model

Beneficial population

An estimate of the total population is required for the value of information analyses, defined as the total population who could benefit from future research that reduces decision uncertainty. The relevant population is all singleton births to nulliparous women in England, excluding those opting for elective CS for reasons other than breech presentation.

NHS Maternity Statistics[202] state that there were 636,401 births in England in FY2016-17. Of these, 91.8% were at ≥37 weeks' gestational age, out of which 33.6% were to nulliparous mothers.[202] The statistics do not disaggregate by reason for elective CS (specifically, whether because of suspected breech position or not). Therefore, this means there were:

$$636{,}401 * 0.918 * \ 0.336 = 196{,}297$$

deliveries in England annually meeting our population definition.

Assuming a 10 year time horizon for the value of information analysis (a proxy for the length of time for which the decision question remains relevant before technological development changes it), an approximately stable number of deliveries per annum and a discount rate of 3.5% yields a beneficial population of 1,689,663.

If our analyses are assumed generalisable to all pregnancies, then the beneficial population is 636,401 per annum, or 5,477,940 over the 10-year horizon (discounted at 3.5%).

Probabilities

Prevalence of SGA, LGA, Breech – nodes A1 & A2

LGA and SGA are defined as a birth weight in the highest and lowest decile of the distribution respectively.[203, 204] The prevalence of each in the population is therefore 10%.

The prevalence of breech at third trimester scan is estimated at 4.6%, based on the POP study, a large prospective cohort study conducted in Cambridge, UK.[10]

Sensitivity and specificity of ultrasound – nodes B, S_B, L_B, B_B

Estimates of the sensitivity and specificity of ultrasound scanning were based on the POP study.[7, 10, 135] Note due to the structure of the model, these figures are not the true sensitivity and specificity of the tests per se, but the probability of detection if everyone is screened ('universal screening') versus the probability of detection with selective screening. The estimates are thus the actual sensitivity and specificities multiplied by the proportion of the population screened. Note we assume the sensitivity and specificity of a positioning scan is 100% as this is an extremely simple procedure, requiring solely the identification of the scull and spinal column to determine orientation of the fetus.

Interventions for Breech – nodes B_ECV, B_ECVs, B_noECV, B_ECVs_rC and B_ECVf_RC

Data on the proportion of mothers accepting ECV, the success rate and reversion rates were extracted from the POP study[7]. Methods and results for this has been published separately[10].

Delivery mode, true negative (AGA babies) node C1

An otherwise healthy baby (i.e. true negative for SGA, macrosomia and breech, node C1) can be delivered via emergency CS or vaginally.

A study of 14,100 singleton live and stillborn infants in French maternity units in 2010 found approximately 19.4% (n=2504/12881) of non-SGA babies were delivered via emergency CS.[21] The POP study (a study of singleton nulliparous pregnancies between Jan 14, 2008, and July 31, 2012 in Cambridge, UK) found 19.9% (735/3689) of non-breech position babies were delivered via emergency CS.[10] A 2018 Cochrane systematic review of IOL versus expectant management in women at or beyond term(Middleton et al., 2018) found an 18.42% (1056/5734) CS rate in the expectant management arm (analysis 1.13[15]).

The most relevant population to this analysis is the POP study.[10] Of the 3689 deliveries, 141 were elective CS. Our defined population excludes elective CSs for indications other than breech therefore we assume 735/3548 = 20.7% of AGA deliveries result in emergency CS (95%CI 19.4%, 22.06%), with 79.3% being delivered vaginally.

We chose to use data from the POP study[10] (a prospective cohort study) for the risk of emergency CS, rather than from Monier et al.[21] (a population based setting), because the study design of the former made the validity of the numbers easier to verify. Compared to a network meta-analysis (NMA), relying on a single study risks potentially overestimating uncertainty, however, due to time constraints conducting a NMA was unfeasible.

Delivery mode, false negatives for SGA and LGA - nodes S_C2, L_C2

If a baby is SGA and this is not spotted (i.e. is a false negative, node S_C2), the relative risk of emergency CS is taken from the French cohort study, which reported an adjusted relative risk of "Caesarean after onset of labour" (assumed to meet the definition of emergency CS) in low risk pregnancies of 1.9 (95%CI 1.4, 2.5, Table 3, Monier et al.[21], figures only reported to 1 decimal place).

If LGA is not spotted (i.e. is a false negative, node L_C2), the odds ratio of emergency CS versus an AGA baby is assumed 1.792 (95%CI 0.718, 4.471). This probability was obtained from a retrospective analysis from 2005 based in the USA that included 241 nulliparous women whose pregnancies were induced and delivered at term (Blackwell et al.[141]). Breech position, stillbirth and pregnancies with other abnormalities were excluded. All underwent estimation of foetal weight via ultrasound prior to labour. 23 of 241 (9.5%) overestimated the EFW by 15% or more. Caesarean delivery rates for labour arrest (assumed to be emergency CS) were 34.8% in the overestimated group and 13.3% in the no-overestimation group. This equates to 8/23 and 29/218 in each group respectively, yielding an odds ratio of 1.792 with a standard error of the log of the odds ratio of 0.466.

Delivery modes – true positives for SGA and LGA– nodes S_C3, L_C3

The relative risk of "Caesarean after onset of labour" (assumed to meet the definition of emergency CS) in true positive SGA babies following induction versus true negatives (i.e. AGA babies) is assumed to be 2.9 (node S_C3). This may be an overestimate as according to the data source (Monier et al.[21]) this is the relative risk of emergency CS for true positive SGAs, whether or not they were induced, and only 27.1% (36/133) were induced <39 weeks.

We could not identify data for how early IOL would affect the risk of emergency CS among true LGA pregnancies. For this reason, we used data from Middleton et al.[15], implicitly assuming the same relative risk reduction for LGA pregnancies as for non-LGA pregnancies. The relative risk for induced versus non-induced LGA pregnancies was 0.92 (95% CI: 0.85, 0.99) and was modelled using lognormal distribution (mean: -0.08, standard error: 0.037).

If the policy for handling macrosomic babies is expectant management (node L_C2), then the emergency CS rate is assumed the same as for a false negative diagnosis.

Delivery modes – false positives for SGA and macrosomia– nodes S_C4, L_C4, L_C1

False positives for SGA will be induced. False positives for macrosomia will be handled depending on the selected management strategy: expectant management or IOL.

A prospective RCT (n=6106) of IOL at 39 weeks in low-risk nulliparous women yielded a relative risk of (emergency) CS of 0.84 (95%CI 0.76, 0.93) associated with induction.[148] Note that the Monier study[21] described above reported a relative risk of emergency CS in false positives for SGA of 1.0 (95%CI 0.5, 2.2). However, as an RCT is generally considered at lower risk than an observational study, we opted for the RCT results[148] and applied this to nodes S_C4 and L_C4, representing the probabilities of emergency CS following IOL for false positive diagnoses of SGA and LGA respectively.

Where the selected management strategy for LGA is expectant management, the risk of emergency CS under a false positive diagnosis (node L_C1) is logically assumed the same as for an AGA baby (node C1).

Delivery modes for breech –false negative and true positive - nodes B_C2, B_C3a-B_C3f

If a baby is breech and is a false negative (i.e. undetected breech, node B_C2), we assume the probability of an emergency CS is 57.7% (95%CI 38.67%, 75.62%). No comparative data were identified for the risk of emergency CS with unidentified breech versus cephalic. However, a retrospective cohort study of the case notes of 131 women in Hong Kong in 1997 found that of those with undiagnosed breech at labour, and excluding those in whom ECV was subsequently attempted, 11 (42.3%) had a vaginal breech delivery and 15 (57.7%) a Caesarean section (Table 2, Leung et al.[163]). Caesarean sections are labelled as the sum of elective and emergencies, but given that these were undiagnosed until labour, we have interpreted these as all emergency CS.

Nodes B_C3a to B_C3f represent delivery modes with and without external cephalic version, taking into account success or failure as well as spontaneous reversion (either to breech or cephalic presentation). All estimates are obtained from the POP study[10] except for node B_C3b, representing delivery modes where ECV was successful but the baby subsequently reverted to breech position. This was due to a lack of relevant observations in the POPS data. We assumed the same distribution as per a false negative diagnosis of breech (57.69% probability of emergency CS, node B_C2).[163] Note we assume this to be an independent probability with the same parameters as node B_C2, rather than taking the exact same value to reflect that this is a different outcome measure to B_C2, but with the same likelihood.

Perinatal morbidity – true negative (AGA babies) - node D1

Node D1 represents the baseline risk of neonatal morbidity from expectant management of an otherwise healthy, non-SGA baby, taken from the POP study (Table 25), (Table 11), and systematic review.[51] Outcomes include no, moderate, and severe neonatal morbidity, and perinatal death. Moderate neonatal morbidity was defined as one or more of the following criteria: a 5 min Apgar score of less than 7, delivery with metabolic acidosis (defined as a cord blood pH <7.1 and base deficit >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). Severe neonatal morbidity was defined as hypoxic ischaemic encephalopathy, use of inotropes, need for mechanical ventilation, or severe metabolic acidosis (defined as a cord blood pH <7.0 and base deficit >12 mmol/L.

Table 25. Prevalence of no, moderate and severe neonatal morbidity in the POP study by fetal size.

	No morbidity	**Moderate morbidity**	**Total**
Non-SGA	3325	198	3523
SGA	298	44	342
Total	3623	242	3865
	Non-severe morbidity	**Severe morbidity**	**Total**
Non-SGA	3501	22	3523
SGA	338	4	342
Total	3839	26	3865

SGA = Small for gestational age

RCOG guideline 20b[12] states that there is a 0.1% risk of perinatal mortality associated with a planned cephalic vaginal delivery. However, this figure includes all stillbirths and neonatal deaths. The relevant figure for the purpose of our model comprises intrapartum stillbirths and neonatal deaths only: deaths prior to this are assumed unrelated to orientation or size of the foetus, and thus do not affect the

results of the incremental analysis. To estimate the risk of stillbirth and perinatal mortality, we used observational data from Moraitis et al.[51], since delivery before 37 week's gestational age was an exclusion criteria for the study. For baseline risk, we used mortality for spontaneous vaginal and assisted vaginal deliveries only. In the study, spontaneous and assisted vaginal deliveries accounted for 88.07% and 59.48% of antepartum stillbirths and delivery-related perinatal mortality, respectively. Data from Tables 2 and 3 showed the risk of stillbirth / perinatal mortality as a function of birth weight. Using this data, we estimated that the total number of stillbirths and perinatal mortality for spontaneous and vaginal deliveries would have been 809.66 and 455.54 if all babies would have been AGA. Multiplying these numbers with the respective proportion of deaths resulting from spontaneous and instrumental vaginal deliveries, we estimated that the total mortality for these categories would have been 984 cases (n = 635,396). Modelling this using a beta distribution, the baseline risk (i.e. for AGA pregnancies delivered vaginally) was 0.155% (95% CI: 0.145%, 165%).

The probabilities of none, moderate or severe morbidity and perinatal death would ideally be modelled as a Dirichlet distribution. However, as these statistics are sourced from different sources they are modelled as independent beta distributions. This may overestimate the uncertainty in morbidity risk. Furthermore, we assume that risk of neonatal morbidity in an AGA baby is independent of delivery mode. A priori, an emergency CS is expected to be associated with a higher risk of perinatal morbidity. However, the relevant population is babies who are neither breech, SGA nor LGA, but who undergo an emergency CS for other reasons. After factoring out these indications for emergency CS the assumption may not be so unreasonable.

Perinatal morbidity – false negative SGA babies – node S_D2

The same sources (POP study and Moraitis et al.[51]) for node D1 report the odds of adverse outcome in SGA babies (i.e. in the bottom decile of the distribution): the odds ratio of moderate and severe morbidity and still birth for SGA compared with AGA babies in the absence of intervention (i.e. induction) are 2.48, 1.88 and 4.89 respectively (node S_D2). Again, we assume the risk of neonatal morbidity in SGA babies is solely a function of their size, and not delivery mode.

Perinatal morbidity – false negative LGA babies – nodes L_D2a & L_D2c

Baselines

Neonatal morbidity for undiagnosed LGA babies (false negatives) were modelled to take account of specific risks for these babies, and therefore modelled as none (no complications), respiratory morbidity, shoulder dystocia, 'other acidosis' or perinatal death. Shoulder dystocia can lead to no long-

term complications, brachial plexus injury (which can be transient or permanent), or acidosis, leading to no long-term complications, severe anoxic brain damage or perinatal mortality. 'Other acidosis' (secondary to other than shoulder dystocia) has the same long-term outcomes as that secondary to dystocia, namely no long-term complications, severe anoxic brain damage or perinatal mortality. The risks of neonatal morbidity (and hence mortality) are related to delivery mode. These are modelled by estimating a baseline risk for each morbidity for the general population and multiplying this by a relevant relative risk. The baseline risks are not used in the model per se as morbidity for otherwise healthy infants is captured via 'none / mild / moderate / perinatal death' (node D1).

The baseline probability of respiratory morbidity was extracted from a study of the influence of timing of elective CS on respiratory morbidity, conducted in Cambridge, UK.[205](Benner et al., 2010) All deliveries between 1985 and 1993 at the centre (n=33,289) were included in the analysis and all cases of respiratory distress syndrome or transient tachypnoea requiring admission to neonatal intensive care were recorded. Of the entire sample, 6955 deliveries occurred at term (week 39 to week 39+6 days) and were delivered vaginally. Of these 22 had respiratory morbidity, reported as 0.32% (95%CI 0.18%, 0.45%). Assigning a beta distribution to these figures yields a similar (but slightly different) 95%CI of (0.20%, 0.46%). This was used as the baseline risk (i.e. risk for AGA babies).

The baseline probability of shoulder dystocia was based on figures quoted in RCOG guidelines on the management of shoulder dystocia.[206] This reported incidences in the literature of between 0.58% and 0.70%. The best quality study informing the estimate was a retrospective analysis by Ouzounian et al.[166]. This reported 1686 cases of shoulder dystocia among 267,228 vaginal births, yielding an incidence of 0.63% (95%CI 0.60%, 0.66%).[166]

The baseline probability of other acidosis (i.e. not secondary to shoulder dystocia) was based on a Cochrane systematic review comparing induction versus expectant management.[15] Analysis 1.4 of the review reported incidence of birth asphyxia, with 5 of 731 pregnancies in the expectant management arm, yielding a base probability of 0.68%.

The baseline risk of perinatal morbidity was assumed to be the same as described above (node D1), i.e. an estimated risk of 0.155% (95% CI: 0.145%, 165%), based upon own estimations using data from Moraitis et al.[51] Since this baseline risk was not specific to fetal size, we used the same baseline risk for SGA and LGA fetuses and distinguished their risk using their respective odds ratios instead.

To estimate the baseline risk of perinatal death, we used observational data from Moraitis et al.[51], since delivery before 37 week's gestational age was an exclusion criteria for the study. For baseline risk, we used mortality for spontaneous vaginal and assisted vaginal deliveries only. In the study, spontaneous and assisted vaginal deliveries accounted for 88.07% and 59.48% of antepartum stillbirths and delivery-related perinatal mortality, respectively. Data from Tables 2 and 3 showed the risk of stillbirth and perinatal mortality as a function of birth weight. Using this data, we estimated that the total number of perinatal deaths for spontaneous and vaginal deliveries would have been 809.66 and 455.54 if all babies would have been AGA. Multiplying these numbers with the respective proportion of deaths resulting from spontaneous and instrumental vaginal deliveries, we estimated that the total mortality for these categories would have been 984 cases (n = 635,396). Modelling this using a beta distribution, the baseline risk (i.e. for AGA pregnancies delivered vaginally) was 0.155% (95% CI: 0.145%, 165%).

Ideally, these mutually exclusive probabilities would be modelled with a Dirichlet distribution. However, as they are from different sources, they are modelled with their respective distributions. This risks generating a set of probabilities that sum to greater than 1. However, given the low absolute percentages, this is highly unlikely. Sampled values were verified in the model code to ensure all were contained within [0,1].

Undetected macrosomia (false negative), vaginal delivery (L_D2a)

No data were available on the relative risk or odds ratio of respiratory morbidity for undetected macrosomia with a vaginal delivery (node L_D2a). Expert opinion estimated that these babies were either at the same or lower risk of respiratory morbidity than AGA babies. We therefore used a point estimate relative risk of 0.75, and assigned a uniform distribution between 0.5 and 1. Note relative risks are more intuitive than odds ratios from an elicitation point of view, we therefore report this as an RR not OR.

The odds ratio of shoulder dystocia in a macrosomic baby delivered vaginally (versus an AGA baby) is assumed 7.18 (95%CI 2.06, 25.00). This is based on a systematic review reporting incidence of shoulder dystocia in all babies with a birth weight >=4000g (Table 2 of Rossi et al.[167]). Two source studies were meta-analysed with a random effects model. Importantly, these data are not disaggregated by delivery method. However, it is reasonable to assume that CS eliminates the risk of shoulder dystocia, and therefore this represents the odds ratio of LGA babies delivered vaginally.

The same table in the review[167] also reported the odds ratio of asphyxia in a macrosomic baby (versus an AGA baby) of 2.88 (95%CI 1.34, 6.22). We assume this meets our definition of 'other acidosis' and apply the figures accordingly, but with the caveat that this is not disaggregated by delivery mode and so may overestimate the risk (for example, asphyxia may be the reason for an emergency CS).

The same table in the review[167] also reported the odds ratio of perinatal death in a macrosomic baby (versus an AGA baby) of 1.77 (95%CI 0.30, 10.34). We apply this to our definition of perinatal mortality, again noting that this is not disaggregated by delivery mode. The rarity of the outcome is also reflected in the wide confidence interval, implying a high degree of uncertainty.

Undetected macrosomia (false negative), emergency Caesarean section (node L_D2c)

The relative risk of respiratory morbidity for a macrosomic baby delivered via emergency CS versus an AGA baby (Table 26) delivered vaginally was taken from the Cambridge cohort described in the baseline probabilities section above (Table 2, Morrison et al.[165]). As stated above, this study was not specific to LGA babies, but the risk of respiratory morbidity is most plausibly associated with intervention to speed delivery rather than the presence of macrosomia. The source table reports the odds ratio of respiratory morbidity with 'CS labour' (assumed to meet the definition of emergency CS) at 39/0 to 39/6 as 3.2 (95%CI 1.4, 7.4) relative to the baseline of vaginal delivery at 40/0 to 40/6. Rebasing relative to vaginal delivery at 39/0 – 39/6 yields an odds ratio of 1.674 (95%CI 1.253, 2.001).

Table 26. Risk of respiratory morbidity from emergency Caesarean section.

	OR	LCL	UCL
CS labour	0.6	0.4	1
Vaginal	3.2	1.4	7.4
Rebased	5.33	3.5	7.4
ln	1.674	1.253	2.001
SE	0.167		

The relative risk of shoulder dystocia for emergency CS was assumed zero.

The relative risk of other acidosis for a macrosomic baby delivered via emergency CS versus an AGA baby (Table 27) was taken from Chongsuvivatwong et al.[168] (as for elective CS described above, and thus the same caveats are attached).

Table 27. Risk of acidosis from emergency Caesarean section.

	n	Severe asphyxia rate/1000	LCL	UCL	Implied n from raw numbers
Vaginal	12591	4.3	3.2	5.6	54
EMCS	4328	8	5.5	11.1	35
	Asphyxia	No asphyxia			
Vaginal	54	12537	12591		
EMCS	35	4293	4328		
	89	16830			
		LCL	UCL		
OR:	1.867	1.217	2.865		
LnOR	0.625				
SE(lnOR)	0.218				

Finally, the relative risk of perinatal mortality for a macrosomic baby delivered via emergency CS versus an AGA baby was taken from the same source[168] (Table 28).

Table 28. Risk of perinatal mortality from emergency Caesarean section.

	n	Dead /1000 dels	LCL	UCL	Implied n from raw numbers
Vaginal	12591	7	5.6	8.6	88
EMCS	4328	12.4	9.3	16.2	54
	dead	alive			
Vaginal	88	12503	12591		
EMCS	54	4274	4328		
	142	16777			
		LCL	UCL		
OR:	1.781	1.266	2.505		
LnOR	0.577				
SE(lnOR)	0.174				

Perinatal morbidity – true positive SGA babies – induction of labour – node S_D3

If an SGA baby is induced, we assume the relative risk of moderate and severe morbidity is 0.7, and for perinatal death 0.33 (node S_D3). These data are based on a systematic review of IOL compared with expectant management in low risk women at or beyond term (approx. 10,000 observations, odds ratios not reported).[15] Critically, this is not the treatment effect with SGA babies, for which we were unable to identify any data, and the relative risk for moderate and severe morbidity was based on data reporting 5-minute Apgar score below 7. However, the central estimates of relative risks (0.7 and 0.33 respectively) were considered plausible by clinical experts (GS, AM), and that the confidence intervals represented plausible summaries of their epistemic uncertainty.

Perinatal morbidity – true positive LGA babies - expectant management and induction of labour – nodes L_D3a and L_D3c

An expectant management policy for true positive diagnoses of LGA (at node MGT_LGA_TP) is identical in management to expectant management under a false negative and the risk of perinatal morbidity is logically the same as under "undetected macrosomia (false negative), spontaneous vaginal" and "undetected macrosomia (false negative), emergency CS" described above. Nodes L_D2a and L_D2c are therefore replicated at this point in the tree (following MGT_LGA_TP >> L_C2).

Under an IOL policy for positive diagnoses of LGA (MGT_LGA_TP >> L_C3a), delivery modes can again be spontaneous vaginal or emergency CS. Where data allow, risks of perinatal morbidity are assumed related to IOL and the presence of macrosomia as well as delivery model (vaginal or emergency CS).

Respiratory complications

A retrospective cross-sectional study of maternal and neonatal outcomes in induced low-risk term pregnancies (N = 131,243) reported neonatal complications by week of delivery comparing IOL with expectant management (Gibson et al.[169]). The adjusted odds ratio of respiratory complications at week 39 is reported as 0.540 (95%CI 0.373, 0.783, Table 4[169]). This was used as odds relative to an AGA baby, whether vaginally delivered or emergency CS (L_D3a and L_D3c respectively). Of note is that these data are not macrosomia-specific.

Shoulder dystocia

A Cochrane systematic review[98] of IOL versus expectant management for suspected fetal macrosomia estimated a relative risk of shoulder dystocia of 0.6 (95%CI 0.37, 0.98) (analysis 1.3 of Boulvain et al.[98]). We therefore applied this relative risk, noting the baseline comparator is MGT_LGA_TP >> L_C2 or MGT_LGA_TA >> L_C3. That is:

P(dystocia | vaginal delivery at node L_D3a) = P(dystocia | vaginal delivery at node L_D2a) * RR

And:

P(dystocia | EmCS at node L_D3c) = P(dystocia | EmCS at node L_D2c) * RR

Data are for 'suspected' macrosomia, not disaggregated by true and false positives. We therefore apply due caution and score the relevance of the data as 'moderate'.

Acidosis

The Boulvain Cochrane review[98] did not report incidence of acidosis or asphyxia. Therefore we sourced data from the Middleton Cochrane review,[15] which compared induction versus expectant

management in all pregnancies at term. Analysis 1.4 reported a relative risk of birth asphyxia of 1.66 (95%CI 0.61, 4.55). We used this to represent the relative risk of 'other acidosis'.

Perinatal mortality

The Cochrane systematic review[98] of IOL versus expectant management for suspected foetal macrosomia observed zero events in the included studies. We therefore used the Middleton Cochrane review, Analysis 1.1,[15] reporting a relative risk of 0.33 (95%CI 0.14, 0.78) compared with not non-induced AGA babies.

The odds ratios and relative risks for node L_D3c are identical to those for L_D3a. However the implied probabilities at the nodes will differ due to the different baseline comparators: for respiratory morbidity, acidosis and perinatal death the ratios are relative to expectant management for AGA babies. For dystocia, macrosomia-specific data were available, comparing induction with expectant management in cases of suspected macrosomia, so the ratio is relative to vaginal delivery or emergency CS for an expectant management policy.

Perinatal morbidity – false positive SGA or LGA babies – induction of labour– node D4

Following an incorrect diagnosis of SGA or following an incorrect diagnosis of LGA under the IOL policy, an AGA baby will be induced. Evidence suggests this reduces the risk of stillbirth, but with the consequence of increasing perinatal complications: a retrospective database analysis of induction versus expectant management at 37 weeks found an odds ratio of 0.15 (95%CI 0.03, 0.68) for perinatal death and 1.92 (95%CI 1.71, 2.15) for admission to neonatal unit or special care baby unit.[162] We assumed admission to these specialist units was a proxy for moderate and severe complications, so applied these ORs to the baseline risks.

Perinatal morbidity – false positive LGA babies – expectant management

Following an incorrect diagnosis of macrosomia, and with an expectant management policy, perinatal outcomes are logically the same as vaginal and emergency CS perinatal outcomes for AGA babies. Therefore, these nodes are labelled as D1.

Perinatal morbidity – breech – false negative and true positive – B_D2a – B_D2c

Perinatal outcomes are assumed dependent on whether the baby is breech at delivery or not. A breech baby who reverts to cephalic positioning either spontaneously or following ECV is assumed as the same risk of perinatal outcomes as an AGA baby.

Vaginal breech delivery (B_D2a) – Perinatal death

RCOG guideline 20b[12] states that vaginal delivery in the breech position is associated with a risk of perinatal mortality of 2/1000, but 0.5/1000 with elective CS, compared with a 1.0/1000 risk for a cephalic vaginal delivery. This is based largely on a Cochrane systematic review of planned CS for term breech delivery,[13] the largest contributor to which was the Term Breech Trial (TBT).[207]

As described above ('perinatal morbidity – true negative (AGA babies) - node D1'), the 1.0/1000 risk of perinatal mortality includes all deaths around the time of delivery. However, our figure of interest is solely intrapartum stillbirth and neonatal death (the implicit assumption is that pre-partum deaths are due to causes other than breech, LGA or SGA). A retrospective cohort study of all term singleton births in delivery units in Scotland between 1992 and 2008 (n = 784,576) observed a mortality rate of 0.04% (234/537745) associated with cephalic vaginal deliveries.[51] The same study reported a mortality rate of 0.29% (5/1719) associated with breech vaginal deliveries, yielding an odds ratio of 6.68 (95%CI 2.75, 16.22).

Vaginal breech delivery (B_D2a) - Moderate & Severe morbidity

We estimate the relative risk of moderate and severe morbidity associated with breech vaginal delivery versus cephalic vaginal delivery at 6.7 (95%CI 5.9, 7.6). This is based on a large retrospective cohort analysis of the Swedish Medical Birth Registry from 1988 to 1997 reporting the odds ratio of 5 minute Apgar below 7.[172] We assume the odds ratios are identical for moderate and severe morbidity. This may be a reasonable assumption: the odds ratio for perinatal death calculated above is 6.68, extremely close to the 6.7 reported here.

Elective Caesarean section delivery (B_D2b) – Perinatal death

A Cochrane systematic review of elective CS versus vaginal delivery for term breech delivery (Hofmeyr et al.[13] Analysis 1.3) observed an overall global relative risk of perinatal death of 0.29 (95%CI 0.10, 0.86).

Elective Caesarean section delivery (B_D2b) - Moderate & Severe morbidity

The same review[13] reports a relative risk of 5 minute Apgar score <7 of 0.43 (95%CI 0.12, 1.47), and Apgar score <4 of 0.11 (95%CI 0.01, 0.87, Analyses 1.4 and 1.5 respectively). We therefore use this as the relative risk of moderate and severe perinatal morbidity respectively associated with elective CS versus planned vaginal breech delivery.

Emergency Caesarean section delivery (B_D2c) – Perinatal death

A study of 32776 breech presentations in Scotland between 1985 and 2004 (Pasupathy et al.[173]) observed 9018 emergency CS deliveries (4108 pre- and 4910 post-labour), of which 14 lead to perinatal and neonatal death (0.16%). As stated above, the Moraitis review[51] reported a mortality rate of 0.29% (5/1719) associated with breech vaginal deliveries. This yields an odds ratio of 0.533 (95%CI 0.192, 1.482). As this odds ratio is based on combining data from different sources, we explore this parameter in greater detail in one-way sensitivity analysis.

Emergency CS delivery (B_D2c) - Moderate & Severe morbidity

In the absence of evidence on the effect of emergency CS versus vaginal breech delivery for the risk of moderate and severe neonatal morbidity, we assumed the odds ratio to be the same as the odds ratio of perinatal death, i.e. 0.533 (95%CI 0.192, 1.482).

Long term outcomes following no, moderate and severe perinatal morbidity (AGA, SGA and breech) – nodes E1 – E3

Long-term outcomes were no complications, Special Educational Needs (SEN), Severe neurological morbidity (SNB), and Neonatal/Infant death. The risks of each were assumed dependent solely on level of perinatal morbidity (where perinatal morbidity is a function of abnormality and delivery management).

A large retrospective cohort study of school children reported the risk of special educational needs (SEN) by 5-min Apgar score inter alia.[174] 4.7% [=18,736/(18736+376,891)] of children with a 5-minute Apgar score at birth of 8-10 required SEN. We used this as the risk of SEN for children with no neonatal complications (node E1). The same study also reported odds ratio for Apgar scores of 4-7 and 0-3, which were used as the increase in risk for moderate and severe neonatal morbidities (nodes E2 and E3).

We used cerebral palsy (CP) as a proxy for severe neurological morbidity. A large retrospective cohort study of births in Sweden analysed the risk of cerebral palsy by 5-minute Apgar score.[175] We calculated the baseline risk of CP as the sum of the number of children with CP with Apgar score >7 divided by the total number of children with Apgar > 7 (=(69+163+674)/(27,664+129,096+1,037,793) = 0.08%, node E1). The study also reported adjusted hazard ratios (HR) by individual Apgar score, rather than grouped categorisations (<4, 4-7 and >7). A weighted geometric mean HR (and 95%CI) was calculated

for each group as per the Table below, and divided by the weighted 7-10 results. We interpreted the hazard ratio as the relative risk. These are different, but related concepts, the former taking account of time whilst the latter assuming all events happen simultaneously. Given the simple structure of our model, and the relative rarity of CP, we felt this was a sufficient approximation.

Infant mortality data were extracted from routine Scottish data from 1992 to 2010.[176] 1,013,363 neonates had a normal Apgar score at birth (defined as >7) (Table 29). Of these there were 628 neonatal (birth to 28 days) and 1446 infant deaths (29 days to 1 year), a total of 0.2%. This was assumed to form the baseline risk of neonatal/infant mortality (node E1). Adjusted relative risks of neonatal and infant mortality were reported in the appendix to the paper. To generate an overall relative risk over 12 months, a weighted geometric mean (and 95%CIs) of the risks reported by Iliodromiti et al.[176] for neonatal and infant mortality was calculated, with weights of 1 and 12 for neonatal and infant mortality respectively (representing the relative length of the time periods, Table 30). Relative risks for Apgar 4-6 and 0-3 were used for moderate and severe neonatal morbidity respectively (nodes E2 and E3).

Table 29. Baseline risk of cerebral palsy by Apgar score (5 min).

5-minute Apgar	By single score			Grouped		
	Number of children	Number with CP	Adjusted HR (95%CI)	Number of children	Number with CP	Adjusted HR (95%CI)
0	136	13	277.7 (154.4, 499.5)	1447	130	145.5 (104, 204.1)
1	215	23	238.2 (153, 371)			
2	388	29	124 (83.8, 183.4)			
3	708	65	148.3 (112.8, 195)			
4	1097	53	75.9 (56.4, 102)	17470	185	10.4 (7.8, 13.9)
5	1830	39	32.6 (23.4, 45.6)			
6	4259	42	15.4 (11.2, 21.2)			
7	10284	51	6.9 (5.1, 9.4)			
8	27664	69	3.8 (3, 4.9)	1194553	906	1 (ref)
9	129096	163	1.9 (1.6, 2.2)			
10	1037793	674	1 (ref)			

Source: Fig 1, Persson et al.[175]

Table 30. Relative risk of cerebral palsy by Apgar score (5 min).

Apgar	Neonatal weight	Adjusted RR	Infant weight	Adjusted RR	Pooled Adjusted RR
0-3	1/13	188.4 (141.7, 250.5)	12/13	55.14 (44.03, 69.06)	60.61 (48.17, 76.26)
06-Apr	1/13	34.16 (23.41, 49.86)	12/13	11.81 (8.64, 16.15)	12.82 (9.33, 17.61)
10-Jul	1/13	1 (ref)	12/13	1 (ref)	1 (ref)

Long term outcomes following LGA birth – nodes L_E1, L_F1, L_G

In our model, LGA babies are at risk of no perinatal complications, respiratory morbidity, shoulder dystocia, other acidosis or perinatal mortality. LGA babies developing shoulder dystocia are at risk of no long-term complications, brachial plexus injury (BPI) or acidosis. BPI can be transient or permanent. Acidosis can lead to no LT complications, SEN, SNM or perinatal mortality. The RCOG Green-top guideline No. 42[206] state that "fewer than 10% resulting in permanent [injuries]", based upon findings from Gherman et al.[208] These figures in turn rely upon the study by Sandmire et al.[171] In total, 8 out of 145 cases of BPI injuries were permanent. We modelled this using a beta distribution, yielding a risk of permanent BPI of 5.5% (95% CI: 2.4, 9.8).

Following no perinatal complications, LGA babies are at background risk of long-term complications, SEN, SNM and neonatal mortality (node E1).

Following respiratory morbidity, we assume infants are at increased risk of long-term complications (SEN, SNM and neonatal/infant mortality) equivalent in severity to severe neonatal morbidity (i.e. node E3).

Shoulder dystocia can lead to no injury to the infant (in which case the background risk of SEN, SNM and neonatal/infant mortality applies), BPI, which can be transient or permanent, or acidosis.
Transient BPI leads to background risk of long-term complications, SEN, SNM and neonatal mortality (node E1).

Permanent BPI leads to baseline risk of long-term complications, SEN, SNM and neonatal mortality, but with a decreased quality of life associated with the injury (node L_G).

Following acidosis, the risk of long-term complications, SEN, SNM and neonatal mortality is assumed in severity to severe neonatal morbidity (node E3).

Costs

Costs of ultrasound scan for fetal size

We obtained the cost of an ultrasound scan for fetal size (and presentation) from the national schedule of reference costs.[177] We used data for 'Ante-Natal Standard Ultrasound scan (NZ21Z)', as reported for outpatient procedures. The reference costs contained the mean as well as lower and upper inter-quartile range (IQR) for costs, listed by every type of service provider. We calculated a weighted average for the mean/inter-quartile ranges based upon the reported numbers of activities over the year for each provider. We then fitted a gamma distribution to the weighted mean/ inter-quartile range, obtaining the parameters $\alpha = 4.6904$ and $\beta = 22.8062$, and yielding a total cost of £107.06 per scan (95% CI: 70.89, 134.92).

Cost of ultrasound scan for fetal presentation only

Estimating a cost for an ultrasound (US) scan for fetal presentation alone is challenging, since this type of US screening is not part of current NHS routine. We theorized that such a scan could be performed by a midwife in conjunction with a standard antenatal visit in primary care, using relatively basic and inexpensive equipment. However, it is uncertain whether implementing such a routine is feasible. For this reason, we estimated the cost of two different type of scenarios for how a US scan for fetal presentation alone could be performed.

Midwife-led screening in primary care setting

We theorized that an ultrasound scan for fetal presentation alone could be provided by a midwife in conjunction with a standard antenatal visit in primary care. While there are NHS reference costs for 'Ante-Natal Standard Ultrasound scan (NZ21Z)'[177], such scans frequently involve assessment of fetal anatomy and/or biometry and since these require much more time and training to assess than fetal presentation alone, we deemed it inappropriate to use this cost as an estimate for the cost of an ultrasound scan for fetal presentation alone.

Following the methodology for Wastlund et al.[10], we estimated the cost of ultrasound scan for fetal presentation as a function of the midwife's time, the equipment cost, and the cost of the room/facilities where the scan would take place.

We obtained the cost of the midwife's time from the Unit Costs of Health and Social Care 2017.[178] We used the total hourly cost for Band 5 nurses, £36; this cost was consistent with the costs reported for

midwives in NHS Staff Earnings 2017.[209] On top of the scan itself, time would be needed to make the woman feel comfortable in the process, and to document the results of the scan, we estimated that the average scan would require 5-10 minutes in total. In the absence of data on how much it would cost to provide ultrasound equipment and sufficient training, we guessed that this could be provided for a total cost between £1-20k. We assumed that the average machine would be operated 400 to 3000 times annually over the 5-year time horizon. We assumed that room costs would be between £4,500 and £6,000 annually[210], and that rooms would be operated 1,573 hours per year[178].

We simulated the total cost per scan using uniform distributions and 100,000 simulations. We then fitted a gamma distribution to the resulting distribution, based upon the mean and inter-quartile range. The resulting parameter estimation was a gamma distribution with alpha = 43.8259, and beta = 0.2159. This resulted in a total cost of ultrasound scan for fetal presentation of £9.46 (95% CI: £6.87, 12.46).

Sonographer-led ultrasound in designated setting

If implementing US assessment in primary care (as part of a standard antenatal visit) would not be possible, the most feasible alternative would be to perform the scan by referral to a designated ultrasonography unit. A scan for fetal presentation alone is much swifter and technically less complicated than the type of scan typically performed as part of a standard antenatal visit. For this reason, we didn't consider 'Ante-Natal Standard Ultrasounds Scan (NZ21Z)' in the NHS reference costs[177] to be a suitable cost estimate. Instead, we used the data for 'Ultrasound Scan with duration of less than 20 minutes, without Contrast (RD40Z)' from the reference costs, diagnostic imaging. The national schedule of reference costs report costs as mean (£52) and inter-quartile range (£37-60) only. To capture the uncertainty of this cost appropriately we fitted a gamma distribution to the mean and inter-quartile range. The resulting parameter estimation was a gamma distribution with alpha = 9.2207, and beta = 5.6395. This resulted in a total cost of ultrasound scan for fetal presentation of £52.00 (95% CI: £24.05, 90.55).

Cost for base-case scenario

Since there is genuine uncertainty over the feasibility of providing midwife-led US screening for fetal presentation only, quantifying the reasonable cost for this parameter was problematic. For the base-case scenario, we used a uniform distribution of costs, ranging between the lower end of the 95% cost interval if midwife-led screening was possible (£6.87) and the upper end of the confidence interval for

sonographer-led screening (£90.55). This way, all plausible cost of ultrasound screening for fetal presentation alone was incorporated into the sensitivity and value of information analysis.

Cost per mode of delivery

We obtained data on costs for different modes of deliveries from the national schedule of reference costs.[177] For a (cephalic) vaginal delivery, we used data for a normal delivery without epidural or assistance. For all modes of deliveries, the reference costs were presented for different levels of complications (CC scores), we calculated a weighted average cost for all levels of these. The reference costs reports the mean, as well as the lower and upper inter-quartile range (IQR) for costs, listed by different types of clinical settings (e.g. elective inpatient, non-elective inpatient, outpatient procedures etc.). We calculated a weighted average for the mean/inter-quartile ranges based upon the reported numbers of activities over the year for each setting. For each of the three modes of deliveries (cephalic vaginal, planned CS and emergency CS), we fitted a gamma distribution to the resulting weighted mean/inter-quartile range. For vaginal delivery, this yielded the parameters α = 7.2606 and β = 252.5824, with a total cost of £1,834.47 (95% CI: £1750.43, 2236.05). The corresponding values for planned CS were α = 11.1212 and β = 307.0169, with a total cost of £3,411.93 (95% CI: £2679.80, 4038.29). For emergency CS the values were α = 14.7329 and β = 318.1354, for a total cost of £4,688.27 (95% CI: £3816.15, 5443.02)

Since the National Schedule of Reference Costs does not list separate costs for vaginal breech deliveries, we made the simplifying assumption that these costs would have the same ratio to the costs of elective caesarean section as reported by Palencia et al. (2006).[180] For that study, the costs were Ca$7,255 and Ca$8,440 for elective caesarean section and vaginal breech delivery, respectively, with a mean cost difference of Ca$1,185 (95% CI: $719, $1663). We fitted a normal distribution (mean = 1.1633, sd = 0.0332) to calculate the relative cost increase from vaginal breech delivery compared to elective CS. This yielded a relative cost increase of 1.1633 (95% CI: 1.0982, 1.2284). To obtain the cost of vaginal breech delivery for our model, we then multiplied the cost of elective CS (as calculated above through the NHS reference costs) with the relative cost increase from vaginal breech delivery.

Cost of External Cephalic Version (ECV)

We obtained the cost of external cephalic version (ECV) from the cost analysis of offering ECV in the UK reported by James et al.[181] The authors provided two different estimates of costs, using low (£186.70) and high (£193.30) staff costs, respectively. To convert to 2017's price level, we used the Hospital & Community Health Services (HCHS) inflation index: compared to baseline, the index was

302.3 for year 2017,[178] and 196.5 for year 2001.[211] The resulting cost per ECV was £287.2 and £297.4 for low and high staff costs, respectively. We interpreted this as the feasible range that costs could assume, and let the model sample from this interval using uniform distribution.

Cost of neonatal unit admission

To capture the cost of admission to neonatal care following delivery we used cost data from the NHS reference costs.[177] We divided neonatal critical care into three levels: 'Intensive care', 'High-dependency', and 'Special care'. For intensive and high dependency care we used currency codes XA01Z and XA02Z, respectively, and for special care we used a weighted average of currency codes XA03Z to XA05Z. We assumed that the proportion of admittance to each level of neonatal care and length of stay was the same as the one reported by Alfirevic et al.[182]. This meant that 19, 7, and 74 percent of admitted neonates went to intensive, high dependency, and special care, and that the length of stay was 2, 1.5, and 2 days, respectively. To capture the uncertainty in the cost of care, we fitted a gamma distribution based upon the mean and inter-quartile values, as reported in the reference costs.

To estimate the number of neonates admitted to neonatal care as a function of neonatal morbidity at delivery, we reanalysed data from the POP study.[7] We used Apgar score (5 min) as a proxy for neonatal morbidity at delivery; Apgar score >7, 4-6, and 0-3 were equivalent to no, moderate and severe neonatal morbidity, respectively. This meant that the risk of admittance was 7.4% (95% CI: 6.6-8.2%) with no morbidity, and 47.4% (95% CI: 31.9-63.1%) with moderate morbidity; we modelled this using the Beta distribution. For severe morbidity, we instead made the simplifying assumption that all neonates with severe morbidity would be admitted to neonatal unit due to the small sample size of severe neonatal morbidity in the POP study. In absence of evidence of how the level of neonatal morbidity at birth affects the chance of ending up in each tier of neonatal care, we assumed that the proportions were constant, and that the level of neonatal morbidity only affected the level of overall admittance.

Cost from Respiratory morbidity

Morrison et al. (1995) reports the incidence and length of stay at hospital for respiratory morbidity.[165] 28% of the morbidities consisted of Respiratory Distress Syndrome (RDS) and the rest of Transient Tachypnea of the Newborn (TTN). The average stay at Neonatal Intensive Care Unit (NICU) was 4 days for RDS and 0.6 days for TTN. The NHS cost of NICU admission is £1,295 per day (inter-quartile range: £1,015-1,541).[177] Given this, the average cost for a case of RDS is £5,180 (IQR: £4,060-6,164), and the

cost for TTN is £777 (IQR: 609-925). Assuming that RDS and TTN makes up 28% and 72% of respiratory morbidities, respectively, the average cost of a case of respiratory morbidity would be £2,010 (IQR: £1,575-2,392). Due to the very low mortality rate from respiratory distress among babies born at term, we made the simplifying assumption that respiratory distress could lead to NICU admission, but would otherwise have no consequences.(Malloy and Freeman, 2000) In order to capture the uncertainty of the cost of respiratory morbidity in one parameter, we fitted a gamma distribution based upon the mean and inter-quartile range. The resulting distributions had parameters alpha = 10.7125 and beta = 187.6316, yielding a total cost of 2011 (95% CI: 993, 3381).

Cost of acidosis without long-term consequences

In the absence of data on the costs associated with short-term acidosis (i.e. acidosis that require neonatal treatment, but resolves without any other health consequences), we made the simplifying assumption that treatment would be required at the neonatal intensive care unit (NICU) for 1-4 days, with equal probabilities. To obtain per-day costs, we fitted a gamma distribution for the unit cost of NICU care using cost data from the NHS reference costs[177], based upon mean and interquartile range. Combining the time and per-day costs, we obtained a total cost distribution. To be able to capture total cost uncertainty in a single parameter, we fitted a gamma distribution to the total cost. The resulting parameter (alpha = 3.6143 and beta = 895.6169) had a total cost of £3,240 (95% CI: £806-7,328).

Cost of transient and permanent BPI

To estimate the costs associated with Brachial Plexus Injury (BPI) we assumed the same resource usage as reported by Culligan et al. (2010)[183]. Transient BPI costs included a hospital consultation by a specialist, weekly physical therapy for 4 months, and one needle electromyography (EMG) test. Permanent BPI costs included the costs from transient BPI but with weekly physical therapy for 3 years instead, plus one outpatient visit to a specialist, and magnetic resonance imaging (MRI) of the shoulder.[183] We obtained costs for the specialist consultations and weekly physiotherapy treatments from the unit costs of health and social care; these were £199 and £87, respectively.[212] The costs for EMG and MRI were taken from the NHS reference costs (AA33D and RD01C)[177]; these were £269.2 and £106.59, respectively. All costs were updated to the price year 2016-17 using the HCHS index.[178] We assumed that all costs except for physiotherapy arose in the first year of life and discounted accordingly; the discount rate was 3.5% as recommended by NICE.[190](National Institute for Health and Clinical Excellence, 2013) The total discounted costs from transient and permanent BPI were £2,066 and £14,133, respectively.

To account for uncertainty, Culligan et al.[183] expanded their cost estimate into a plausible range of costs, which ranged between 50-200% of the point estimate. However, directly incorporating this plausible range into our own estimation (after adjusting for cost differences) by using uniform distribution would have been inappropriate, since this would overestimate costs. Instead, we interpreted the plausible range as a 95% confidence interval (CI) for total costs, and then fitted a log-normal distribution to the appropriate mean and CI range. This way, the lower and upper 95% CI was still 50% and 200% of the point estimate, respectively, but in this case following a log-normal distribution. For transient BPI, the resulting distribution had a logged standard error of 0.3536, and the total costs were £2,066 (95% CI: £1033 – 4132). The corresponding figures for permanent BPI were a logged standard error of 0.3536, and a total cost of £14,133 (95% CI: £7067-28264).

Cost of perinatal death

We used the cost of stillbirth as a proxy for the cost of perinatal death. The direct costs of stillbirth were obtained from Mistry et al.[184] The authors estimated that the costs would be between £1,242 (core investigation and counselling only), up to £1,804 depending on the clinical scenario surrounding the stillbirth and what tests were needed. The authors choose not to present a most plausible estimate within this estimate, but instead just reported these costs as the full range of costs for stillbirth. For this reason, we interpreted these costs as the upper and lower boundaries that the cost of perinatal death could reasonably assume. We updated these costs to the price year of 2016-17 (The original source used price year 2010) using the HCHS index[178], and used a uniform distribution.

Cost of special educational needs

We obtained the cost of special educational needs (SEN) from Barrett et al.[185], using the difference in costs to typically developing groups. The cost difference was £6,315 (95% CI: £3798, 8832) These costs were estimated for the cost year of 2007-08, hence we inflated this to the value of price year 2016-17 using the HCHS index[178]; resulting in a cost difference of £7,428 (95% CI: £4467, 10389). This cost was applied annually for years 6-17 of life (the typical school years) and discounted using a discount rate of 3.5% as recommended by the National Institute for Health and Care Excellence (NICE)[190].

The cost of severe neurological morbidity

We used cerebral palsy (CP) as a proxy for severe neurological morbidity. In the absence of English cost data detailed enough to provide an annual cost for the relevant payer perspective, we instead obtained the annual cost of CP from Cerebral Palsy Australia[186]. We used total per capita cost for the health system, as well as indirect costs (e.g. program services, aids, and home modifications), but we

omitted productivity losses, dead weight losses from financial transactions, and costs for informal carers. The annual average cost per case of CP in 2005 was 5,362 AUD. We converted this to sterling pound (£) by the exchange rate of 31-12-2005, and updated to the price level of 2016/17 using the hospital & community health services (HCHS) index.[178] This meant a total annual cost of £2,929.6. Since the data was derived from the nationwide population of people with CP, this average annual cost is applicable for any year of life.

Capturing the uncertainty in these costs were problematic as costs are not easily transferable between different healthcare systems. Further, Cerebral Palsy Australia did not provide any estimates of cost uncertainty. For this reason, we chose to assume that English costs could reasonably fluctuate between half and double those quoted in Australia. We interpreted this as a 95% confidence interval stretching between £1465 and £5859, and fitted a log-normal distribution to this interval.

Quality of life

Baseline long-term Quality-Adjusted Life Years

In the absence of neonatal morbidity at birth, lifetime quality-adjusted life years (QALYs) were calculated using survival and Quality of Life (QoL) weights for a general UK population. Survival rates were obtained from the Office for National Statistics.[188] These were adjusted using age-specific QoL data from EuroQol. The QoL for each age group was modelled using a normal distribution with mean and standard errors as provided by EuroQol for the UK using the time trade-off (TTO) method.[187] We finally limited the total QALYs to the model's time horizon and discounted these QALYs, using a discount rate of 3.5% as recommended by NICE.[190]

Quality of life for brachial plexus injury (BPI)

We obtained the estimated quality of life following brachial plexus injury (BPI) from Culligan et al.[183] These data were estimated as a plausible range by an expert panel, and the authors used a uniform distribution within the plausible range. The authors provided separate estimates for different complexity levels of BPI. We assumed that long-term BPI in the context of our model would be equivalent to either 'Permanent brachial plexus injury (mild to moderate)', or 'Permanent brachial plexus injury (severe) and uncomplicated delivery'. We therefore chose to consider the plausible range to stretch between 0.30 (the lower boundary for severe BPI) and 0.70 (the upper boundary for mild to moderate BPI).

Long-term health outcomes following severe neurological morbidity

To get an estimate of the long-term consequences from severe neurological morbidity, we constructed a model based upon the work by Leigh et al.[189], using cerebral palsy as a proxy for severe neurological morbidity. Analogous to Leigh et al., we divided all cases of cerebral palsy into five levels according to the Gross Motor Function Classification System (GMFCS), system describing the ambulatory functionality of people with CP.[213] We obtained the GMFCS-specific quality of life (QoL) by letting the model sample values from the gamma distribution provided by Leigh et al., subtracting these values from 1 (highest possible QoL) to provide utility weights. A benefit of using these QoL weights was that they were derived using EQ-5D[214], facilitating comparison to the QoL of the general population. We let QoL decrease over time at the same rate as Leigh et al. hence indirectly assuming that ageing has no greater effect on QoL for those with CP than otherwise healthy members of the UK.

Since cerebral palsy affect mortality as well as QoL, we had to adjust the model for survival. We calculated GMFCS-specific survival rates, using the average mortality rates provided by Leigh et al. for each GMFCS and age group (0-10 years, 11-20 years, and 21-30 years). Unlike for Leigh et al., our model was not probabilistic in regards to survival; parameter uncertainty was restricted to QoL only. In the absence of evidence on GMFCS-specific mortality rates beyond 30 years of life, we made the conservative assumption that the mortality rate for those born with severe neurological morbidity would mimic the general population in the UK after this age.

We obtained the distribution of GMFCS states from Young et al.[215] and captured the parameter uncertainty of the distribution by letting the model sample input values from the data; we sampled using Dirichlet distribution.

Combining QoL with survival, we obtained expected lifetime QALYs for neonates born with severe neurological morbidity. We finally limited the total QALYs to the model's time horizon and discounted these QALYs, using a discount rate of 3.5% as recommended by NICE.[190]

Appendix 8. 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 (don't want scan)	Disagree	Neither agree nor disagree	Agree	Strongly agree (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 (eg 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