Diagnostic accuracy of biochemical tests of placental function versus ultrasound assessment of fetal size for stillbirth and small-for-gestational-age infants

Heazell, A; Hayes, DJL; Whitworth, M; Takwoingi, Yemisi; Bayliss, Susan; Davenport, Clare

DOI:
10.1002/14651858.CD012245

Document Version
Publisher's PDF, also known as Version of record

Cochrane Library

Cochrane Database of Systematic Reviews

Diagnostic accuracy of biochemical tests of placental function versus ultrasound assessment of fetal size for stillbirth and small-for-gestational-age infants (Protocol)

Heazell AEP, Hayes DJL, Whitworth M, Takwoingi Y, Bayliss SE, Davenport C

Heazell AEP, Hayes DJL, Whitworth M, Takwoingi Y, Bayliss SE, Davenport C.
Diagnostic accuracy of biochemical tests of placental function versus ultrasound assessment of fetal size for stillbirth and small-for-gestational-age infants.
Cochrane Database of Systematic Reviews 2016, Issue 6. Art. No.: CD012245.
DOI: 10.1002/14651858.CD012245.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Protocol]

Diagnostic accuracy of biochemical tests of placental function versus ultrasound assessment of fetal size for stillbirth and small-for-gestational-age infants

Alexander EP Heazell[1], Dexter JL Hayes[2], Melissa Whitworth[1], Yemisi Takwoingi[3], Susan E Bayliss[4], Clare Davenport[4]

[1]Maternal and Fetal Health Research Centre, University of Manchester, Manchester, UK. [2]Department of Maternal and Fetal Health, University of Manchester, Manchester, UK. [3]Institute of Applied Health Research, University of Birmingham, Birmingham, UK. [4]Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham, UK

Contact address: Alexander EP Heazell, Maternal and Fetal Health Research Centre, University of Manchester, 5th floor (Research), St Mary's Hospital, Oxford Road, Manchester, M13 9WL, UK. alexander.heazell@manchester.ac.uk. alex_heazell@talk21.com.

Editorial group: Cochrane Pregnancy and Childbirth Group.
Publication status and date: New, published in Issue 6, 2016.

Citation: Heazell AEP, Hayes DJL, Whitworth M, Takwoingi Y, Bayliss SE, Davenport C. Diagnostic accuracy of biochemical tests of placental function versus ultrasound assessment of fetal size for stillbirth and small-for-gestational-age infants. *Cochrane Database of Systematic Reviews* 2016, Issue 6. Art. No.: CD012245. DOI: 10.1002/14651858.CD012245.

ABSTRACT

This is the protocol for a review and there is no abstract. The objectives are as follows:

The primary objective of this review is to assess and compare the diagnostic accuracy of ultrasound assessment of fetal growth and placental biomarkers alone and in any combination used after 24 weeks of pregnancy in the identification of placental dysfunction as evidenced by either stillbirth or born small-for-gestational age (SGA). Accuracy is described by the proportion of fetuses who are subsequently stillborn or who have a SGA baby detected by a positive test result (the presence of placental dysfunction) (sensitivity) and by the proportion of fetuses that have an uncomplicated pregnancy following a negative index tests result (absence of placental dysfunction) (specificity).

We will investigate the effect of clinical (patient and test characteristics) and methodological factors (study design, threshold used to define SGA) on test performance. The clinical factors include patient group (low-risk or high-risk pregnancies), gestation at measurement, ethnicity, maternal age and method of testing. With regard to methodological variation, studies may include an intervention (delivery or additional fetal surveillance for test positive cases) which will impact on the outcome; therefore we will assess whether this is a source of heterogeneity.

BACKGROUND

Stillbirth affects 2.6 million pregnancies worldwide each year (Cousens 2011). Although the majority of cases occur in low- and middle-income countries, stillbirth places a significant burden in high-income countries (HICs) with the UK and the US reporting rates above the mean for HICs (Flenady 2011b). In HICs, the most frequently reported association with stillbirth is placental dysfunction, which may be clinically evident as fetal growth restriction (FGR), small-for-gestational-age (SGA) infants, placental abruption or hypertensive disorders of pregnancy. Placen-

tal abnormalities are noted in 11% to 65% of stillbirths (Ptacek 2014). Identification of FGR is difficult in utero and even after birth, with SGA being most commonly used as a surrogate measure (Worton 2014). The degree of SGA is associated with the likelihood of FGR; 30% of infants with a birthweight <10th centile are thought to be FGR, while 70% of infants with a birthweight < third centile are thought to be FGR. Critically, SGA is the most significant antenatal risk factor for a stillborn infant (Flenady 2011a; Gardosi 2013; McCowan 2007). Importantly, correct identification of SGA infants is associated with a reduction in the perinatal mortality rate (Gardosi 2013). However, currently used tests, such as measurement of symphysis-fundal height, have a low reported sensitivity and specificity for the identification of SGA infants (RCOG 2014).

Due to the importance of the placenta there is growing interest in antenatal placental evaluation in an attempt to identify pregnancies at increased risk of stillbirth or fetal compromise (Heazell 2015a). A systematic review of biochemical tests of placental function found insufficient evidence to conclude whether these interventions had any effect on perinatal mortality or fetal compromise (Heazell 2015b). In contrast, a single trial of placental grading assessed by ultrasound demonstrated reduced perinatal mortality (Proud 1987). Systematic reviews of other methods employed to identify fetal compromise such as ultrasound assessment of fetal growth or umbilical artery Doppler (measurement of blood flow through the umbilical artery) in late pregnancy have also found insufficient evidence to conclude whether these interventions reduce perinatal mortality in a low-risk maternity population (Alfirevic 2015; Bricker 2015), although both are effective in women deemed to be at high risk of pregnancy complications (Alfirevic 2013). The efficacy of umbilical artery Doppler in high-risk populations may be due to its prognostic accuracy; a systematic review found this test predicted SGA infants with a positive likelihood ratio of 3.76 and stillbirth with a positive likelihood ratio of 4.37 (Morris 2011).

Two components are necessary to reduce perinatal mortality and minimise unwarranted intervention. Firstly, the test must accurately identify fetal compromise and secondly, the intervention must be effective in preventing the adverse outcome. There is now strong evidence that planned delivery (by induction of labour) after 37 weeks of pregnancy is associated with a reduction in perinatal mortality (Stock 2012). Therefore, the most accurate test to identify fetal compromise needs to be determined so that it may be combined with planned delivery.

Target condition being diagnosed

The target condition of interest is placental dysfunction - which describes the condition in which the placenta does not meet the demands of the fetus (Heazell 2015a). As with other organ dysfunction, there are multiple pathways that can result in placental dysfunction including vascular, inflammatory, infective and genetic disorders. These various processes may lead to changes in placental structure and/or function that may lead to two clinical outcomes i) stillbirth or ii) an SGA infant. As placental dysfunction cannot easily be quantified this review will use these two clinical outcomes as the target conditions of interest.

Index test(s)

This review will evaluate tests used in late pregnancy (after 24 weeks) to identify pregnancies that have placental dysfunction informing decisions to continue with the pregnancy or institute intervention. The tests to be evaluated include assessment of placental structure and biochemical function by ultrasound scan or measurement of placental products in maternal blood (plasma or serum) or urine.

Biochemical tests of placental function measure placental products (proteins, peptides, metabolites) in maternal biofluids (serum, plasma, urine); it is hypothesised that levels of such products in maternal fluids reflect endocrine and metabolic functions of the placenta. Many placental products can be detected in maternal biofluids including protein hormones: human chorionic gonadotrophin (hCG), human placental lactogen (hPL), human placental growth hormone (hPGH), placental growth factor (PlGF), placental protein-13 (PP-13), pregnancy specific glycoproteins and steroid hormones including oestrogens and progesterone with their related metabolites. Ultrasonography has been used to measure the size, shape and echotexture of the placenta; the majority of such studies have used 2D ultrasound to evaluate placental morphology, although newer studies have utilised 3D techniques. We have not included umbilical artery Doppler in this analysis as a systematic review and meta-analysis has been conducted (Morris 2011).

Clinical pathway

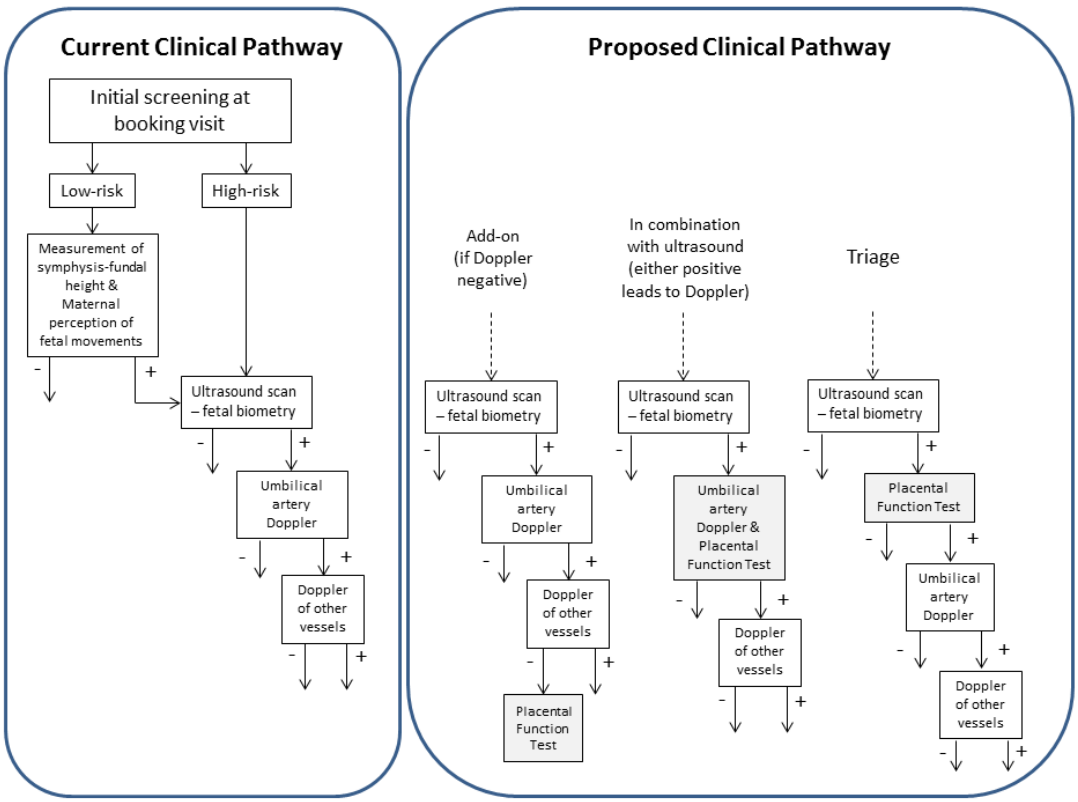
Antenatal care differs between countries, the clinical pathway described here applies to the UK and follows guidance from the Royal College of Obstetricians and Gynaecologists (RCOG 2014) and the National Institute for Health and Social Care Excellence (NICE 2008).

Prior test(s)

Currently, in the UK women are grouped into high risk and low risk for SGA in early pregnancy at the booking-in visit by assessing a woman's past medical history, obstetric history and risk factors for an SGA infant (RCOG 2014). All woman are offered screening for Down's syndrome (which is currently based on measurement of nuchal translucency by ultrasound scan and measurement of serum analytes between 11 and 13^{+6} weeks of pregnancy) and for fetal anomaly (by ultrasound scan from 18 to 20^{+6} weeks).

In clinical practice, placental dysfunction is suspected by identification of an SGA infant. However, subsequent testing for SGA will depend upon the risk status of the woman (RCOG 2014). The National Institute for Health and Care Excellence do not recommend routine measurement of fetal growth by ultrasound scan in late pregnancy (NICE 2008). Fetal growth is assessed in women deemed to be at low risk of an SGA infant by measurement of symphysis-fundal height with a tape measure (RCOG 2014). Women at increased risk of SGA are recommended to have a uterine artery Doppler (to assess blood flow through both uterine arteries) at 20 weeks' gestation and regular scans to measure fetal biometry with assessment of liquor volume and umbilical artery Doppler. Umbilical artery Doppler is the most frequently employed test to predict fetal outcome; the relationship between umbilical artery Doppler indices and placental function is not clear. In addition to recommendations for the diagnosis and management of an SGA fetus, ultrasound assessment of fetal growth, liquor volume and umbilical artery Doppler are recommended following maternal presentation with reduced fetal movements, as this may be a symptom of placental insufficiency (RCOG 2011). The current clinical pathway is shown in Figure 1.

Figure 1. Current clinical pathway and three proposed uses of a placental function test. Currently, women are screened for a small for gestational age fetus (as a proxy for placental dysfunction) using symphysis-fundal height and maternal awareness of fetal movements. Women deemed to be an increased risk are screened using ultrasound measurement of fetal biometry. We propose three different clinical pathways for placental function tests. Firstly, they could be used as an additional test when Doppler measurements were normal. They could be used in combination with currently used tests and finally they could be used as a triage test to differentiate infants who are constitutionally small from those with placental dysfunction. Although treatment decisions would be tailored to individual cases, a positive test would be expected to lead to increased surveillance or intervention (planned delivery) and negative test would lead to continuing with the pregnancy.



When an SGA infant is identified by tests, clinical management is dependent upon gestation. At ≥ 37 weeks delivery is offered (RCOG 2014). Prior to this gestation, fetal compromise is assess by measurement of Doppler waveforms primarily in the umbilical artery, but may also include the middle cerebral artery and ductus venosus. Delivery is recommended when fetal compromise is identified (RCOG 2014).

There are currently no routinely used measures of placental function after 16 weeks of pregnancy. There is evidence that measurement of placental analytes as part of screening for aneuploidy may identify fetuses at high risk of early-onset FGR (Smith 2002; Smith 2006). Assessment of these analytes is incorporated into the current clinical pathway (RCOG 2014); women with low pregnancy-associated plasma protein A (PAPP-A) levels are managed as high risk for SGA. Therefore, we wish to focus on placental tests performed in late pregnancy (after 24 weeks' gestation).

Role of index test(s)

Due to the established use of ultrasound in obstetric practice, we envisage that additional tests of placental function would most likely be added to an ultrasound measurement of fetal size rather than replacing it (Figure 1); this is certainly true of the intervention trials of placental assessment (by biochemical tests) that have been conducted (Duenholter 1976; Heazell 2013; Sharf 1984). It is hypothesised the addition of a placental function test to an ultrasound scan would improve identification of an SGA infant and consequently focus intervention on pregnancies at greatest risk of stillbirth or fetal compromise, thereby reducing the burden of perinatal mortality and morbidity. It is also possible that a placental function test could be used to triage infants who were SGA to identify which were constitutionally small and which had placental dysfunction. This would allow the pregnancy to continue in otherwise healthy constitutionally small infants, reducing unnecessary intervention. If placental function tests were used in this way a false positive test would lead to increased intervention, either surveillance or delivery, and a false negative test would allow a potentially compromised baby to remain in utero without increased surveillance. From the perspective of reducing perinatal mortality and morbidity, a false negative test would be more harmful than a false positive test.

Alternative test(s)

Presently, there are no tests in widespread clinical use that directly assess placental biochemical function.

Rationale

There are several tests of placental structure and function. Systematic reviews of the measurement of biochemical placental factors and the effectiveness of ultrasound in late pregnancy found that few tests of placental structure or function have been evaluated in robust intervention studies (Bricker 2015; Heazell 2015a). This review aims to identify and evaluate tests of placental structure and function, not restricted to those evaluated in intervention studies, to determine which measurement(s) have the greatest diagnostic accuracy for detection of placental dysfunction leading to stillbirth and SGA. The most accurate test(s) can then be taken forward into intervention studies to determine whether performing investigations can reduce perinatal morbidity or mortality.

OBJECTIVES

The primary objective of this review is to assess and compare the diagnostic accuracy of ultrasound assessment of fetal growth and placental biomarkers alone and in any combination used after 24 weeks of pregnancy in the identification of placental dysfunction as evidenced by either stillbirth or born small-for-gestational age (SGA). Accuracy is described by the proportion of fetuses who are subsequently stillborn or who have a SGA baby detected by a positive test result (the presence of placental dysfunction) (sensitivity) and by the proportion of fetuses that have an uncomplicated pregnancy following a negative index tests result (absence of placental dysfunction) (specificity).

Secondary objectives

We will investigate the effect of clinical (patient and test characteristics) and methodological factors (study design, threshold used to define SGA) on test performance. The clinical factors include patient group (low-risk or high-risk pregnancies), gestation at measurement, ethnicity, maternal age and method of testing. With regard to methodological variation, studies may include an intervention (delivery or additional fetal surveillance for test positive cases) which will impact on the outcome; therefore we will assess whether this is a source of heterogeneity.

METHODS

Criteria for considering studies for this review

Types of studies

Presently, there are no effective interventions to reverse placental dysfunction *in utero*. This means that once detected, an intervention cannot be employed to reverse the small-for-gestational age (SGA) phenotype following a positive test result. Delivery may be indicated, although at earlier gestations this does not affect perinatal mortality (GRIT 2003). Therefore, we will include prospective and retrospective cross-sectional or cohort studies in which all women receive one or more index tests and the outcome of their pregnancy is known. Case-control studies will be excluded.

We will include studies which measure index tests on one occasion (cross-sectional design).
We will exclude studies where it is not possible to derive a 2 x 2 table of the number of true positives, false positives, false negatives and true negatives, or studies that report preliminary experimental findings, i.e. laboratory-based studies.

Participants

We will include studies of pregnant women after 24 weeks' gestation that record relevant outcomes of pregnancy (live birth/stillbirth; SGA infant).
We will include studies of pregnant women of any reproductive age, who are deemed to be low or high risk for complications (e.g. who have pre-existing medical disorders or previous stillbirth) or studies of mixed populations (of low and high risk for complications).
We will exclude pregnancies complicated by fetal abnormalities, as they often have a higher risk of stillbirth from non-placental causes.
We will exclude studies of women with multi-fetal pregnancies.

Index tests

We will include, but not be restricted to, the following index tests of placental biochemical function, placental structure or assessment of fetal biometry to identify an SGA infant:

- human placental lactogen (hPL) in maternal urine/blood;
- oestriol in maternal urine/blood;
- placental growth factor in maternal blood;
- ultrasound assessment of placental echogenicity;
- ultrasound assessment of fetal size.

With regard to biochemical tests, we will include assays that have been performed using different techniques, including: immunoassay, enzyme-linked immunosorbent assay, chromatography or point of care test in any combination and at any threshold used to determine test positivity. Examples of current commercially available tests are listed in Appendix 1.

Target conditions

The target conditions are stillbirth and delivery of a SGA infant as clinical manifestations of placental dysfunction.

Reference standards

The outcome of pregnancy is considered as the reference standard. A "positive" result will be either i) a stillbirth - an infant born with no signs of life after 24 weeks' gestation, or ii) a birthweight classified as SGA. A "negative" result will be a live birth after 24 weeks' gestation or a birthweight classified as appropriate for gestational age.
The classification of SGA will be determined according to the definition used in the study. Where possible the definition of an infant with a birthweight $\leq 10^{th}$ centile using a customised birthweight calculator will be used (Clausson 2001). Where this is not possible, the definition of SGA from the manuscript will be used and recorded. If there are sufficient studies, the effects of different definitions of SGA will be addressed as a potential source of heterogeneity.

Search methods for identification of studies

We will conduct a comprehensive search for existing systematic reviews and primary studies relevant to the prevention of adverse pregnancy outcome in women at increased risk of stillbirth by detecting placental dysfunction. A scoping search was undertaken in the bibliographic databases MEDLINE, MEDLINE In Process, Embase, the Cochrane Library (CDSR, DARE, HTA, NHS EED and Central Register of Controlled Trials (CENTRAL) databases), HTA and relevant web sites in order to identify existing reviews and to gauge the nature and number of relevant studies to inform the protocol.

Electronic searches

We will develop full search strategies based on the scoping searches, expert advice, and consultation with the Cochrane Pregnancy and Childbirth Group's Information Specialist. They will include a combination of text words and index terms. Methodological search filters for diagnostic test accuracy will be avoided if possible as they have been shown to miss relevant studies (Whiting 2011a). We will not apply any language or date restrictions. We plan to search the following sources:

- bibliographic databases - MEDLINE, MEDLINE In Process and Embase via Ovid, Cochrane (Wiley) CENTRAL, Science Citation Index (Web of Science), CINAHL (EBSCO) with search strategies adapted for each database as required;
- ISRCTN Registry, UK Clinical Trials Gateway, WHO International Clinical Trials Portal (ICTRP) and ClinicalTrials.gov for ongoing studies;
- specialist abstract and conference proceeding resources (British Library's ZETOC and Web of Science Conference Proceedings Citation Index).

A sample search strategy for MEDLINE is provided in Appendix 2.

Searching other resources

We will check citation lists of included studies and relevant reviews. We will examine grey literature by searching websites of companies producing biochemical tests of placental function (Alere 2015; Perkin Elmer 2015; Roche 2015). We will also undertake consultation with experts in the field to access relevant unpublished data.

Data collection and analysis

We will use the methods described in the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (http://dta.cochrane.org/handbook-dta-reviews).

Selection of studies

Two review authors will independently screen the titles and abstracts of all studies identified by the search strategy. We will obtain full-text versions of all potentially relevant studies. Two review authors will independently assess studies for inclusion using prespecified inclusion criteria stated earlier. We will include studies of pregnant women after 24 weeks' gestation that record relevant outcomes of pregnancy (live birth/stillbirth; SGA infant), and present data to construct a 2 x 2 table. We will resolve any disagreement between the two review authors or by discussion with a third party if needed. Reasons for study exclusion will be documented.

Data extraction and management

We will develop a customised form to ensure reproducible collection of data items. Data collection will be piloted on five manuscripts then reviewed by the review authors. Data will be extracted independently by two review authors. We will resolve discrepancies, where they occur, through discussion of if required we will consult a third author drawing on clinical and methodological expertise in the team as appropriate to the content of the query. We will extract characteristics of participants, index tests or test combinations (including thresholds used), and details of the reference standard in terms of pregnancy outcome (live birth or stillbirth) and whether the infant was SGA. For studies that report data at multiple thresholds for a test, we will extract a 2 x 2 table at each reported threshold. Where possible we will record the frequency of obstetric intervention and infant admission to neonatal intensive care. If reported, we will also record data on outcomes including harms of testing, need for further testing, and the effects of the test. We will not address women's experiences of testing, caregiver's satisfaction with testing or economic evaluation of testing as this is beyond the scope of this review.

We will attempt to contact the authors of included studies where information considered key to assessment of methodological quality, investigation of heterogeneity, or completion of a 2 x 2 table is unclear or missing. Studies published only as conference abstracts will be followed up to identify whether a subsequent full paper has been published.

Assessment of methodological quality

We will use the QUADAS-2 tool (Whiting 2011b) to assess the risk of bias and applicability of included studies. We will tailor the tool to our review question using the operational criteria detailed in Appendix 3 to answer signalling questions and make the overall judgement of risk of bias and applicability concerns for each domain of the tool. Two review authors will assess each included study separately. We will resolve differences in assessment through discussion and if required, by discussion with a third person. We will assess each criteria in QUADAS-2 as "yes", "no" or "unclear" and summarise the results graphically or in tables (Appendix 3).

We will include all domains of QUADAS-2 assessment including the time interval between testing and the outcome and any intervention as these may alter the outcome. We have operationalised the domains of the QUADAS-2 tool for the clinical context of this review. For example, the domain concerning patient selection was amended to ensure to allow appropriate exclusion criteria for studies as placental function tests may be altered in women with multiple pregnancies or with fetal abnormalities. However, other criteria that would not be expected to alter tests of placental function (e.g. ethnicity, maternal age and income) would be inappropriate exclusions. It is essential that the sample is generalisable to those in the review question, whereas studies may be restricted to specific high-risk groups, e.g. maternal hypertension, which will reduce the applicability. Studies may also use varied measures of SGA (the reference standard); some of which are unrelated to gestation, e.g. low birthweight (< 2.5 kg) which are less accurate and may reduce study quality. Studies using a threshold which alters with gender and gestation, e.g. individualised birthweight centile will be rated more highly than those which do not.

Statistical analysis and data synthesis

For each test and type of biofluid, estimates of sensitivity and specificity from each study will be plotted in receiver operating characteristic (ROC) space and forest plots for preliminary investigations of the data. We anticipate that studies will use different thresholds to dichotomise tests measured on a continuous scale. Therefore, we plan to perform meta-analyses using the hierarchical summary ROC (HSROC) model (Rutter 2001) to estimate SROC curves in RevMan 2014. Where a study reports multiple thresholds for a test, we will select one threshold at random or the threshold most frequently reported across studies so that only one 2 x 2 table is included in the meta-analysis. Methods that allow joint synthesis of sensitivities and specificities at multiple thresholds have been proposed, but are not used in practice and require further evaluation before they can be used in Cochrane reviews (Macaskill 2010). In separate analyses, where studies report common thresholds (e.g. for placental growth factor (PlGF) < 12 pg/mL, Alere 2015), we will also estimate summary sensitivities and specificities using functions of HSROC model parameters.

The main test comparison will be an indirect comparison pooling all relevant studies that assessed at least one of the index tests. In secondary analyses, we will perform direct comparisons by restricting the analyses to only studies that have compared tests head-to-head in the same study population. This analytical strategy was adopted because of the paucity of comparative studies of diagnostic accuracy (Takwoingi 2013). If we identify a large number of tests, because of potential model complexity and additional num-

ber of model parameters, we will limit the indirect comparison to only those tests that provide adequate data, for example, at least four studies. For direct comparisons, we will perform pair-wise comparisons of tests. The test comparisons will be performed by adding a covariate for test type to the HSROC model to estimate differences in accuracy, threshold, and/or shape of SROC curves. We will assess the statistical significance of differences between tests using likelihood ratio tests comparing models with and without the covariate terms. The NLMIXED procedure in the SAS software package (version 9.3; SAS Institute, Cary, NC, USA) will be used for meta-analyses.

Investigations of heterogeneity

We will initially examine heterogeneity between studies by visually inspecting forest plots of sensitivity and specificity and summary ROC plots. Where a sufficient number of studies assess the same index test, potential sources of heterogeneity will be separated into clinical (e.g. population studied, test type) and methodological (as appropriate) sources. We will define whether these potential sources of heterogeneity are dichotomous or continuous variables and will perform appropriate meta-regression by adding the potential source of heterogeneity as a covariate to the hierarchical model. We will test for statistical difference in observed variation by comparing shapes of the SROC curve for different variables (e.g. type method, threshold used).

Sensitivity analyses

If there are sufficient studies of an individual index test, we will perform the following sensitivity analyses by restricting analyses to studies:

- without an intervention that may have altered outcome;
- at low risk of bias in each of the four domains of the QUADAS-2 tool (Appendix 3);
- that specifically describe histological evidence of placental insufficiency.

Assessment of reporting bias

We will not undertake any formal assessment of reporting bias in our review due to current uncertainty about how to assess reporting bias in diagnostic test accuracy reviews, especially in the presence of heterogeneity (Macaskill 2010).

ACKNOWLEDGEMENTS

Alexander Heazell is supported by funding from Tommy's - the baby charity and by a Clinician Scientist Fellowship (CS-2013-009) from the National Institute for Health Research (NIHR).

As part of the pre-publication editorial process, this protocol has been commented on by three peers (an editor and two referees who are external to the editorial team) and a member of the Pregnancy and Childbirth Group's international panel of consumers.

This project was supported by the National Institute for Health Research, via Cochrane Infrastructure funding to Cochrane Pregnancy and Childbirth. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, NHS or the Department of Health.

REFERENCES

Additional references

Alere 2015
Alere. Alere Triage® PlGF Test. http://www.alere.com/ww/en/product-details/triage-plgf-test.html (accessed 27th August 2015).

Alfirevic 2013
Alfirevic Z, Stampalija T, Gyte GML. Fetal and umbilical Doppler ultrasound in high-risk pregnancies. *Cochrane Database of Systematic Reviews* 2013, Issue 11. [DOI: 10.1002/14651858.CD007529.pub3]

Alfirevic 2015
Alfirevic Z, Stampalija T, Medley N. Fetal and umbilical Doppler ultrasound in normal pregnancy. *Cochrane Database of Systematic Reviews* 2015, Issue 4. [DOI: 10.1002/14651858.CD001450.pub4]

Bricker 2015
Bricker L, Medley N, Pratt JJ. Routine ultrasound in late pregnancy (after 24 weeks' gestation). *Cochrane Database of Systematic Reviews* 2015, Issue 6. [DOI: 10.1002/14651858.CD001451.pub4]

Clausson 2001
Clausson B, Gardosi J, Francis A, Cnattingius S. Perinatal outcome in SGA births defined by customised versus population-based birthweight standards. *BJOG: an international journal of obstetrics and gynaecology* 2001;**108** (8):830–4.

Cousens 2011
Cousens S, Blencowe H, Stanton C, Chou D, Ahmed S, Steinhardt L, et al. National, regional, and worldwide estimates of stillbirth rates in 2009 with trends since 1995: a systematic analysis. *Lancet* 2011;**377**(9774):1319–30.

Duenholter 1976
Duenhoelter JH, Whalley PJ, MacDonald PC. An

analysis of the utility of plasma immunoreactive estrogen measurements in determining delivery time of gravidas with a fetus considered at high risk. *American Journal of Obstetrics and Gynecology* 1976;**125**(7):889–98.

Flenady 2011a

Flenady V, Koopmans L, Middleton P, Frøen JF, Smith GC, Gibbons K, et al. Major risk factors for stillbirth in high-income countries: a systematic review and meta-analysis. *Lancet* 2011;**377**(9774):1331–40.

Flenady 2011b

Flenady V, Middleton P, Smith GC, Duke W, Erwich JJ, Khong TY, et al. Stillbirths: the way forward in high-income countries. *Lancet* 2011;**377**(9778):1703–17.

Gardosi 2013

Gardosi J, Madurasinghe V, Williams M, Malik A, Francis A. Maternal and fetal risk factors for stillbirth: population based study. *BMJ* 2013;**346**:f108.

GRIT 2003

GRIT Study Group. A randomised trial of timed delivery for the compromised preterm fetus: short term outcomes and Bayesian interpretation. *BJOG: an international journal of obstetrics and gynaecology* 2003;**110**(1):27–32.

Heazell 2013

Heazell AE, Bernatavicius G, Roberts SA, Garrod A, Whitworth MK, Johnstone ED, et al. A randomised controlled trial comparing standard or intensive management of reduced fetal movements after 36 weeks gestation--a feasibility study. *BMC Pregnancy Childbirth* 2013;**13**:95. [DOI: 10.1186/1471-2393-13-95]

Heazell 2015a

Heazell AE, Worton SA, Higgins LE, Ingram E, Johnstone ED, Jones RL, et al. IFPA Gábor Than Award Lecture: Recognition of placental failure is key to saving babies' lives. *Placenta* 2015;**36**(Suppl 1):S20–8.

Heazell 2015b

Heazell AEP, Whitworth MK, Duley L, Thornton JG. Use of biochemical tests of placental function for improving pregnancy outcome. *Cochrane Database of Systematic Reviews* 2015, Issue 11. [DOI: 10.1002/14651858.CD011202.pub2]

Macaskill 2010

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi YA. Analysing and presenting results. *Cochrane Handbook of Diagnostic Test Accuracy Reviews*. 1.0. Cochrane Diagnostic Test Accuracy Group, 2010.

McCowan 2007

McCowan LM, George-Haddad M, Stacey T, Thompson JM. Fetal growth restriction and other risk factors for stillbirth in a New Zealand setting. *Australian & New Zealand Journal of Obstetrics & Gynaecology* 2007;**47**(6):450–6.

Morris 2011

Morris RK, Malin G, Robson SC, Kleijnen J, Zamora J, Khan KS. Fetal umbilical artery Doppler to predict compromise of fetal/neonatal wellbeing in a high-risk population: systematic review and bivariate meta-analysis. *Ultrasound in Obstetrics & Gynecology* 2011;**37**(2):135–42.

NICE 2008

National Institute for Health and Clinical Excellence. *Clinical Guideline 62 - Antenatal Care*. National Institute for Health and Clinical Excellence, 2008.

Perkin Elmer 2015

Perkin Elmer. The potential of PP13 in pre-eclampsia early detection. http://www.perkinelmer.co.uk/Content/TechnicalInfo/TCH˙PP13NonUS.pdf (accessed 27th August 2015).

Proud 1987

Proud J, Grant AM. Third trimester placental grading by ultrasonography as a test of fetal wellbeing. *BMJ* 1987;**294**(6588):1641-4.

Ptacek 2014

Ptacek I, Sebire NJ, Man JA, Brownbill P, Heazell AE. Systematic review of placental pathology reported in association with stillbirth. *Placenta* 2014;**35**(8):552–62.

RCOG 2011

Royal College of Obstetricians and Gynaecologists. *Reduced Fetal Movements: Guideline 57*. Royal College of Obstetricians and Gynaecologists, 2011.

RCOG 2014

Royal College of Obstetricians and Gynaecologists. *The Investigation and Management of the Small-for-Gestational-Age Fetus - Guideline 31*. Royal College of Obstetricians and Gynaecologists, 2014.

RevMan 2014 [Computer program]

The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Roche 2015

Roche. Elecsys® Preeclampsia (sFlt-1 & PlGF). http://www.cobas.com/home/product/clinical-and-immunochemistry-testing/elecsys-preeclampsia-assays-sFlt-1-PlGF.html (accessed 27th August 2015).

Rutter 2001

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

Sharf 1984

Sharf M, Eibschitz I, Hakim M, Degani S, Rosner B. Is serum free estriol measurement essential in the management of hypertensive disorders during pregnancy?. *European Journal of Obstetrics, Gynecology, and Reproductive Biology* 1984;**17**(6):365–75.

Smith 2002

Smith GC, Stenhouse EJ, Crossley JA, Aitken DA, Cameron AD, Connor JM. Early pregnancy levels of pregnancy-associated plasma protein a and the risk of intrauterine growth restriction, premature birth, preeclampsia, and stillbirth. *Journal of Clinical Endocrinology and Metabolism* 2002;**87**(4):1762–7.

Smith 2006
Smith GC, Shah I, Crossley JA, Aitken DA, Pell JP, Nelson SM, et al. Pregnancy-associated plasma protein A and alpha-fetoprotein and prediction of adverse perinatal outcome. *Obstetrics and Gynecology* 2006;**107**(1):161–6.

Stock 2012
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* 2012;**344**:e2838.

Takwoingi 2013
Takwoingi Y, Leeflang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. *Annals of Internal Medicine* 2013;**158**(7):544–54.

Whiting 2011a
Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *Journal of Clinical Epidemiology* 2011;**64**(6):602–7.

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

Worton 2014
Worton S, Sibley CP, Heazell AEP. Understanding the placental aetiology of fetal growth restriction; could this lead to personalized management strategies?. *Fetal and Maternal Medicine Review* 2014;**25**(2):95–116.

* *Indicates the major publication for the study*

APPENDICES

Appendix 1. Examples of placental function tests available for diagnostic use (compiled 29*th* March 2016)

Placental growth factor (PlGF)
Triage PlGF (Alere, San Diego) - point of care fluorescence immunoassay (http://www.plgf.com/home/proposed-clinical-use-of-plgf/alere-triage-plgf.html)
Elecsys™ Preeclampsia (sFlt-1 & PlGF) - automated immunoassay performed on Roche platform (http://www.cobas.com/home/product/clinical-and-immunochemistry-testing/elecsys-preeclampsia-assays-sFlt-1-PlGF.html)
Oestriol (E3)
AutoDELFIA Unconjugated Estriol (Perkin Elmer) - automated fluorescence immunoassay performed on Perkin-Elmer platform. (http://www.perkinelmer.co.uk/product/autodelfia-unconjugated-estriol-ue3-ki-b083-301)
Beckman Coulter - automated immunoassay performed on Beckman Coulter platform (https://www.beckmancoulter.com/wsrportal/bibliography?docname=DS14764A%20Access%20Unconjugated%20Estriol%20US%20Data%20Sheet.pdf)
Elecsys™ Estradiol - automated immunoassay performed on Roche platform (http://www.cobas.com/content/dam/cobas˙com/pdf/lists/parameter-list-swa.pdf)

Appendix 2. Example Search Strategy

Database: Ovid MEDLINE(R) 1946 to July Week 1 2015
Search Strategy:
---
1 Placental insufficiency/
2 ((placenta$ or fetoplacental or uteroplacental) adj2 (insufficien$ or fail$ or function$)).ti,ab.
3 fetal movement/
4 fetal growth retardation/
5 ((reduc$ or decline$) adj2 fetal movement).ti,ab.
6 (stillborn or stillbirth).ti,ab.
7 Stillbirth/
8 ((fetal or intrauterine or intra-uterine) adj2 (growth or death$ or loss$)).ti,ab.

9 IUGR.ti,ab.
10 (small adj2 gestational age).ti,ab.
11 ((neonatal or perinatal or fetal or birth$ or deliver$) adj2 outcome$).ti,ab.
12 fetal move$.ti,ab.
13 or/1-12
14 oestradiol.ti,ab.
15 estradiol.ti,ab.
16 exp Estradiol/
17 oestriol.ti,ab.
18 exp progesterone/
19 progesterone.ti,ab.
20 exp pregnenolone/
21 pregnenolone.ti,ab.
22 exp Chorionic Gonadotropin/
23 human chorionic gonadotrophin.ti,ab.
24 hCG.ti,ab.
25 placental lactogen/
26 hPL.ti,ab.
27 human placental lactogen.ti,ab.
28 human placental growth hormone.ti,ab.
29 placental protein 13.ti,ab.
30 placental growth factor.ti,ab.
31 plasma placental protein.ti,ab.
32 pregnancy specific glycoprotein$.ti,ab.
33 Pregnancy-Specific beta 1-glycoproteins/
34 schwangerschaft protein 1.ti,ab.
35 pregnancy specific beta 1-glycoprotein.ti,ab.
36 exp ultrasonography, Prenatal/
37 (sonograph$ or ultraso$).ti,ab.
38 Grannum grading.ti,ab.
39 biomarkers/
40 biomarker$ or marker$.ti,ab.
41 or/14-40
42 13 and 41
43 (animals not (humans and animals)).sh.
44 42 NOT 43

Appendix 3. QUADAS 2 tool for assessing methodological quality of included studies

Domain	Signalling question	Signalling question	Signalling question	Risk of bias	Concerns about applicability
Patient selection	Was a consecutive or random sample of patients enrolled?	Was a case-control design avoided?	Did the study avoid inappropriate exclusions?	Could the selection of patients have introduced bias?	Are there concerns that the included patients and setting do not match the review question?

(*Continued*)

	Yes *if participants were consecutively enrolled or if all eligible participants were enrolled or participants were randomly sampled.* No *if participants were selected from those eligible.* Unclear *if participant selection was not clear from the report.*	Yes *if a case control design was avoided.* No*if a case control design was used.* Unclear *if the study design could not be determined from the report.*	Yes *if the study avoided inappropriate exclusions (e.g. only excluded multiple pregnancy, congenital abnormalities).* No *if participants were excluded inappropriately (e.g. ethnicity, age, income).* Unclear *if appropriateness of exclusions could not be assessed from report.*	Low risk *if yes to all of the signalling questions.* High or unclear risk *if "no" or "unclear" was reported for at least one signalling question.*	Low concern *if the sample of pregnant women represent the women indicated by the review question and if inappropriate exclusions were avoided.* High concern *if the sample of pregnant women are different from those indicated in the review question.* Unclear concern *if insufficient information was available.*
Index test - test of placental function	Were the index test results interpreted without knowledge of the results of the reference standard?	If a threshold was used was it pre-specified?		Could the conduct or interpretation of the index test have introduced bias?	Are there concerns that the index test, its conduct or its interpretation differ from the review question?
	Yes *if the result(s) of the test of placental function was interpreted without knowledge of the reference standard.* No *if the result(s) of the test of placental function was interpreted with knowledge of the reference standard.* Unclear *if this was not clear in the report.*	Yes *if the criteria for a positive result of the placental function test were pre-specified.* No *if the criteria for a positive result were not pre-specified or deviated from that specified.* Unclear *if this was not clear from the report.*		Low risk *if yes to all of the signalling questions.* High or unclear risk *if "no" or "unclear" was reported for at least one signalling question.*	Low concern *if the placental function test was performed as described in the review question (e.g. after 24 weeks of pregnancy to assess placental function).* High concern *if the placental function test was performed in a different way to that described in the review question.* Unclear concern *if insufficient information was available.*
Reference standard and target condition	Is there reference standard likely to correctly classify the target condition?	Were the reference standard results interpreted without knowledge		Could the reference standard, its conduct or interpretation have intro-	Are there concerns that the target condition as defined by the reference stan-

(*Continued*)

		of the results of the index test?		duced bias?	dard does not match the question?
	Yes *if an acceptable reference standard was used (e.g. SGA = birthweight < 10th centile, Stillbirth = baby born with no signs of life after 24 weeks' gestation).* No *if pregnancy outcome was not classified by an acceptable reference standard (e.g. low birthweight < 2.5 kg).* Unclear *if this was not clear from the report.*	Yes *if pregnancy outcome (live or stillbirth), and a diagnosis of a small for gestational age infant was made without the knowledge of results of the placental function test.* No *if pregnancy outcome and a diagnosis of a small for gestational age infant were made with the knowledge of the results of the placental function test.* Unclear *if this was not clear from the report.*		Low risk *if yes to all of the signalling questions.* High or unclear risk *if "no" or "unclear" was reported for at least one signalling question.*	Low concern *if acceptable reference standards were used and if the reference standard was interpreted without the knowledge of the placental function test.* High concern *if an acceptable reference standard was not used or the results were interpreted with knowledge of the result of the placental function test.* Unclear concern *if insufficient information was available.*
Flow and Timing	Was there an appropriate interval between the index test and reference standard?	Did all patients receive the same reference standard?	Were all patients included in the analysis?	Could the patient flow have introduced bias?	
	Yes *If acquisition of the index test occurred prior to birth (reference standards both determined after birth).* No *if sample acquired after delivery of the infant (i.e. known reference standard).* Unclear *if this was not clear from the report.*	Yes *if all participants had the outcome of pregnancy and birthweight recorded.* No *if some participants do not have the outcome of pregnancy and birthweight recorded.* Unclear *if this was not clear from the report.*	Yes *if all participants recruited to the study were included in the final analysis.* No *if all participants were not included in the final analysis.* Unclear *if this was not clear from the report.*	Low risk *if yes to all of the signalling questions.* High or unclear risk *if "no" or "unclear" was reported for at least one signalling question.*	

Appendix 4. Glossary

Fetal growth restriction (FGR) - a condition where a fetus fails to attain its growth potential, i.e. is smaller than expected for its genetic potential.
Small-for-gestational-age infant (SGA infant) - the condition where the fetal weight or birthweight is beneath a specific threshold, generally considered to be the 10th centile.
hPL - human placental lactogen - a protein made by the trophoblast layer of the placenta.
PlGF - placental growth factor- a protein made by the trophoblast layer of the placenta.
SROC - summary receiver operator characteristic - a graphical representation of different estimates of test accuracy.
Umbilical artery Doppler - a measurement of fetal blood flow through the umbilical artery using Doppler ultrasound
Uterine artery Doppler - a measurement of maternal blood flow through the uterine artery using Doppler ultrasound

CONTRIBUTIONS OF AUTHORS

AEP Heazell (AEPH) conceived the idea for the systematic review. All authors contributed to the design of the review and writing the protocol. AEPH is the guarantor for the review.

DECLARATIONS OF INTEREST

Susan Bayliss: none known.

Clare Davenport's employer (The University of Birmingham) received funding for her participation in this review as part of an NIHR clinical fellowship awarded to Alex Heazell (the lead author and contact person)

Dexter Hayes: none known.

Yemisi Takwoingi's employer (The University of Birmingham) received funding for her participation in this review as part of an NIHR clinical fellowship awarded to Alex Heazell (the lead author and contact person)

Melissa Whitworth: none known.

Alexander Heazell has received research grants from Alere (UK) and Action Medical Research to investigate placental factors in maternal serum in women with reduced fetal movements. He is also a Supervisor for a Clinical Research Fellowship from Action Medical Research which incorporates projects to detect placental factors in maternal serum. In addition, he holds a Clinician Scientist Award from National Institute of Health Research (NIHR) (CS-2013-13-009) and this review is part of that programme of work. The views expressed here are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. He has also received payment from Solutions for Public Health for consultancy work as the obstetric lead for the Confidential Enquiry into Stillbirths and Neonatal Deaths in Cumbria from 2009-2010.

SOURCES OF SUPPORT

Internal sources

- No sources of support supplied

External sources

- National Insitute of Health Research, UK.

Dr Alexander Heazell is funded by a Clinician Scientist Award from the National Institute of Health Research (CS-2013-009). This protocol presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.