PREDICTION AND PREVENTION OF FETAL GROWTH RESTRICTION AND COMPROMISE OF FETAL WELLBEING. SYSTEMATIC REVIEWS AND META-ANALYSES WITH MODEL BASED ECONOMIC EVALUATION

By Rachel Katherine Morris

A thesis submitted to the University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

College of Medical and Dental Sciences

The University of Birmingham

September 2010

Volume I

SYNOPSIS

THIS THESIS SYSTEMATICALLY REVIEWED THE LITERATURE ON TESTS AND TREATMENTS FOR FETAL GROWTH RESTRICTION AND COMPROMISE AND INCORPORATED THIS INFORMATION IN A DECISION-ANALYTIC MODEL

Abstract

Restriction of fetal growth and compromise of its wellbeing remain significant causes of perinatal death and childhood disability. There is a lack of scientific consensus about the best strategies for predicting these conditions before birth and thus there is uncertainty about the best management of pregnant women who might have a growth-restricted baby. This health technology assessment thesis used state of the art methods to review 337 studies including 472,544 women. It determined : 1. The accuracy of available tests for predicting small for gestational age infants (SGA) and 2. Compromise of fetal wellbeing and 3. Summarised the effectiveness of available treatments for these conditions. To allow translation of these results into patient care, the diagnostic and therapeutic information was integrated in a model based economic evaluation. This thesis has demonstrated that the tests reviewed have a limited use in screening/diagnosis for SGA/compromise of fetal and neonatal wellbeing when used in isolation. The quality of primary research was variable with recommendations being made particularly for the use of standardised and relevant outcome measures. The decision model and economic analysis identified that an effective, affordable and safe intervention applied to all mothers without prior testing is likely to be the most cost-effective strategy in the prevention of these conditions.

Executive Summary

Background

Restriction of fetal growth and compromise of its wellbeing remain significant causes of perinatal death and childhood disability. At present, there is a lack of scientific consensus about the best strategies for predicting these conditions before birth. Therefore, there is uncertainty about the best management of pregnant women who might have a growth-restricted baby. This is likely to be due to a dearth of clear, collated information from individual research studies drawn from different sources on this subject. This thesis contains health technology assessment of test treatment strategies for prevention of fetal growth restriction and compromise of fetal wellbeing to guide clinical practice and future research in this area.

Objectives

This thesis undertook health technology assessment for a range of tests and interventions for SGA and compromise of fetal/neonatal wellbeing. The objectives were to a) obtain summary estimates of effects of available tests and treatments for restriction of fetal growth and compromise of its wellbeing; b) To evaluate the effect of study reporting and methodological quality on test accuracy; c) to integrate the summarised diagnostic and therapeutic information using decision-analytic modelling.

Methods

A health technology assessment was performed based on prospective protocols using contemporary methods. The following methods were employed: systematic review and meta-analyses of test accuracy, reviews of effectiveness of interventions and economic evaluation using a decision tree model.

For the test accuracy reviews, literature was identified from electronic sources, contact with experts and checking of reference lists. Inclusion criteria were studies performed in pregnant women at any gestation using an appropriate test and reference standard where 2x2 data could be calculated. Methodological quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) criteria. Meta-analyses were performed with the bivariate approach. Assessment of study reporting and methodological quality of included accuracy studies and assessment of their impact on accuracy was also performed.

The effectiveness review was a review of systematic reviews of effectiveness. Literature was identified from electronic sources, reference lists and contact with experts. Included reviews had to be systematic reviews of randomised controlled trials comparing an intervention for prevention of fetal growth restriction or compromise of fetal wellbeing to placebo, no intervention or usual care. Quality was assessed using a checklist based on the Critical appraisal skills programme (CASP) checklist. Data were presented as relative risk with 95% confidence intervals.

The economic evaluation employed a decision tree model. The perspective was that of the National Health Service (NHS) with inputs to the model derived from the systematic reviews performed and published literature. Deterministic and probabilistic sensitivity analyses were performed. The main outcome measure was cost-effectiveness per case of fetal growth restriction avoided.

Results

Main findings of test accuracy reviews

There were 10,107 citations identified as being potentially relevant for this work. Of these, 1,157 papers were read in full with 337 included in the reviews including 472,544 women tested. The following tests were reviewed; five Down's syndrome screening markers (alpha feto-protein, human chorionic gonadotrophin, estriol, inhibin A, pregnancy associated plasma protein A) and four Doppler measurements (uterine artery, umbilical artery, middle cerebral artery and ductus venosus). The median number of women included was 33,292 (interquartile range 13,273-40,637). The median number of studies per test was 60 (interquartile range 31-86). The tests overall for prediction of small gestational age infants and adverse perinatal outcome demonstrated low predictive accuracy with no tests having a positive likelihood ratio LR>5 and a negative LR<0.5.

Main findings of effects of study quality on test accuracy

A total of 195 studies were included in this work. The overall reporting quality of included studies was poor (adequate reporting >50% of the time for 62.1% (18/29) of the items). The overall methodological quality was poor (>50% compliance with 57.1%

of quality items). There was a positive correlation ($p<0.0001$) between study sample size and reporting quality but not with methodological quality. No correlation with geographical area of publication and compliance with quality criteria could be demonstrated. Meta-regression analysis showed that no individual quality item had a significant impact on accuracy. There was an association between reporting and methodological quality ($r=0.51$ $p<0.0001$). This work demonstrated that the reporting and methodological quality of papers in Obstetrics is improving but that there is still considerable scope for improvement.

Main findings of review of systematic reviews of effectiveness of interventions

This work included 71 systematic reviews with a total of 733 randomised controlled trials reporting on 42 different interventions. After considering the results and the quality of evidence, antiplatelets and multiple micronutrient supplements were the interventions that were found to be effective in preventing the small for gestational age fetus and suitable for use in all pregnant women. For high risk pregnant women the following were found to be effective: antiplatelets, multiple micronutrient supplements, smoking cessation interventions and progesterone therapy. For prevention/reduction of perinatal mortality antiplatelets and antenatal corticosteroids were the interventions shown to be effective.

Main findings from health economic evaluation and decision analytic modelling

The model used an outcome of cost per case of fetal growth restriction avoided and incorporated first and second trimester tests from the test accuracy reviews and all relevant interventions identified from the effectiveness reviews. Costs of tests and

treatments and outcomes were identified from local data and the literature. Testing prior to intervention was not shown to be the most cost-effective strategy in the analyses for all pregnant women. Anti-platelet therapy, without prior testing, was highlighted as potentially cost-effective in preventing fetal growth restriction in this population. In high risk women, testing with serum human chorionic gonadotrophin followed by anti-platelet therapy in those that test positive was a potentially cost-effective strategy. Threshold analysis revealed that for a test to be considered as an option prior to treatment in unselected pregnant women it would have to have high levels of accuracy and be relatively cheap (£5). This is likely to be due to the fact that the majority of treatments available are themselves relatively cheap (£2.60 for aspirin) and thus from a cost point of view it will always be preferential to apply treatment to all rather than to test first. This has to be interpreted in light of the limitations of the model, importantly the lack of inclusion of adverse effects of treatment.

Conclusions

This thesis has demonstrated that the tests reviewed have limited use in screening/diagnosis for SGA baby/compromise of fetal and neonatal wellbeing when used in isolation. The main implications of this work are thus not for recommendations for practice but for future research. Further research in this area needs to consider the use of tests in combination and the role that other diagnostic tools, such as risk factor assessment and clinical features, add to the clinical decision making process. This research needs to be robustly designed, include primary test evaluation strategies with reference to relevant quality criteria and include a sample size calculation to ensure that results have sufficient power. There is a particular need for researchers in the area of

fetal growth restriction to determine the most appropriate reference standards/outcome measures to be used that truly identify the growth restricted baby. This will ensure that primary research is not only directed at the fetuses/pregnancies at risk but will facilitate future systematic reviews and meta-analysis.

To ensure that the results of any future economic analysis and decision model analysis can be translated into recommendations for practice there will be a need for models, and the primary research that informs them, to be able to compare both directly and indirectly all combinations of tests and treatments with consideration of side effects. There will also need to be further primary research to determine accurate costs of the outcomes. This research will also need to be directed to look at the impact of these clinical management strategies on multiple outcomes e.g. pre-eclampsia, pre-term birth and fetal growth restriction to ensure that a truly comprehensive clinical management pathway that is applicable to a general pregnant population within the NHS can be devised.

Dedication page

This thesis is dedicated to my husband Simon and daughter Amélie.

Acknowledgements

To enable this thesis to be completed and submitted has required the help and support of everyone I have the fortune and pleasure to call colleagues. These colleagues work within the academic department, fetal medicine, university departments such as finance, administration, accounts, library services, the list is endless and I thank you all. In particular to the girls (and occasionally some boys who join us) of the research fellows' office I thank you for your patience, for lending a helpful ear and the cups of tea.

I would like to thank Dr Tara Selman for her lovely welcome on my first day, for providing a great travelling partner to Madrid and for all her support during the last five years.

Professor Khalid Khan, my supervisor; for all his support, encouragement, amazing knowledge and ideas, for knowing when to push for more and when to give advice or leave me alone. I thank him for being an excellent supervisor but also a mentor and friend.

Dr Catherine Meads, my co-supervisor; again for giving me excellent advice and stimulating ideas and helping to make this thesis a coherent piece of work and for the excellent training received through her department in health technology assessment.

I thank Professor Mark Kilby for bringing me to Birmingham, for the mentorship and guidance; for the ideas and opportunities that will hopefully continue in the future.

I owe a large debt to Dr Javier Zamora for not only helping me to understand statistics but also for making them enjoyable; thank you for all the e-mails and phone conversations and of course your hospitality in Madrid.

Professor Tracy Roberts and Angelos Tsourapas for their help with the economic analysis and decision analytic modelling and helping to make the world of health economics enjoyable and understandable for a clinician. Dr Pelham Barton for advice with the modelling.

I thank Dr Jeltsje Cnossen and her colleagues at the Amsterdam Medical Center (AMC) for the excellent training in systematic reviewing that I received with them and for helping with the data extraction and analysis of the uterine artery Doppler and Down's syndrome screening review. Thank you for your hospitality during my time working with you in Amsterdam. I also thank Faridi van Etten at the AMC for help with the electronic searches.

Professor Steve Robson and Dr Jos Kleijnen who were involved in the Wellbeing of Women project grant that enabled me to start my research career and that led to this work.

There were also many colleagues not already mentioned who helped by being second reviewers on some of the systematic reviews or by translating papers; Dr Gemma Malin, Dr Vicky Hodgetts; Dr Marloes Langejans; Dr Beccy Say and Dr Stephen Ong. I confirm

that for all chapters where a publication involved other authors that I was involved at all stages for design, data collection, analysis, interpretation and writing of the papers.

I was responsible for the design of the protocol for this work and for obtaining the funding from the Medical Research Council. For the diagnostic accuracy reviews, I performed the literature searches, obtained the articles, performed literature selection, data extraction and quality assessment. I performed all the statistical analyses with support for the meta-regression and bivariate meta-analysis from Dr Javier Zamora. The interpretation of results was my own.

For the methodological work on quality assessment (chapters 11 and 12) I performed the data extraction and analysis with support for the meta-regression from Dr Javier Zamora. The interpretation of results was my own.

For the effectiveness reviews, I performed the literature searches, study identification, data extraction, statistical analysis and interpretation.

For the economic analysis, I performed all the data collection around costs of interventions and tests and the work to determine the cost of the outcome. I built the tree and performed all the economic and decision model analysis with the support of Angelos Tsourapas and supervision of Professor Tracy Roberts and Dr Pelham Barton. The interpretation of results was my own.

I would like to thank the charity Wellbeing of Women and the Medical Research Council for funding this work.

Finally I must thank my family for their continued support but in particular my husband Simon, for all the sacrifices that he has made for me and for being my greatest supporter.

TABLE OF CONTENTS

VOLUME 1

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AC	Abdominal circumference
AFP	Alpha feto-protein
APO	Adverse perinatal outcome
AUC	Area under curve
BMI	Body mass index
BNF	British National Formulary
BPD	Biparietal diameter
BW	Birth weight
BWH	Birmingham Women's Hospital NHS Foundation Trust
CASP	Critical appraisal skills programme
CEAC	Cost effectiveness acceptability curve
CI	Confidence interval
CONSORT	Consolidated Standards of Reporting of Randomised Controlled Trials
CTG	Cardiotocograph
DHAS	Dehydroepiandrosterone sulphate
DOR	Diagnostic odds ratio
DV	Ductus venosus
ECG	Electrocardiogram
EFW	Estimated fetal weight
FGR	Fetal growth restriction

FL	Femur length
HC	Head circumference
HCG	Human chorionic gonadotrophin
HRG	Health resource groups
HSROC	Hierarchical summary receiver operating characteristic
HTA	Health technology assessment
ICER	Incremental cost effectiveness ratio
IPD	Individual patient data
LR	Likelihood ratio
MCA	Middle cerebral artery
MeSH	Medical subject heading
MoM	Multiples of the median
MOOSE	Meta-analyses of observational studies in epidemiology
NA	Not applicable
NHS	National Health Service
NICE	National Institute of Clinical Excellence
NICU	Neonatal intensive care unit
NNT	Number needed to test
NNTreat	Number needed to treat
NPV	Negative predictive value
OR	Odds ratio
PAPP-A	Pregnancy associated plasma protein A
PE	Pre-eclampsia
PI	Pulsatility index

PICOS	Population, intervention, comparator, outcome, study design
PP10	Placental protein 10
PPV	Positive predictive value
PSA	Probabilistic sensitivity analysis
PSV	Peak systolic velocity
QUADAS	Quality assessment of diagnostic accuracy studies
QUOROM	Quality of reporting of meta-analyses
RCOG	Royal College of Obstetricians and Gynaecologists
RCT	Randomised controlled trial
RDOR	Ratio of diagnostic odds ratio
RI	Resistance index
ROC	Receiver operating characteristic
RR	Relative risk
SD	Standard deviation
SD	Systolic diastolic
SFH	Symphseal fundal height
SGA	Small for gestational age
STARD	Standards for reporting of diagnostic accuracy
UA	Uterine artery
Umb	Umbilical

PUBLICATIONS FROM THIS THESIS

Morris R, Khan KS, Coomarasamy A, Robson S, Kleijnen J. The value of predicting restriction of fetal growth and compromise of its wellbeing: Systematic quantitative overviews (meta-analysis) of test accuracy literature. *BMC Pregnancy and Childbirth* 2007;7:3.

Morris R K, Cnossen JS, Langejans M, Robson SC, Kleijnen J, Ter Riet G, Mol BW, van der Post JA, Khan KS. Serum screening with Down's Syndrome markers to predict pre-eclampsia and small for gestational age: Systematic review and meta-analysis. *BMC Pregnancy Childbirth 2008 Aug 4;8(1):33.*

Cnossen JS, Morris RK, Mol BWJ, ter Riet G, van der Post JAM, Bindels PJE, Robson S, Kleijnen J, Coomarasamy A, Khan KS. Uterine artery Doppler to predict fetal growth restriction: a systematic review and bivariate meta-analysis. *Canadian Medical Association Journal 2008;178(6):701-11*

Morris RK, Malin GL, 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. *Archives of Disease in Childhood Fetal and Neonatal* 2010;95(suppl 1): Fa13-Fa14.

Morris RK, Malin GL, 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. Epub 2010 *Ultrasound in Obstetrics and Gynaecology*.

Morris RK, Say R, Robson SC, Kleijen J, Khan KS. Systematic review of middle cerebral artery Doppler to predict fetal growth restriction/compromise of fetal wellbeing. *Archives of Disease in Childhood: Fetal and Neonatal Edition* 2008; 93 (supplement 1): Fa 31-36.

R.Katie Morris, Tara J Selman, Meenakshi Verma, Stephen C Robson, Jos Kleijnen, Khalid S Khan. Systematic review and meta-analysis of the test accuracy of ductus venosus Doppler to predict compromise of fetal/neonatal wellbeing in high risk pregnancies with placental insufficiency. *Eur J Obstet Gynecol Reprod Biol* 2010 Epub.

Selman TJ, Morris RK, Zamora J, Khan KS. The quality of reporting of primary test accuracy studies in Obstetrics and Gynaecology: application of the STARD criteria. *BMC Women's Health 2010* (In press).

Morris RK, Selman TJ, Zamora J, Khan KS. Methodological quality of test accuracy studies included in systematic reviews in obstetrics and gynaecology: sources of bias. *BMC Women's Health 2010* (In press).

Morris RK, Malin GL, Tsourapas A, Roberts TE, Khan KS. An economic evaluation of alternative test-intervention strategies to prevent fetal growth restriction in singleton pregnancies. Archive of Disease in Childhood Fetal and Neonatal 2010;95(suppl 1): Fa1

CHAPTER 1: INTRODUCTION

1.1 Aim and objectives of thesis

The aim of this thesis was to perform an health technology assessment (HTA) in Obstetrics through evaluation of a range of tests and interventions for small for gestational age (SGA) fetuses and compromise of fetal/neonatal wellbeing. The main objectives were as follows:

1. To obtain summary estimates of accuracy of available tests for SGA fetuses and compromise of fetal wellbeing.
2. To evaluate the effect of study reporting and methodological quality on test accuracy.
3. To obtain summary estimates of effects of available treatments for SGA fetuses and compromise of fetal wellbeing.
4. To integrate summarised diagnostic and therapeutic information using decision-analytic modelling.

1.2 Outline of thesis

The work performed has been divided into two volumes:

Volume I

PART A: SYSTEMATIC REVIEWS OF TEST ACCURACY

Systematic reviews of the existing evidence on the accuracy of tests to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing (chapters 4-10).

PART B: METHODOLOGICAL RESEARCH INTO THE EFFECTS OF REPORTING AND METHODOLOGICAL QUALITY OF SYSTEMATIC REVIEWS OF TEST ACCURACY

The quality of reporting of primary test accuracy studies in Obstetrics reviewed in this thesis: application of the STARD criteria (chapter 11). Methodological quality of test accuracy studies included in systematic reviews in Obstetrics reviewed in this thesis: Sources of bias (chapter 12).

PART C: REVIEW OF SYSTEMATIC REVIEWS OF THE EVIDENCE ON EFFECTIVENESS OF AVAILABLE INTERVENTIONS FOR PREVENTION OF SMALL FOR GESTATIONAL AGE AND COMPROMISE OF FETAL/NEONATAL WELLBEING

Review of systematic reviews of the existing evidence on the effectiveness of available interventions for prevention of small for gestational age fetuses and compromise of fetal/neonatal wellbeing (chapter 13).

PART D: COST EFFECTIVENESS ANALYSIS WITH ECONOMIC MODELLING

Cost effectiveness analysis with economic modelling to assess test and treatment strategies for the management of the small for gestational age fetus and compromise of fetal/neonatal wellbeing (chapter 14).

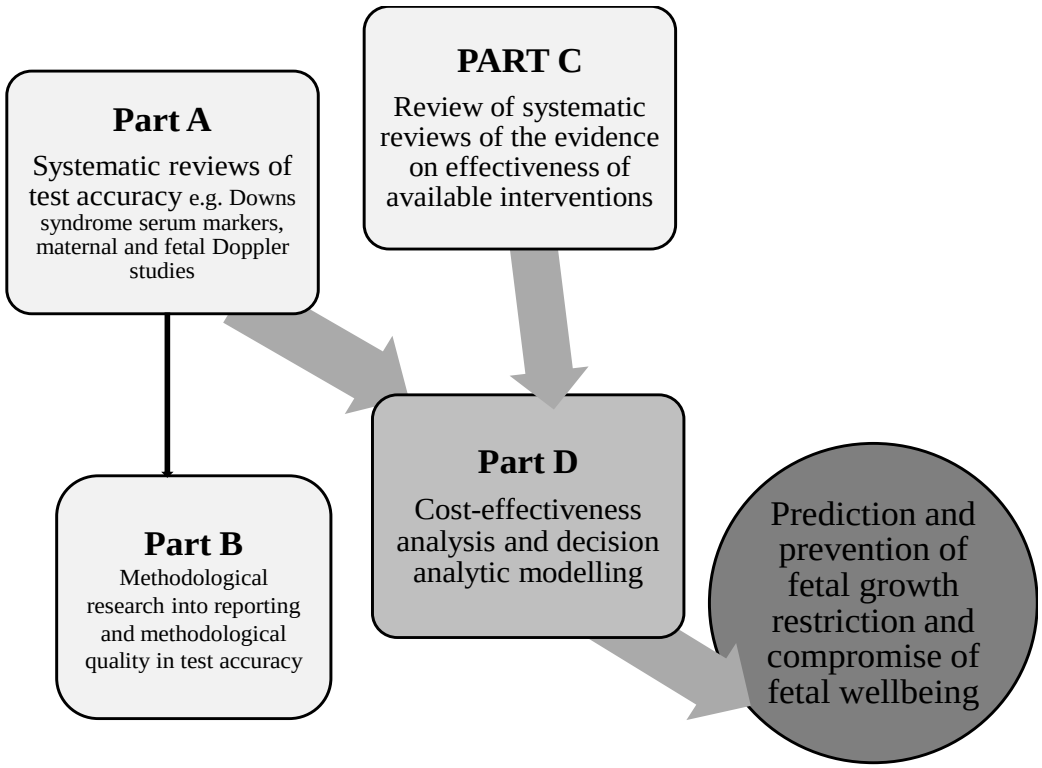
Volume 2

Appendices and references.

1.3 Methods

The thesis employed systematic reviews and decision-analytic modelling based on prospective study protocols. Primary importance is given to reviews of test accuracy looking at individual tests, and included methodological research looking at the impact of study quality on test accuracy. Research was also performed to determine the effectiveness of available interventions for SGA fetuses and fetal/neonatal compromise. Finally, the results from the systematic reviews of test accuracy and effectiveness were combined in a decision tree model to allow economic based evaluation, an essential part of health technology assessment. Figure 1.1 summarises the HTA conducted in this PhD thesis.

Figure 1.1: Process of health technology assessment for the prediction and prevention of fetal growth restriction and compromise of fetal/neonatal wellbeing



CHAPTER 2: BACKGROUND

2.1 Health Technology Assessment

Health services research is concerned with the relationship between provision, effectiveness and efficient use of health services and the health needs of the population. It aims to produce reliable and valid research data on which to base effective, efficient and acceptable health services. "Health technologies" are broadly defined as all interventions to promote health, prevent and treat disease, and improve rehabilitation and long-term care[1]. Health technology assessment (HTA) involves the systematic appraisal and evaluation of health technologies through primary research, systematic reviews and model based synthesis of available evidence. It asks four fundamental questions[1]:

- Does the technology work?
- For whom does it work?
- What is the cost?
- How does it compare with the alternatives?

The steps in an HTA are[2]:

1. Clearly define the question
2. Search for available information
3. Generate a "decision tree"
4. Find the evidence
5. Sort and appraise the evidence
6. Search for cost information

7. Extract and summarise/synthesise data
8. Perform an economic evaluation
9. Consider the wider ethical, legal and social implications
10. Write an HTA report

2.1.1 HTA of diagnostic technologies

Timely prediction of SGA and compromise of fetal wellbeing is of essence in antenatal care. Without accurate prediction, clinicians are handicapped and unable to institute appropriate management. Wrong or delayed prediction puts the baby at risk of an adverse outcome whereas correct prediction provides an opportunity to optimise care. If high-risk groups are accurately and efficiently identified, they could benefit from monitoring of wellbeing and appropriate interventions such as steroid administration and timely delivery. However, decision-making is hampered due to lack of precise information on estimates of risk.

Obstetrics has seen rapid growth in the development of new tests in the area of fetal growth restriction (FGR) particularly advances in ultrasound imaging and first trimester screening[3] (table 2.1). A key aspect of research on these is presented in the form of test accuracy studies[4], which generate a comparison of measurements made by an index test against those of an accepted reference standard test – the "gold standard". These comparisons enable an assessment of the accuracy of an index test, which are often expressed as sensitivity, specificity, likelihood ratios (LRs), diagnostic odds ratio (DOR), positive predictive value (PPV) and negative predictive value (NPV) or area under receiver operator characteristic curve (AUC)[5]. These measures of test accuracy take into

Table 2.1: Table of available tests for prediction of restriction of fetal growth and compromise of fetal wellbeing.

Tests	Tests for restriction of fetal growth	Tests for compromise of wellbeing
History and Examination	◦ Clinical risk scoring ◦ Palpation of abdomen to assess size ◦ Symphyseal fundal height (SFH) measurement	◦ Clinical risk scoring ◦ Fetal movement counting
Ultrasound	**Biometry (anthropometric measures)** ◦ Single measures: Abdominal circumference (AC), Head circumference (HC), Biparietal diameter (BPD), Femur length (FL), Thoracic diameter, Abdominal diameter, Abdominal area, Chest area, Liver size, Thigh circumference, Subcutaneous fat ◦ Ratio measures: FL/AC, HC/AC, FL/HC, Head area/abdominal area, FL/ thigh circumference ◦ Composite measures: Estimated fetal weight (EFW), Fetal ponderal index, Total intrauterine volume, Trunk area x Crown Rump Length, ◦ Growth velocity measurements **Doppler** ◦ Uterine artery ◦ Umbilical artery ◦ Middle cerebral artery ◦ Descending aorta ◦ Internal carotid artery **Other** ◦ Amniotic fluid volume ◦ Placental grade	**Doppler** ◦ Uterine artery ◦ Umbilical artery ◦ Middle cerebral artery ◦ Venous Doppler ◦ Uteroplacental **Other** ◦ Amniotic fluid volume ◦ Biophysical profile
Biochemical and Haematological	◦ Oestriols, Human placental lactogen, Plasma fibronectin, Alpha Feto Protein, Human chorionic gonadotrophin, Beta-1 glycoprotein, Placental protein 10, Pregnancy Associated Plasma Protein A, dehydroepiandrosterone sulphate loading test, Epidermal growth factor, Amniotic fluid C-amino peptide, Serum cystine aminopeptidase, Schwangerschafts protein 1, Serum alpha 2 – macroglobulin, Maternal leukocyte zinc level, Form stability index	
Other tests	◦ Customised growth charts of SFH and ultrasound EFW	◦ Cardiotocography (CTG) ◦ Fetal ECG ◦ Fetal magnetocardiography

account the false positive and false negative results from a test and the relationship between these two values thus helping to determine the usefulness of a test in clinical practice. It is thus important that these summary measures of accuracy are obtained through systematic reviews of the highest quality of evidence available.

2.1.2 HTA of therapeutic interventions

The ability to accurately predict or diagnose disease is only one step in the clinical pathway. Following testing there needs to be effective, safe and acceptable intervention with the purpose of either preventing development of the condition, improving or maintain the current health status or avoiding further deterioration, or in some instances to provide palliation. Evaluation of a therapy must therefore include an assessment of its efficacy usually within the context of a randomised controlled trial (RCT), investigation of its safety, assessment of the economic costs, assessment of its acceptability to patients and consideration of the wider social, legal and ethical implications. Once a therapy has been found to be effective, safe and acceptable to parents it must then be proven to be cost-effective within a wider population before it can be considered as part of the clinical pathway. The gold standard for assessment of the efficacy of interventions is the systematic review of RCTs e.g. Cochrane reviews of effectiveness.

2.1.3 Decision modelling for economic evaluation of health technologies

Any diagnostic tool will have false positive and false negative results and all treatments will have a cost attached to them. When assessing a health technology, it is thus important that the tests and treatments are assessed as part of a cost-effectiveness analysis. Resource use within the NHS cannot only be dictated by effectiveness but must also take into

account cost and be able to assess new technologies, treatments and management pathways in a comparative manner[6]. The National Institute for Health and Clinical Excellence (NICE) uses economic evaluation to provide guidance to the NHS on the introduction of new technologies.

To enable the diagnostic and therapeutic information to be integrated, a model is used to provide a framework for all potential levels of effects and associated costs. Advantages of using a model are that all possible combinations of outcomes may be considered and advanced statistical techniques, such as Monte Carlo simulations may be used. Even with slight values of effectiveness for available treatments, as can be the case in Obstetrics, model-based decision analysis provides the most objective way to assess whether a test and a treatment should be employed. This also allows extrapolation of the data beyond the time scale of any original studies. Finally, an appropriately structured model-based economic analysis can also help decide which research to invest in in the future via value of information analysis. This aims to quantify the total uncertainty in terms of the value of removing that uncertainty via probabilistic sensitivity analysis. The model can thus provide a hierarchy of most promising test/treatment combinations and identify areas in which additional data collection, and hence the reduction of uncertainty, would be of most value.

2.2 Description of the underlying health problem

2.2.1. The importance of small for gestational age fetuses and compromise of fetal/neonatal wellbeing in Obstetrics

Restriction of fetal growth and compromise of its wellbeing remain significant causes of perinatal death and childhood disability[7-9]. The most recent confidential enquiry into perinatal deaths reported a neonatal mortality rate of 32.5 per 1,000 live births for babies with birth weight < 2500g (compared to 0.8 per 1,000 live births for birth weight > 2500g) and determined that 37.9% of the unexplained stillbirths were small for gestational age (birth weight less than 10th centile for gestation)[10]. These babies on reaching adulthood are at greater risk of developing cardiovascular disease, hypertension, and non-insulin dependent diabetes[11;12].

2.2.2 Definitions of the growth restricted baby

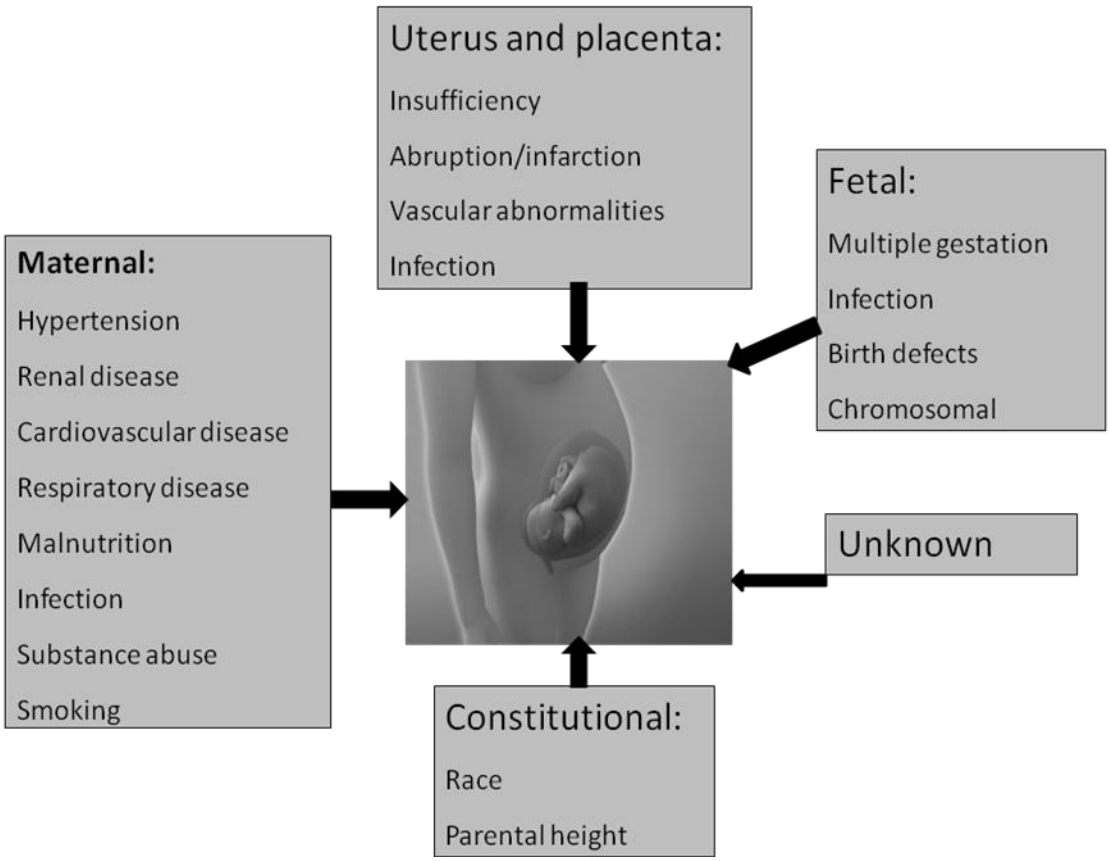
There are various centile based thresholds (2.5th, 3rd, 5th, 10th, 15th and 25th) for defining SGA and absolute birth weight thresholds for defining low birth weight (<2500g, <1750g, <1500g, mean < 2 standard deviations). The most commonly accepted and used standard is the 10th centile[13]. However, by definition this standard will represent 10% of the population being assessed. This will thus constitute a very heterogeneous population including both the constitutionally small baby, the abnormal baby with, for example, chromosomal defects and the truly growth restricted baby (FGR). In this context FGR refers to a fetus that has failed to achieve its genetic growth potential usually because of placenta-mediated disease restricting nutrient supply.

Where a fetus is assessed as small for gestational age, up to 50-70% of these fetuses are constitutionally small[14;15], not truly growth restricted. Studies have shown that perinatal mortality is eight times higher when birth weight is below the 10th percentile[16]. If stricter criteria, such as birth weight less than the 3rd centile are used then this is associated with a twenty times increase in mortality[16]. Customised growth charts that are adjusted for sex, gestational age, parity, maternal weight and height and ethnicity, have been shown to improve the detection of at risk of stillbirth[17]. Neonatal indices, such as skin fold thickness, have been shown to identify the malnourished infant at risk of peripartum asphyxia[16] and long term neurological sequelae[18].

2.2.3 Aetiology

The aetiology of FGR remains unclear (figure 2.1) but can be split into two main areas: 1) fetal factors causing the fetus not to achieve its full growth potential e.g. chromosomal defects, fetal infections, structural abnormalities and 2) extrinsic factors affecting the supply of nutrients and oxygen via the placenta to the fetus. This "placental insufficiency" may be related to placental factors such as inadequate trophoblast invasion of the spiral arteries as is seen in conditions such as pre-eclampsia (PE)[19] or related to maternal factors limiting the supply. This may be systemic maternal diseases, such as cardiac and renal disease, or social factors such as severe maternal malnutrition, smoking or drug abuse[20].

Figure 2.1: Aetiology of fetal growth restriction



2.3 Current service provision

2.3.1 The current investigation of pregnancies at risk of a small for gestational age fetus or compromise of fetal/neonatal wellbeing and the current evidence on accuracy of diagnostic tools in this area

Screening and diagnosis of FGR and prediction and monitoring for compromise of fetal wellbeing in a clinical setting includes a combination of patients' characteristics, symptoms, physical signs and tests, which form the basis of clinical care[21]. For instance, methods employed to screen for and detect FGR might include obtaining previous history of small babies, recording symphyseal fundal height on a customised growth chart and estimating fetal weight with ultrasound[21] . Similarly, current history of fetal movements,

abdominal palpation to assess liquor volume, ultrasound amniotic fluid index, Doppler flow velocimetry and cardiotocography might be used to assess fetal wellbeing[21]. Tests of wellbeing are aimed at predicting fetal acidaemia, which is perceived, at least in the model of chronic placental failure, to lead ultimately to organ damage and death. Data from cord blood sampling studies confirm there is a correlation between cord pH and neurodevelopmental outcome in small fetuses[22-24]. This implies that the accuracy of tests for FGR need to be assessed separately to those used for assessment of fetal wellbeing, but existing reviews often do not make this distinction.

Prior to the commencement of this work, formal searches (MEDLINE 1966-2003) were undertaken to identify existing systematic reviews and evidence based guidelines in this area to avoid duplication. This search revealed numerous non-systematic reviews and non-evidence based guidelines. There were nine relevant publications consisting of seven systematic reviews[13;25-30] and two evidence based guidelines[21;31]. Assessment of the methods of these reviews with standard checklists revealed five main deficiencies: i) reviews covered a limited number of tests, ii) search strategies were limited, iii) scientific strategies to limit bias were not employed, iv) with a few exceptions, meta-analyses were not employed to summarise the findings, v) there was a lack of clinically meaningful measures of test accuracy, such as likelihood ratios[32].

In 2002 the Royal College of Obstetricians and Gynaecologists (RCOG) published an evidence based guideline on "The Investigation and Management of the Small for Gestational Age Fetus"[21]. This guideline was developed using robust guideline methodology however, the recommendations were limited due to a lack of systematic

collation of diagnostic information on the subject. The current guideline recommends the use of customised fundal height charts for prediction of SGA fetuses; abdominal circumference and estimated fetal weight <10th centile on customised charts to diagnose SGA and the use of umbilical artery Doppler as the primary surveillance tool[21].

The variation in the design of research on accuracy of tests for identification of growth restriction and compromise of wellbeing, the scatter of this research across many databases and languages, and the dearth of clear collated up-to-date summaries of this literature contribute to the uncertainty about the best diagnostic and monitoring strategies[21]. The role of systematic reviews of test accuracy in this area is thus at the forefront of research. The Cochrane collaboration have initiated such reviews and published guidelines on the methods for test accuracy reviews[33]. The last decade has seen many improvements in the methods of these reviews with the introduction of checklists for reporting[34] and methodological quality[35;36] and advances in the statistical methods used in meta-analysis[37].

A comprehensive systematic review of the literature on available tests, using contemporary methods, will improve the ability to identify those pregnancies at greatest risk of developing clinically relevant intra-partum and neonatal consequences of impaired fetal growth. NICE have also recommended further prospective research to evaluate the diagnostic value and effectiveness of predicting small for gestational age fetuses[38].

2.3.2 Current interventions available for pregnancies at risk of a small for gestational age fetus or compromise of fetal/neonatal wellbeing.

The potential for a therapy to be effective depends on the nature of the underlying aetiology (section 2.2.3). A thorough assessment of the fetus and mother must be performed to identify those causes that will not be amenable to therapy e.g. chromosomal anomalies, and in those cases where therapy might be an option, to determine which may be the most appropriate. The major difficulties in this area are however the lack of accurate predictive and diagnostic tests for the growth restricted fetus and the potential for there to be more than one contributory cause e.g. pre-term labour and fetal growth restriction or fetal growth restriction and pre-eclampsia.

Previous reviews of effectiveness of interventions for fetal growth restriction have concluded that there are few interventions that are likely to be beneficial and that further high quality research is required[39]. The same conclusion was reached in the RCOG guideline (2002)[21]which states that smoking cessation programmes, particularly behavioural strategies, can be effective for a small minority of smokers in increasing birth weight but there are no data to suggest that this intervention improves perinatal outcome. Further trials are needed to assess the value of aspirin in the treatment of FGR; there is not enough evidence to assess the value of oxygen therapy, nutrient therapy, hospitalisation and bed rest, betamimetics, calcium channel blockers, hormonal therapy and plasma volume expansion in treating growth restriction[21].

Since this guideline was written there has been further research in this area including the updating of many of the reviews of the Cochrane Pregnancy and Childbirth group in

2009[40]. When evidence is spread across many databases, and in the case of fetal growth restriction evidence may be related to other conditions such as pre-eclampsia, it can be difficult to access appropriate up to-date robust evidence for clinical decision making. Systematic reviews provide a technique to allow individual pieces of research to be collected and if appropriate subjected to meta-analysis[41]. It is essential that these reviews are performed with rigorous methods and include an assessment of study quality of they are to have valid inferences and produce usable summaries to guide medical practice[41]. A review of systematic reviews of effectiveness for interventions for fetal growth restriction and compromise of fetal wellbeing will thus help summarise the most up to-date evidence and allow an assessment of the quality of the evidence.

2.4 Development of this thesis

The charity Wellbeing of Women funded an evidence synthesis project to systematically review the accuracy data for available tests for fetal growth restriction. The author worked on this project grant performing the systematic reviews. This work was then developed into a Clinical Research Training Fellowship funded by the Medical Research Council. This fellowship developed the original work to incorporate systematic reviews of effectiveness and decision analytic modelling.

2.5 Nomenclature used in thesis

As discussed in section 2.2.2 there is a lack of a standardised definition for FGR thus throughout the literature the terms SGA and FGR are used interchangeably. To try and ensure consistency through this thesis, the term SGA has been used when referring to infants/fetuses determined as small by being below a predetermined centile e.g. 10th and

FGR when describing those that may be SGA and/or definitely FGR from postnatal measuremnts.

As the objective of this thesis is to assess antenatal tests and make recommendations for obstetric practice the term SGA fetus has been employed as the tests are assessing the fetus not an infant. It is recognised that the eventual diagnosis is only made once the "fetus" has been delivered and thus the tests are in effect predicting/diagnosing SGA infants. As the primary literature assessed in this thesis uses the term SGA or SGA fetus, the decision was made to employ this terminology in this thesis.

CHAPTER 3: RESEARCH QUESTIONS ADDRESSED IN THE THESIS

3.1 Research questions addressed in this thesis by systematic review

3.1.1 Questions addressed in thesis by reviews of diagnostic test accuracy

1. What is the accuracy of Down's syndrome serum screening markers to predict small for gestational age fetuses?
2. What is the accuracy of uterine artery Doppler to predict small for gestational age fetuses?
3. What is the accuracy of umbilical artery Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing?
4. What is the accuracy of middle cerebral artery Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing?
5. What is the accuracy of ductus venosus Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing?
6. What is the summary of the evidence reviewed of test accuracy for prediction of small for gestational age fetuses and compromise of fetal wellbeing?
7. What is the reporting quality of primary studies of test accuracy in Obstetrics reviewed in this thesis and how has this changed over time?
8. What are the methodological quality of primary test accuracy studies in Obstetrics reviewed in this thesis and the sources of bias and variation and have these changed with the introduction of quality standards?

3.1.2 Questions addressed in the thesis by systematic review of reviews of effectiveness

9. How effective are the available treatments for preventing small for gestational age fetuses and compromise of fetal/neonatal wellbeing?

3.2 Question addressed in the thesis by decision analytic model based economic evaluation

10. What is the cost effectiveness of the antenatal tests and various treatment combinations to prevent small for gestational age fetuses and compromise of fetal/neonatal wellbeing?

PART A: SYSTEMATIC REVIEWS OF TEST ACCURACY

CHAPTER 4: METHODS FOR SYSTEMATIC REVIEWS OF DIAGNOSTIC TEST ACCURACY

4.1 Introduction

Chapters 5-10 of this thesis evaluates the test accuracy of antenatal tests to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing in pregnant women at varying levels of risk for these conditions. Systematic reviews of the available evidence were performed to assess test accuracy using a common methodology. This chapter provides an overview of the methods employed, where adaptations were necessary these are detailed in the chapters relevant to each test.

All reviews were performed with reference to the existing recommended methods and guidelines[42-46] and based on a prospective protocol[47]. The reviews were performed using the following steps[48]: (i) Framing the question, (ii) Study identification, (iii) Study quality assessment, (iv) Data synthesis and (v) Interpreting the findings.

This work has been published Morris R, Khan KS, Coomarasamy A, Robson S, Kleijnen J. The value of predicting restriction of fetal growth and compromise of its wellbeing: Systematic quantitative overviews (meta-analysis) of test accuracy literature. *BMC Pregnancy and Childbirth* 2007;7:3.

4.2 Framing the question

To ensure that the systematic review is correctly designed and ensure that the question is fully answered the research questions must be formulated appropriately. A clearly defined research question has four key components relating to the population under study, the test or intervention, the reference standard or comparator and the type of study designs to be included. When posing the question for the review the reviewer must construct the question so that all these components are included and thus all problems that need to be answered are identified. A comprehensive research question will then help determine the components of the search strategy. The questions posed in the systematic reviews of diagnostic test accuracy included in this thesis have common components as summarized below:

Population: Pregnant women in any health care setting, at any level of risk.

Index test: Down's syndrome serum screening markers - alpha feto protein (AFP), human chorionic gonadotrophin (HCG), unconjugated oestriol, inhibin A, pregnancy associated plasma protein A (PAPP-A). Doppler ultrasound – uterine artery (UA), umbilical artery (Umb), middle cerebral (MCA), ductus venosus (DV).

Reference standard: Any measurement of birth weight or nutritional status of newborn performed postnatally.

Study designs: Randomised controlled trials or observational studies where the results of the index test are compared with the results of the reference standard, allowing generation of 2x2 tables of accuracy.

4.2.1 Population

For each review the population was pregnant women. To ensure full assessment of the included tests as screening methods was made there were no restrictions made on healthcare setting nor level of risk of the women for complications. These factors were however considered in sub-group analyses.

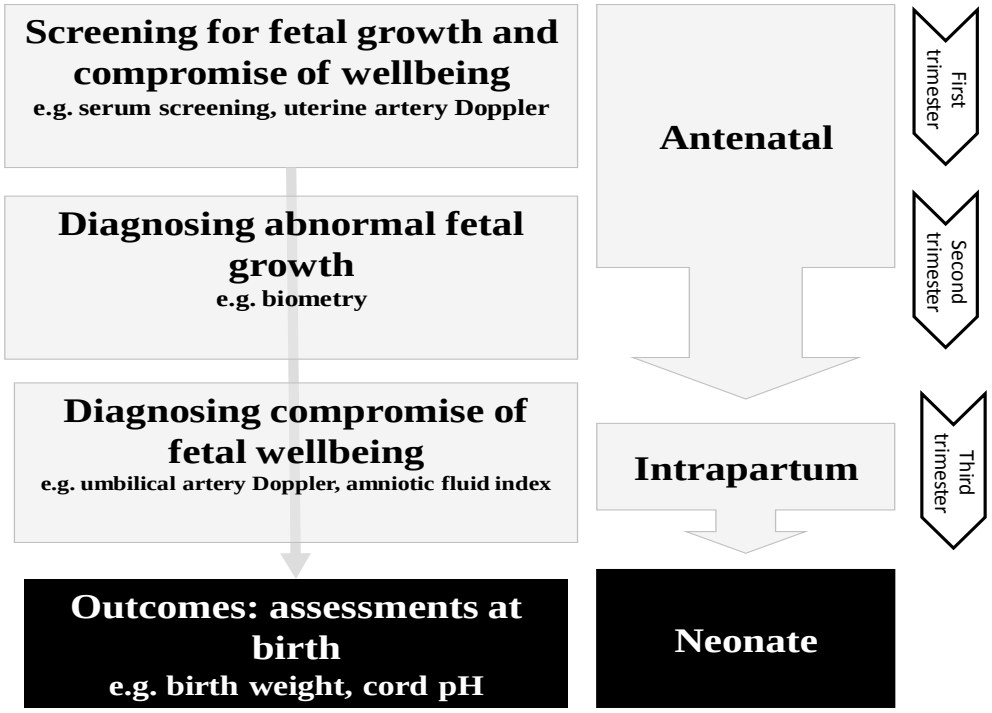
4.2.2 The tests

The tests to be investigated were prioritised on the basis of clinical relevance after consultation with experts in the field and investigation of the available evidence in a preliminary search (appendix 1) and consideration of their use in pregnancy (figure 4.1). For all the tests investigated all methods of analysis, imaging modalities and thresholds reported by the authors of the included studies were assessed. Sub-group meta-analyses of individual techniques and cut off values were employed in this thesis to give a full assessment of the tests under investigation.

4.2.3 Reference standards

As discussed in the background of this thesis (section 2.2.2), reference standards for SGA and FGR vary with no consensus on the best outcome measure to use for prediction of short term and long term mortality and morbidity in the infant. For the purpose of the test accuracy reviews the reference standards used were any reported in the included studies (birth weight centiles, absolute birth weight cut-offs and neonatal anthropometric measures) with meta-analysis only employed with studies using the same outcome measure.

Figure 4.1: Flow diagram of testing in pregnancy



Reference standards for neonatal wellbeing were any outcome measure performed after birth relating to neonatal wellbeing reported by the study authors. In many studies a composite outcome measure, adverse perinatal outcome (APO) was used. Composite outcomes bring together two or more events that are considered as a single outcome. For neonatal wellbeing this was usually a combination of outcome measures such as birth weight, Apgar scores, cord pH values etc. Where the authors did not report outcomes for APO then results were constructed using a hierarchy of outcomes, this technique maximised the number of events that could be included in the analyses. One problem with composite outcome measures is the assumption that the significance of the result applies to all components[49]. To address this issue a separate analysis was also performed using the component outcomes of the composite outcome measure. When the composite

outcome measure was used, care was taken to ensure that each individual was only counted once in each analysis. The use of composite outcome measures is an accepted technique in systematic reviews as long as the direction of effect for each of the included outcomes is in the same direction and separate analysis is performed looking at the individual components[50].

4.2.4 Study design

Acceptable study designs were RCTs and observational test accuracy studies (cohorts, case-control prospective) allowing generation of 2x2 tables of accuracy. Case series <10 cases and case-control studies defined by reference standard outcome (birth weight measurement) were excluded as these study designs have been shown to be associated with bias as sampling based on diseased (cases) and non-diseased (controls) can introduce spectrum effects[51].

4.3 Identification of the literature

The search protocol was designed with the aim of identifying literature concerning diagnostic tests to predict or diagnose small for gestational age fetuses/fetal growth restriction or compromise as using the elements of the framed question (section 4.2). Pilot searches were performed to ensure that the search strategies gave an acceptable level of specificity without compromising sensitivity.

The first database search performed used terms for FGR/SGA and combined them with methodological filters for identification of aetiologic and diagnostic test studies[52;53] (appendix 2). This search was to identify all tests performed for prediction and diagnosis

of the condition and was used for the reviews of serum markers (chapter 5) and uterine artery Doppler (chapter 6) as these reviews were performed in collaboration with researchers in Amsterdam who were looking at prediction of pre-eclampsia[54]. This search was performed by an experienced clinical librarian from the Amsterdam Medical Center (see Acknowledgements). Databases were searched from inception to April 2006 (see appendix 2 for databases).

The final search strategies (performed by the author for chapters 7, 8, 9) used relevant medical subheadings (MeSH), text words and word variants for FGR/SGA or fetal wellbeing and combined these with terms for the index tests using the AND operator. The individual search strategies for each review are detailed with each review chapter and databases were searched from inception to 2009.

Literature was identified via the following sources:

(a) General bibliographic databases including MEDLINE and EMBASE.

(b) Specialist computer databases – DARE, MEDION (a database of diagnostic test reviews set up by Dutch and Belgian researchers), the Cochrane Library and relevant specialist registers of the Cochrane Collaboration, particularly the Pregnancy and Child Birth Group;

(c) Contact with individual experts and those with an interest in this field to uncover grey literature;

(d) Contact with manufacturers of tests;

(e) Hand-searching of selected specialist journals;

(f) Checking of reference lists;

(g) SCISEARCH and Web of Science to identify frequently cited articles and conference abstracts.

All searches were made without language restrictions. A comprehensive database of articles relevant to each test was constructed using Reference Manager 11.0 software.

The titles and abstracts of the citations were scrutinised by the author, copies of full manuscripts of the citations that were likely to meet the selection criteria were obtained. The author then selected the studies, which met predefined and explicit criteria regarding populations, tests, reference standards and study design using a checklist, the items of which were based on selection criteria related to the question as detailed above in section 4.2 (population, diagnostic test, reference standard and study design). This checklist was piloted to ensure that all eligible studies were included and that the process was reproducible and reliable. Ideally whenever possible this process was repeated by a second reviewer independently (see acknowledgments). When disagreements occurred, the two reviewers met and if a consensus could not be reached the opinion of a third reviewer (Professor Khalid Khan) was sought. In the case of duplicate publications the most recent or up to date manuscript was selected. All foreign language papers were translated (see acknowledgements).

4.4 Assessment of the quality of the literature

The papers meeting the inclusion criteria were assessed for quality using well developed and validated tools for diagnostic research[34-36] by two independent reviewers (including

author). Each manuscript was assessed for reporting quality using the Standards for Reporting of Diagnostic Accuracy (STARD)[34] checklist (appendix 3). Methodological quality was defined as the confidence that the study design, conduct and analysis had minimized biases in addressing the research question, thereby focusing on the internal validity (i.e. the degree to which the results of an observation are correct for the patients being studied). Methodological quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS)[35] checklist (appendix 4).

In the assessment of study quality for the population, consecutive or random recruitment of pregnant women was considered ideal. Prospective recruitment was considered to introduce less bias than retrospective recruitment. The description of the population was considered ideal if the there was sufficient information about the pregnant women given to assign a level of obstetric risk, ideally this risk level was stated by the authors in the study's methods. The incidence of FGR was calculated for each study (reported in the table of study characteristics for each review) and used as a check for the authors' quantification of the risk category of the population.

Assessing the quality of performance and reporting of the index standard was individualised for each review enabling the assessment to look at individual aspects of each test that might introduce bias. Further details on the assessment of quality relating to individual tests can be found in the respective chapters. For the reference standard, any representation of birth weight or nutritional status of the newborn was considered acceptable. Information was collected on method of determination of reference standard, execution and blinding.

Ideal study design were trials or cohort studies, case-control studies were included but wherever possible, when the number of studies allowed it, they were excluded from meta-analysis due to the risk of introducing bias as discussed in section 4.2.4.

Verification bias was assessed using a flow chart for each study which documented the number of eligible women for the study, the number of women subjected to the index test, the number of women receiving the reference standard and the number of exclusions, withdrawals and uninterpretable results. Due to the large number of included studies these flow charts cannot be reported individually thus an assessment of verification was made with ideal verification when all women could be accounted for and the number of eligible women progressing to the reference standard was >90%.

The assessment of quality was represented by a bar chart. No attempt was made to apply a quality score as this has been shown to have little validity with the possibility of obscuring the strengths and weaknesses of a study[55] and quality was not used as an aspect in inclusion/exclusion of studies. Instead, an individual assessment was made of the most important quality items for each individual test under review and studies defined as high or low quality. This definition was used in the sub-group analysis. If the number of studies allowed then meta-regression analysis based on quality items was performed.

4.5 Data extraction

The extraction of a study's findings was conducted using a pre-designed and piloted data extraction form to avoid any errors. Data were recorded on an Excel spreadsheet. In at least 10% of studies the data extraction was repeated by another independent reviewer

(see acknowledgements). The form was used to extract clinical, methodological (QUADAS and STARD) and statistical data and varied only slightly during the reviews according to the test under review (see individual review chapters for data extraction forms). Any disagreements between reviewers were resolved by consensus or arbitration by a third reviewer (Professor Khalid Khan). Where multiple publications of the same study were identified, the most complete, relevant and up to-date study was included to avoid duplication.

4.6 Quantitative data synthesis

4.6.1 Summary measures for test accuracy

From the 2x2 tables, the true positive rate (sensitivity), false positive rate (1-specificity) and likelihood ratios (LRs) were calculated for each study along with their 95% confidence intervals (CIs). Where 2x2 tables contained zero cells, 0.5 was added to each cell to enable calculations[56]. MetaDisc[57] was used for calculations and STATA 10 (StataCorp, College Station, Tx, USA) for bivariate meta-analysis, meta-regression and drawing receiver operating characteristic curves (ROC). Statsdirect (Statsdirect Ltd.) was used to draw Forest plots.

4.6.2 Exploration of heterogeneity

Heterogeneity of results between studies was assessed graphically by looking at the distribution of sensitivities and specificities in the ROC space and LRs as measurement of accuracy size using a forest plot. Extreme values, outliers and threshold phenomena were explored. The loglikelihood and X^2 test were used to assess for heterogeneity statistically. The reasons for heterogeneity were explored using meta-regression and

sub-group analyses planned *a priori*. This was carried out using factors considered to be important beforehand, including:

- Variations in population – high and low risk defined by prevalence of FGR within the population.
- Variations in index test – e.g. type of test parameter, cut-off used
- Variations in reference standard – test used, threshold used
- Study quality
- Study design – cohort studies only.

4.6.3 Meta-analysis

Where sensitivity and specificity were independent of each other, meta-analysis was used. Pooled summary estimates were produced in the form of the summary LR as this is the measure which is most applicable clinically, in keeping with recommendations from Evidence-based Medicine Groups[4;58]. The LRs allow estimation of the probability of FGR or neonatal compromise with a specific test result. To generate the practical application of these LRs the post test probability of having the disease was generated (for either a positive or negative test result) using Bayes' theorem and the following formula: post test probability = likelihood ratio x pre-test probability/ [1-pre-test probability x (1-likelihood ratio)][59]. Estimates of pre-test probability were made using reports from previous studies and taking into account the risk rates for the population in question. The range of uncertainty was calculated using the 95% confidence intervals of the LRs for each test.

The bivariate meta-regression model[37;60;61] was used to meta-analyse estimates of sensitivity, specificity and LRs. The model assumes a bivariate normal distribution for the

logit transformed sensitivity and specificity values across studies by directly analysing the logit transformed sensitivity log (sens/(1-sens)) and specificity log (spec/(1-spec)) of each study in a single model. It preserves the two dimensional nature of the data produced in test accuracy studies and incorporates the inherent correlation that exists between sensitivity and specificity due to threshold effect. The model also accounts for heterogeneity beyond chance due to clinical or methodological differences in studies, employing a random effects model. In addition, the model acknowledges the difference in precision by which sensitivity and specificity have been measured in each study. This means that studies with a larger number of patients with the target condition receive more weight in the calculation of the summary estimate of sensitivity, while studies with more patients without the target condition are more influential in the pooling of specificity.

Sensitivity analysis was performed to check the robustness of all results. A p value of <0.05 was used throughout for statistical significance.

4.6.4 Publication bias

Analysis for assessing the risk of publication bias was carried out by producing funnel plots of log diagnostic odds ratios verses inverse of variance[62]. When no publication bias is present the plots will be shaped like a funnel because studies of smaller size are expected to have increased variation in the estimates of accuracy. The bigger the study variance, the lower the weighting of the study and the less information it provides. This means that in addition to small sample size of included primary studies, those studies reporting very high accuracy will also have a relatively big variance and thus be weighted less. The interpretation of the funnel plots took into account the debate surrounding their

use in diagnostic reviews[63]. The Harbord regression test for asymmetry based on the efficient score and its variance, Fisher's information, was performed[64].

4.6.5. Clinical application

The clinical impact of estimates of accuracy for a screening test depend on how the results of the test alter the patient's pre-test probability of disease, based on disease prevalence. The post-test probability can then be combined with estimates of effectiveness for known treatments[65]. From this data the number of women needed to be tested (number needed to test- NNTest) can be calculated, using a particular test, to prevent one case of an SGA fetus with a particular treatment and the number needed to treat (NNTreat), the number of test positive women needed to be treated to prevent one case of SGA. In the reviews of Down's syndrome serum markers (chapter 5) and uterine artery Doppler (chapter 6) (i.e. first and second trimester tests) clinical application was assessed using aspirin as this treatment is accepted as of potential use in SGA fetuses[39;66].

4.7 Description of data

For each test, information on individual studies was summarised as follows:

- **Table with methodological and reporting characteristics of included studies.** The table states the number of women tested in each study, the incidence of fetal growth restriction (based on the number of analysed cases divided by the total number of women at baseline (cohort studies and nested case-control studies)) and maternal age (given as mean (± SD) for the whole group unless otherwise stated).

- **Summary of quality and reporting items of the included studies.**

  Results were presented as 100% stacked bars, where figures in the stacks represent the number of studies.

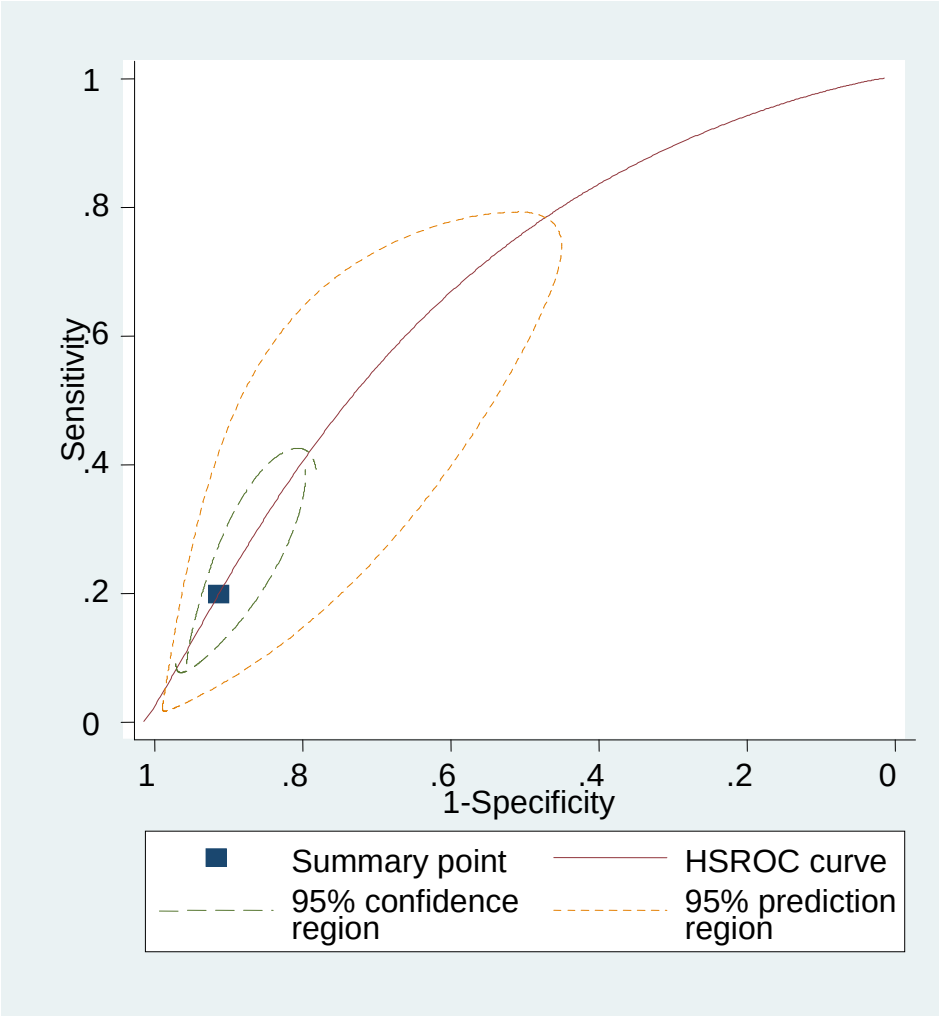
- **Forest plots of sensitivities (%), specificities (%) and LRs with their corresponding 95% CIs.**

  Numbers of women analysed are true positives/(true positives + false negatives) for sensitivity and true negatives/(false positives + true negatives) for specificity. Positive likelihood ratio = (LR+ve) sensitivity/(1-specificity) and negative likelihood ratio = (LR-ve) (1-sensitivity)/specificity. In chapter 6 the data for uterine artery Doppler is presented in tabular form to allow sensitivity and likelihood ratio data to be demonstrated.

- **HSROC (hierarchical summary ROC) curve.**

  A summary HSROC curve was drawn (according to the bivariate model) and example of which is shown in figure 4.2. In the summary ROC curve the vertical axis shows sensitivity, while the horizontal axis shows 1-specificity. The square represents the summary point of accuracy. The closer the index values are to the upper left corner, the greater the accuracy of the test. The ellipses represent the region containing likely combinations of mean values of sensitivity and specificity for the summary point (95% confidence region) and the likely value for the true operating point in a single future study (95% prediction region) . In chapter 6 the results are presented as an ROC plot to allow the results for the different Doppler indices to be presented.

Figure 4.2: Example of a hierarchical summary receiver operating characteristic curve drawn using the bivariate method.



- **Table with subgroup analyses.** (If applicable.)
- **Significance of publication bias.**

  The significance of publication bias is demonstrated using Harbord regression test of asymmetry p value.

CHAPTER 5: SYSTEMATIC REVIEW OF ACCURACY OF DOWN'S SYNDROME SERUM MARKERS TO PREDICT SMALL FOR GESTATIONAL AGE FETUSES

5.1 Abstract

5.1.1 Background

The purpose of this systematic review was to determine the accuracy of five serum analytes used in Down's syndrome serum screening for the prediction of small for gestational age fetuses.

5.1.2 Methods

These included searching of electronic data sources (inception to February 2007), hand searching of relevant journals, reference list checking of included articles and contact with experts. Articles in which the accuracy of an analyte used in Down's syndrome serum screening before the 25th gestational week was associated with the occurrence of small for gestational age fetuses were selected. Two authors independently extracted data on study characteristics, quality and results.

5.1.3 Results

Five serum screening markers were evaluated. 86 studies, testing 382,005 women (20,339 FGR cases) met the selection criteria. The results showed low predictive accuracy overall. For small for gestational age fetuses the best predictor was AFP>2.0MoM to predict birth weight <10th centile with birth <37 weeks; positive likelihood ratio 27.96 (8.02,97.48) and negative likelihood ratio 0.78 (0.55,1.11) (single

study). A potential clinical application using aspirin as a treatment is given as an example. There were methodological and reporting limitations in the included studies thus studies were heterogeneous giving pooled results with wide confidence intervals.

5.1.4 Conclusion

Down's syndrome serum screening analytes have low predictive accuracy for small for gestational age fetuses. They may be a useful means of risk assessment or of use in prediction when combined with other tests.

5.1.5 Publications arising from this work

Morris R K, Cnossen JS, Langejans M, Robson SC, Kleijnen J, Ter Riet G, Mol BW, van der Post JA, Khan KS. Serum screening with Down's Syndrome markers to predict pre-eclampsia and small for gestational age: Systematic review and meta-analysis. *BMC Pregnancy Childbirth 2008 Aug 4;8(1):33.*

5.2 Introduction

Second trimester serum screening for Down's syndrome is routinely offered to women in the United Kingdom and United States, either with the triple test (alpha-fetoprotein (AFP), human chorionic gonadotrophin (HCG) and unconjugated oestriol) or with the addition of inhibin A as the quadruple test. More recently first trimester screening with fetal nuchal translucency, HCG and pregnancy associated plasma protein A (PAPP-A) has provided an earlier, more effective screening method[67]. Due to their origin and sites of metabolism these biochemical markers may be useful in the prediction of SGA fetuses, there are however conflicting reports in the literature. Maternal serum levels of these analytes have been shown to be associated with adverse outcome[68;69] with low levels of PAPP-A having been suggested as a marker for impaired placental function and placentation[70]. There are studies however reporting contrasting views[71].

The purpose of this review was to investigate the accuracy of serum biochemical markers used in first and second trimester Down's syndrome serum screening in predicting SGA fetuses.

5.3 Methods

The methods used are outlined in chapter 4 with those specific to this review detailed below.

5.3.1 Data sources and searches

Electronic searches were performed with the assistance of experienced clinical librarians targeting the prediction of SGA fetuses. Medline, Embase, the Cochrane

Library (2006;4) and Medion from inception until February 2007 were searched. The search strategy is detailed in appendix 2.

5.3.2 Study selection

Criteria for included studies were those that reported on singleton pregnancies at any level of risk in any healthcare setting using any serum biochemical test used in Down's syndrome serum screening before the 25th week of gestation.

5.3.3 Data extraction and Study Quality Assessment

The data extraction form for this review can be found in appendix 5. For the index test any methods of laboratory analysis at any threshold was accepted (centiles and multiples of the median (MoM)). Acceptable reference standards for SGA fetuses included birth weight < 10th centile adjusted for gestational age and based on local population values and absolute birth weight threshold < 2500g. Severe SGA was defined as birth weight < 5th or < 3rd centile or < 1750g or and preterm SGA for SGA leading to delivery < 37 weeks. Neonatal ponderal index < 10th centile, skin fold thickness, and mid-arm circumference/head circumference were also assessed.

Items considered important for a good quality paper were prospective design with consecutive recruitment, full verification of the test result with reference standard (>90%), adequate description of the index test, use of appropriate reference standard, application of any preventative treatments, whether cases of pre-eclampsia were excluded from the results, whether fetuses with chromosomal and structural anomalies

were excluded and whether stillbirths and intrauterine deaths were excluded from the results. Further explanation of the quality assessment can be found in appendix 6.

5.3.4 Data synthesis and analysis

Results were pooled among groups of studies with similar characteristics, the same threshold for the index test, same reference standard threshold for SGA and the same trimester for testing. Bivariate meta-analysis was used to produce overall summary results.

Sub-groups were defined at the start of the review based on clinical criteria known to affect prognosis, method of index test or study quality: level of risk of population (high or low based on authors assessment and calculated incidence rates from results); type of assay used for index test; whether babies with chromosomal anomalies were excluded from the results; use of preventative treatment; quality of study. Sub-group analyses were performed where there were at least 3 studies with similar characteristics within that group. Funnel plots and the regression test for asymmetry were used to assess for publication bias.

5.3.5 Clinical application

In this review clinical application was assessed using aspirin as this treatment is accepted as of potential use in SGA fetuses[39;66] as detailed in section 4.6.5.

5.4 Results

5.4.1 Literature identification and study characteristics.

Figure 5.1 summarises the process of literature identification and selection. The references of the included papers are listed in appendix 7. Tables detailing the individual study characteristics of the included studies are available in appendix 8.

There were 86 included studies for SGA, reporting on 382,005 women (20,339 cases of SGA fetuses, incidence 5.32%). Among these studies, there were 61 cohort studies and 25 case control studies. Thirty-one studies were prospective, 17 retrospective and 38 of unclear design. Calculated incidence rates of SGA correlated well with the threshold used in 78 of studies and poorly in 8, incidence range for birth weight <10th centile was 1.2-63%. Three of the studies were performed in high risk populations, whereas the remainder were performed in low risk or screening populations. Due to the inclusion criteria of the studies the majority of tests were performed between 15 to 20 weeks. There were ten studies reporting on first trimester screening. Fifty studies reported on birth weight <10th centile, 13 on birth weight <5th centile, 27 on birth weight <2500g, 1 on birth weight <1500g, 1 on birth weight < 15th centile and 12 reported no threshold. The twelve studies with not threshold for SGA were excluded from the meta-analysis.

5.4.2 Study Quality

The quality assessment of included studies for SGA revealed deficiencies (figure 5.2). Only 40 studies contained an adequate description of the performance of the index test. None of the studies reported clearly on the performance of the reference standard. Blinding of the reference test was also poorly reported as was the use of any treatment

Figure 5.1: Study selection process for systematic review of accuracy of Down's syndrome serum markers to predict small for gestational age fetuses

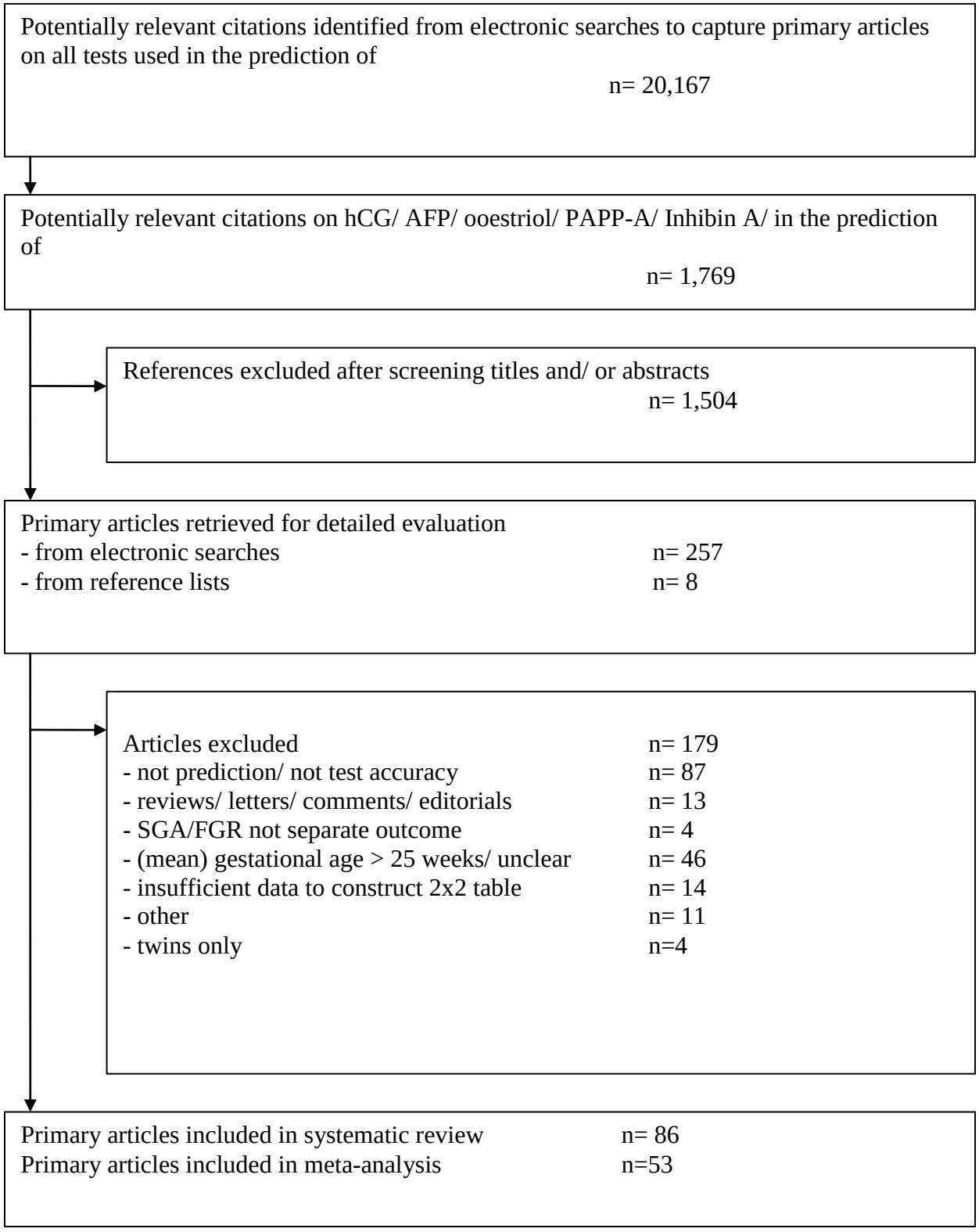
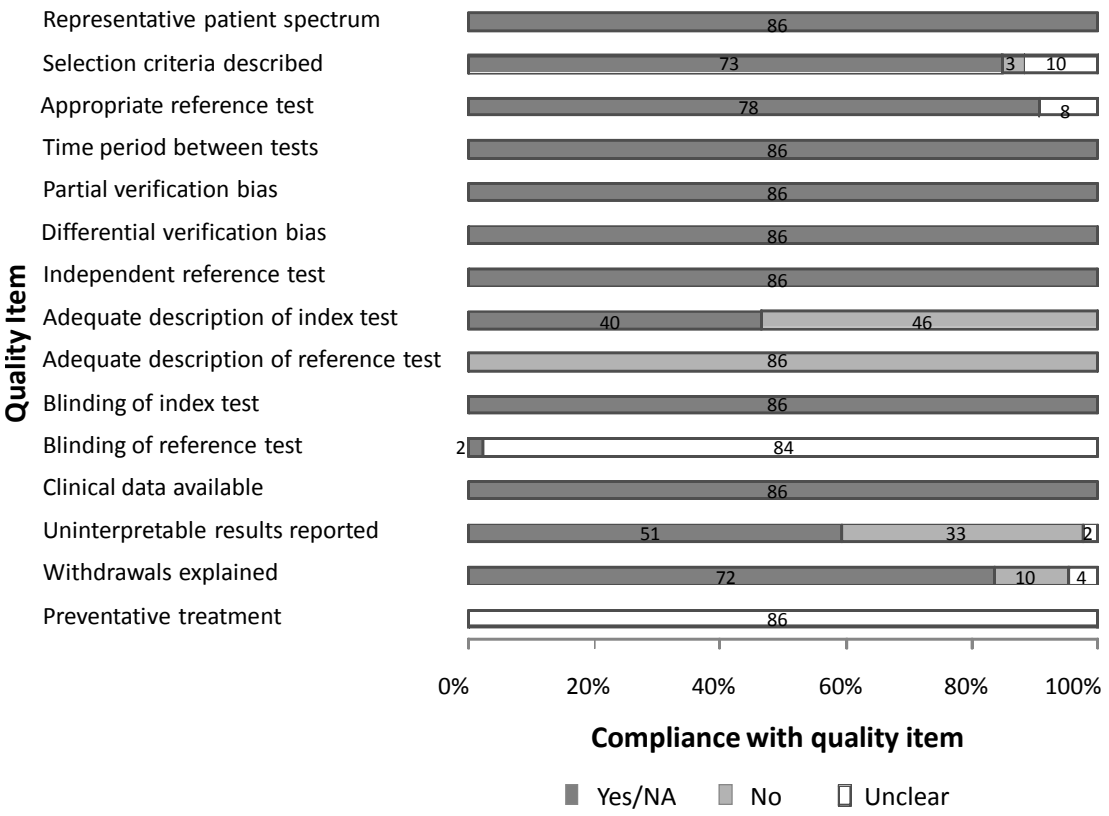


Figure 5.2: Bar chart illustrating the compliance with quality items for included studies in a systematic review of accuracy of Down's syndrome serum markers to predict small for gestational age fetuses. (Numbers in bars represent actual number of studies compliant).



in between the index test and reference standard. These items of quality of study design are important in diagnostic accuracy reviews.

Four papers only distinguished between SGA with PE and SGA alone; intrauterine deaths and stillbirths were excluded from the results for SGA in only 16 papers, in the remainder it was unclear; chromosomal and structural anomalies were excluded from 62

studies, unclear in 24. Twenty-five case control studies and eight studies in which thresholds for SGA were not defined were excluded from the final meta-analysis, leaving 53 studies.

5.4.3 Data analysis

The results for all serum markers are summarised in forest plots in figure 5.3 and summary receiver operating characteristic curves in figure 5.4.

Maternal serum alpha fetoprotein (AFP)

The results for AFP are summarized in appendix 9. All studies were performed in the second trimester. There were thirty studies included in the meta-analysis. The commonest threshold used were >2.0MoM (10 studies) and >2.5MoM (five studies) to predict birth weight <10^{th} centile. The best predictor for birth weight <10^{th} centile was AFP<10^{th} centile; LR+ 8.80 (5.57, 13.91), LR- 0.02 (0.00,0.34), this was a single study. For birth weight<5^{th} centile and birth weight <2500g, AFP>3.0MoM was the most accurate predictor. The most accurate predictor overall was AFP>2.0MoM to predict severe SGA (birth weight<10^{th} centile with birth <37 weeks): LR+ 27.96 (8.02, 97.48), LR- 0.78 (0.55, 1.11).

Maternal serum human chorionic gonadotrophin (HCG)

The results for HCG are summarized in appendix 10. There were 22 included studies in the meta-analysis, five looked at testing in the first trimester. The commonest thresholds used were HCG>2.0MoM (seven studies) and HCG>2.5MoM (four studies) for birth weight <10^{th} centile. The most accurate predictor for birth weight <10^{th} centile was

HCG>2.0MoM; LR+ 1.74 (1.48,2.04), LR- 0.95 (0.93,0.96). For birth weight <5th centile HCG>2.0MoM in the second trimester was the most accurate and for birth weight <2500g HCG>2.5MoM.

Maternal serum unconjugated Oestriol

The results for unconjugated oestriol are summarized in appendix 11. All studies were performed in the second trimester. There were seven included studies, the commonest threshold was oestriol<0.75MoM (2 studies) for birth weight <10th centile. The most accurate predictor for birth weight <10th centile was oestriol<0.75MoM; LR+ 2.54 (1.54, 4.19), LR- 0.75 (0.63,0.89). For birth weight <5th centile there were two studies for oestriol<0.5 MoM; LR+ 6.54 (0.98, 43.91), LR- 0.59 (0.03,13.28).

Maternal serum pregnancy associated plasma protein A (PAPP-A)

The results for PAPP-A are summarized in appendix 12. There were 10 included studies, seven were performed in the first trimester, the commonest thresholds were PAPP-A <5th centile (four studies), PAPP-A<10th centile (five studies) for birth weight<10th centile. The most accurate predictor for birth weight <10th centile was PAPP-A<1st centile; LR+ 3.50 (2.53, 4.82), LR- 0.98 (0.97,0.99). For birth weight <5th centile, the most accurate predictor was again PAPP-A<1st centile; LR+ 4.36 (3.27, 5.80), LR- 0.97 (0.96, 0.98).

Figure 5.3 Forest plot summarising the accuracy results of Down's syndrome serum markers to predict small for gestational age fetuses. Open triangles represent overall pooled result, shaded triangles represent sub-groups.

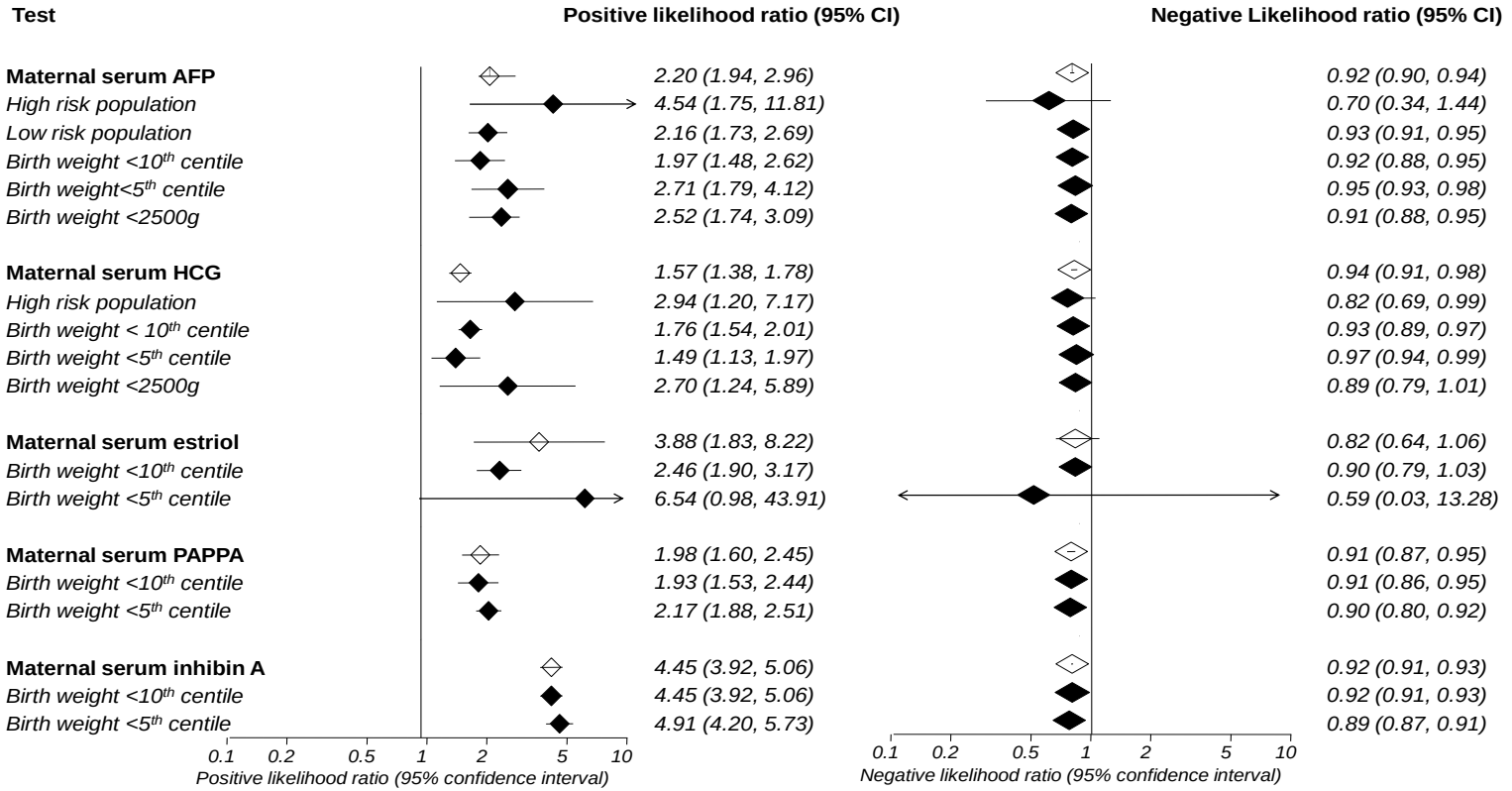
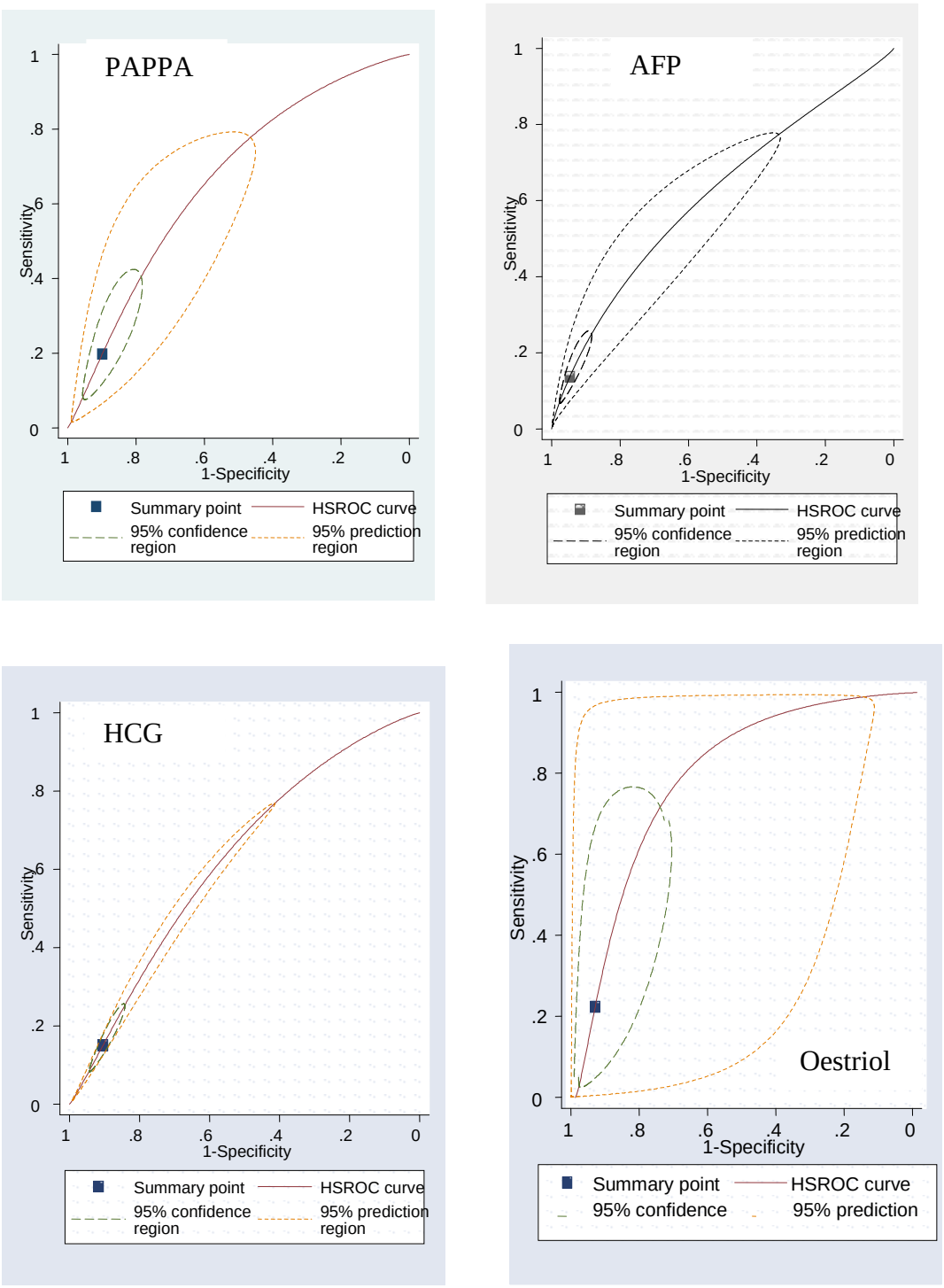


Figure 5.4: Summary receiver operating characteristic curves for Down's syndrome serum markers to predict small for gestational age fetuses produced using the bivariate method.



Maternal serum inhibin A

The results for inhibin A are summarized in appendix 13. There was only one study, looking at second trimester testing, using a cut-off of inhibin A>2.0MoM, the results for prediction of birth weight<10^{th} centile were LR+ 4.45 (3.92, 5.06), LR- 0.92 (0.91,0.93) and birth weight <5^{th} centile; LR+ 4.91 (4.20,5.73), LR- 0.89 (0.87,0.91). As there was only one study for inhibin A there is no corresponding HSROC curve.

Triple test (serum AFP, HCG and unconjugated oestriol)

There were two studies, second trimester testing, with different cut-offs for prediction of birth weight<10^{th} centile: triple test >1:190 LR+ 1.07 (0.60, 1.91), LR- 0.98 (0.82,1.17) and triple test>1:250 LR+ 2.71 (1.77,4.17), LR- 1.19 (0.01,2.47).

Gestation at which testing performed

The table in appendix 14 shows the different results achieved where testing was performed in both the first and second trimester. Overall for HCG, testing in the second trimester had improved positive likelihood ratios.

Sub-group and sensitivity analysis.

For sub group analysis, a sub-group had to include at least three studies within each analyte and threshold and thus it was only possible to conduct a sub-group analysis for calculated incidence of disease (incidence >10% or incidence ≤ 10%). The results for sub-group analysis are shown in appendix 15. There was no significant difference between the subgroups. Most of the studies included in the review excluded fetuses with other structural or chromosomal anomalies from the results and included live births

only, thus subgroup analysis could not be performed in these areas. Sensitivity analysis including only those studies with these characteristics showed no significant difference. The same was true for the assessment of study quality i.e. most studies were of an overall similar quality to make sub-group analysis impossible but sensitivity analysis showed no difference when extremely low quality studies were excluded.

Publication bias

Funnel plots (not shown) and the regression test for asymmetry showed no significant publication bias (p=0.6).

Clinical application with aspirin

The results for clinical application with aspirin for SGA fetuses are shown in table 5.1. The results show that by testing with inhibin A in a low risk population the number of women needed to treat to prevent one case of SGA can be reduced from 90 to 30, having to test 909 women.

5.5 Discussion

This review evaluated the accuracy of five serum screening markers used in Down's syndrome screening and a composite triple test. The results showed low predictive accuracy overall. The best predictor for SGA overall for birth weight <10^{th} centile was AFP<10^{th} centile while AFP>3.0MoM was the best predictor of birth weight <5^{th} centile. These results were both based on single studies. AFP showed improvement in

predictive accuracy when looking at severe disease (birth weight <3rd centile). HCG showed improved prediction when comparing second trimester to first trimester testing.

The strength of this review and validity of its findings lie in the methodological strengths used. It complied with existing guidelines for the reporting of systematic reviews[43] and also guidelines specific to the reporting of systematic reviews of observational studies[42].

Extensive literature searches without language restrictions were performed. Careful attention to assessment of quality of study design and reporting was made.

Previously published reviews in this area are restricted to a systematic review evaluating predictive tests for pre-eclampsia[72]. The review by Conde Agudelo concluded that the tests investigated had a low predictive value. The methods used in this review have however been criticized[73] and it was restricted in the thresholds and tests it reviewed. To our knowledge there are no previously reported systematic reviews in this area for SGA.

This review primarily reported likelihood ratios, as discussed in section 4.6.3, as they are thought to be more clinically meaningful than sensitivities and specificities. Recent research suggests that independently pooled likelihood ratios should be interpreted with caution as positive and negative likelihood ratios are related statistics (just like sensitivity and specificity)[74]. Bivariate analysis[37] was employed to account for this, as was sensitivity analysis with pooled sensitivity and specificity. This found no difference in the interpretation of the results.

Table 5.1: Serum screening among pregnant women and number of women needed to be tested and treated with aspirin to prevent one case of SGA fetus (birth weight <10th centile).

Test result	Prevalence SGA (%)	Probability of SGA after testing positive (%)	Risk of SGA after treatment*	Probability of SGA after treatment	NNTest[1]	NNTreat[2]
No test, no treatment[3]	10.0	10.0	-	10.0	-	-
No test, treat all[3]	10.0	-	0.90	9.0	-	90
Alpha feto-protein>2.0MoM: Sensitivity 60%; Specificity 98%						
Test all, treat test positives	10.0	28.3	0.90	25.4	167	35
Human chorionic gonadotrophin>2.0MoM: Sensitivity 12%; Specificity 94%						
Test all, treat test positives	10.0	16.2	0.90	14.6	833	62
Unconjugated oestriol<0.75MoM: Sensitivity 37%; Specifcity 88%						
Test all, treat test positives	10.0	22.0	0.90	19.8	270	45
Pregnancy associated plasma protein A (PAPP-A)<1st centile: Sensitivity 3%; Specificity 99%						
Test all, treat test positives	10.0	28.0	0.90	25.2	3333	36
Inhibin A>2.0MoM: Sensitivity 11%; Specificity 98%.						
Test all, treat test positives	10.0	33.1	0.90	29.8	909	30

Alpha feto-protein>2.0MoM to predict severe FGR: Sensitivity 22%, Specificity 99%						
Test all, treat test positives	1.0	22.0	0.90	19.8	454	45

* RR 0.90 (95% CI 0.83-0.98) Askie et al[66].

[1] NNTest is number needed to test and treat with aspirin to prevent one case of SGA calculated by 1/ (proportion true positives (TP) – (proportion TP * RR)).

[2] NNTreat is number need to treat if only treat test positives with aspirin calculated by 1/ (probability after testing positive – probability after treatment).

[3] Numbers are equal for all tests regardless of threshold, sensitivity and specificity.

MoM multiples of median

SGA small for gestational age

The assessment of study quality was hindered by a lack of clear reporting, which is a common problem in diagnostic reviews as standards for quality and checklists for assessing it are fairly new. It has been previously reported that poor study design and conduct can affect the estimates of diagnostic accuracy[51;75]. However, it is not entirely clear how individual aspects of quality may affect this and to what magnitude, particularly in the area of Obstetrics. Application of quality scores has been shown to be of little value on diagnostic reviews[55]. However, due to the lack of clear reporting, it was not possible to perform sub-group analysis based on individual quality criteria.

One of the areas in which reporting was uniformly poor was in the details provided regarding performance of the reference standard. For FGR there is still no convincing evidence as to which is the best definition of the condition at birth nor which is the best predictor of future infant and childhood morbidity and mortality for term. Population-based birth weight standards were the most commonly used, however it is important to realize that these do not distinguish between the small healthy infant and the compromised infant. Customised growth charts that are adjusted for sex, gestation, parity, maternal weight and height and ethnicity, have been shown to improve the detection of at risk of stillbirth[17] while neonatal indices have been shown to identify the malnourished infant at risk of peripartum asphyxia[16]. Unfortunately these were rarely used as outcome measures in the included studies.

Confounding factors in the measurement of serum screening markers, but mainly AFP, is the association with these markers with intrauterine death, preterm labour and chromosomal and structural anomalies[76-78]. Ideally all the included papers in this review

should have included only women with live births and fetuses with no other chromosomal or structural anomalies, this however was not always clearly reported. Sensitivity analysis, including only studies that did report exclusion of these subjects, showed no significant difference in estimates of test accuracy.

In this review it was assumed that the markers act independently but this may not be the case. The relationship between PE and SGA must also be taken into account. For HCG measurement the risk of SGA has been shown by logistic regression to be dependent on the presence of PE[79]. Ideally, included cases of SGA for this review would have been those where there was no PE but this was again poorly reported.

When assessing the clinical relevance of these tests it is important to look at severe disease as this causes the majority of maternal, fetal and neonatal complications, and thus prediction and prevention of this form of disease would have the greatest health impact. For the studies included in the meta-analysis there were only three that had results for severe SGA and these were insufficient to make an accurate assessment of the prediction of this form of disease.

The calculations of NNTreat and NNTest show that the number of women needed to treat with aspirin to prevent one case of a SGA fetus can be reduced by testing with a serum screening marker and then only treating the test positives. As aspirin is not routinely used as a treatment, these calculations serve to contextualize the predictive value of these markers as individual tests. The costs of introducing aspirin as a treatment would need to be balanced against the costs of the test, costs of failing to treat

the women with a false negative result that then go on to develop disease and any patient costs in terms of anxiety from screening and over treatment in the false positive category. To thus calculate the true clinical effectiveness of these tests these results would need to be incorporated in to a full cost-effectiveness analysis. Before any treatment is introduced into clinical practice an assessment of its side effects must be made e.g. gastric bleeding for the mother, risk of placental abruption.

As SGA is a condition with relatively low prevalence a clinically useful test would need to have a high positive LR (>10) and low negative LR (<0.10)[59]. From the results of this review it is unlikely that any one serum screening marker in isolation will provide this. Future research should thus concentrate in two areas. The first should be to address the limitations within the primary literature as identified by this review; poor reporting, exclusion of intrauterine deaths and chromosomal and structural anomalies from the results, separation of PE and SGA and prediction of severe disease. This may not necessarily require further primary research as there are sufficient large, cohort studies available, but meta-analysis based on individual patient data to address the deficiencies in reporting could be performed. Secondly future research should focus on combinations of markers as predictors and combinations of tests such as serum screening markers and uterine artery Doppler[80] to improve the predictive accuracy to a clinically useful value.

As Down's serum screening is routinely performed in many developed countries the cost of implementing use of these results as a predictive test for SGA would be small. However as aspirin is the only preventative treatment with any proven benefit in these

conditions and has minimal adverse events this cost has to be compared to that of implementing aspirin treatment to all pregnant women.

5.6 Conclusion

Down's serum screening analytes have low predictive accuracy for small for gestational age fetuses. They may be a useful means of risk assessment or of use in prediction when combined with other tests.

CHAPTER 6: SYSTEMATIC REVIEW OF ACCURACY OF UTERINE ARTERY DOPPLER TO PREDICT SMALL FOR GESTATIONAL AGE FETUSES

6.1 Abstract

6.1.1 Background

Alterations in uterine artery waveforms are associated with development of preeclampsia and small for gestational age fetuses, which are important causes of maternal and perinatal morbidity and mortality. The purpose of this review was to evaluate the accuracy of uterine artery Doppler to predict small for gestational age fetuses.

6.1.2 Methods

Searches in Medline, Embase, Cochrane Library, MEDION (all from inception to April 2006), reference lists of eligible articles, and contact with experts. Without language restrictions, all studies on uterine artery Doppler in first and second trimester that allowed 2x2 table construction were selected. Multiple reviewers independently selected studies, extracted data on participants, Doppler indices, and outcomes, and assessed study validity. Bivariate meta-analysis of sensitivity and specificity was conducted and likelihood ratios were calculated.

6.1.3 Results

There were 61 studies testing 41,131 women (3,723 cases) for SGA fetuses. Second trimester testing performed better than first trimester testing. Increased pulsatility index (PI) with notching best predicted SGA in low-risk populations and increased PI (with notching) best predicted severe SGA fetuses, range of LR+ 9.1 to 14.6, LR- 0.34 to 0.89. Most Doppler indices showed low predictive accuracy. Estimates vary across population risk and severity of outcome.

6.1.4 Conclusions

Abnormal uterine artery waveforms show low predictive accuracy overall for small for gestational age fetuses with moderate predictive accuracy when predicting more severe forms. Pulsatility index alone or in combination with (bilateral) notching, are the most predictive Doppler indices. The use of these indices in clinical practice should not hamper execution of meta-analysis based on individual patient data or of new methodological high quality primary studies combining Doppler with other tests.

6.1.5 Publications arising from this work

Cnossen JS, Morris RK, Mol BWJ, ter Riet G, van der Post JAM, Bindels PJE, Robson S, Kleijnen J, Coomarasamy A, Khan KS. Uterine artery Doppler to predict fetal growth restriction: a systematic review and bivariate meta-analysis. *Canadian Medical Association Journal 2008;178(6):701-11*

6.2 Introduction

Pregnancies affected by SGA fetuses are characterized by a failure of the second wave trophoblast invasion (at 16-22 weeks) of the endometrio-myometrial vasculature[19]. This results in abnormal uteroplacental blood flow, which has led to the idea of using Doppler assessment of the uterine artery blood flow velocity waveforms as part of routine ultrasound screening[81]. Non-pregnant and first-trimester uterine artery blood flow velocity waveforms are characterized by low end-diastolic velocities and an early diastolic notch. Persistence of a diastolic notch (beyond 24 weeks' gestation) or abnormal flow velocity ratios are associated with inadequate trophoblast invasion[82]. However, the results of uterine artery Doppler studies show considerable variation. Several factors, for example variation in design and population, have been reported in relation to conflicting results[82].

The objective of this review was to investigate the accuracy of all uterine artery Doppler indices in predicting small for gestational age.

6.3 Methods

The methods used are outlined in chapter 4 those specific to this review are detailed below.

6.3.1 Data Sources

The electronic searches were performed from inception to April 2006. The electronic search strategy consisted of MeSH or keyword terms related to the topic (SGA/FGR)

combined with methodological filters for identification of studies on diagnostic tests and aetiology[53]. The search is detailed in appendix 3.

6.3.2 Study selection

Criteria for included studies were that they reported on singleton pregnancies at any level of risk in any healthcare setting using uterine artery Doppler screening before the 25th week of gestation.

6.3.3 Data extraction and Study quality assessment

The data extraction form for this review can be found in appendix 16. For description of the index test for uterine artery Doppler, the paper had to clearly state the route of measurement (transvaginal or transabdominal), the site of measurement, the measurement parameter used (e.g. pulsatility index) and the cut-off point used. For site of measurement, only the papers that used the main uterine artery branch were included, thus papers reporting on uteroplacental, placental bed, arcuate or spiral arteries were excluded. Ideally the site of measurement in the main uterine artery was further defined as the apparent crossover of the uterine artery with the internal iliac artery for trans-abdominal scanning, as this is the point of measurement that is considered the most reproducible thus reducing inter-operator variation (used as an assessment of quality not for inclusion/exclusion). The corresponding point for trans-vaginal measurement was the internal cervical os.

For this review the following quality items, assessing quality of the study design, were considered to not be applicable due to the nature of the test; time period between tests

(review was using test for prediction not diagnosis), partial verification bias (reference standard is equal to outcome), differential verification bias (non invasive reference test), incorporation bias (uterine artery Doppler is always independent of the measurement of birth weight). When assessing quality of reporting of the study, the following were considered not applicable; adverse events from performing index test or reference standard due to the non-invasive nature of the tests being reviewed. Acceptable reference standards for SGA/FGR included birth weight <10th, <5th, or 3rd centile adjusted for gestational age and based on local population values; neonatal ponderal index <10th centile; skin fold thickness; mid-arm circumference/head circumference and absolute birth weight thresholds. Further explanation of the quality assessment can be found in appendix 17.

6.3.4 Data synthesis and analysis

Results were pooled among groups of studies with similar Doppler indices (table 6.1), similar outcome, and according to risk using a bivariate regression model. In cases where there were several reported thresholds for a particular Doppler index, the most commonly reported threshold was selected for meta-analysis. Subgroup analyses as defined *a priori* was performed: outcome (severe; mild/overall), risk (low risk/unselected population; high-risk) and gestational age at testing (before and after 16 weeks). Sensitivity analyses were performed for application of preventative treatment (yes; no/unclear) and high quality studies. Studies were considered of high quality when they scored positive on at least four out of the following items: prospective design with consecutive recruitment, appropriate reference standard, adequate description of the index test, follow-up > 90%, and reporting of preventative treatment.

Table 6.1: Explanation of Doppler indices.

A/C ratio	peak systolic / early diastolic ratio (A/C)
Any notching	presence of early diastolic notching, unilateral or bilateral not specified
Bilateral notching	presence of early diastolic notching in both main uterine arteries
D/S ratio	diastolic/ systolic ratio
D/S or notching	D/S ratio with or without presence of unilateral of bilateral early diastolic notching
Notch (Depth) Index	(notch – early diastolic flow) / notch ((D-C)/D)
Pulsatility index (PI)	(peak systolic flow – end diastolic flow) / mean flow ((A-B)/M)
PI and notching	pulsatility index combined with presence of unilateral or bilateral early diastolic notching
PI or notching	pulsatility index with or without presence of unilateral or bilateral early diastolic notching
Resistance index (RI)	(peak systolic flow – end diastolic flow) / peak systolic flow ((A-B)/A)
RI and notching	resistance index combined with presence of unilateral or bilateral early diastolic notching
RI or notching	resistance index with or without presence of unilateral or bilateral early diastolic notching
S/D ratio	peak systolic/ late diastolic ratio (also known as A/B ratio)
S/D or notching	S/D ratio with or without presence of unilateral of bilateral early diastolic notching
Unilateral notching	presence of early diastolic notching in one main uterine artery

6.3.5 Clinical Application

Clinical application using aspirin therapy was performed as for the review of Down's syndrome serum screening described in section 4.6.5.

6.4Results

6.4.1 Literature identification and study characteristics

Figure 6.1 summaries the process of literature identification and selection. The electronic search generated 18,871 citations, screening on title and abstract identified 311 potentially relevant manuscripts and reference list checking identified a further eight, the full papers of all these articles were obtained. After reading the full manuscripts a total of sixty papers were included in this review, the references of the included papers are listed in appendix 18. Tables detailing the individual study characteristics are available in appendix 19 (low risk and unselected population) and appendix 20 (high risk population). The sixty included studies tested 40,637 pregnant women (4067 SGA cases). Fifty-seven cohort studies and three randomised trials were included. Forty-eight studies were prospective, ten were retrospective, and two were unclearly designed. Calculated incidences of SGA fetuses correlated poorly with the thresholds (birth weight in centiles) based on local population charts. The mean incidences for an unselected, a low risk, and high risk population with a birth weight threshold < 10^{th} centile were 9.6%, 8.2% and 20.7% respectively. Doppler measurements were mainly performed between 18 and 24 weeks at a routine prenatal care scan. Ten studies reported data on testing prior to 16 weeks. Thirty-five studies reported data on more than one Doppler index or a combination of Doppler indices.

6.4.2 Study Quality

The quality assessment of included studies for SGA fetuses showed deficiencies in some areas (figure 6.2). Over 70% of studies met the following QUADAS items: avoidance of partial and differential verification, independent reference test, and blind assessment of index test. Many studies scored poorly on the following items: adequate descriptions of selection criteria and reference test, blind assessment of the reference test, and availability of clinical data. Application (or not) of preventative treatment was reported in 18 publications. Preventative treatment was applied in the three included randomized trials after performing Doppler tests in a series of women.

6.4.3 Data Analysis

Table 6.2 and 6.3 show the estimates for prediction of SGA fetuses in low risk/unselected populations and high risk populations respectively, these are demonstrated as tables rather than forest plots as in the other reviews to allow more data to be presented. Overall an increased PI with notching in the second trimester best predicted SGA (LR+ 9.1 (95% CI 5.0-16.7); LR- 0.89 (95%CI 0.85-0.93)) in low-risk populations. Severe SGA in low risk populations was best predicted by an increased PI (LR+ 13.7 (95% CI 10.3-16.9); LR- 0.34 (95%CI 0.23-0.48)) or an increased PI with notching (LR+ 14.6 (95% CI 7.8-26.3); LR- 0.78 (95%CI 0.68-0.87)) in the second trimester. Doppler testing to predict SGA in high-risk populations showed low estimates of accuracy. An increased RI (> 0.58 or > 90th centile) in the second trimester best predicted severe SGA (LR+ 10.9 (95% CI 10.4-11.4); LR- 0.20 (95%CI 0.14-0.26)). Figure 6.3 shows the ROC plots for (pooled) results of Doppler testing in the second trimester according to risk and for severe disease.

Figure 6.1: Process from initial search to final inclusion for uterine artery Doppler studies to predict SGA fetuses/FGR. FGR fetal growth restriction; SGA small for gestational age.

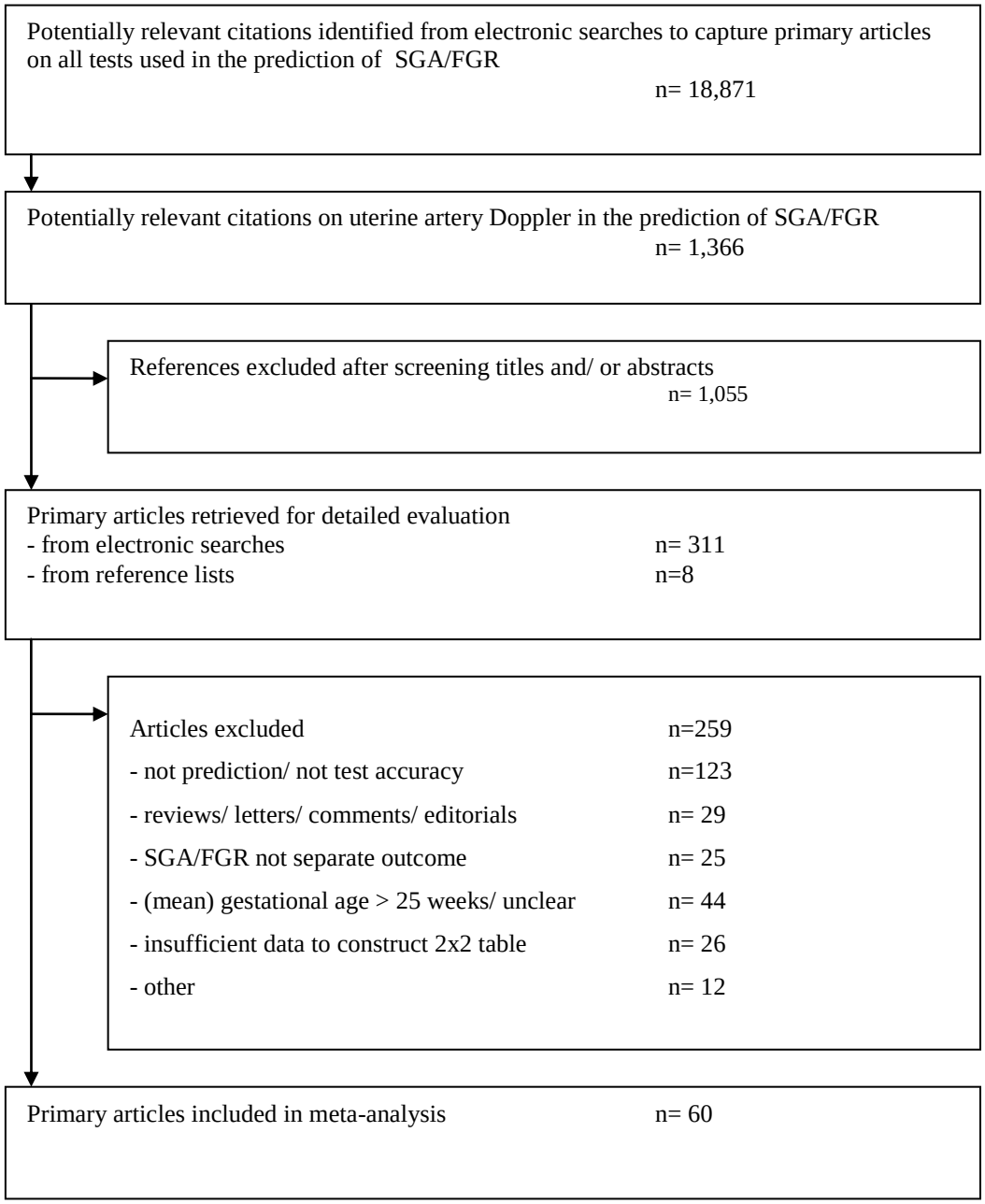
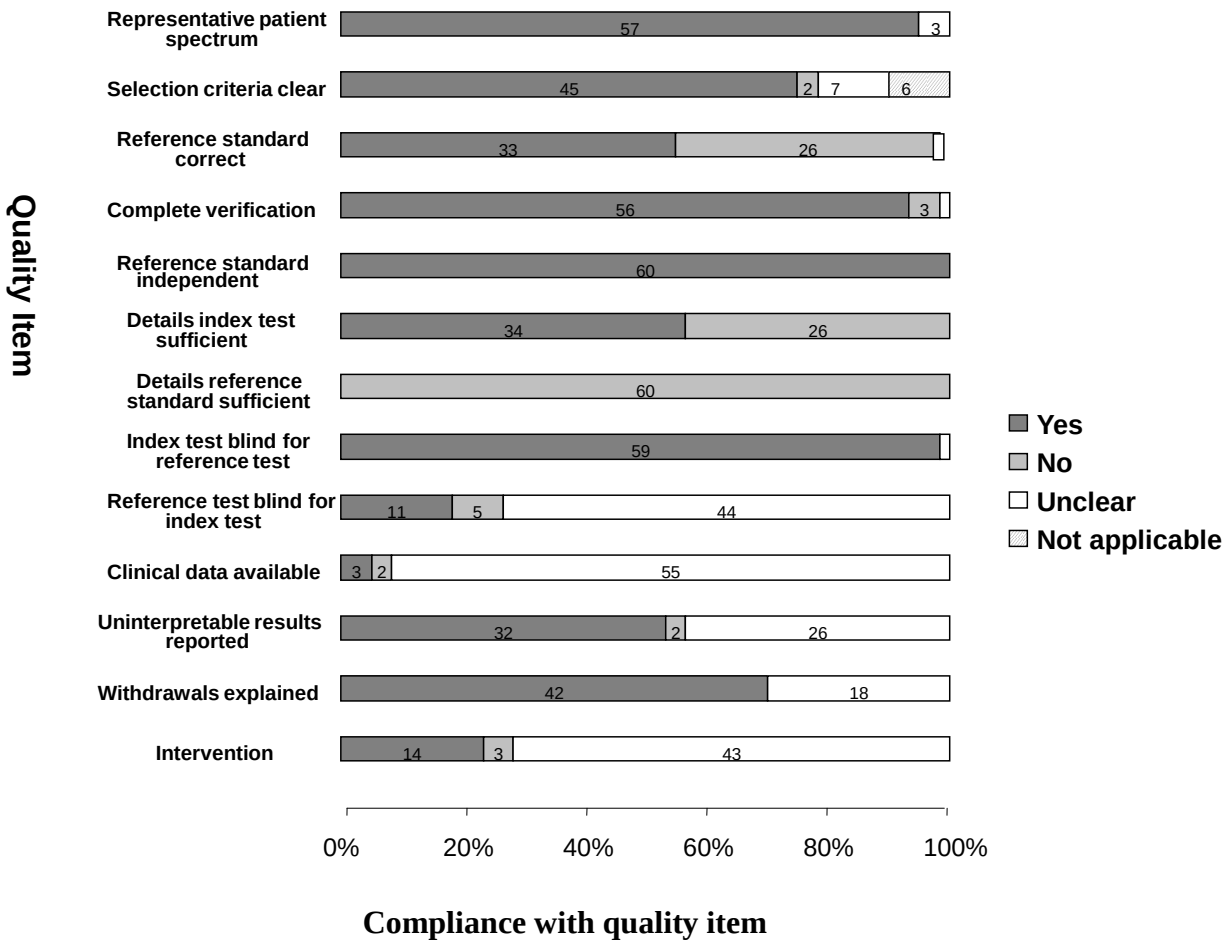


Figure 6.2: Bar chart illustrating the compliance with quality items for included studies in the systematic review of accuracy of uterine artery Doppler to predict small for gestational age fetuses. (Numbers in bars represent actual number of studies compliant).



Sub-group and sensitivity analysis

Sensitivity analysis, excluding studies that applied preventative treatment, did not improve the predictive accuracy (appendix 21). When the criteria for high quality studies were applied results showed low to moderate accuracy estimates in both low-risk (appendix 22) and high-risk populations (appendix 23).

Publication bias

Funnel plots (not shown) and the regression test for asymmetry showed no significant effect of publication bias (p=0.67).

Clinical application with aspirin

Table 6.4 shows the impact in clinical practice with the use of aspirin in women at risk of a SGA fetus. The absolute effect of aspirin depends on the risk of fetal growth restriction. The higher the risk, the lower the number of women needed to treat to prevent one case of fetal growth restriction and vice versa. As shown in the table, if aspirin were to be used for all pregnant women (prevalence 10%) without Doppler testing then 91 women would need to be treated with aspirin to prevent one case of a SGA fetus. If only those women with a positive test result were treated (pulsatility index, sensitivity 23%, specificity 91%) then 41 women would need to be *treated*, a number considerably lower than that without testing. However, to obtain this 395 women need to be *tested* to prevent this one case. Using a test such as resistance index and notching, which has a much higher sensitivity (40%) but same specificity (91%) will lead to a lower number needed to *treat* (28), and a similar number needed to *test* (227).

Table 6.2: Pooled and single estimates for uterine artery Doppler predicting fetal growth restriction in low risk/ unselected populations.

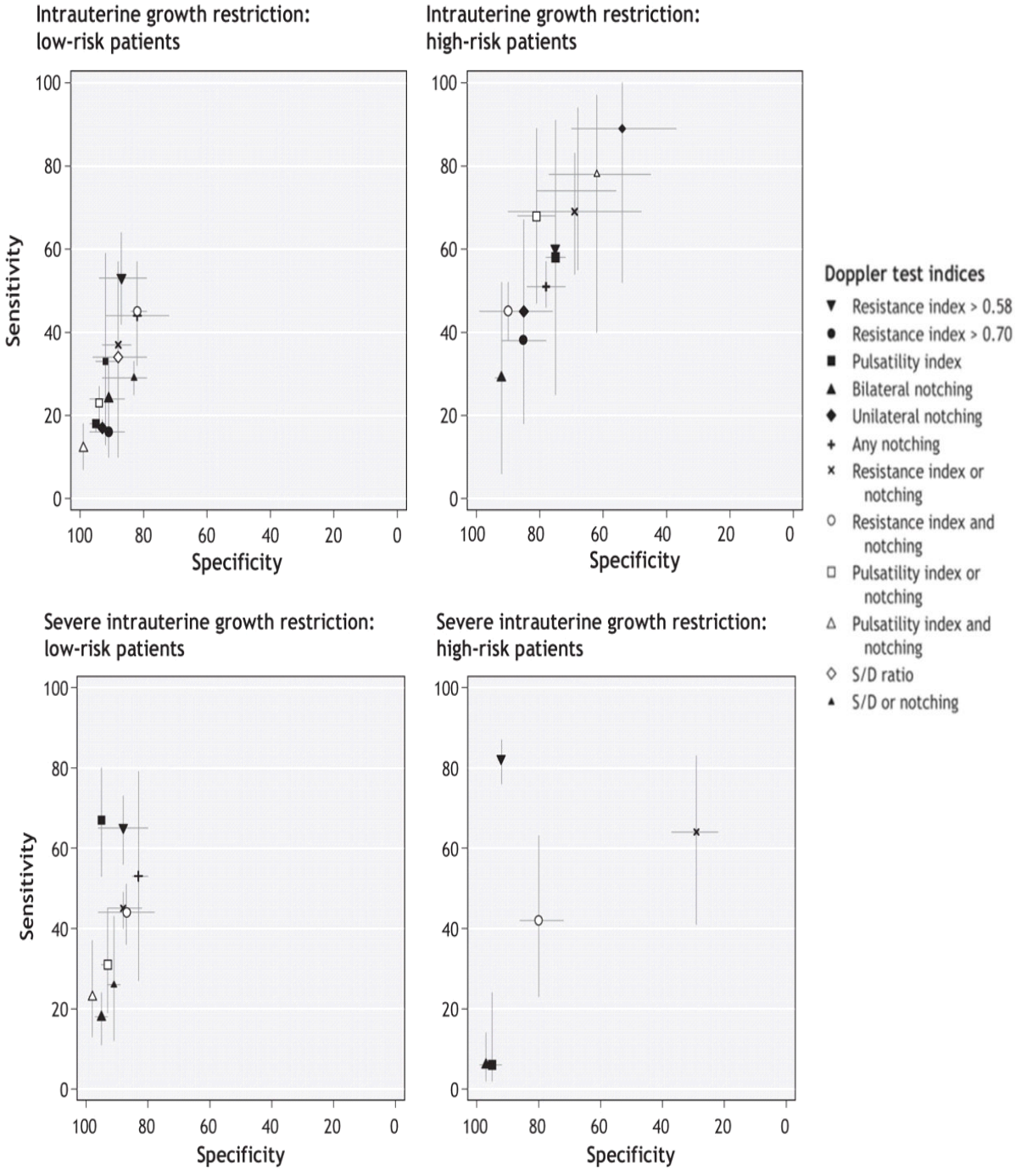
Doppler Index	No. of studies	No. of women	Sensitivity % (95% CI)	Specificity % (95% CI)	LR positive (95% CI)	LR negative (95% CI)
Birth weight < 10th centile or < 2500g/ 2nd trimester Doppler testing						
RI (0.58 or 90th)	9	3304	53 (42-64)	87 (79-94)	4.0 (0.68-23.1)	0.54 (0.27-1.1)
RI (0.7 or 95th)	2	665	16 (10-23)	91 (86-97)	1.9 (0.5-3.3)	0.92 (0.83-1.0)
PI	3	12097	18 (16-19)	95 (92-97)	3.4 (1.7-5.1)	0.87 (0.84-0.90)
Bilateral notching	11	10229	24 (14-34)	91 (86-97)	2.8 (0.26-30.0)	0.83 (0.53-1.3)
Unilateral notching	2	3819	18 (12-24)	91 (90-92)	2.4 (1.6-3.7)	0.89 (0.84-0.94)
Any notching	4	2162	44 (32-57)	82 (72-92)	1.8 (0.61-5.6)	0.73 (0.45-1.2)
RI or notching	5	5043	37 (33-40)	89 (81-96)	3.3 (0.72-15.2)	0.71 (0.59-0.87)
RI and notching	1	946	45 (37-53)	82 (79-84)	2.4 (1.9-3.0)	0.68 (0.58-0.77)
PI or notching	2	2116	12 (8-16)	94 (93-95)	3.9 (3.0-4.7)	0.82 (0.77-0.87)
PI and notching	1	1757	12 (7-18)	99 (98-99)	9.1 (5.0-16.7)	0.89 (0.85-0.93)
S/D ratio	3	1661	34 (10-57)	88 (79-96)	2.7 (0.53-13.9)	0.76 (0.42-1.4)
Notch index	1	288	33 (13-59)	92 (88-95)	4.3 (1.9-8.4)	0.72 (0.49-0.91)
S/D or notching	3	2173	29 (25-33)	83 (79-93)	2.1 (1.0-3.2)	0.82 (0.74-0.90)
Birth weight < 10th centile or < 2500g/ 1st trimester Doppler testing						
RI (0.70 or 95th)	1	1008	67 (35-90)	75 (72-78)	2.7 (1.6-3.5)	0.44 (0.18-0.81)
PI	1	3045	12 (8-16)	96 (95-96)	2.7 (1.9-3.8)	0.92 (0.88-0.96)
Bilateral notching	3	1420	74 (55-93)	42 (0-84)	1.3 (0.35-4.7)	0.62 (0.08-4.7)
Any notching	(H) 2	866	85 (80-89)	50 (49-50)	1.7 (0.60-4.8)	0.31 (0.07-1.4)
Birth weight < 5th centile, < 3rd centile or < 1750g/ 2nd trimester Doppler testing						
RI (0.58 or 90th)	(H) 3	1551	66 (64-67)	88 (82-94)	5.6 (0.86-36.5)	0.39 (0.21-0.73)

PI	1	1757	67 (53-80)	95 (94-96)	13.7 (10.3-16.9)	0.34 (0.23-0.48)
Bilateral notching	(H) 2	2657	22 (17-27)	96 (95-96)	4.9 (0.66-37.0)	0.82 (0.47-1.4)
Any notching	1	890	53 (27-79)	83 (80-85)	3.1 (1.7-4.4)	0.57 (0.30-0.85)
RI or notching	(H) 3	3650	45 (31-59)	88 (79-98)	3.8 (0.96-14.9)	0.63 (0.52-0.76)
RI and notching	(H) 2	1404	44 (36-52)	87 (85-89)	3.4 (0.86-13.3)	0.64 (0.52-0.79)
PI or notching	1	1757	31 (19-45)	93 (92-95)	4.7 (2.9-7.0)	0.74 (0.60-0.86)
PI and notching	1	1757	23 (13-37)	98 (98-99)	14.6 (7.8-26.3)	0.78 (0.68-0.87)
S/D or notching	1	768	26 (12-43)	91 (89-93)	2.9 (1.5-4.9)	0.82 (0.64-0.95)
Birth weight < 5th centile, < 3rd centile or < 1750g/ 1st trimester Doppler testing						
PI	1	999	24 (12-41)	95 (94-97)	5.3 (2.8-9.5)	0.79 (0.64-0.91)

Table 6.3: Pooled and single estimates for uterine artery Doppler predicting fetal growth restriction in high risk populations.

Doppler Index	No. of studies	No. of women	Sensitivity % (95% CI)	Specificity % (95% CI)	LR positive (95% CI)	LR negative (95% CI)
Birth weight < 10th centile or < 2500g/ 2nd trimester Doppler testing						
RI (0.58 or 90th)	6	885	74 (55-94)	68 (56-81)	2.4 (0.72-7.8)	0.37 (0.07-2.0)
RI (0.7 or 95th)	4	527	38 (18-58)	85 (78-92)	2.6 (0.78-8.6)	0.73 (0.41-1.3)
PI	2	445	60 (33-86)	74 (69-78)	2.3 (1.1-4.8)	0.55 (0.19-1.6)
Bilateral notching	4	588	29 (6-52)	92 (91-94)	3.8 (0.7-7.0)	0.77 (0.51-1.0)
Unilateral notching	2	151	45 (23-67)	85 (76-94)	3.0 (0.72-12.8)	0.65 (0.25-1.6)
Any notching	10	989	55 (42-69)	78 (72-83)	2.5 (0.93-6.7)	0.57 (0.25-1.3)
RI or notching	4	629	69 (54-83)	69 (48-90)	2.2 (0.55-9.1)	0.45 (0.17-1.2)
RI and notching	4	444	62 (23-100)	89 (80-98)	5.7 (0.87-37.2)	0.42 (0.07-2.7)
PI or notching	2	138	68 (47-89)	81 (75-87)	3.6 (2.0-5.1)	0.40 (0.14-0.65)
D/S ratio	1	48	78 (40-97)	62 (45-77)	2.0 (1.1-2.7)	0.36 (0.10-0.94)
D/S or notching	1	48	89 (52-100)	54 (37-70)	1.9 (1.1-2.2)	0.21 (0.04-0.85)
Birth weight < 10th centile or < 2500g/ 1st trimester Doppler testing						
PI	3	785	34 (31-37)	76 (73-80)	1.5 (0.43-4.8)	0.86 (0.47-1.6)
Bilateral notching	1	72	75 (19-99)	41 (29-54)	1.3 (0.50-1.7)	0.61 (0.11-1.8)
Birth weight < 5th centile, < 3rd centile or < 1750g/ 2nd trimester Doppler testing						
RI (0.58 or 90th)	(H) 2	362	82 (76-87)	92 (92-93)	10.9 (0.7-168.9)	0.20 (0.11-0.35)
PI	1	351	6 (2-14)	95 (92-97)	1.2 (0.47-3.2)	0.99 (0.92-1.03)
Bilateral notching	1	351	6 (2-14)	97 (95-99)	2.3 (0.79-6.7)	0.97 (0.92-1.0)
RI or notching	1	182	64 (41-83)	29 (22-37)	0.90 (0.61-1.2)	1.2 (0.64-2.1)
RI and notching	1	170	42 (23-63)	80 (72-86)	2.1 (1.2-3.4)	0.72 (.50-0.95)

Figure 6.3: Receiver operating characteristics plots for pooled and single accuracy estimates with 95% confidence intervals for second trimester Doppler indices predicting small for gestational age fetuses (SGA) according to risk and for predicting severe SGA. Note: x-axis shows reversed specificity.



When looking at a more severe outcome, that of fetal growth restriction leading to preterm delivery, sensitivity and specificity of the test improve and thus only 142 women are needed to undergo the test and 15 treated with aspirin to prevent one case of severe FGR.

6.5 Discussion

This review evaluates the accuracy of 15 uterine artery Doppler indices used to predict SGA fetuses. An increased pulsatility index alone or in combination with notching predicts (severe) SGA best in low-risk populations, whereas in high-risk populations the best predictor is an increased resistance index. Other Doppler indices show low to moderate accuracy estimates.

Strengths and weaknesses of the review

These are as discussed in the previous chapter (section 5.5) and in the summary of test accuracy reviews chapter (chapter 10). Due to the lack of clear reporting it was not possible to perform multivariate subgroup analysis based on individual quality criteria. Therefore this review reports the overall results. Areas in which reporting was uniformly poor in this review were in the details regarding performances of the index test and the reference standard.

Table 6.4: Doppler testing among pregnant women and number of women needed to be tested and treated with aspirin to prevent one case of small for gestational age fetus.

Test result	Prevalence SGA (%)	Probability of SGA after testing positive (%)	Risk of SGA after treatment*	Probability of SGA after treatment	NNTest[1]	NNTreat[2]
No test, no treatment[3]	10.0	10.0	-	10.0	-	-
No test, treat all[3]	10.0	-	0.90	8.9	-	91
Pulsatility Index: Sensitivity 23%; Specificity 91%						
Test all, treat test positives	10.0	22.1	0.90	19.7	395	41
Resistance Index and notching: Sensitivity 40%; Specificity 91%						
Test all, treat test positives	10.0	33.1	0.90	29.4	227	28
Pulsatility Index, delivery < 34 weeks: Sensitivity 64%; Specificity 95%						
Test all, treat test positives	10.0		0.90		142	15

* Treatment benefit, relative risk (RR) 0.90 (95% CI 0.83-0.98) Askie et al[66]. [1] Number needed to test calculated by 1/ (proportion true positives (TP) – (proportion TP * RR)).

[2] Number needed to treat calculated by 1/ (probability after testing positive – probability after treatment).

[3] Numbers are equal for all tests regardless of threshold, sensitivity and specificity.

Strengths and weaknesses in relation to other studies

Previously published reviews concluded that the tests investigated had limited predictive accuracy for SGA fetuses[25;26]. The reviews were restricted in the thresholds and Doppler indices they reviewed. Two reviews reported on preeclampsia, FGR and perinatal death and were both restricted to Medline. One[26] was based on a systematic search until January 1997, the other[25] included only unselected populations until 2001. Since these reviews, substantial new evidence has emerged in particular on some Doppler indices, e.g. pulsatility index, allowing for more robust and specific inferences for clinical practice.

Unanswered questions, future research and implications

When considering whether a predictive test should be applied in clinical practice the following must be considered: the prevalence of the disease and the predictive accuracy of the test, the cost and patient acceptability of the test and the treatments available for the disease in question. SGA is a disease with relatively low prevalence and a clinically useful test would thus need to have a high positive LR (>10) and low negative LR (<0.10)[59]. From the results of this review pulsatility index and (bilateral) notching are the most promising Doppler indices to provide this and thus these are the indices that should be used in daily clinical practice. However, it should be recognized that the results vary according to population risk (tables 6.2 and 6.3). Uterine artery Doppler is a non-invasive test and thus acceptable to patients. It is a specialized test both in terms of the equipment required and the expertise of the operator. In Western countries uterine artery Doppler could be fairly easily performed at the time of the detailed anomaly scan, for developing countries it would be a difficult test to introduce into routine antenatal

practice. For mothers being identified as “at risk” of SGA from an antenatal test can cause considerable anxiety. At present there is no pharmacological treatment or management strategy (e.g. regular ultrasound scanning, early delivery) that has been shown to effectively prevent the development of these diseases or ameliorate their complications. However, research into aspirin as a treatment has shown a small preventative effect [RR 0.90 (0.83-0.98)][66] in the absence of any serious side effects, it is a cheap and readily available treatment. In this instance, a false negative test result is potentially more harmful than a false positive test result and this must be considered in the future when looking at predictive test accuracy and test/treatment combinations in SGA. The clinical impact of the estimates of accuracy that are produced depends on how the resultant changes in probabilities due to Doppler testing alter therapeutic effectiveness in decision making[65]. This impact can be illustrated with an example of decision making in clinical practice about the use of aspirin in women at risk of preeclampsia (table 6.4). The absolute effect of aspirin depends on the risk of a SGA fetus. The higher the risk, the lower the number of women needed to treat to prevent one case of SGA and vice versa. The risk, and hence the therapeutic benefits, depends on the post-test probabilities, calculated from sensitivity and specificity, of SGA associated with Doppler testing. To calculate the real effectiveness of such strategies, economic and patient costs and benefits of both test(s) and treatment should be incorporated.

The results of this meta-analysis are limited by the quality of the primary included studies. This poor quality may be due to poor quality of the design and conduct of the original study or due to poor reporting. This quality issue may be resolved in two ways; in the first instance a meta-analysis based on individual patient data (IPD) should be

performed[83]. IPD meta-analysis can overcome problems due to poor reporting and study heterogeneity. If however, following an IPD meta-analysis it is found that the design and conduct of the individual studies is poor then a large prospective primary research study must be recommended. Any future research should also concentrate on the application of combinations of tests e.g. biochemical tests and uterine artery Doppler; this is the diagnostic process that is used in clinical care and may also help improve the predictive accuracy of the tests to clinically important values.

6.6 Conclusion

Abnormal uterine artery waveforms show low predictive accuracy overall for small for gestational age fetuses with moderate predictive accuracy when predicting more severe forms. Pulsatility index alone or in combination with (bilateral) notching, are the most predictive Doppler indices. The use of these indices in clinical practice should not hamper execution of meta-analysis based on individual patient data or of new methodological high quality primary studies combining Doppler with other tests.

CHAPTER 7: SYSTEMATIC REVIEW OF UMBILICAL ARTERY DOPPLER TO PREDICT SMALL FOR GESTATIONAL AGE FETUSES AND COMPROMISE OF FETAL/NEONATAL WELLBEING

7.1 Abstract

7.1.1 Background

Alterations in umbilical artery waveforms are associated with the small for gestational age fetus and the fetus at risk of compromise, which are important causes of perinatal morbidity and mortality. The purpose of this review was to evaluate the accuracy of umbilical artery Doppler to predict the fetus at risk of compromise.

7.1.2 Methods

Electronic searches of Medline, Embase, Cochrane library, Medion (inception-March 2009), hand searching of relevant journals, reference list checking of included articles, contact with experts to identify relevant literature. Two reviewers independently selected relevant articles without language restrictions. The proportion of initially identified studies that met the selection criteria was 3.5%. Data were extracted on study characteristics, quality and results to construct 2x2 tables. Likelihood ratios for positive and negative test results, sensitivity and specificity were generated for different Doppler indices at various thresholds and the different reference standards using bivariate meta-analysis.

7.1.3 Results

One hundred and forty studies met the selection criteria. 11 were in a low-risk population (8042 fetuses); 104 in a high-risk population (19,191 fetuses); 15 in a mixed risk population (4350 fetuses) and 10 studies exclusively in multiple pregnancies (1709 fetuses). Umbilical artery Doppler showed better prediction of small for gestational age [pooled positive likelihood ratio 3.76 (2.96, 4.76), pooled negative likelihood ratio 0.52 (0.45, 0.61)] and for compromise of fetal/neonatal wellbeing [pooled positive likelihood ratio 3.41 (2.68, 4.34), pooled negative likelihood ratio 0.55 (0.48, 0.62)] in a high risk population. Sub-group analysis in a high risk population showed clinically useful accuracy for prediction of intra-uterine death, acidosis and admission to neonatal intensive care.

7.1.4 Conclusions

Umbilical artery Doppler is a moderately useful test in a high risk population to predict the fetus/neonate at risk of compromise. Further research should concentrate on its use within a series of tests e.g. with biophysical profile and other fetal Doppler indices.

7.1.5 Publications arising from this work

Morris RK, Malin GL, 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. *Archives of Disease in Childhood Fetal and Neonatal* 2010;95(suppl 1): Fa13-Fa14. In press with *Ultrasound in Obstetrics and Gynaecology*.

7.2 Introduction

The majority of the research in this area has been to find an antenatal test that can distinguish between the normal SGA fetus and the fetus at risk that may benefit from intervention. The most investigated technique is umbilical artery Doppler which at present forms the mainstay of risk assessment in this area[21;84]. Previous systematic reviews and meta-analyses have reported varying results in different populations[27;85-88], most strikingly in a high risk population a series of Cochrane reviews have culminated in an odds ratio (OR) of 0.71 (95% confidence interval 0.50-1.01) for perinatal mortality[85;87]. The lack of observed effect in randomized controlled trials (RCT) could be due to poor accuracy of tests as the RCT design evaluates both the effectiveness of the test and any intervention[89].

The purpose of our review was to investigate the accuracy of umbilical artery Doppler in all pregnant populations in predicting SGA fetuses and compromise of fetal/neonatal wellbeing.

7.3 Methods

The methods used are outlined in chapter 4 those specific to these review are detailed below.

7.3.1 Data Sources

Electronic searches were performed targeting the prediction of FGR/SGA and fetal/neonatal compromise. The databases searched were Medline, Embase, the

Cochrane Library (2009;1) and Medion from inception until March 2009. The search strategies are detailed in appendix 24.

7.3.2 Study selection

Criteria for selection were studies that reported on pregnancies at any level of risk in any healthcare setting using umbilical artery Doppler at any gestation. Test accuracy studies allowing generation of 2x2 tables were included. Acceptable reference standards for SGA fetuses included birth weight < 10^{th} centile adjusted for gestational age and based on local population values, birth weight mean less than two standard deviations and absolute birth weight threshold < 2500g. Severe FGR was defined as birth weight < 5^{th} or < 3^{rd} centile or < 1500g. Neonatal ponderal index < 10^{th} centile, skin fold thickness, and mid-arm circumference/head circumference were also assessed. Reference standards for wellbeing were any outcome measure performed after birth relating to neonatal wellbeing that was reported by the study authors. Any outcome measure was accepted as there is no consensus as to which is the best measure at birth to predict long term morbidity and mortality. A composite outcome measure adverse perinatal outcome (APO) was used by some authors (see section 4.2.3). For multiple/duplicate publication of the same data set, the most recent and/or complete study was included only.

7.3.3 Data Extraction and Study Quality Assessment

The data extraction form for this review can be found in appendix 25. See appendix 26 for further explanation of the quality assessment. As there were a sufficient number of cohort studies, case control studies were excluded from the statistical analysis as this

type of design in diagnostic test accuracy studies has been shown to be associated with bias[51].

7.3.4 Data Synthesis and analysis

Results were pooled among groups of studies with similar characteristics; population (high, low, mixed risk or multiple gestations based on investigators description) and the same reference standard threshold for SGA fetus/compromise of wellbeing. To allow the maximum number of studies to be included in the meta-analysis two composite outcome measures were used. The measure SGA utilized birth weight<10^{th} centile where possible, where this was not reported the nearest measure of birth weight reported was used. For wellbeing the outcome APO was used where reported. For studies that did not report this measure, outcomes were used in a hierarchy using the severest form of neonatal compromise reported.

Heterogeneity was assessed graphically by looking at the distribution of the sensitivities and specificities in the receiver operating characteristic (ROC) space and LRs as a measurement of accuracy size using a Forest plot. The X^2 test and Cochrane Q test were used to assess for heterogeneity statistically. The reasons for heterogeneity were explored using meta-regression and subgroup analyses. Sub-groups were defined at the start of the review based on clinical criteria known to affect prognosis (SGA population versus pre-eclamptic (PE) population), method of index test, study quality, gestation, singleton gestations only, whether babies with chromosomal anomalies were excluded from the results and timing of test used for analysis to delivery. Sub-group analyses were performed where there were at least 3 studies with similar characteristics

within that group. Studies were considered to be high quality when they reported positively on 4 or more of the items as listed previously. Threshold effect (when bivariate meta-analysis was not possible, less than 4 studies) was explored by observing the ROC curves and by calculating Spearman correlation coefficients[46].

7.4 Results

7.4.1 Literature identification and study characteristics

Figure 7.1 summarises the process of literature identification and selection. Tables detailing the citations of the included studies and the individual study characteristics are available in appendices 27 and 28 respectively.

High risk population

There were 104 included studies investigating 19,191 fetuses. These studies generated 444 2x2 tables; 59 reported on SGA fetuses of which 26 reported on SGA alone, 78 reported on wellbeing (45 wellbeing alone). There were 102 cohort studies and 2 cross-sectional. Thirty-eight studies used prospective recruitment, 14 retrospective and 52 unclearly designed. Fifty-three studies reported on singleton pregnancies only. Fifty-three were performed in the third trimester, five in the second, 23 mixed, seven post-term and 19 unclearly reported.

Low risk population

There were 11 included studies reporting on 8042 fetuses, generating 44 2x2 tables. Four studies looked at both SGA and wellbeing, six SGA alone and one wellbeing alone. Among these 11 studies, there was one RCT and 10 cohort studies. There were

three prospective studies, one retrospective and eight where the study design was unclear. Eight studies reported on singleton pregnancies only. Five studies were performed in the third trimester, one in the second trimester, three mixed gestation and two gestation not reported.

Mixed, unselected, unreported population

There were 7 studies in an unselected population, 5 mixed risk and 3 unreported risk; 4350 fetuses were investigated generating 55 2x2 tables. Nine studies reported on SGA, 5 on SGA alone; 10 on wellbeing, 6 on wellbeing alone. There was one RCT, 14 cohort studies of which 7 reported prospective recruitment, 1 retrospective and 7 unclearly designed. Eleven studies reported on singleton pregnancies only. 13 studies were performed in the third trimester, 2 in a mixed gestation population.

Multiple pregnancies

There were 10 studies that included multiple pregnancies only, 1709 fetuses were investigated generating 31 2x2 tables. Ten studies reported on SGA, 3 on SGA alone and 7 studies on wellbeing combined with SGA. All studies were cohort studies, one reported prospective recruitment. Six studies were performed in the third trimester, 4 in mixed gestation.

7.4.2 Study Quality

The quality assessment of included studies is summarized in figure 7.2. There was poor reporting of description of index and reference tests and blinding of the reference test. Only twenty studies reported clearly whether preventative treatment had been used.

Figure 7.1: Process from initial search to final inclusion for umbilical artery Doppler studies to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing (up to March 2009).

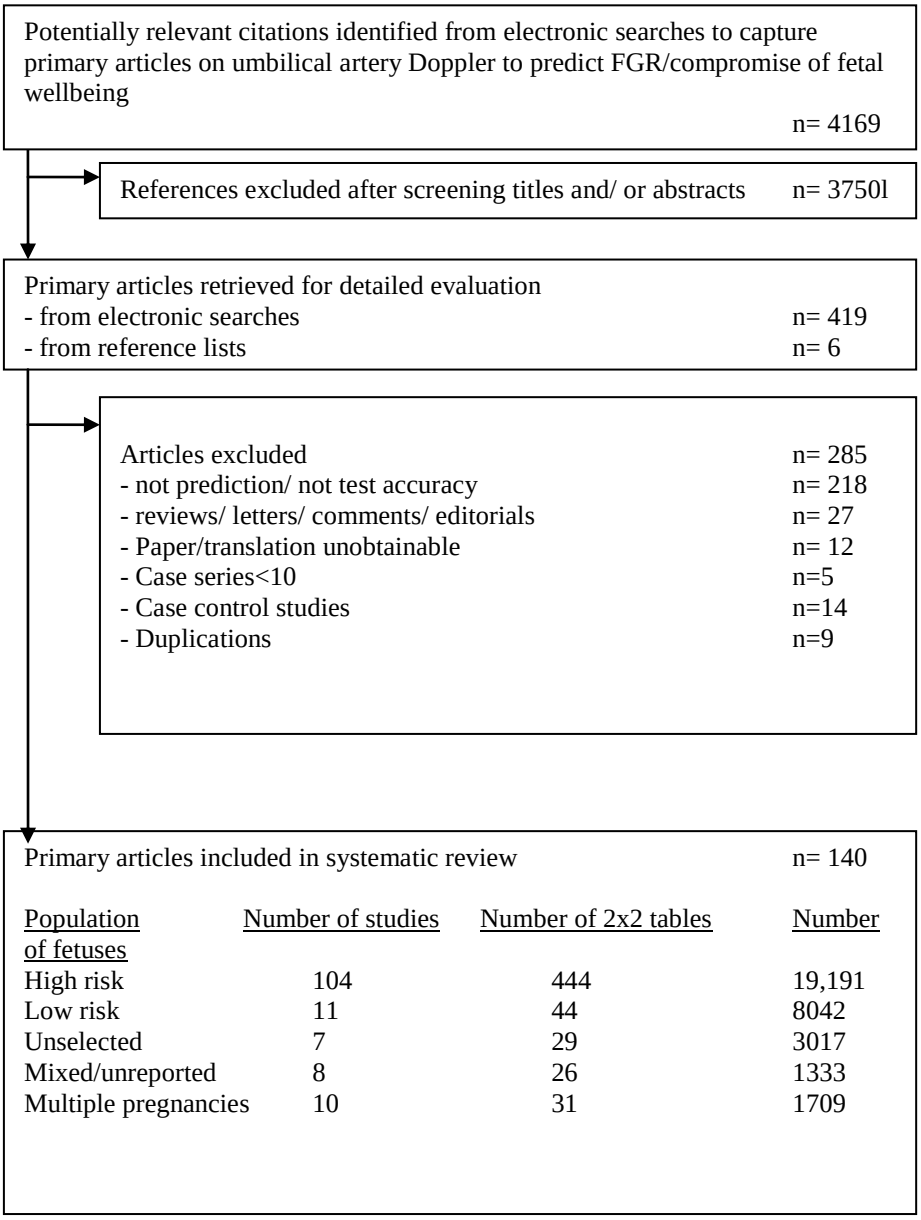
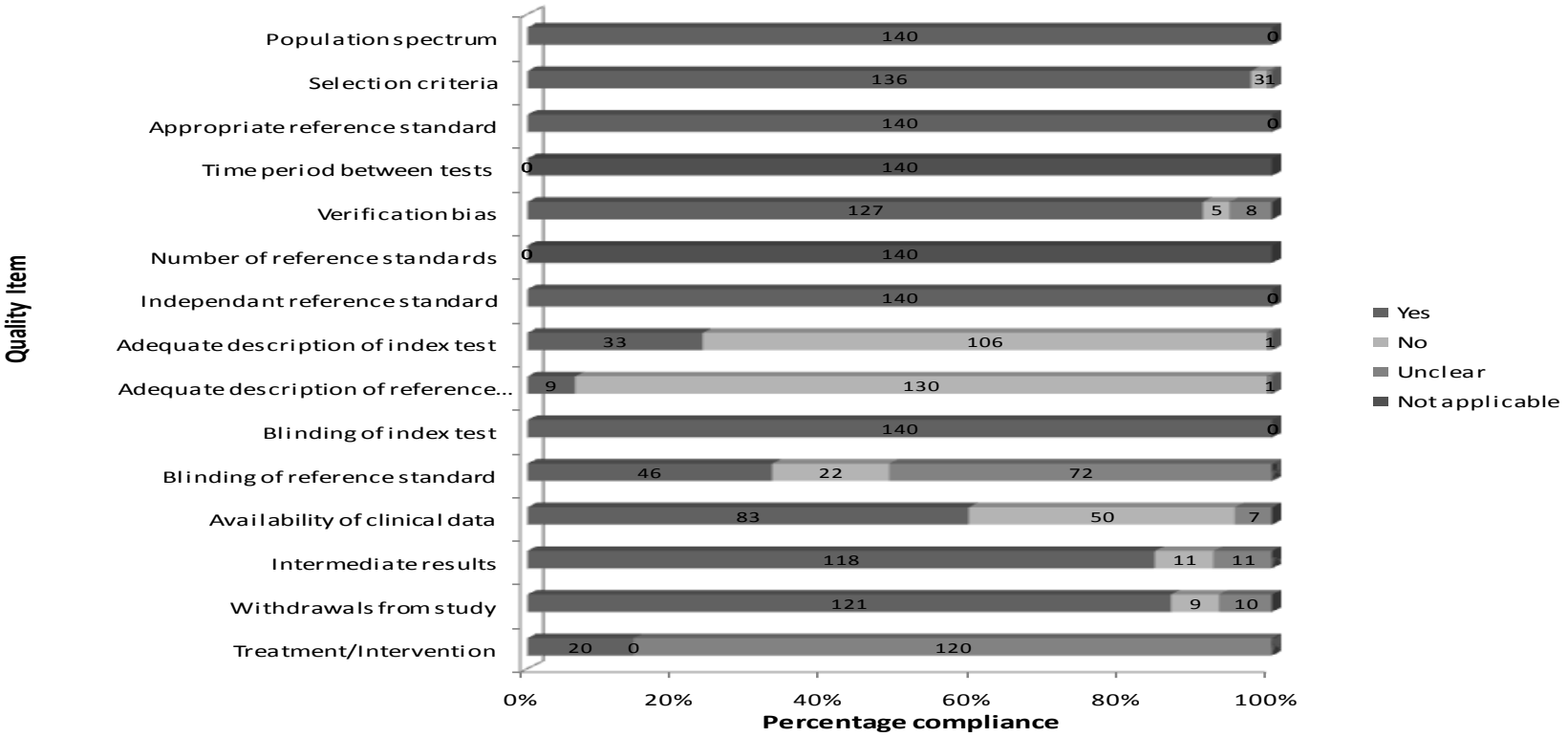


Figure 7.2: Bar chart illustrating the compliance with quality items for included studies in the systematic review of accuracy of umbilical artery Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing. (Numbers in bars represent actual number of studies compliant).



7.4.3 Data Analysis

Table 7.1 summarises the results from bivariate meta-analysis for small for gestational age fetuses and compromise of fetal/neonatal wellbeing respectively in the different populations. For small for gestational age there was improved prediction in the high risk population and for the severe measures of SGA (e.g. birth weight<3rd centile). For prediction of APO the best results were in the mixed population for APO overall and a low risk population for prediction of apgar at 5 minutes<7. However, these meta-analyses involved only three studies and there were wide confidence intervals suggesting imprecision of results. As there was significant heterogeneity the ROC curves are presented in figures 7.3 and 7.4 as the summary measures of accuracy.

Univariable bivariate meta-regression was performed for the high risk population for both prediction of SGA fetuses and for compromise of fetal/neonatal wellbeing using a composite outcome measure, APO, as there were sufficient studies. Meta-regression was also performed looking at the effect of study quality on accuracy. The covariates used were the same as for the planned sub-group analysis and the QUADAS items respectively. No variable was significantly associated with accuracy however, there was still significant statistical heterogeneity. Thus subgroup analysis according to reference standard for the high risk population was performed and this accounted for a significant amount of the heterogeneity. Figure 7.5 summarises the results of this analysis where there was a significant/clinically relevant result.

Table 7.1: Subgroup analysis according to reference standard for umbilical artery Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing.

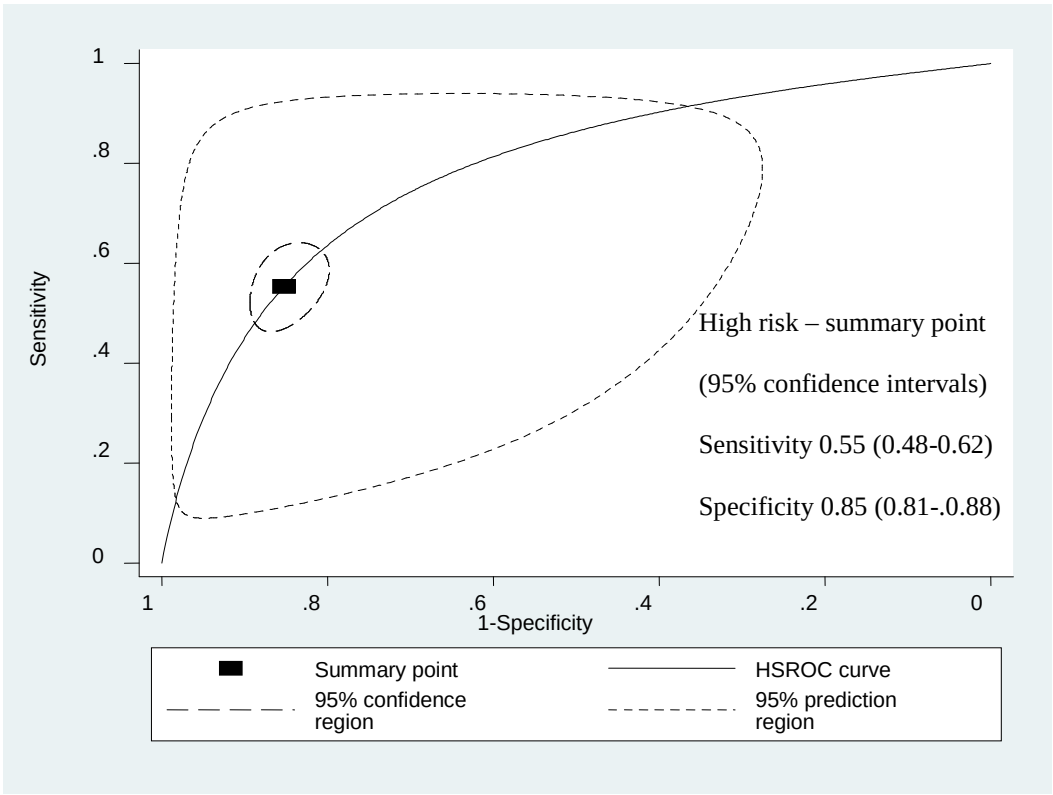
	Population definition							
	Low risk		High risk		Mixed risk		Unselected	
	Likelihood ratio (LR) (95% Confidence interval)							
	Positive LR	Negative LR	Positive LR	Negative LR	Positive LR	Negative LR	Positive LR	Negative LR
Reference standard								
Subgroup								
SGA all	**2.39 (1.74-3.28)**	**0.79 (0.69-0.90)**	**3.76 (2.96-4.76)**	**0.52 (0.45-0.61)**	**3.78 (1.96-7.27)[a]**	**0.73 (0.43-1.23)[a]**	**3.58 (2.49-5.14)**	**0.77 (0.66-0.91)**
Birth weight<10th centile	2.12 (1.65-2.73)	0.87 (0.80-0.95)	4.13 (3.08-5.56)	0.52 (0.44-0.62)	3.78 (1.96-7.27)[a]	0.73 (0.43-1.23)[a]	3.58 (2.49-5.14)	0.77 (0.66-0.91)
Birth weight<5th centile	-	-	3.06 (1.87-5.01)	0.54 (0.39-0.76)	-	-	-	-
Birth weight<3rd centile	-	-	4.91 (3.41-7.07)[a]	0.58 (0.49-0.69)[a]	-	-	-	-
Birth weight<2 sd mean	-	-	4.37 (3.16-6.05)[a]	0.43 (0.28-0.65)[a]	-	-	-	-
Wellbeing all	**3.11 (0.48-20.0)**	**0.81 (0.45-1.46)**	**3.41 (2.68-4.34)**	**0.55 (0.48-0.62)**	**5.98 (1.73-20.61)[a]**	**0.35 (0.02-7.75)[a]**	**3.93 (2.33-6.61)[a]**	**0.65 (0.44-0.96)[a]**
Apgar 1 minute<7	-	-	2.42 (1.68-3.50)	0.62 (0.47-0.81)	-	-	2.67 (1.17-6.06)[a]	0.83 (0.51-1.37)[a]
Apgar at 5 minutes<7	9.97 (3.24-30.69)[a]	0.62 (0.24-1.63)[a]	2.34 (1.83-3.01)	0.53 (0.38-0.74)	-	-	3.93 (2.33-6.61)[a]	0.65 (0.44-0.96)[a]
Admission to neonatal intensive care	-	-	3.35 (2.58-4.36)	0.46 (0.39-0.55)	-	-	-	-
Perinatal mortality	-	-	2.50 (1.88-3.31)	0.26 (0.10-0.67)	-	-	-	-
Acidosis	-	-	2.75 (1.48-5.11)	0.58 (0.36-0.94)	-	-	-	-
Necrotising enterocolitis	-	-	1.54 (1.03-2.29)	0.62 (0.35-1.09)	-	-	-	-
Intra-uterine death	-	-	4.37 (0.88-21.8)	0.25 (0.07-0.91)	-	-	-	-
Neonatal death	-	-	1.88 (1.34-2.41)	0.61 (0.43-0.87)	-	-	-	-
Ventilation	-	-	2.38 (1.41-4.00)	0.03 (0.00-13.77)	-	-	-	-
Respiratory distress	-	-	2.50 (1.63-3.81)	0.54 (0.38-0.75)	-	-	-	-
Neonatal morbidity	-	-	3.05 (1.68-5.52)	0.53 (0.36-0.80)	-	-	-	-
Intra-ventricular/cranial haemorrhage	-	-	1.65 (1.19-2.30)	0.60 (0.26-1.38)	-	-	-	-

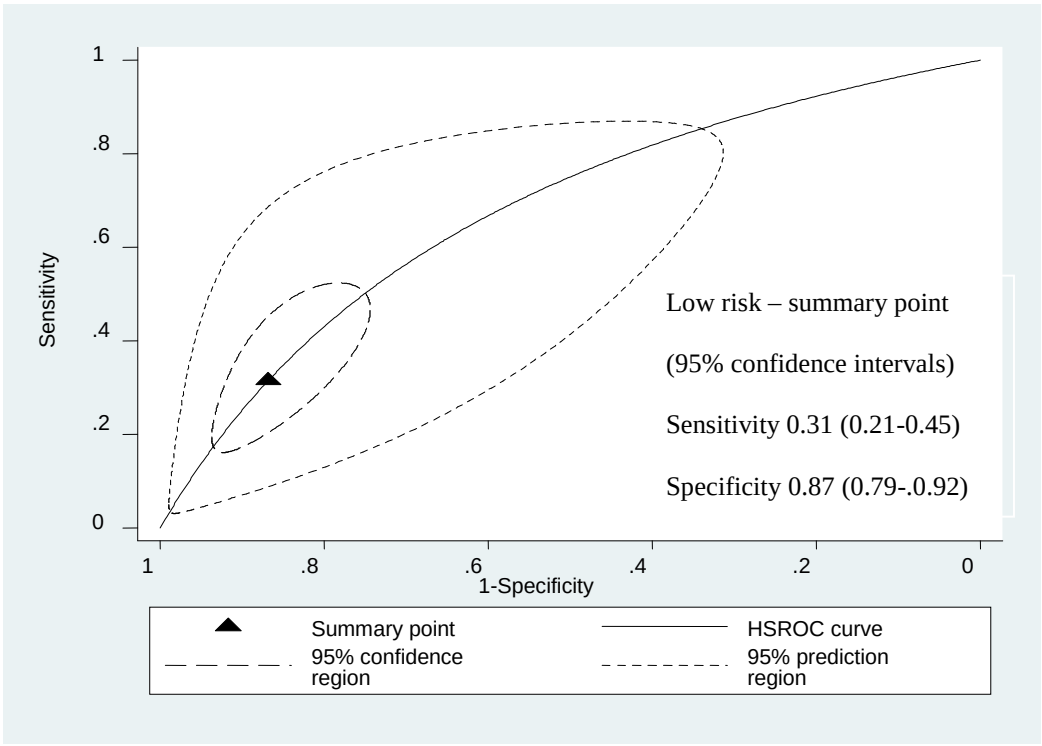
SGA small for gestational age

Sd standard deviations

[a] *univariate meta-analysis*

Figure 7.3: Bivariate analysis of the accuracy of umbilical artery Doppler to predict small for gestational age fetuses in different populations according to risk.





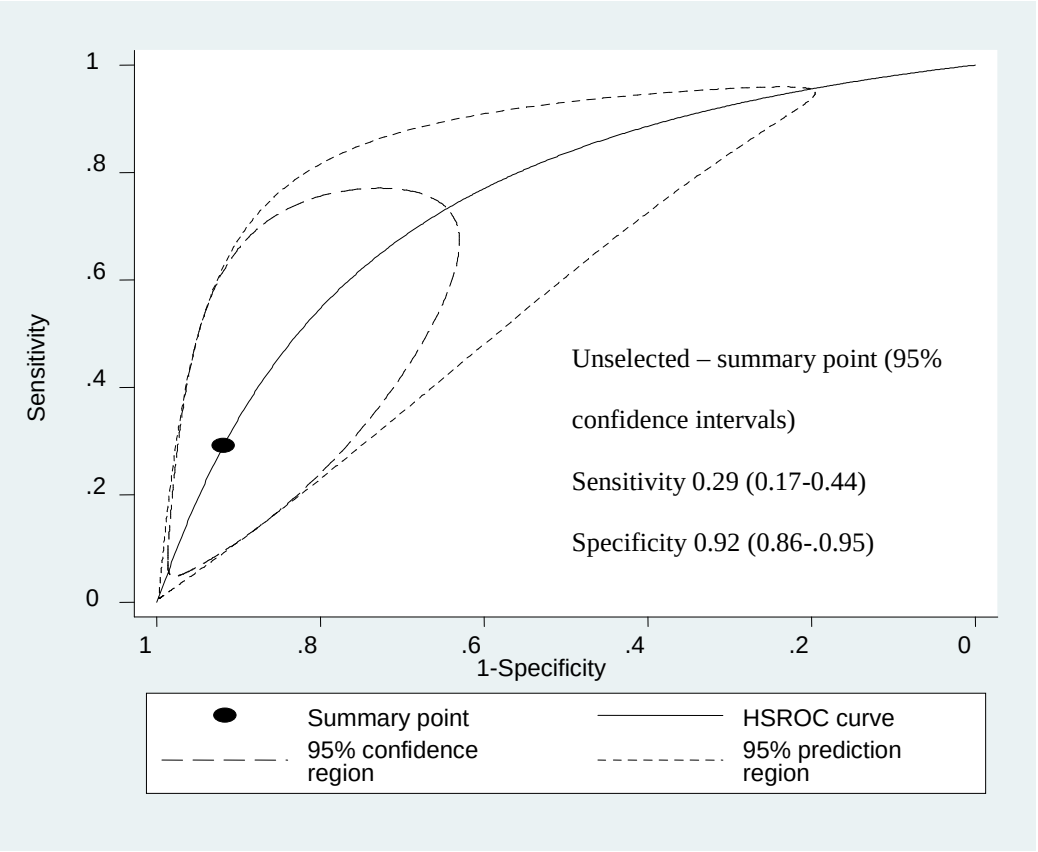
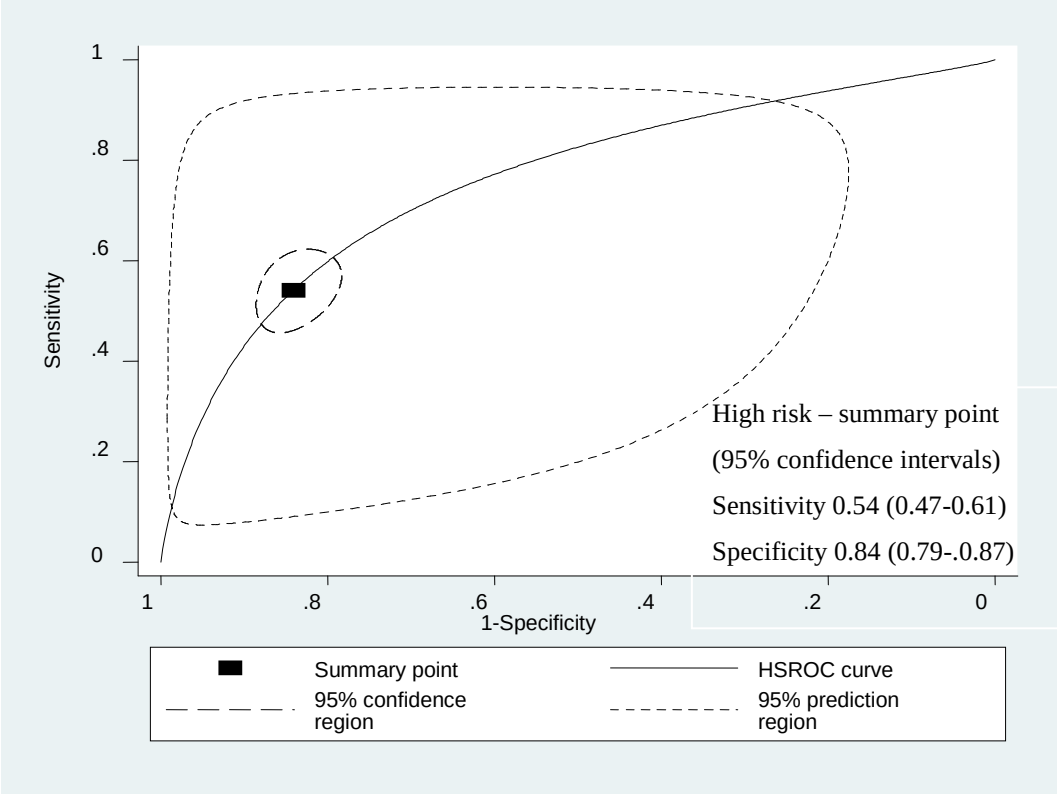

Sensitivity
1
.8
.6
.4
.2
0
1
.8
.6
.4
.2
0
1-Specificity
Unselected – summary point (95% confidence intervals)
Sensitivity 0.29 (0.17-0.44)
Specificity 0.92 (0.86-.0.95)
Summary point
HSROC curve
95% confidence region
95% prediction region

Figure 7.4: Bivariate analysis of the accuracy of umbilical artery Doppler to predict adverse perinatal outcome in different populations according to risk.



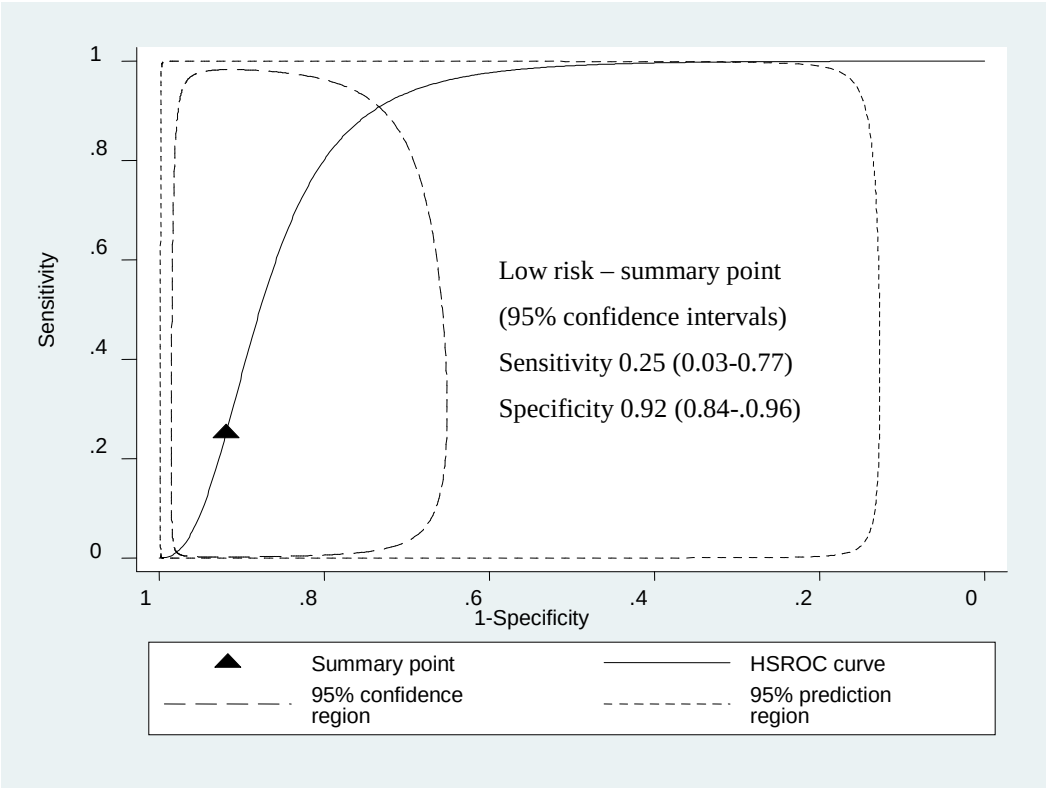
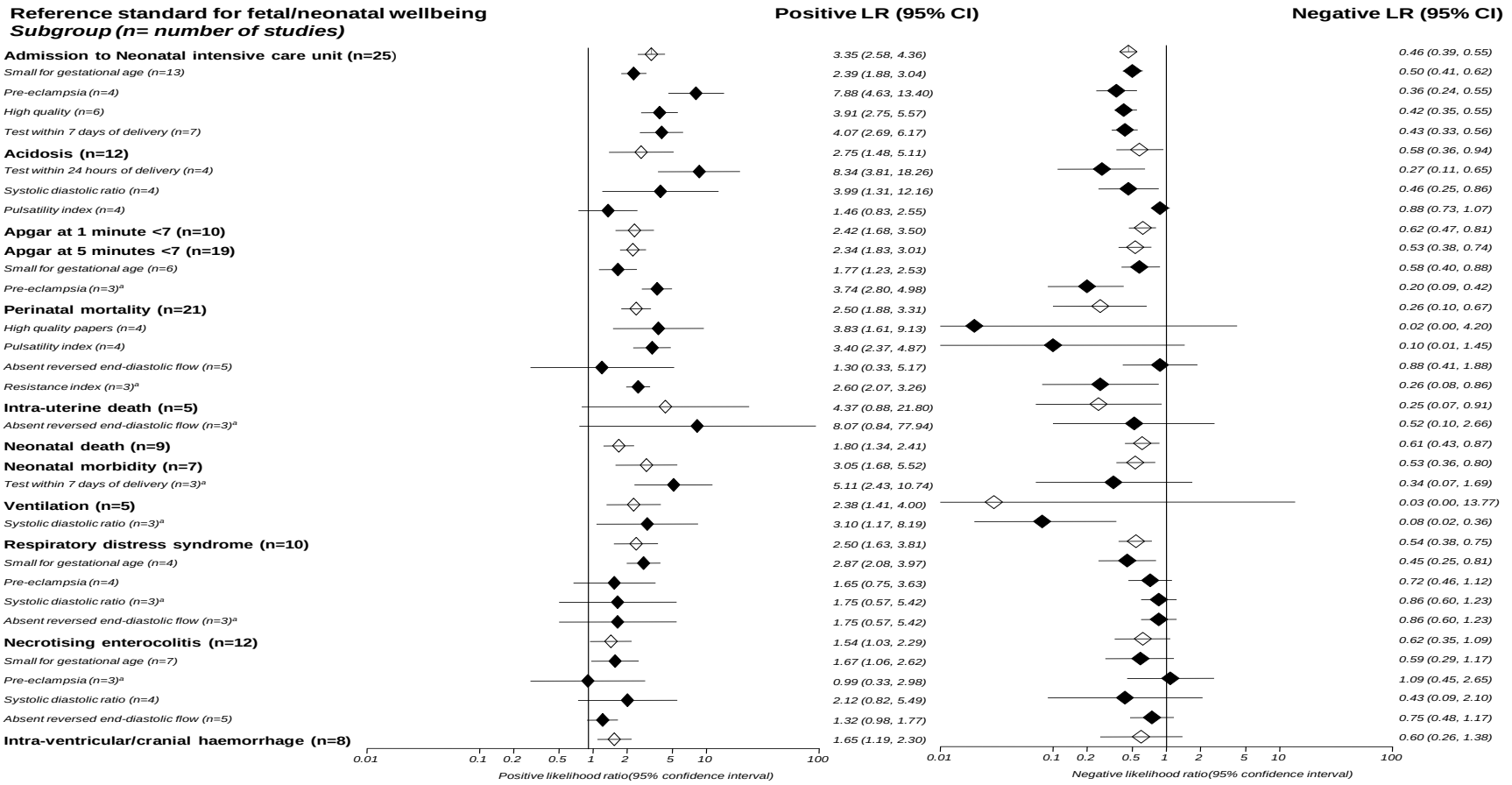


Figure 7.5: Forest plot of likelihood ratios (LR) with 95% confidence intervals (CI) for umbilical artery Doppler to predict compromise of fetal/neonatal wellbeing in a high risk population, bivariate meta-analysis according to reference standard with subgroup analysis ([a] meta-analysis using univariate method).



In a high risk population subgroup analysis showed clinically relevant results for prediction of admission to neonatal intensive care unit (NICU) and acidosis and prediction of intra-uterine death (figure 7.5). For admission to NICU prediction was better in pregnancies affected by PE rather than those with suspected SGA, in high quality papers and when the test was performed within 7 days of delivery. For acidosis prediction was greatly improved when the test was performed within 24 hours of delivery and a systolic diastolic (SD) ratio was more accurate than the pulsatility index (PI). For an Apgar score less than 7 at 5 minutes prediction was improved in PE pregnancies versus SGA. For perinatal mortality there was improved prediction with high quality papers and the use of the PI. For respiratory distress syndrome and necrotizing enterocolitis there was improved prediction in SGA pregnancies versus PE. For prediction of intra-uterine death absent reversed end-diastolic flow showed improved prediction.

In multiple pregnancies the results for SGA were LR+ve 3.37 (2.18-5.22), LR-ve 0.65 (0.59-0.72) and for adverse perinatal outcome LR+ve 8.08 (2.63-24.77), LR-ve 0.58 (0.49-0.68).

Funnel plots and the regression test for asymmetry for SGA as an outcome were symmetrical/not significant in all populations (p=0.3). When assessing papers reporting on wellbeing as an outcome there was asymmetry across all populations however, when the tests were restricted to sub-groups according to outcome the plots were symmetrical suggesting the asymmetry was due to use of a combined reference standard introducing heterogeneity rather than publication bias.

7.5 Discussion

This review evaluated the accuracy of umbilical artery Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing. In a high risk population the test showed moderately useful results for prediction of severe forms of SGA and intra-uterine death, and in certain sub-groups clinically useful results for prediction of admission to NICU and acidosis. The results showed low predictive accuracy overall for a low risk/unselected population. In multiple pregnancies the results suggest the test may be clinically useful to predict adverse perinatal outcome.

This review complies with existing guidelines for the reporting of systematic reviews[43] and also guidelines specific to the reporting of systematic reviews of observational studies[42]. Extensive literature searches were performed without language restrictions, used validated methods for quality assessment, investigated for potential sources of heterogeneity and employed new advanced statistical techniques all planned a priori. In 1994 Neilson et al were the first to report a systematic review in this area in a high risk population as a Cochrane review, reporting a 49% reduction in perinatal mortality[90], this was later updated in 1995 by Alfirevic et al and the reduction found to be 38%[85]. Later, with the exclusion of a small study, the Cochrane group reported that this reduction had fallen to 29%, OR 0.71 (95% confidence interval 0.50-1.01)[91]. The conclusion of the final update was that umbilical artery Doppler in a high risk population appeared to improve a number of obstetric care outcomes and appeared promising in reducing the number of perinatal deaths. The authors made a case for the need for a larger trial[91]. In unselected and low risk populations the previously published evidence has not supported the routine use of umbilical artery Doppler[86;87]. There has been one published

test accuracy study, looking at the use of intrapartum umbilical artery Doppler and concluded that it was a poor predictor of adverse outcome[27]. As discussed previously, RCTs assess the effectiveness of both the test and any subsequent intervention on outcome. This review shows that umbilical artery Doppler is useful clinically in a high risk population as a predictive test for neonatal/fetal compromise.

The assessment of study quality and reporting in diagnostic reviews has advanced over the last ten years with the development and validation of the QUADAS and STARD checklists. It has been previously reported that poor study design and conduct can affect the estimates of diagnostic accuracy[51;75]. However, it is not entirely clear how individual aspects of quality may effect this and to what magnitude particularly in the area of Obstetrics. Application of quality scores has been shown to be of little value on diagnostic reviews[55]. The assessment of study quality in this review was hindered by lack of clear reporting, which is a common problem in diagnostic reviews. The areas in which reporting was uniformly poor were in the details provided regarding performance of the index test and reference standard, blinding of the assessors of the reference standard and whether interventions were used e.g. early delivery. Meta-regression showed no effect of study quality on results in a high risk population however, in light of the poor reporting the true effect of study quality cannot truly be assessed. As interventions were poorly reported there is the possibility of a treatment paradox (where the application of a treatment affects the outcome/reference standard and thus the test accuracy results) being introduced however, as there as so few interventions that may be used for these conditions and that reporting was globally poor the likelihood is that any paradox will be uniform.

This review was also limited by the reference standards used as discussed in section 5.5 and chapter 10.

In clinical practice tests are not applied in isolation, the clinician makes an assessment of the likelihood of an outcome/disease based on history, examination, test results and their own experience/beliefs. A test alone cannot alter the clinical course of a pregnancy but provides information for the clinician to make a decision on clinical management based on the analysis of this information. To truly assess the value of a test a clinician thus needs to know where it fits into a diagnostic pathway and what management decisions may be made based on its results. This can be modelled in the form of a decision tree and combined with economic analysis to give a true assessment of the tests value in clinical practice. Future research thus needs to address the limitations in primary research as identified in this review and the need to consider tests in combination with other investigations and as part of the management pathway with interventions.

7.6 Conclusion

Umbilical artery Doppler is a moderately useful test in a high risk population to predict the fetus/neonate at risk of compromise. Further research should concentrate on its use within a series of tests e.g. with biophysical profile and other fetal Doppler indices.

CHAPTER 8: SYSTEMATIC REVIEW OF MIDDLE CEREBRAL ARTERY DOPPLER TO PREDICT SMALL FOR GESTATIONAL AGE FETUSES AND COMPROMISE OF FETAL/NEONATAL WELLBEING

8.1 Abstract

8.1.1 Background

The accuracy of fetal middle cerebral artery (MCA) Doppler for prediction of the fetus at risk of compromise of wellbeing is not known. The purpose of this review was to determine the accuracy of MCA Doppler to predict SGA fetuses and compromise of fetal/neonatal wellbeing.

8.1.2 Methods

The following electronic databases were searched: Medline, Embase, Cochrane library, Medion (inception to May 2009), hand searching of journal and reference lists, contact with experts. Two reviewers independently selected articles in which the results of middle cerebral artery Doppler were associated with the occurrence of compromise of fetal/neonatal wellbeing. There were no language restrictions applied. Data were extracted on study characteristics, quality and results to construct 2x2 tables. Likelihood ratios for positive and negative test results, sensitivity, specificity and their 95% confidence intervals were generated for the different indices and thresholds.

8.1.3 Results

Thirty one studies, testing 3337 fetuses met the selection criteria. Meta-analysis showed low predictive accuracy. The best result was for the prediction of need for neonatal intensive care with a positive likelihood ratio 4.00 (2.13, 7.50) and negative likelihood ratio 0.61 (0.50, 0.75). For prediction of adverse perinatal outcome and perinatal mortality the results were positive likelihood ratios 2.79 (1.61, 3.07) and 1.36 (1.10, 1.67) and negative likelihood ratios 0.56 (0.43, 0.72) 0.51 (0.29, 0.89) respectively.

8.1.4 Conclusion

Abnormal middle cerebral artery Doppler showed limited predictive accuracy for compromise of fetal/neonatal wellbeing. High quality primary research or individual patient data meta-analysis looking at this test in combination with other tests is required.

8.1.5 Publications arising from this work

Morris RK, Say R, Robson SC, Kleijnen J, Khan KS. Systematic review of middle cerebral artery Doppler to predict fetal growth restriction/compromise of fetal wellbeing. *Archives of Disease in Childhood: Fetal and Neonatal Edition* 2008; 93 (supplement 1): Fa 31-36.

8.2 Introduction

Animal studies have shown that in response to hypoxia, there is a redistribution of cardiac output with preferential flow to the heart, brain and adrenal glands[92;93]. Doppler studies of the human fetal circulation have shown similar results[94-97]. Cerebral vasodilatation is a manifestation of the increase in cerebral diastolic flow, a sign of the 'brain-sparing effect' of chronic hypoxia, and results in decreases in Doppler indices of the middle cerebral artery (MCA) such as the pulsatility index (PI)[94]. Previous studies looking at the value of MCA Doppler in the detection of the at risk fetus have conflicting conclusions, some report poor predictive value[98;99] while others report that MCA Doppler may be a useful tool[95;100].

The purpose of this review was to investigate the accuracy of MCA Doppler used in predicting compromise of fetal/neonatal wellbeing.

8.3 Methods

The methods used are outlined in chapter 4 with those specific to this review detailed below.

8.3.1 Data Sources

Electronic searches were performed targeting citations on the prediction of SGA/FGR and fetal/neonatal compromise. Medline, Embase, the Cochrane Library (2009;2) and Medion database were searched from inception until May 2009. The search strategies are detailed in appendix 29.

8.3.2 Study Selection

Criteria for inclusion were studies that reported on singleton and multiple pregnancies at any level of risk in any healthcare setting undergoing MCA Doppler at any gestation. Test accuracy studies allowing generation of 2x2 tables were included. The reference standards used were those reported by the authors. Reference standards for wellbeing were any outcome measure performed after birth relating to neonatal wellbeing that was reported by the study authors. The outcome measure APO was used as detailed in sections 4.2.3.

8.3.3 Data Extraction and Study Quality Assessment

The data extraction form for this review can be found in appendix 30. Quality assessment was performed as described in section 4.4. See appendix 26 for further explanation of the quality assessment.

8.3.4 Data Synthesis and Analysis

This was performed as detailed in section 7.3.4.

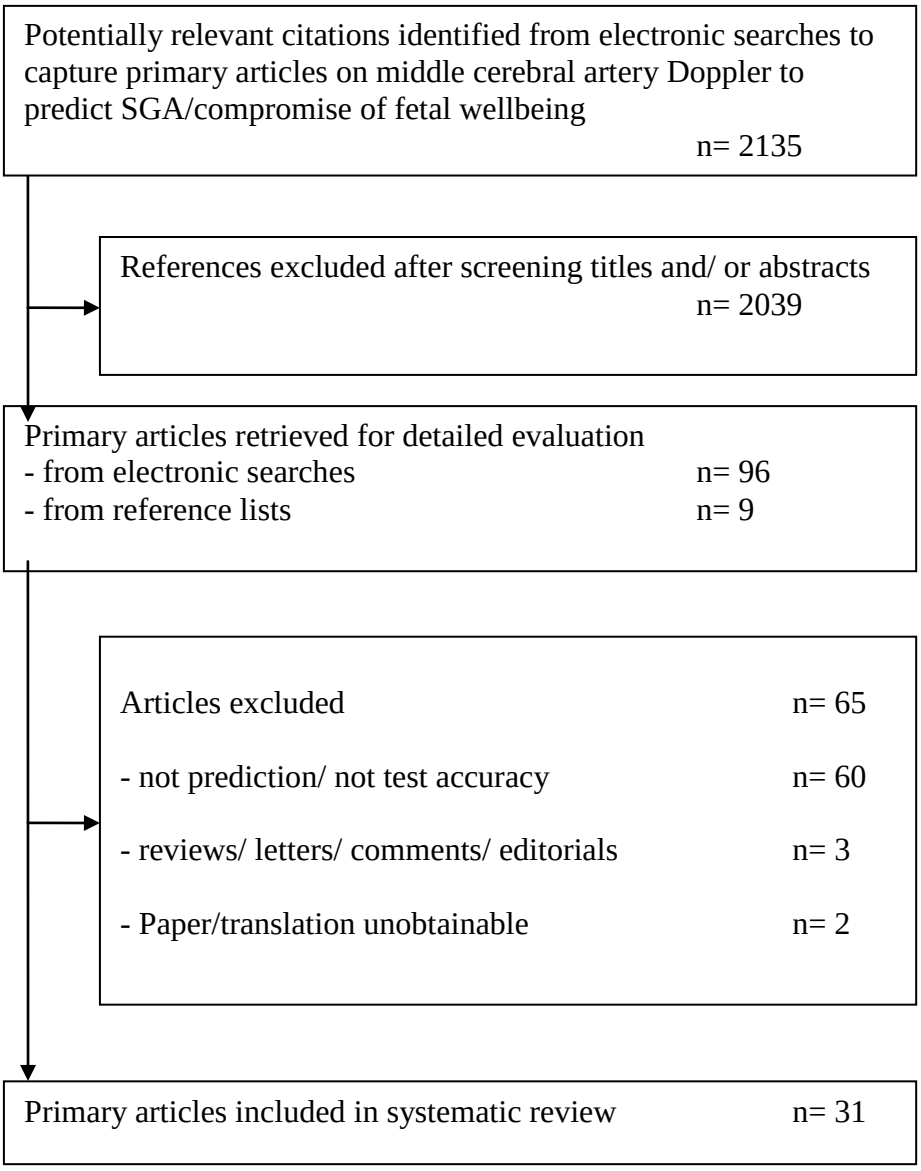
8.4 Results

8.4.1 Literature Identification and Study Characteristics

Figure 8.1 summarises the process of literature identification and selection. Tables detailing the citations of the included studies and the individual study characteristics are available in appendices 31 and 32 respectively. There were 31 studies included overall, testing 3337 fetuses and producing 85 2x2 tables. The majority of studies were

performed in the third trimester (28) with 10 reporting on second trimester testing and five reporting data on post-dates pregnancies. In 87.1% (27/31) of included studies the population under investigation was classified by the author as high-risk, one low risk, one unselected, one mixed and one paper had no classification. In the high risk populations 14 included patients with suspected FGR, four with hypertensive diseases and five included patients with both risk factors. 18 papers reported on singleton pregnancies only and one paper on multiple pregnancies only, in the remaining 12 papers the authors did not state whether multiple pregnancies were excluded. Only 55% (17/31) papers excluded fetuses with structural and chromosomal anomalies from the results. Twenty-seven of the included studies reported measures of fetal/neonatal wellbeing. There were 13 studies for SGA (11 reported birth weight<10^{th} centile, two birth weight <5^{th} centile and one birth weight <3^{rd} centile). None of the included studies used anthropometric measurements other than birth weight.

Figure 8.1: Process from initial search to final inclusion for middle cerebral artery Doppler to predict small for gestational age/compromise of fetal wellbeing (up to May 2009). (SGA small for gestational age).



8.4.2 Study Quality

The quality assessment of included studies revealed deficiencies (figure 8.2). Only 12 studies used prospective data collection, 21 were of a cohort design and only nine studies used both. Only seven studies contained an adequate description of the performance of the index test and only four reported clearly on the performance of the reference standard. Blinding of the reference test was also poorly reported (5/31). Availability of clinical data were reported in 16 studies and 20 reported on intermediate results and withdrawals. Only one study reported adequately on all these quality items. One reported adequately on the use of any treatment in between the performance of the MCA Doppler and delivery, stating the use of betamethasone.

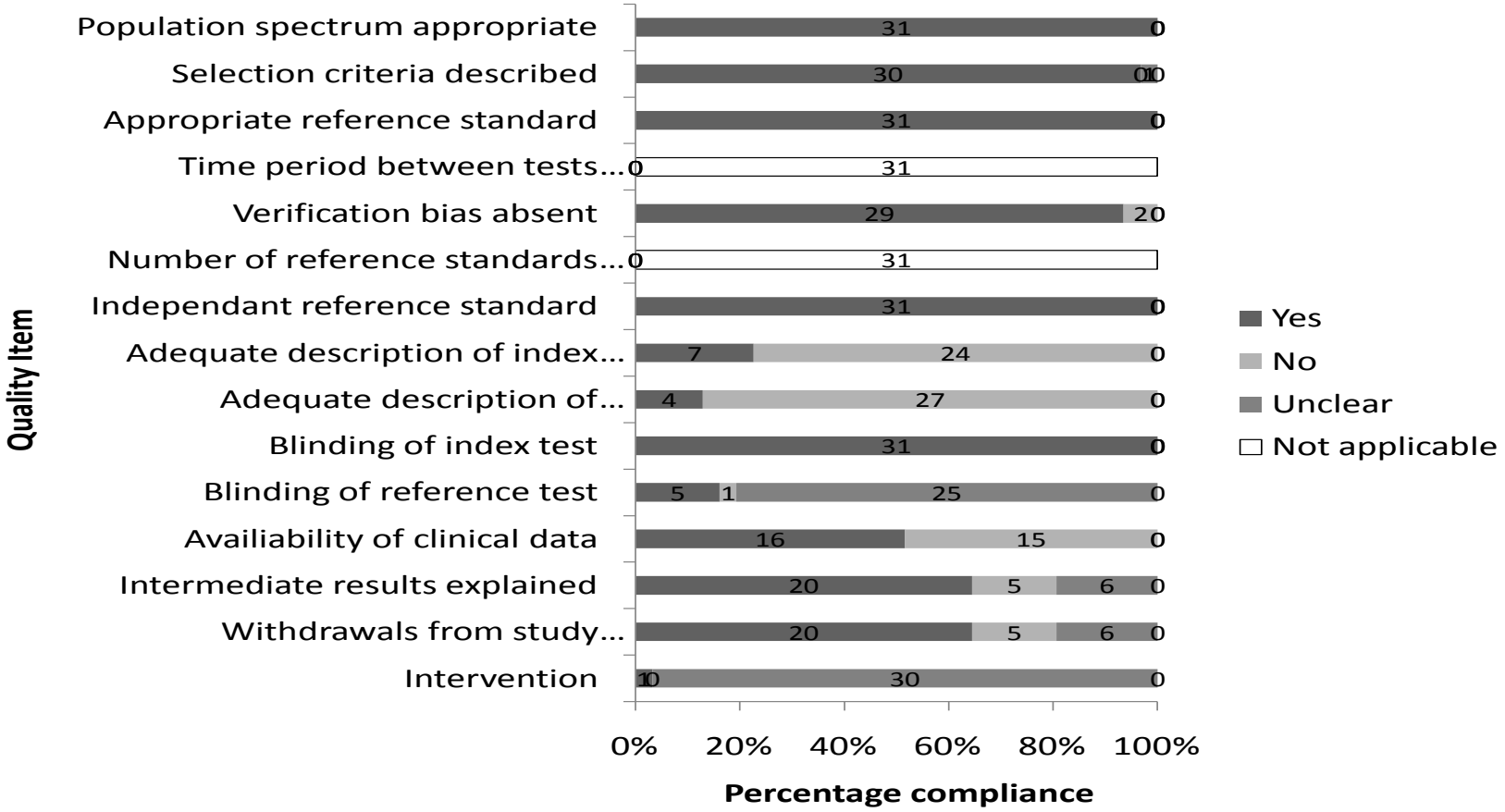
8.4.3 Data Analysis

Fetal/Neonatal Compromise

Statistical analysis could be performed for the following outcome measures: adverse perinatal outcome (figure 8.3), Apgar at 1 minute and 5 minutes <7 (appendix 33), cord blood gas analysis/requirement for neonatal resuscitation (appendix 34), admission to neonatal intensive care unit (NICU)/neonatal complications (appendix 35) and perinatal morbidity and mortality (appendix 36).

The best result obtained was for prediction of admission to NICU LR+ 4.40 (2.13, 7.50) LR-ve 0.61 (0.50, 0.75). Disappointingly meta-analysis using a composite outcome measure, adverse perinatal outcome, did not show good predictive accuracy LR+ve 2.79 (1.61, 3,07) LR –ve 0.56 (0.43, 0.72) and there was significant heterogeneity X^2=37.96 (p=0.00).

Figure 8.2: Bar chart illustrating the compliance with quality items for included studies in the systematic review of accuracy of middle cerebral artery Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing. (Numbers in bars represent actual number of studies compliant).



When looking at one of the most important outcome measures, perinatal mortality, the results were again disappointing, LR+ve 1.36 (1.10, 1.67) and LR-ve 0.51 (0.29, 0.89).

Sub-group analysis for these measures did account for some of the heterogeneity and allow an improvement in prediction. For neonatal cord blood acidosis, restricting the threshold to pH<7.20 and a high risk population [LR+ve 2.29 (0.74, 7.11) LR-ve 0.66 (0.40, 1.10)]; and for adverse perinatal outcome use in an unselected population [LR+ve 5.53 (2.88, 10.64) LR-ve 0.68 (0.54, 0.65)] and use of pulsatility index<5th centile [LR+ve 3.66 (1.04, 12.96) and LR-ve 0.48 (0.34, 0.70)].

Where ROC curves were plotted the area under the curve (AUC) were all between 0.70-0.80, and assessed as a moderate test. Summary receiver operating characteristic curves are shown in figure 8.5 for APO and SGA (other outcomes are shown in appendix 37).

Small for gestational age fetuses (SGA)

The results for SGA fetuses are summarized in figure 8.4. For prediction of birth weight <10th centile, there was significant heterogeneity in results. Sub-group analysis was performed based on population risk, Doppler index parameter used, singletons only and MCA within 2 weeks of delivery. The most predictive test was a systolic/diastolic ratio<10th centile in any risk population LR+ve 9.32 (3.91, 22.19) LR-ve 0.53 (0.43, 0.65), X^2 = 1.91 (p=0.17). There did not appear to be an improvement in prediction with the more severe forms of SGA i.e. birth weight <5th or <3rd centile.

Sub-group and sensitivity analysis.

Sensitivity analysis including only those studies which excluded chromosomal and structural anomalies showed no significant difference. When assessing study quality, sub-group analysis based on study quality could only be performed for the meta-analyses with a large number of included studies; sensitivity analysis showed no difference when extremely low quality studies were excluded.

Publication bias

Funnel plots (not shown) and the regression test for asymmetry showed no significant publication bias (p=0.1).

8.5 Discussion

This is the only published systematic review and meta-analysis of the value of MCA Doppler to predict perinatal wellbeing. Disappointingly MCA Doppler was found to have a low predictive accuracy overall. For fetal/neonatal compromise the best predictor was any Doppler parameter in any risk population to predict need for neonatal intensive care. For SGA fetuses the best predictor was SD ratio<10^{th} centile in any risk population.

Figure 8.3: Forest plot of positive and negative likelihood ratios for middle cerebral artery Doppler to predict compromise of fetal wellbeing (adverse perinatal outcome). Single studies are represented by a filled box, pooled results by an open diamond and subgroup analysis by a filled diamond.

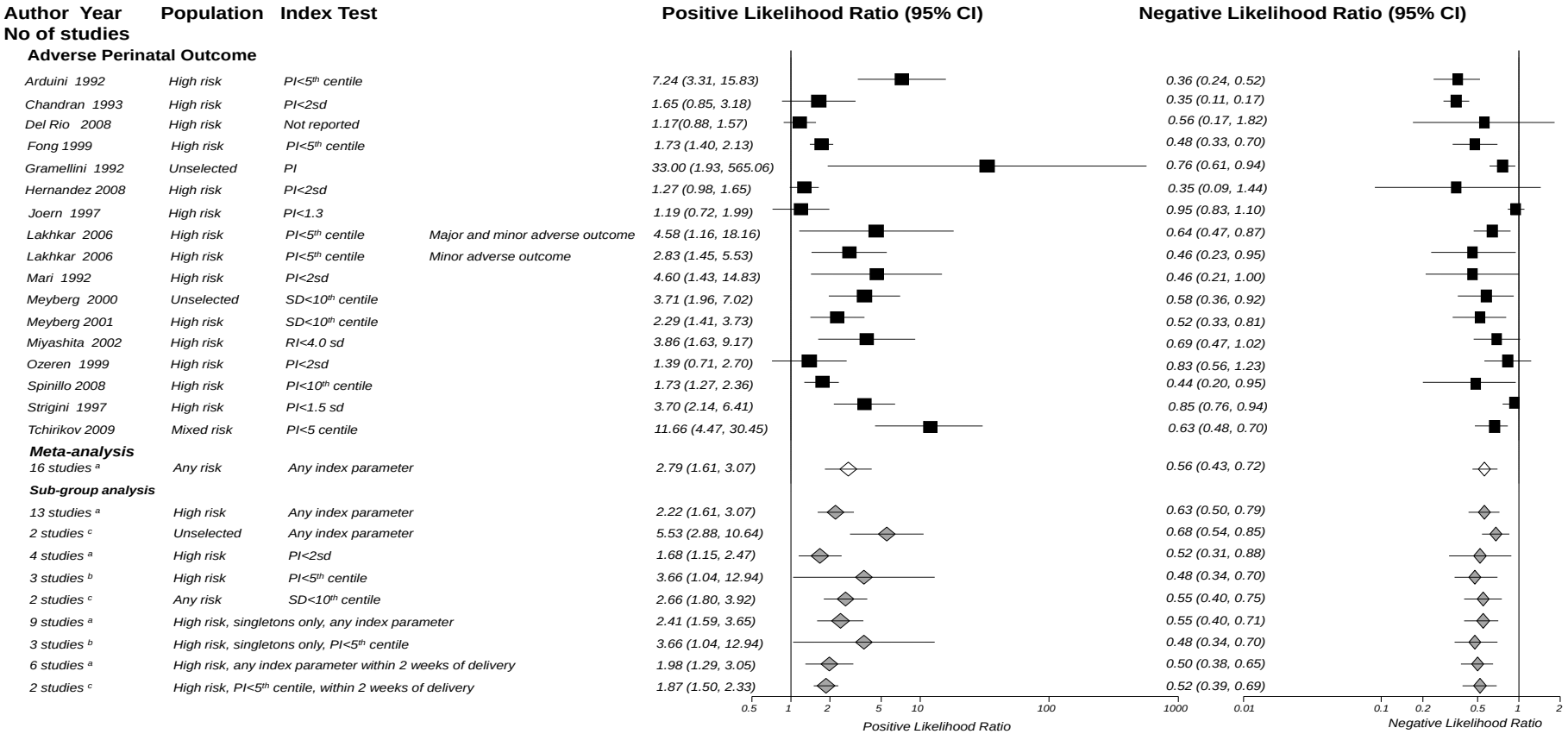


Figure 8.4: Forest plot of positive and negative likelihood ratios for middle cerebral artery Doppler to predict small for gestational age fetuses (Birth weight <10th/5th/3rd centile). Single studies are represented by a filled box, pooled results by an open diamond and subgroup analysis by a filled diamond.

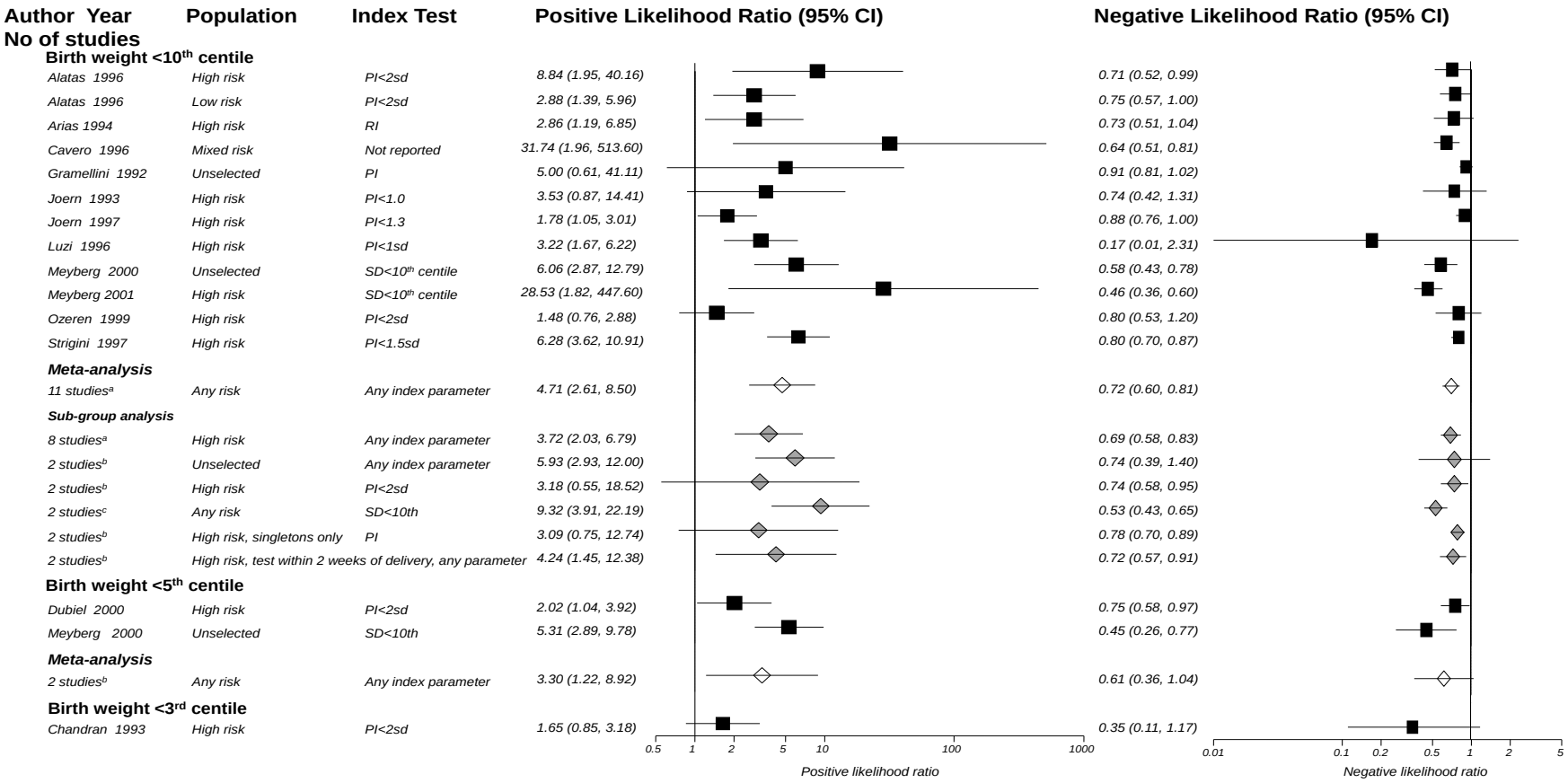
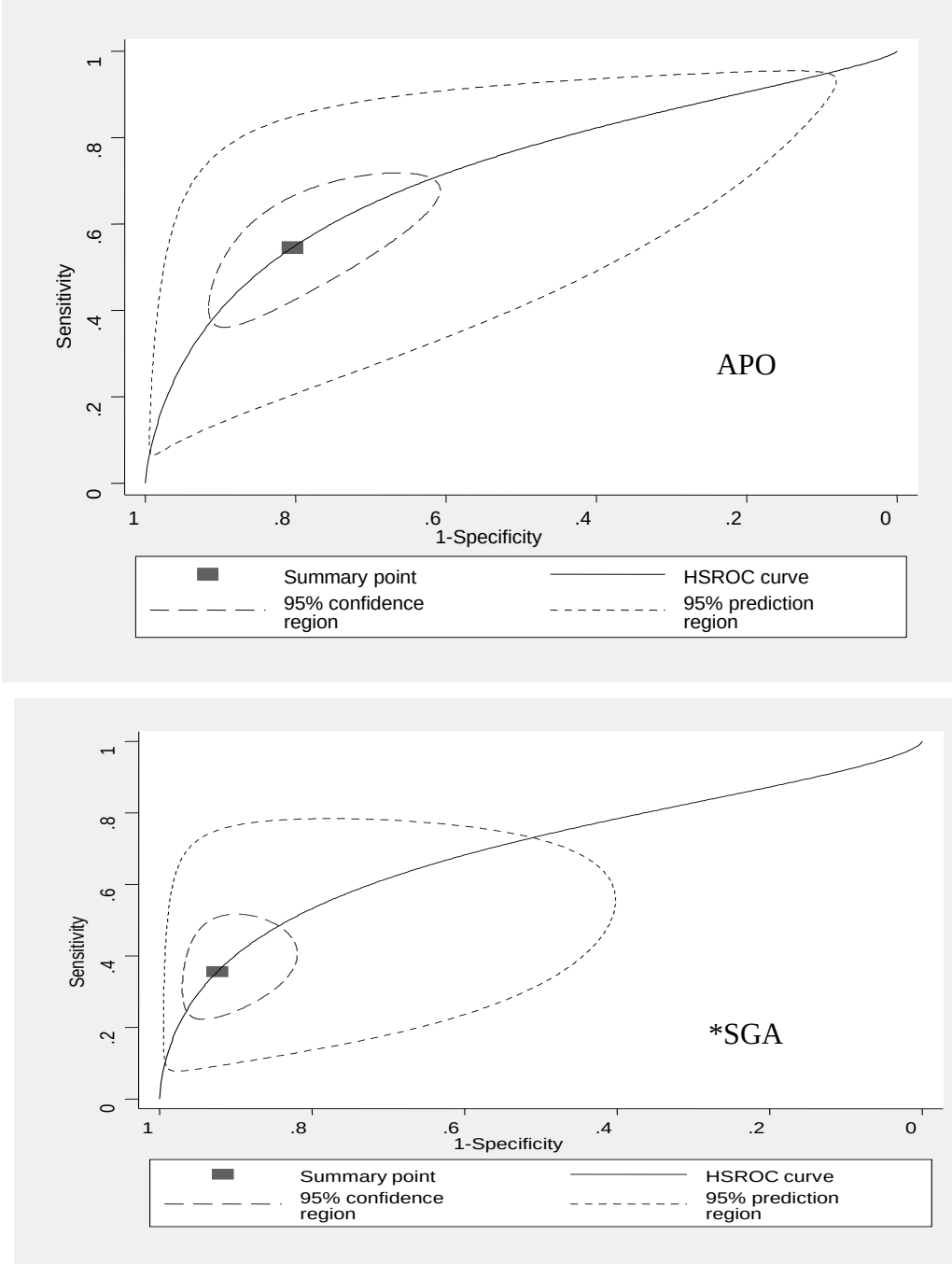


Figure 8.5: Summary receiver operating characteristic curves for middle cerebral artery Doppler to predict small for gestational age fetuses (SGA) and adverse perinatal outcome (APO) produced using the bivariate method. (*SGA birth weight <10th centile)



The use of pooled likelihood ratios as summary measures has been discussed in section 5.5. When analyzing the results, pooled sensitivity and specificity were used in a sensitivity analysis and no difference in the interpretation of the results was found. The bivariate technique was used in the meta-analysis to overcome the concerns about pooling related statistics.

In this review, sub-group analysis using the aspects of study quality that are best reported (study design, recruitment, description and blinding of index and reference standard) showed no significant difference in results when excluding the "low quality" studies. Areas of study design where reporting was uniformly poor were in the description of the index test and reference standard, blinding of the reference standard and use of any intervention between the index test and reference standard.

In this review results are analysed and reported according to absolute cut-offs for MCA Doppler indices. Sub-group analysis was performed where possible looking at individual indices as it is known that changes for instance in the MCA PSV and PI occur at different stages in the progression of fetal compromise[101]. Although some of the included papers investigated trends in MCA Doppler in an individual fetus the test accuracy data for this was not reported. MCA Doppler is also used with umbilical artery Doppler as the cerebroplacental ratio, although outside the scope of this review there were 11 included papers that reported on cerebroplacental ratio, this test showed greatly improved accuracy for prediction of adverse perinatal outcome LR+ve 4.42 (1.88, 10.37) and LR-ve 0.36 (0.22, 0.60).

In the included studies it was disappointing that only population based birth weight standards were used as these do not distinguish between the small healthy infant and the compromised infant. Thus although this review set out to evaluate the accuracy of MCA Doppler to predict FGR, it is accepted that the results reported in this review can only be considered to be predictive of SGA fetuses rather than FGR. It is important that future research in this area uses customised growth charts that are adjusted for sex, gestation, parity, maternal weight and height and ethnicity and neonatal indices of malnutrition as the former have been shown to improve the detection of at risk of stillbirth[17] and the latter to identify the malnourished infant at risk of peripartum asphyxia[16].

As FGR and severe compromise of fetal wellbeing are diseases with relatively low prevalence a clinically useful test would need to have a high positive LR (>10) and low negative LR (<0.10)[59]. At present the results of this review show low predictive accuracy. Future research should concentrate on addressing the limitations as already identified within the primary literature in particular in the choice of reference standards for FGR and should utilize individual patient data meta-analysis. This research should also look at the use of antenatal tests in combination e.g. umbilical and MCA Doppler to improve predictive accuracy and thus clinical value. It must also be recognized that when considering implementing MCA Doppler as a clinical test, researchers will need to take into account the cost of implementing such a specialized test and the lack of effective interventions in this area.

8.6 Conclusion

Abnormal middle cerebral artery Doppler showed limited predictive accuracy for compromise of fetal/neonatal wellbeing. High quality primary research or individual patient data meta-analysis looking at this test in combination with other tests is required.

CHAPTER 9: SYSTEMATIC REVIEW OF DUCTUS VENOSUS DOPPLER TO PREDICT SMALL FOR GESTATIONAL AGE FETUSES AND COMPROMISE OF FETAL/NEONATAL WELLBEING

9.1 Abstract

9.1.1 Background

The accuracy of ductus venosus Doppler for prediction of the fetus at risk of compromise of wellbeing is not known. The purpose of this review is to determine the accuracy of ductus venosus Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing.

9.1.2 Methods

Electronic searches of the following databases were performed: Medline, Embase, Cochrane library, Medion (inception to May 2009), hand searching of journal and reference lists, contact with experts. Two reviewers independently selected articles in which the results of ductus venosus Doppler were associated with the occurrence of compromise of fetal/neonatal wellbeing. There were no language restrictions applied. Data were extracted on study characteristics, quality and results to construct 2x2 tables. Likelihood ratios for positive and negative test results, sensitivity, specificity and their 95% confidence intervals were generated for the different indices and thresholds.

9.1.3 Results

Twenty studies, testing 13,273 fetuses met the selection criteria. Meta-analysis showed moderate predictive accuracy. The best result was for the prediction of perinatal mortality, positive likelihood ratio 4.21 (1.98, 8.96) and negative likelihood ratio 0.43 (0.30, 0.61). For prediction of adverse perinatal outcome the results were positive likelihood ratio 3.15 (2.19, 4.54) and negative likelihood ratio 0.49 (0.40, 0.59).

9.1.4 Conclusion

Abnormal ductus venosus Doppler showed moderate predictive accuracy for compromise of fetal/neonatal wellbeing and perinatal mortality in high risk pregnancies with placental insufficiency.

9.1.5 Publications arising from this work

R.Katie Morris, Tara J Selman, Meenakshi Verma, Stephen C Robson, Jos Kleijnen, Khalid S Khan. Systematic review and meta-analysis of the test accuracy of ductus venosus Doppler to predict compromise of fetal/neonatal wellbeing in high risk pregnancies with placental insufficiency. *Eur J Obstet Gynecol Reprod Biol* 2010 In press.

9.2 Introduction

Systematic Doppler application, can predict placental dysfunction in the form of FGR, preeclampsia or both[102]. The importance of venous Doppler (especially ductus venosus (DV)) in FGR stems from its dual capacity to evaluate cardiac function. As FGR worsens, rising afterload affects cardiac systolic and diastolic function. Forward venous flow is normally reduced during right atrial contraction (depicted as the a-wave in the DV waveform). A retrograde a-wave signifies the onset of overt fetal cardiac failure. While arterial Doppler evaluation provides important detail about placental function and its impact on regional fetal circulation, venous Doppler is essential to a complete understanding of fetal condition by quantifying fetal cardiovascular compromise[102]. DV is perceived to be the optimal test to predict fetal acidaemia and is used by many as the test on which to base decisions regarding delivery in preterm FGR[103].

The purpose of this review was to investigate the accuracy of ductus venosus Doppler to predict compromise of fetal/neonatal wellbeing.

9.3 Methods

The methods used are outlined in chapter 4 those specific to these review are detailed below.

9.3.1 Data Sources

Systematic searches were performed in Medline (Ovid), Embase (Ovid), the Cochrane Library (2009; 2) and MEDION from inception until May 2009. Search terms relating to SGA/FGR and fetal/neonatal compromise were combined with methodological filters for

identification of all primary articles reporting on the accuracy of ductus venosus Doppler (appendix 38).

9.3.2 Study Selection

Test accuracy studies reporting on ductus venosus Doppler allowing generation of 2x2 tables of accuracy (true positives, false positives, false negative, true negatives) were included. Criteria for inclusion were studies that reported on pregnancies at any level of risk in any healthcare setting using the ductus venosus Doppler to predict compromise of fetal/neonatal wellbeing. Case series <10 cases were excluded. The outcome measures used were those reported by the authors. Reference standards for wellbeing were any outcome measure performed after birth relating to neonatal wellbeing that was reported by the study authors. The outcome measure APO was used as detailed in sections 4.2.3.

9.3.3 Data Extraction and Study Quality Assessment

The data extraction form for this review can be found in appendix 39. Quality assessment was performed as described in section 4.4. See appendix 26 for further explanation of the quality assessment.

9.3.4 Data Synthesis and Analysis

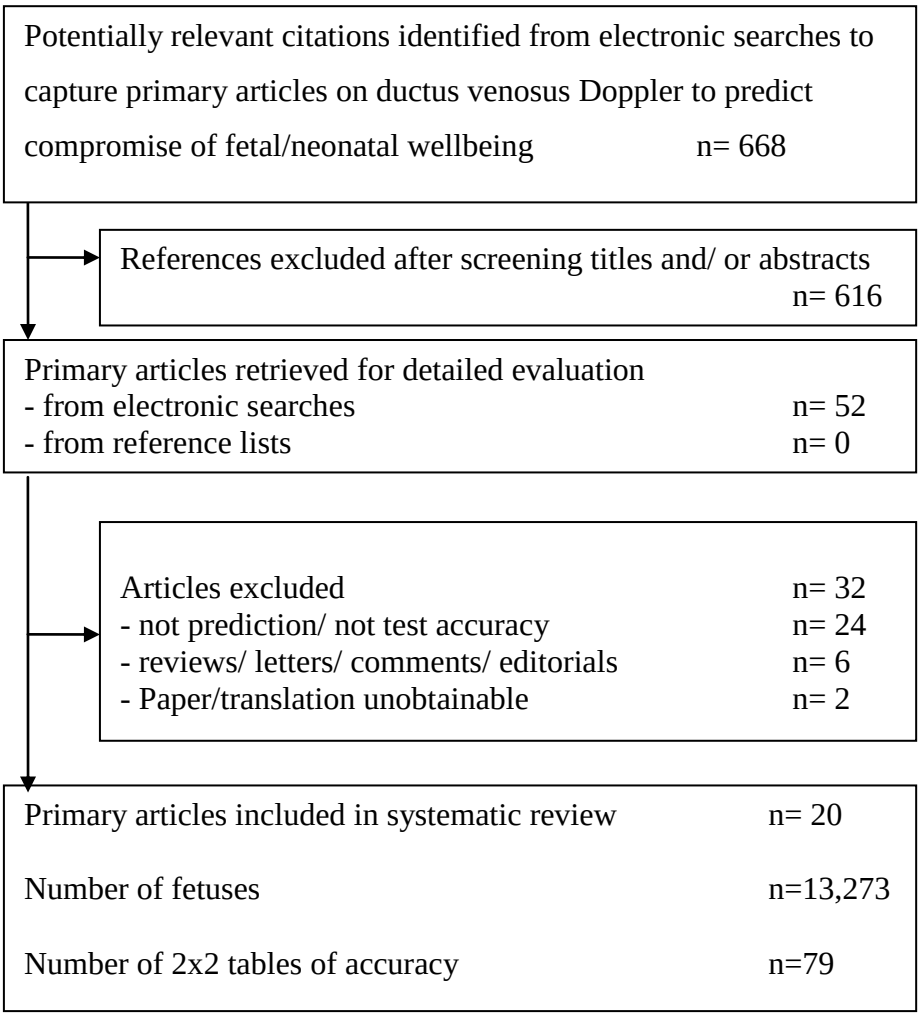
This was performed as detailed in section 4.6 and 7.3.4.

9.4 Results

9.4.1 Literature Identification and Study Characteristics

Figure 9.1 summarises the process of literature identification and selection. The references for the included studies and the details of the individual study characteristics can be found in appendix 40 and 41 respectively. There were 20 studies included overall for ductus venosus Doppler, reporting on 13,273 pregnancies and producing 79 2x2 tables. The Doppler results used for analysis were performed within two weeks of delivery with a range of gestational age from 11-41 weeks. All but one of the studies was performed in a high risk population (suspected placental insufficiency) and only one study reported exclusively on multiple pregnancies. In the remaining studies, 15 reported exclusively on singleton pregnancies and in four studies it was unclear whether multiple pregnancies were excluded. Fourteen studies excluded fetuses with chromosomal and structural anomalies. There were 18 cohort studies and two cross-sectional. There were 10 prospective, one consecutive, four retrospective, and six studies of unclear design. Five studies were identified as being from the same research group and had potential overlap between patients, despite contact with the authors it was not possible to determine the exact nature of this overlap thus each meta-analysis only included the most appropriate study ensuring patients were only counted once[102;104-107]. One paper reported exclusively on multiple pregnancies (*Maiz et al 2009* appendix 40*)* and one on first trimester testing (*Maiz et al 2008* appendix 40). These were excluded from the main meta-analysis as it was felt that this gave a more homogenous population and the first trimester ductus venosus testing represents a different pathological process. This left 18 studies, testing 2267 pregnancies all performed in high risk pregnancies between 20-41 weeks.

Figure 9.1: Process from initial search to final inclusion for ductus venosus Doppler to predict compromise of fetal/neonatal wellbeing (up to May 2009).



9.4.2 Study Quality

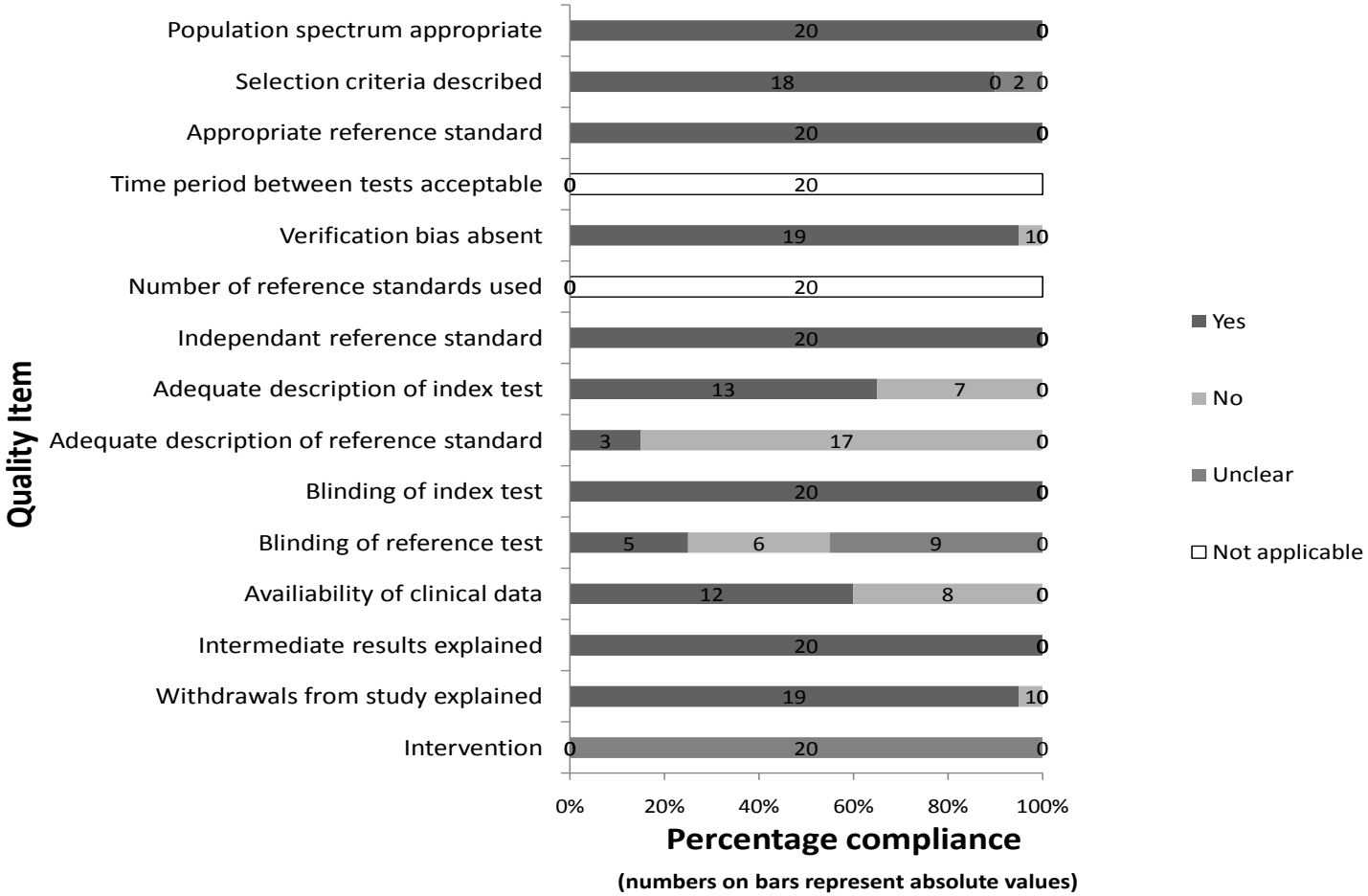
Figure 9.2 shows a summary of the quality assessment of included studies which revealed deficiencies. Only 13 studies contained an adequate description of the performance of the index test and three on the outcome measure. Blinding of the assessors of the outcome measure to the results of the Doppler was also poorly reported (5/20 studies). No studies reported on the use of any treatment in between the Doppler

and delivery. These items of quality of study design are important in diagnostic accuracy reviews. Verification bias was minimized as the number of eligible women progressing to the reference standard in included studies was >90% in 19/20.

9.4.3 Data Analysis

The results are summarized in figure 9.3 and table 9.1. An abnormal ductus venosus waveform predicted adverse perinatal outcome with a pooled positive likelihood ratio (LR+) of 3.15 (95% CI 2.19-4.54) and negative likelihood ratio (LR-) of 0.49 (95% CI 0.40-0.59), there was significant statistical heterogeneity. Sub-group analysis was performed based on for the following outcome measures: acidaemia, Apgar at 1 minute and 5 minutes <7, neonatal resuscitation, perinatal morbidity and mortality (table 9.1). This did account for some of the heterogeneity. Sub-group analysis according to type of waveform index could only be performed for absent or reversed a-wave and pulsatility index for veins>95th centile (table 9.1). To investigate the use of ductus venosus Doppler in preterm pregnancies sub-group analysis was performed using those studies including fetuses <37 weeks (table 9.1), sub-group analysis was not possible for severe preterm (<32 weeks) growth restriction (only 2 studies). Accuracy for prediction of acidaemia also improved when the Doppler was performed within 48 hours of delivery LR+ 4.25 (95% 1.01-16.47), LR- 0.64 (95% CI 0.46-0.88). The measures of adverse perinatal outcome, abnormal ductus venosus best predicted perinatal mortality with a LR+ 4.21 (95% CI 1.98-8.96), LR- 0.43 (95% CI 0.30-0.61). The receiver operating characteristic curve for this is shown in figure 9.4.

Figure 9.2: Bar chart showing quality of evidence on ductus venosus Doppler to predict compromise of fetal/neonatal wellbeing.



When assessing study quality, sub-group analysis could not be performed as there were too few studies of a high quality. For the purpose of this review it was felt to be important to look particularly at those studies in which ductus venosus Doppler had not been used in the management of the pregnancies (n=7 studies), that had blinded the investigators to the Doppler results (n=5 studies) and those studies that looked at preterm fetuses only/ did not use Doppler in management (n=4 studies). These results (table 9.1) show that test accuracy did decrease for adverse perinatal outcome [(LR+ve 2.10, LR-ve 0.49 for exclusion from management) (LR+ve 2.60, LR-ve 0.49 for studies with blinding) (LR+ve 2.31, LR-ve 0.41 for exclusion from management/preterm)].

Sensitivity analysis including only those studies which categorically stated they excluded chromosomal and structural anomalies or where the authors stated that singleton pregnancies only were included showed a slight improvement in results (table 9.1).

One paper (*Maiz et al 2009* appendix 40) reported exclusively on multiple pregnancies and thus was excluded from the meta-analysis. The results were very imprecise; for prediction of single fetal death in monochorionic twins LR+ 74.57 (95% CI 3.33-1672); LR- 0.79 (0.54-1.16) and in dichorionic twins LR+ 169.7 (95% CI 7.4-3889), LR- 0.83 (0.62-1.12).

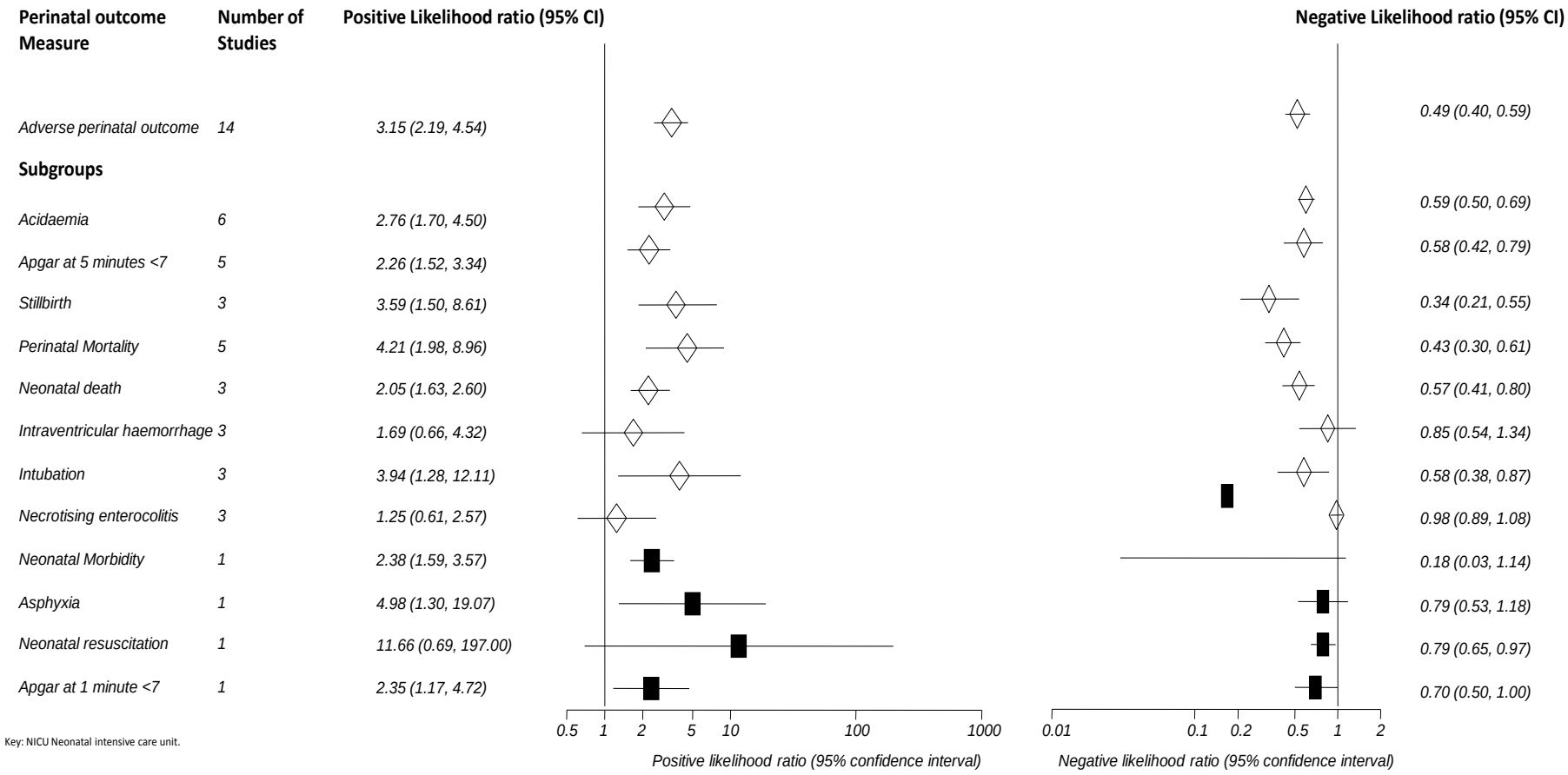
Sub-group and sensitivity analysis.

Sensitivity analysis including only those studies which excluded chromosomal and structural anomalies or where the authors stated that singleton pregnancies only were included showed a slight improvement in results (table 9.2). When assessing study quality, sub-group analysis based on study quality could not be performed as there were too few studies of a high quality.

Publication bias

Funnel plots (not shown) and the regression test for wellbeing as an outcome showed asymmetry (p=0.02). However, when the tests were restricted to sub-groups according to outcome, the plots were symmetrical suggesting the asymmetry was due to use of a combined reference standard introducing heterogeneity rather than publication bias (p=0.2).

Figure 9.3: Forest plot of abnormal ductus venosus Doppler in second/third trimester to predict compromise of fetal/neonatal wellbeing with subgroup analysis. Diamonds represent pooled results, squares represent individual studies.



Key: NICU Neonatal intensive care unit.

Figure 9.4: Bivariate analysis of the accuracy of ductus venosus Doppler to predict perinatal death.

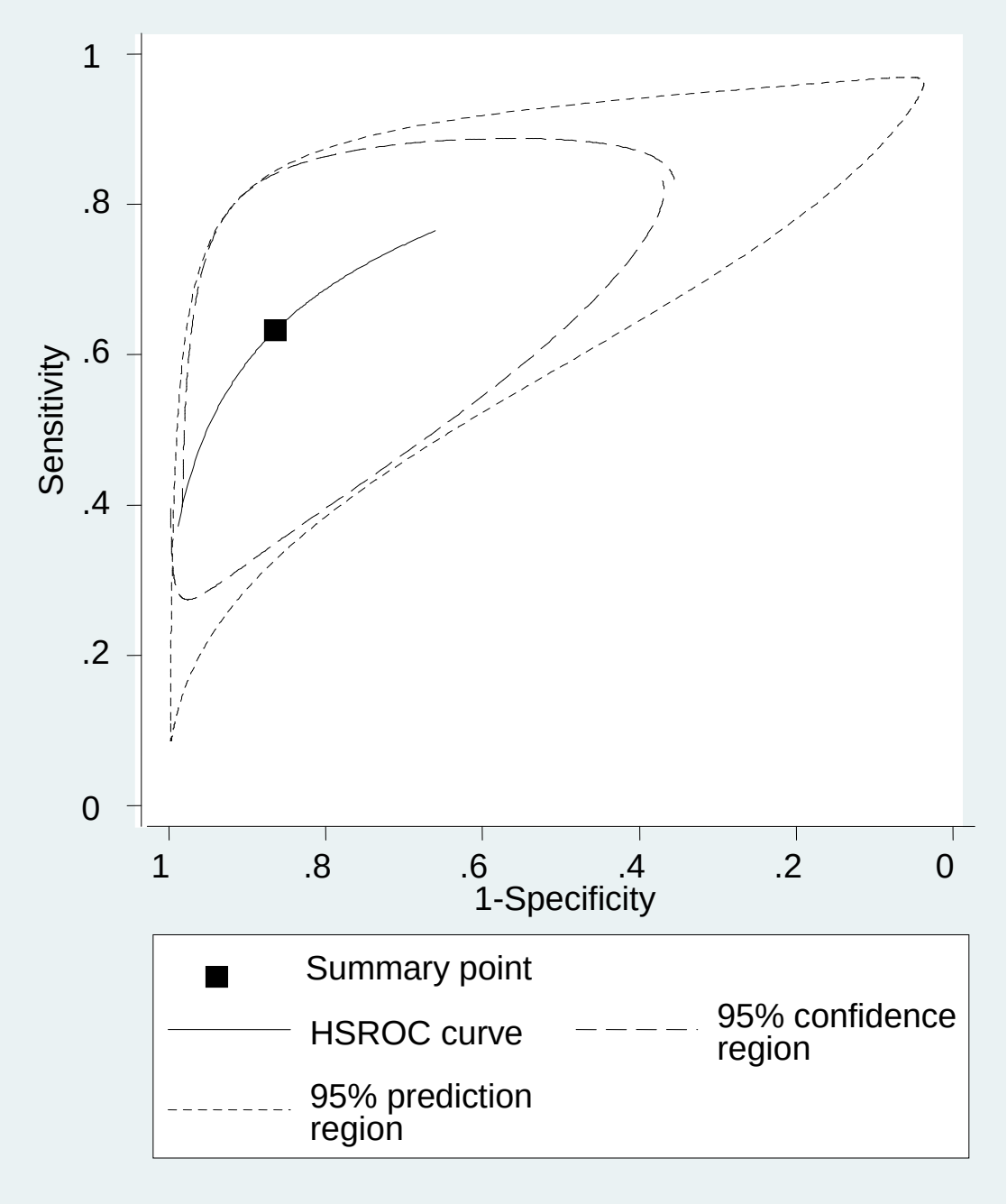


Table 9.1: Sub-group analysis for ductus venosus Doppler to predict compromise of fetal/neonatal wellbeing.

Outcome measure *Subgroup n=number of studies*	LR+ (95% CI)	LR- (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Adverse perinatal outcome n=14	**3.15 (2.19-4.54)**	**0.49 (0.40-0.59)**	**0.61 (0.50-0.70)**	**0.81 (0.70-0.88)**
Congenital abnormalities excluded n=12	3.75 (2.48-5.67)	0.47 (0.38-0.59)	0.60 (0.49-0.70)	0.84 (0.74-0.91)
Singleton pregnancies only n=10	3.93 (2.32-6.66)	0.41 (0.29-0.58)	0.66 (0.50-0.79)	0.83 (0.69-0.92)
Preterm delivery only n=11	3.79 (2.18-6.60)	0.51 (0.43-0.60)	0.57 (0.47-0.67)	0.85 (0.72-0.93)
Test to delivery within 24 hours n=5	3.14 (1.46-6.74)	0.45 (0.33-0.61)	0.64 (0.48-0.78)	0.79 (0.54-0.93)
Absent or reversed a-wave n=7	3.46 (1.67-7.16)	0.46 (0.35-0.61)	0.62 (0.46-0.75)	0.82 (0.61-0.93)
Pulsatility index for veins>95th centile n=3	3.74 (1.49-9.39)	0.57 (0.4-0.82)	0.54 (0.45-0.63)	0.87 (0.81-0.92)
DV Doppler not used in management/blinding and preterm n=4	2.31 (1.25-4.28)	0.41 (0.28-0.60)	0.72 (0.57-0.83)	0.69 (0.44-0.86)
Acidaemia n=6	**2.76 (1.70-4.50)**	**0.59 (0.50-0.69)**	**0.53 (0.45-0.60)**	**0.81 (0.69-0.89)**
Congenital abnormalities excluded n=6	3.10 (1.67-5.77)	0.55 (0.46-0.66)	0.54 (0.46-0.63)	0.82 (0.67-0.91)
Singleton pregnancies only n=4	3.87 (1.70-8.79)	0.51 (0.36-0.72)	0.57 (0.40-0.72)	0.85 (0.67-0.94)
Test to delivery within 48 hours n=3	4.25 (1.10-16.47)	0.64 (0.46-0.88)	0.45 (0.34-0.56)	0.89 (0.81-0.94)
Umbilical cord pH<7.20 n=5	4.39 (1.26-15.30)	0.60 (0.50-0.72)	0.53 (0.44-0.62)	0.76 (0.70-0.81)
Apgar score at 5 minutes<7 n=5	**2.26 (1.52-3.34)**	**0.58 (0.42-0.79)**	**0.57 (0.42-0.70)**	**0.75 (0.64-0.83)**
Congenital abnormalities excluded n=3	2.81 (1.37-5.76)	0.54 (0.37-0.78)	0.61 (0.45-0.76)	0.68 (0.63-0.72)
Singleton pregnancies only n=3	2.44 (1.31-4.55)	0.60 (0.43-0.84)	0.59 (0.43-0.74)	0.67 (0.63-0.72)
Absent or reversed a-wave n=3	3.17 (1.42-7.05)	0.64 (0.39-1.06)	0.52 (0.31-0.72)	0.78 (0.71-0.85)
Perinatal mortality n=5	**4.21 (1.98-8.96)**	**0.43 (0.30-0.61)**	**0.63 (0.47-0.77)**	**0.85 (0.66-0.94)**
Congenital abnormalities excluded n=4	5.48 (2.17-13.84)	0.44 (0.29-0.66)	0.61 (0.42-0.77)	0.89 (0.70-0.96)
Singleton pregnancies only n=4	3.80 (1.76-8.19)	0.44 (0.30-0.65)	0.63 (0.44-0.79)	0.83 (0.62-0.94)
Absent or reversed a-wave n=3	5.18 (0.97-27.61)	0.60 (0.41-0.86)	0.53 (0.36-0.70)	0.78 (0.69-0.85)
Neonatal death n=3	**2.05 (1.63-2.60)**	**0.57 (0.41-0.80)**	**0.59 (0.49-0.68)**	**0.67 (0.63-0.71)**

LR+ positive likelihood ratio; LR- negative likelihood ratio, CI confidence interval, DV ductus venosus

9.5 Discussion

For the prediction of compromise of fetal/neonatal wellbeing overall ductus venosus Doppler was found to have moderate predictive accuracy in a high risk pregnancy due to placental insufficiency. The best predictor was abnormal ductus venosus Doppler to predict perinatal mortality. Restricting the Doppler waveform to absent or reversed a wave and to tests performed within 24 hours of delivery showed improvement in accuracy. This review suggests that ductus venosus Doppler is a useful test in the management of the pregnancy at risk of fetal/neonatal compromise.

The strengths of this review lie in the methodology used which complies with existing guidelines for the reporting of systematic reviews[43] and also guidelines specific to the reporting of systematic reviews of observational studies[42]. The literature searches were extensive and designed to be sensitive rather than specific and were performed without language restrictions. Careful attention was paid to assessment of quality of study design and reporting. There are no previously published systematic reviews looking at ductus venosus Doppler in this area. There is debate as to whether pooled LRs or pooled estimates of sensitivity and specificity should be used due to the correlation between the paired statistics[74]. To account for this the bivariate method was utilised in the meta-analysis and sensitivity analysis with pooled sensitivities and specificities performed, no significant difference was found.

The limitations due to lack of clear reporting have been previously discussed (section 5.5 and chapter 10). Due to the smaller number of primary studies included in this review it was not possible to apply meta-regression thus sub-group analysis using the

aspects of study quality deemed important for this review was performed and no significant difference was found when analysis was restricted to "high" quality studies. Areas of study design where reporting was uniformly poor were in the description of the outcome measure and use of any intervention between the performance of the Doppler and delivery and whether decision to deliver was based on the results of the ductus venosus Doppler. Subgroup analysis using only those studies that blinded the results of the investigation did show a significant reduction in test accuracy. An argument could be made for not performing meta-analysis due to the significant heterogeneity. The decision to perform meta-analysis was based on the belief that there were sufficient similarities in the included populations, performance of the index test and outcome measure used to make this appropriate. Rigorous sub-group analysis was performed using those characteristics of the included studies that may have a significant impact on accuracy and shown varying results. These limitations do allow potential for bias within this review.

As FGR and severe compromise of fetal wellbeing are diseases with relatively low prevalence a clinically useful test would need to have a high positive LR (>10) and low negative LR (<0.10)[59]. At present the results of this review show moderate predictive accuracy however, it is important to stress the limitations of the meta-analysis in light of the significant heterogeneity. It is still important to report these results as it can help to determine those areas of study design that will be important for future research. In the future research in this area should take into account the limitations identified in this review; suitable populations; use of appropriate well defined outcome measures, utilization of treatments in between test and outcome and consider the use of tests in combination and the way they interact in the individual patient. Recent publications

have highlighted the importance of using Doppler measurements in combination and the use of an appropriate threshold[108]. A randomised controlled trial (TRUFFLE www.truffle.org) is currently recruiting to determine the use of ductus venosus Doppler in the timing of delivery of preterm growth restricted and to determine which is the most appropriate threshold to ensure that delivery is timed to minimise not only mortality but also neurological morbidity. The need for such a threshold must be taken into account in future systematic reviews of accuracy in this area as the best test/threshold may not necessarily therefore be the most accurate. Implementation of any testing/screening strategy within the pregnant population will need to take into account the cost of performing such a test and the availability of acceptable interventions including a consideration of side effect profiles and patient acceptability.

9.6 Conclusion

Abnormal ductus venosus Doppler showed moderate predictive accuracy for compromise of fetal/neonatal wellbeing and perinatal mortality in high risk pregnancies with placental insufficiency.

CHAPTER 10: SUMMARY OF REVIEWS OF TEST ACCURACY FOR PREDICTION OF SMALL FOR GESTATIONAL AGE FETUSES AND NEONATAL/FETAL COMPROMISE

10.1 Abstract

10.1.1 Background

In 2002 the ROCG published an evidence based guideline on "The Investigation and Management of the Small for Gestational Age Fetus". This guideline was developed using robust guideline methodology however, the recommendations were limited due to a lack of systematic collation of diagnostic information on the subject. Since this time Obstetrics has seen rapid growth in the development of new tests in the area of FGR particularly advances in ultrasound imaging and first trimester screening. The aim of this review is to summarise the systematic reviews performed with the objective of improving our understanding of the accuracy of the tests available to identify pregnancies at greatest risk of developing clinically relevant intrapartum and neonatal consequences of impaired fetal growth.

10.1.2 Methods

The accuracy of nine tests for prediction of fetal growth restriction and compromise of fetal wellbeing was evaluated. Tests were reviewed for test accuracy according to

prospective protocols including assessment of a study's methodological and reporting quality and bivariate meta-analysis to synthesise data. The main outcome measures were positive and negative likelihood ratios with 95% confidence intervals.

10.1.3 Results

In total 1,157 papers were read in full with 337 included in the reviews with 472,544 women tested. The median number of women included was 33,292 (interquartile range 13,273-40,637). The median number of studies per test was 60 (interquartile range 31-86). The quality of studies was variable as shown in figure 10.1 with the overall quality being poor. The main deficiencies were in the areas of description of the index test and reference standard, blinding of the reference standard and reporting of any intervention in between the index test and reference standard. The tests overall for prediction of small gestational age and adverse perinatal outcome demonstrated low predictive accuracy with no tests having a positive LR>5 and a negative LR<0.5.

10.1.4 Conclusion

The results show that the tests reviewed have a limited use in screening/diagnosis for these conditions when used in isolation but that special consideration has to be given to the prediction of particular conditions and in particular at risk groups.

10.2 Introduction

As discussed in chapter 2, fetal growth restriction is one of the commonest complications to affect pregnancies and represents a major cause of perinatal morbidity and mortality[7;8]. Birth weight remains a significant predictor of perinatal outcome even when other factors such as maternal obstetric complication and gestation are taken into account[109]. There are various definitions for fetal growth restriction including absolute birth weight thresholds and centile thresholds to classify the small for gestational age baby however both definitions are associated with an increased risk.

By definition however, babies born with a birth weight less than the 10th centile will constitute 10% of the population and not all these babies will have an increased risk of morbidity and mortality. The aetiology of fetal growth restriction is also diverse with intrinsic fetal causes, maternal causes and placental causes. The challenge is thus to determine which tests available to Obstetricians antenatally will help to determine the fetus that will be small for gestational age but also compromised either *in utero* or at birth affording the option for intervention using effective treatment or timely delivery.

Currently screening and diagnosis of fetal growth restriction and prediction and monitoring for compromise of fetal wellbeing in a clinical setting includes a combination of patients' characteristics, symptoms, physical signs and tests, which form the basis of clinical care[21]. For instance, methods employed to screen for and detect FGR might include obtaining previous history of small babies, recording symphyseal fundal height on a customised growth chart and estimating fetal weight with ultrasound[21]. Similarly, current history of fetal movements, abdominal palpation to

assess liquor volume, ultrasound amniotic fluid index, Doppler flow velocimetry and cardiotocography might be used to assess fetal wellbeing[21] . Tests of wellbeing are aimed at predicting fetal acidaemia, which is perceived, at least in the model of chronic placental failure, to lead ultimately to organ damage and death. Data from cord blood sampling studies confirm there is a correlation between cord pH and neurodevelopmental outcome in small fetuses[22-24].

In 2002 the ROCG published an evidence based guideline on "The Investigation and Management of the Small for Gestational Age Fetus"[21]. This guideline was developed using robust guideline methodology however, the recommendations were limited due to a lack of systematic collation of diagnostic information on the subject. Since this time Obstetrics has seen rapid growth in the development of new tests in the area of FGR particularly advances in ultrasound imaging and first trimester screening[3].

The aim of this chapter is to summarise the systematic reviews performed with the objective of improving our understanding of the accuracy of the tests available to identify pregnancies at greatest risk of developing clinically relevant intrapartum and neonatal consequences of impaired fetal growth.

10.3 Methods

The methods for the systematic reviews are detailed in chapter 4 with details pertaining to the individual reviews in the relevant chapters (chapter 5-9). In all reviews bivariate meta-analysis and sub-group analysis was employed. For the reviews of tests performed in the first and/or second trimester [Down's syndrome serum markers (chapter 5) and

uterine artery Doppler (chapter 6)] a clinical application with aspirin was explored. In the review on umbilical artery Doppler (chapter 7), there were sufficient studies to allow meta-regression analysis to explore heterogeneity.

10.4 Results

10.4.1 Summary of literature identification and study selection

Table 10.1 summarises the process from search to inclusion of papers in the five systematic reviews. In total 1157 papers were read in full with 337 included in the reviews with 472,544 women tested. The median number of women included was 33,292 (interquartile range 13273-40637). The median number of studies per test was 60 (interquartile range 31-86).

10.4.2 Summary of study quality

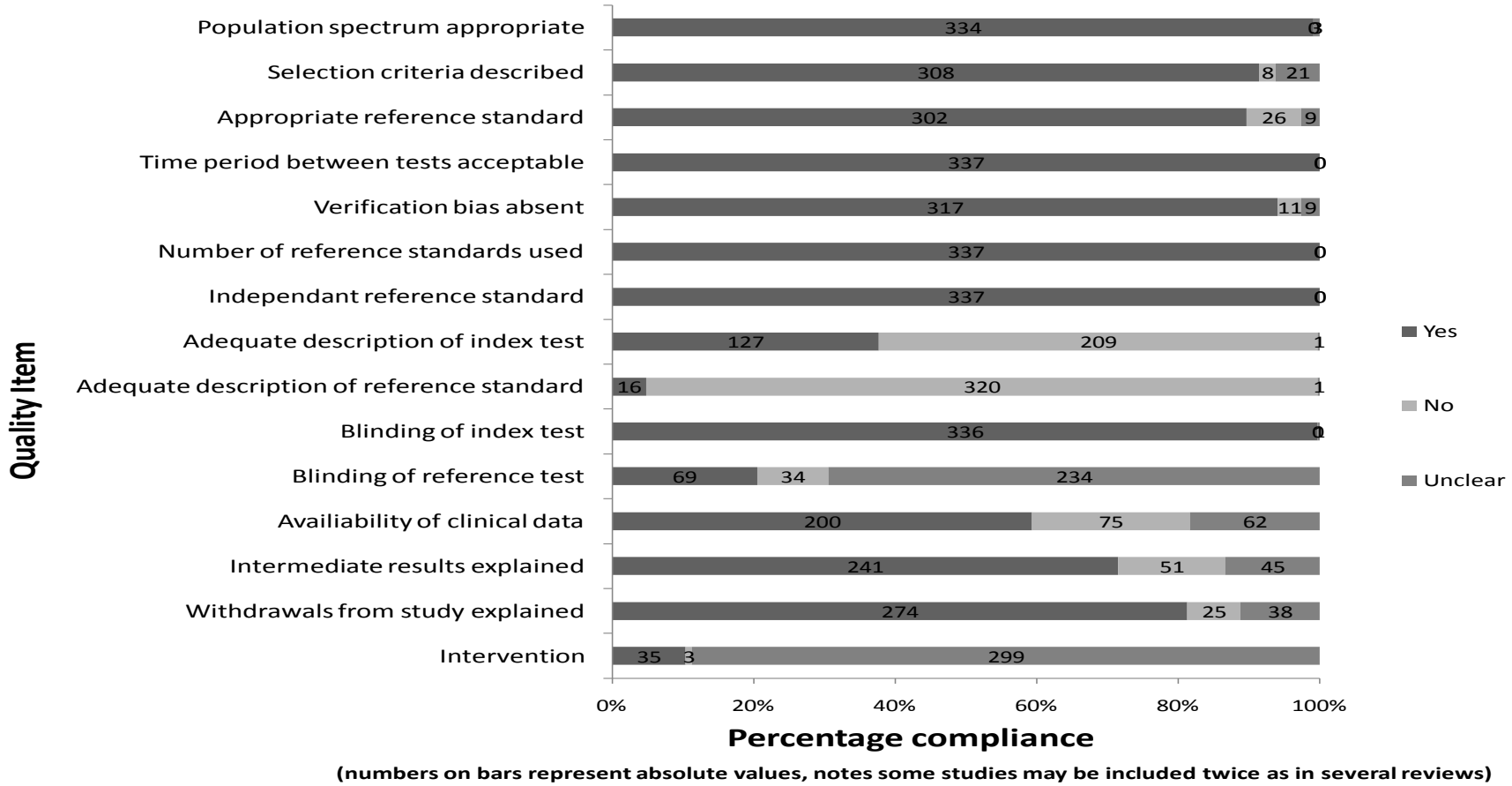
The quality of studies was variable as shown in figure 10.1 with the overall quality being poor. The main deficiencies were in the areas of description of the index test and reference standard, blinding of the reference standard and reporting of any intervention in between the index test and reference standard.

Table 10.1: Process from initial search to final inclusion for accuracy reviews

		Papers retrieved for detailed evaluation			Reasons for exclusion							
Index Test	**Total no. Of citations**	**No. from electronic searches**	**No. from reference lists**	**Total no. of full text papers excluded**	**Not prediction and/or not test accuracy**	**Reviews/letters/editorials/comments**	**Composite outcome (combined prediction of FGR and PET/PIH)**	**Mean gestational age at time of testing>25 weeks or unclear**	**Insufficient data to construct 2x2 table**	**Other[e]**	**Total no. of papers included**	**Total number of foetuses/women**
Uterine artery Doppler	1366[a]	311	8	259	123	29	25	44	26	12	60	40,637
Umbilical artery Doppler	4169[b]	419	6	285	218*	27	0	NA	-	40	140	33,292
Middle cerebral artery Doppler	2135[c]	87	9	65	60*	3	0	NA	-	2	31	3337
Ductus venosus Doppler	668[c]	52	0	32	24*	6	0	NA	-	2	20	13273
Downs syndrome screening markers Oestriol Alpha fetoprotein Human chorionic gonadotrophin Pregnancy associated plasma protein A Inhibin A	1769[d]	257	8	179	87	13	4	46	14	15	86	382,005

[a] Search until April 2006
[b] Search until March 2009
[c] Search until May 2009
[d] Search until February 2007
[e] Other reasons for exclusion include duplicate publication, paper or translation unobtainable or case control studies for some reviews, no threshold SGA
* Where marked this includes papers excluded as no 2x2 table could be constructed
Key: FGR fetal growth restriction, PET pre-eclampsia, PIH pregnancy induced hypertension

Figure 10.1: Quality of all tests reviewed for prediction of small for gestational age fetuses and compromise of fetal/neonatal wellbeing.



Areas where methodology was of a high quality were in the selection of patients, use of an appropriate reference standard and in achieving >90% verification. In all reviews the meta-analysis was hampered by the lack of clear reporting and methodological quality reducing the number of studies that could be included in the analysis.

One major criticism of the included studies in all reviews was the choice of reference standards used for determination of fetal growth. The protocol for the test accuracy reviews (chapter 4) allowed the inclusion of results from any reference standard as long as there was an appropriate population, index test and extractable 2x2 data. However, the majority of included studies used population based indices and usually birth weight <10th centile, no studies used customised charts. Indices that are said to be more indicative of nutritional status such as ponderal index were rarely reported and were not reported often enough to be used in any of the meta-analyses.

10.4.3 Summary of test accuracy findings

Figures 10.2 and 10.3 show the final summary results for the tests reviewed according to reference standard and population risk for prediction of small for gestational age fetuses and compromise of fetal/neonatal wellbeing respectively. These were the results that were considered for the decision analytic model (chapter 14). The tests overall for prediction of small gestational age fetuses and adverse perinatal outcome demonstrated low predictive accuracy with no tests having a positive LR>5 and a negative LR<0.5.

Small for gestational age fetuses

For Down's syndrome serum screening markers the predictive accuracy was low with improved prediction for severe forms of SGA (birth weight <5th and 3rd centiles). For uterine artery Doppler the predictive accuracy was again low with improved prediction for more severe forms and when looking at individual Doppler indices (pulsatility index with or without notching and resistance index). Umbilical artery Doppler showed improved prediction in a high risk population and for birth weight <3rd centile. Middle cerebral artery Doppler showed improved prediction in unselected versus high risk populations.

Adverse perinatal outcome

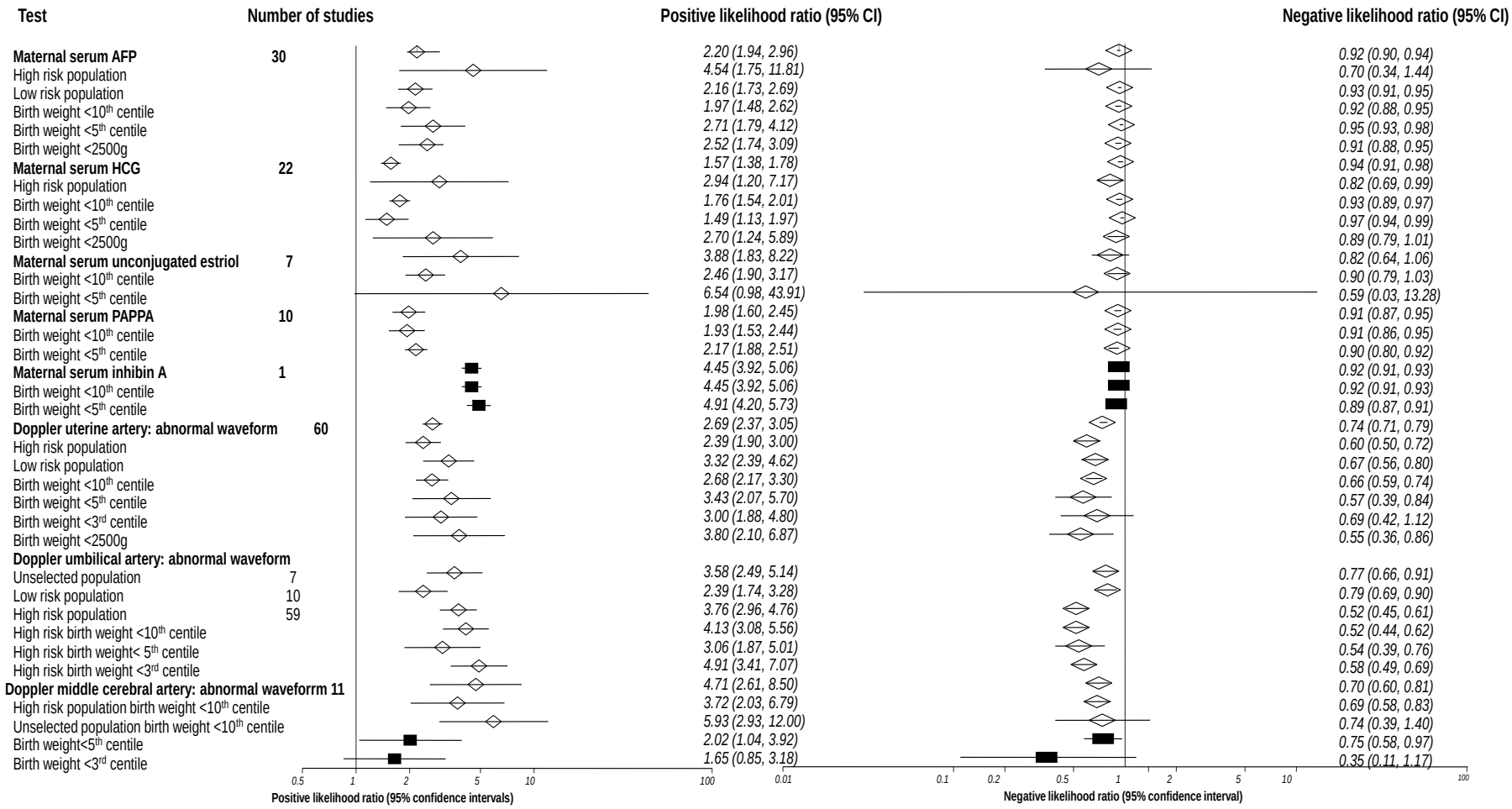
Umbilical artery Doppler showed moderate predictive accuracy in a high risk population with moderate prediction of intra-uterine death, acidosis and admission to neonatal intensive care unit. Middle cerebral artery Doppler showed low predictive accuracy with improved prediction in unselected populations. Ductus venosus Doppler showed moderate predictive accuracy in a high risk population in particular for prediction of perinatal mortality.

Despite sub-group analysis there was significant unexplained heterogeneity in most cases this affected the interpretation of the results and confidence in the predictive ability of the tests under review. In all of the reviews sub-group analysis could be performed based on population risk thus allowing an assessment of test accuracy across different populations. This has been translated into different decision analytic models according to population risk (chapter 14) and allowed recommendations for clinical practice.

The significant unexplained heterogeneity led to the decision to re-analyse all data using the bivariate method as this accounts for this type of heterogeneity in its random effects model.

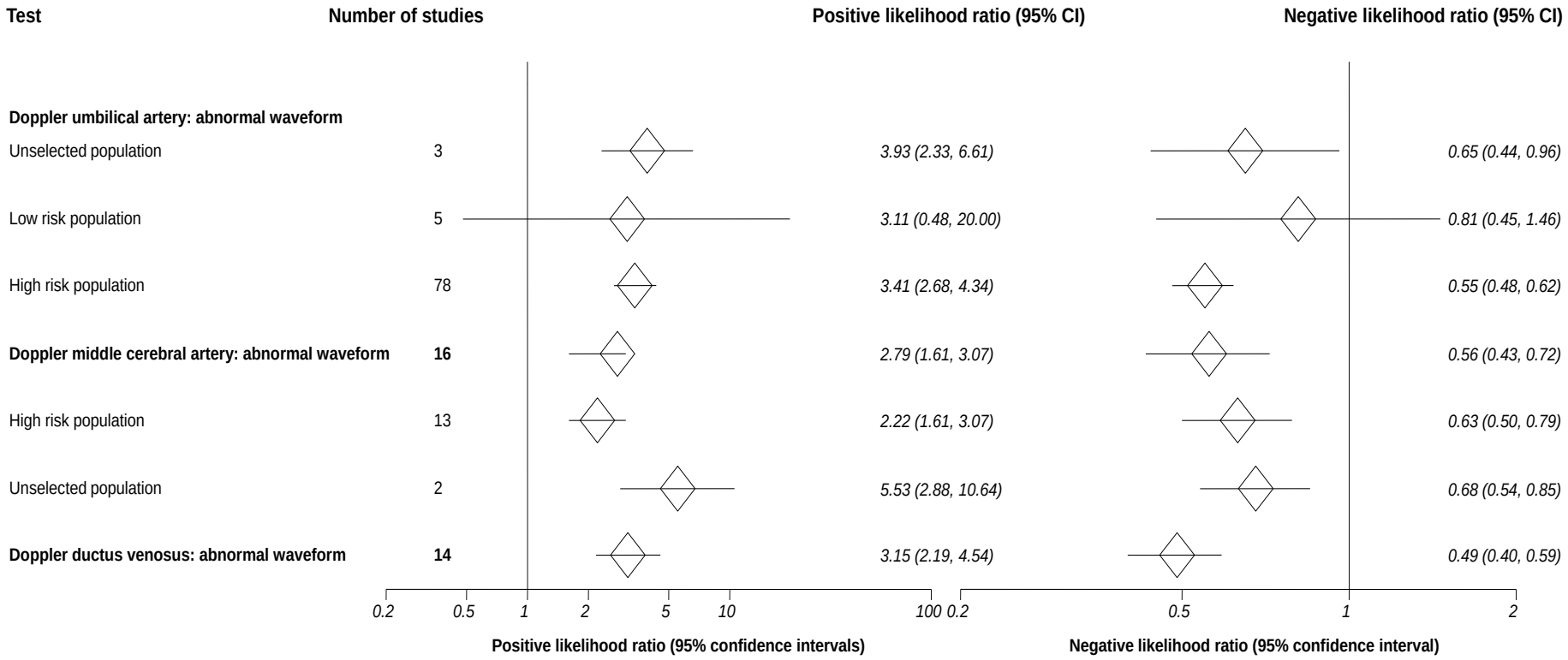
Tests performed in the first and second trimester to screen for SGA (Down's syndrome serum markers and uterine artery Doppler) showed high specificities (>90% for some indices) and very low sensitivities (<25%). There was wide variation in the precision of the estimates. Tests performed in later pregnancy for the diagnosis of SGA (umbilical artery and middle cerebral artery Doppler) showed improvement in the sensitivity estimates (0.29-0.55 depending on risk of population) but with a decrease in the specificity estimates (0.74-0.92). Tests performed in later pregnancy to diagnose compromise of fetal wellbeing (umbilical, middle cerebral and ductus venosus Doppler) showed better sensitivities (0.25-0.65) but with a decrease in specificities (0.8-0.84).

Figure 10.2: Forest plot of accuracy estimates from all tests reviewed for prediction of small for gestational age fetuses (diamonds represent pooled results, squares individual studies)



CI confidence interval; AFP alpha feto-protein; HCG human chorionic gonadotrophin

Figure 10.3: Forest plot of accuracy estimates from all tests reviewed for prediction of fetal/neonatal compromise (adverse perinatal outcome) (diamonds represent pooled results, squares individual studies.



CI confidence interval

10.5 Discussion

The results show that the tests reviewed have a limited use in screening/diagnosis for these conditions when used in isolation but that special consideration has to be given to the prediction of particular conditions and in particular at risk groups. The strength of the evidence lies in the methodology and rigorous statistical analysis using contemporary methods.

Limitations arising from problems with primary data

The interpretation of the results presented must take into account the deficiencies identified in study quality and the association between design quality and diagnostic performance[51;75]. There was rigorous assessment of methodological and reporting quality for all included studies and appropriate assessment to take into account the unique situation of screening for conditions in pregnancy that cannot be verified by a reference standard until the end of the pregnancy e.g. blinding of index test assessment from results of reference standard was always considered to have occurred. Thus when assessing whether a study was of high quality special consideration was given to those aspects of design/reporting that were felt to have potentially more impact on test accuracy e.g. blinding of reference standard, description of index test and reference standard. Despite research being performed to study the effect of study methodology and reporting on accuracy estimates there has been no research done in the specific area of obstetrics[51;75]. The decisions as to which items of quality were the most important had to be made based on clinical and epidemiology experience. This deficiency in the published literature has been addressed in chapters 11 and 12.

Due to lack of clarity in the description of included populations it was not always possible to adequately identify populations as high or low risk and this has implications for any future recommendations for practice. The same can be said for the lack of adequate description of index and reference standards. One significant deficiency was the lack of use of reference standards such as ponderal index, this leads to the conclusion therefore that the results as shown in figure 10.2 relate to prediction of small for gestational age not fetal growth restriction. This is a major limitation in the use of these tests in clinical practice as the ideal test would discriminate between those fetuses that are small for gestational age from those that are growth restricted allowing intervention in the latter group only. To determine which fetuses might benefit from intervention the clinician would need to look at the tests performed in later pregnancy for diagnosis (umbilical, middle cerebral artery and ductus venosus Doppler) of adverse perinatal outcome. In many studies there was also a lack of description of the use of any treatment throughout the course of the pregnancy e.g. giving aspirin to test positive patients with abnormal uterine artery Doppler or early elective delivery in those with abnormal ductus venosus Doppler. This meant that it was impossible to assess for the risk of treatment paradox, this however has to be interpreted in the light of the data presented in chapter 13 showing that treatments in this area are not proven to be effective.

Sub-group analysis was limited due to the number of included studies in each of the reviews limiting the number of patients within each sub-group. While the majority of heterogeneity could be accounted for by sub-group analysis looking at population risk or particular reference standard it was not possible to account for all possible confounding factors and often the author of the primary papers definition of risk had to be relied upon

as there was insufficient information about the included populations to discriminate. Ideally meta-regression analysis as performed in the review of umbilical artery Doppler (chapter 9) would have been performed in all reviews to explore the heterogeneity but the number of studies with each variable precluded this and would have meant that the meta-regression was underpowered[110].

The data analysed within these reviews was also limited due to potential diagnostic confounding by other diagnostic information e.g. patient history, obstetric risk profile, and other test results. This information from the clinical history as well as other test results all contains diagnostic information as well as that obtained from the test under investigation. Confounding occurs when there is a relationship between the predictive/diagnostic capabilities of the different measures so that it is difficult to assess the actual independent predictive value of the test under investigation. The issue of diagnostic confounding can be dealt with in two ways – multivariable analysis of the primary study data or individual patient data meta-analyses (IPD)[111]. These techniques require considerable extra resources and time which was outside the scope of this work. In an attempt to counteract this potential confounding, patient groups that were as homogenous as possible for these characteristics were created for the analysis.

The included studies reported many different test thresholds which limited the summaries of test accuracies that could be generated. To account for this the bivariate method was employed which estimates the correlation in sensitivity and specificity due to threshold effect as well as accounting for unexplained statistical heterogeneity. Many studies provided estimates of more than one diagnostic indicator for each individual patient (e.g.

uterine artery Doppler results for resistance index and pulsatility index) this meant that a valid statistical comparison of the accuracy of the different diagnostic indicators within tests as the compared study samples were not statistically independent.

Limitations arising from review methods

As demonstrated in table 2.1, chapter 2 there is a wide range of tests available in the literature for prediction of fetal growth restriction and compromise of fetal wellbeing. While there is published evidence relating to these tests the majority of them are not used in clinical practice. For the purpose of this thesis the tests chosen for review of test accuracy were selected on the basis of opinion of the research team and expert clinical opinion from experts within the field of Fetal medicine known to the researchers. There was then consideration of the evidence already published to determine the tests where the evidence either did not exist or it needed updating (appendix 1). Ideally a Delphic survey of practice would have been performed. This technique involves the collection and aggregation of expert opinion using questionnaire rounds, feed-back responses and the opportunity for participants to modify their responses and anonymity of responses[112]. Whilst the scientific merit and validity of this technique has been questioned it can be useful where there is no conclusive evidence available by relying on and sharing expert opinion[113;114].

To increase the number of studies that could be included in analysis a composite outcome measure was employed as discussed in 4.2.3. It is recognised that one of the hazards of composite outcome measures is the assumption that the significance of the result applies to all components[49]. To address this issue a separate analysis was also performed using

the component outcomes of the composite outcome measure with care taken to ensure that each individual was only counted once in each analysis. While the use of composite outcome measures is an accepted technique in systematic reviews as long as the direction of effect for each of the included outcomes is in the same direction and separate analysis is performed looking at the individual components[50], it must be accepted that ideally individual reference standards would be used to reduce heterogeneity due to the varying components of the composite.

Limitations arising from things not done

As demonstrated in appendix 1 there were some tests where too few studies were identified to make a review worthwhile considering the restraints on time and resources. Conversely in some of reviews the relatively large number of studies identified (uterine and umbilical artery Doppler) meant that these reviews required a lot of time to extract all the data and ensure that appropriate meta-analysis was performed. This meant that two reviews that were planned to be completed as part of this thesis had to be postponed to be completed at a later date as it was felt that to assign the correct amount of time to complete them would have been to the detriment of the other parts of this thesis namely the effectiveness reviews and the decision model analysis. These two reviews looked at amniotic fluid measurements and the biophysical profile and after the searches had been performed and inclusion of papers had been completed there were 115 and 62 papers included respectively. It is recognised that these are two important tests that are performed in clinical practice where there is suspected FGR or fetal compromise and that any guidelines on management of these clinical problems must include the evidence available on the use of these tests[115;116]. Ideally if time had allowed a test accuracy review

looking at combination testing using Down's syndrome serum markers and uterine artery Doppler in combination in the same patient would also have been performed.

As discussed above ideally the analyses would have been limited to high quality studies and techniques such as meta-regression and multivariable analyses would have been performed but the small number of studies per test/variable and the poor reporting/methodological quality of included studies meant that these methods could not be employed.

There have also been developments in meta-analytic techniques since these reviews were performed such as the calculation of the estimated predictive interval which relates to the effect of a new study that would be eligible for inclusion in the meta-analysis and therefore allows the full uncertainty around inferences to be calculated, including both magnitude and consistency[117].

Findings in the light of these limitations

The reviews performed present the best available evidence for the value of these tests in the prediction of small for gestational age fetuses and compromise of fetal/neonatal wellbeing at the time of completing the work. Although substantial limitations can be identified, mainly in the quality and quantity of the available primary evidence it can be concluded that overall the tests reviewed have limited value in the prediction of small for gestational age and compromise of fetal wellbeing. However, the true value of a test has to be assessed in light of its use in the clinical pathway – is it to be used for screening/prediction or diagnosis? What actions will be taken on the basis of a positive

result – will further testing be offered or will treatment be implemented? What is the effectiveness of available treatments? (chapter 13). These questions will be considered in the decision-analytic model in chapter 14.

Recommendations for research

As no one test had proved to be accurate for either screening or diagnosis there is a need for further primary research to look at new markers and tests for FGR and fetal compromise. Once identified from primary laboratory studies these tests should be investigated with robustly designed diagnostic accuracy studies that must include a sample size calculation to ensure sufficient statistical power to estimate test sensitivity and specificity/likelihood ratios precisely in study groups as a whole and in clinically relevant subgroups[118;119]. Future research should look at the place of these tests in the clinically pathway and assess them for their added value to clinical data obtained from history or examination (risk profile) using appropriate statistical analysis.

Future systematic reviews in this area should consider the use of a Delphic survey of practice to identify the tests to be examined. These reviews should also ideally use individual patient data meta-analysis to help overcome the problems of unclear reporting, composite outcome measures, appropriate sub-group analysis and diagnostic confounding. There is a need for systematic review of the evidence for combination testing in this clinical area. It is likely however that any such review will be hampered by the same limitations and potential for bias as the test accuracy reviews of individual tests performed as part of this work.

Considerations for future research

A test with perfect accuracy would have a sensitivity of 100% and a specificity of 100% i.e. no false negatives nor false positives. In reality these perfect tests do not exist and tests are designed with an accepted error rate based on the nature of the test, the disease under investigation and the consequences of a false positive or false negative test result. This error will take into account the nature of the correlation between sensitivity and specificity i.e. that as sensitivity increase the specificity will decrease and vice versa.

For FGR and compromise of fetal wellbeing one must consider the implications of a false negative and false positive result to determine what levels of sensitivity and specificity might be acceptable. In current clinical practice a false positive result will lead to increased surveillance of the pregnancy (e.g. growth scans, Dopplers) and the potential interventions of aspirin and early delivery. While there may be an increase in anxiety for the mother and inconvenience due to the enhanced surveillance the treatments are not associated with serious side effects. There may potentially be a side effect to the fetus of iatrogenic prematurity from early delivery with false positive tests for wellbeing later in pregnancy. With a false negative result, women and clinicians are falsely reassured that the fetus is not at risk of being growth restricted or its wellbeing compromised and thus the increased surveillance and treatments are not implemented. This fetus is thus at increased risk of perinatal morbidity and mortality as well as the longer term implications for its infant and adult life. Thus it can be argued on balance that for FGR and compromise of fetal wellbeing a test should have a level of accuracy that has an error rate with a preference for a low number of false negative results i.e. a high sensitivity. The relationship between preference for low false positives and negatives and whether

sensitivity or specificity most reduces error rate is also influenced by the frequency of disease. The ideal requirements of a test for a particular disease/condition can be further refined by considering the costs of the tests and treatments that might be instituted via modelling (chapter 14). Any new test can thus be assessed for its added value to the existing antenatal care system and how its implementation will reduce the number of unwanted screening errors.

10.6 Conclusion

The results show that the tests reviewed have a limited use in screening/diagnosis for these conditions when used in isolation but that special consideration has to be given to the prediction of particular conditions and in particular at risk groups. The main implications are thus not for recommendations for practice but for future research.

PART B: METHODOLOGICAL RESEARCH INTO THE EFFECTS OF REPORTING AND METHODOLOGICAL QUALITY OF SYSTEMATIC REVIEWS OF TEST ACCURACY

CHAPTER 11: THE QUALITY OF REPORTING OF PRIMARY TEST ACCURACY STUDIES IN OBSTETRICS REVIEWED IN THIS THESIS: APPLICATION OF THE STARD CRITERIA.

11.1 Abstract

11.1.1 Background

In obstetrics there has been a rapid growth in test accuracy studies. It is important that the reporting of these studies is transparent so that a valid assessment of the reported results can be made. The purpose of this study was to determine the quality of reporting in diagnostic test accuracy studies in obstetrics using the Standards for Reporting of Diagnostic Accuracy – STARD checklist.

11.1.2 Methods

The included studies of seven systematic reviews were assessed for reporting quality using the STARD checklist. The compliance with each of the reporting criteria was assessed. Using appropriate statistical tests it was investigated whether there was an improvement in reporting quality since the introduction of the STARD checklist, whether a correlation existed between study sample size, country of origin of study and reporting quality.

11.1.3 Results

A total of 195 studies were included. The overall reporting quality of included studies to the STARD criteria was poor. The studies reported adequately >50% of the time for 62.1% (18/29) of the items. There was a positive correlation ($p<0.0001$) between study sample size and reporting quality. No correlation between geographical area of publication and compliance with the reporting criteria could be demonstrated.

11.1.4 Conclusions

The reporting quality of papers in Obstetrics is improving. This may be due to initiatives such as the STARD checklist as well as historical progress in awareness among authors to accurately report studies. There is however considerable scope for further improvement.

11.1.5 Publications arising from this work

Selman TJ, Morris RK, Zamora J, Khan KS. The quality of reporting of primary test accuracy studies in Obstetrics and Gynaecology: application of the STARD criteria. *BMC Women's Health 2010* (In press).

11.2 Introduction

In obstetrics there has been a rapid growth in the development of new tests and primary studies of their accuracy. These studies generate a comparison of the result from an index test against an accepted reference standard or outcome measure[4]. The accuracy of the index test is usually expressed as sensitivity and specificity or other measures like the diagnostic odds ratio (DOR), likelihood ratio (LR) or area under a receiver-operator characteristics curve[5]. These allow clinicians to judge the usefulness and suitability of testing in clinical practice. It is imperative that such studies are reported with transparency allowing the detection of any potential bias that may invalidate the results[120;121]. Guidelines for the reporting of other study types have widely been accepted e.g. CONSORT[122] for randomised control trials. There has been a format for reporting evaluations of tests called Standards for Reporting of Diagnostic Accuracy – STARD[34], introduced in 2003.

The object of the STARD initiative is to improve the reporting of test accuracy studies to allow for the detection of potential bias in a study and to make a judgement on the applicability of the index test results. One of the benefits of using the STARD initiative is to develop a consistent reporting format across all types of tests. The STARD group identified 33 previously published checklists for diagnostic research. From an initial 75 point check list a consensus meeting formulated a 25 point list that could be employed to accuracy studies. This list was designed to help readers judge the studies and to act as a study design tool for authors. Points were specifically chosen on evidence supporting their ability to show variations in measures of diagnostic accuracy[34]. Further supplementing the checklist was flow diagram which aids the assessment of the

study population, the recruitment method and indicates the numbers receiving the index test, those excluded and those compared with the reference standard at different stages of the study. STARD should allow a reader to critically appraise the study design, analysis and results.

Previous studies have looked at the impact of STARD in specific clinical areas[123-127] with varying outcomes and the overall quality of reporting of studies which was generally found to be poor. There is no published research looking at the impact of STARD in Obstetrics.

This chapter aims to assess the reporting quality of test accuracy studies in obstetrics that form part of this thesis and the impact of the STARD statement.

11.3 Methods

A protocol to assess the impact of STARD on studies included in seven systematic reviews performed over the period 2005-2007, as part of this thesis, was developed. The studies covered the time period 1977-2007. The included reviews assessed the accuracy of Down's syndrome serum screening markers and uterine artery Doppler to predict small for gestational age fetuses in obstetrics (chapters 5 and 6)[128;129]. These reviews were chosen as they had all been performed by the author who had received training in use of the STARD checklist and had reached a consensus *a priori* as to how compliance with the checklist should be assessed for each review. The STARD checklist was applied to each of the studies included in all the reviews with the reporting item being determined as either present, absent, unclear or not applicable (appendix 3). All studies

were assessed in duplicate by the author and a second reviewer (Dr Tara Selman) who had received the same training and been involved in the discussions as to how compliance would be assessed, where there was disagreement consensus was achieved following assessment by a third reviewer (Professor Khalid Khan). In the event that several tests had been applied to the same patient, the results including the largest number of patients were used in the study or where there was no difference, one index test was selected at random, this ensured patients were only included once.

The following questions were addressed: Has the introduction of STARD improved reporting quality?; does study size correlate with reporting quality?; is there a geographical pattern to reporting quality?; is there a relationship between reporting quality and methodological quality?. The percentage compliance of studies with STARD items was compared before and after the introduction of STARD and over time using the unpaired t test to assess the effect of STARD on the reporting quality of studies. With the publication of STARD in 2003 the assumption was made that all studies published pre 2004 were published without the benefit of this directive.

The relationship between sample size and compliance with STARD was examined using Spearman's rank correlation coefficient (Rho). Kruskal Wallis was used to investigate any relationship between geographical distribution and reporting quality. The country of origin of a study was determined by the country of the corresponding author. Where a significant result was found, pairways comparison was made using Conover Inman. Countries were grouped depending on the number of articles published and the mean journal impact factor and adjusted for gross domestic product and

population, based on a previous publication[130]. Where there was a large disparity in number of studies per geographical area, some studies were re grouped to avoid large differences in group size and potentially spurious results. The geographical areas were Oceania, USA, Canada, Asia, Japan, Africa, Eastern Europe and Western Europe.

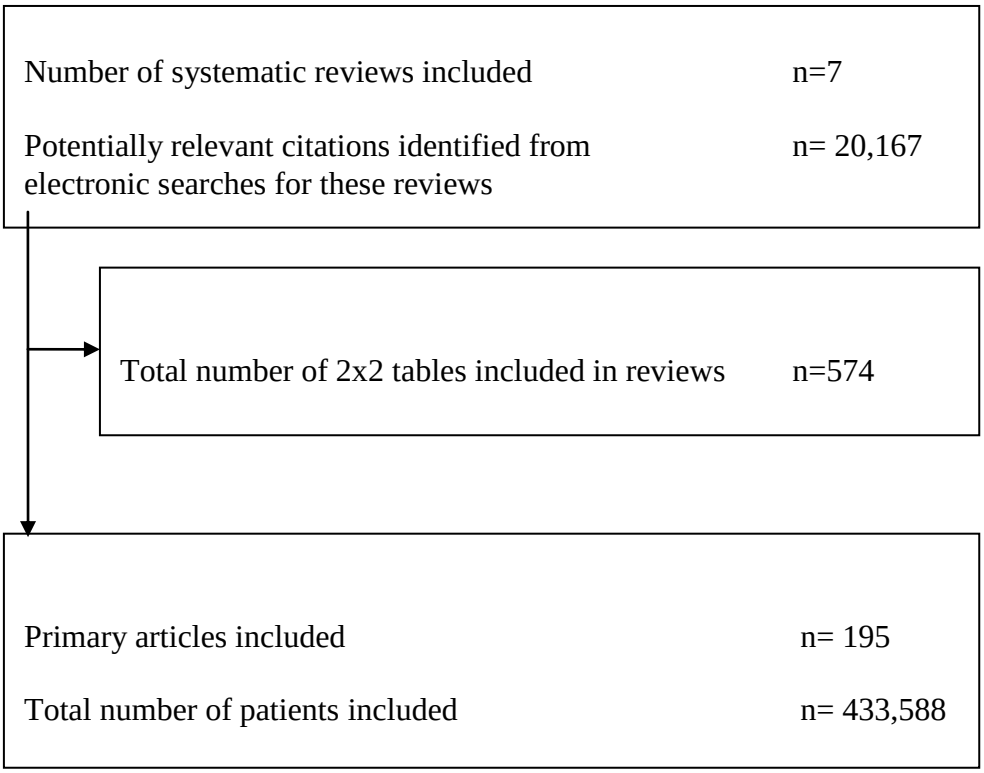
In the initial analysis those reporting items coded as unclear and not applicable were excluded. For all of the above analysis, due to the uncertainty of whether reporting items coded as unclear represented methodological failure, sensitivity analysis was performed excluding this code and adding it to the not reported group for all comparisons. Similarly sensitivity analysis was also performed to assess the effect of those items assessed as not applicable, with their initial exclusion to the analysis and then addition as if they were reported so as not to penalise studies which had a larger number of not applicable items and would therefore potentially have a seemingly lower compliance with STARD.

11.4 Results

A total of 195 studies were identified and included in this analysis (figure 11.1). 82% (160/195) were published prior to the STARD initiative. The overall percentage compliance with individual reporting items is shown in table 11.1. The included studies reported adequately >50% of the time for 62.1% (18/ 29) of the items as assessed in this review. Items where reporting was uniformly poor (<50%) were participant sampling, description of technique of reference standard, description of expertise of people performing index and reference standard, blinding of results of index test to those

interpreting reference standard, assessment of test reproducibility, tabulation of results and description of adverse events.

Figure 11.1: Process from initial search to final inclusion for papers from obstetric systematic reviews for assessment of compliance with STARD.



There was significant improvement in the reporting quality of studies after the introduction of STARD (p=0.0004). Figure 11.2 shows the trend in compliance with the STARD criteria over time. Analysis of the correlation between sample size and compliance with STARD revealed a positive correlation (Rho = 0.37, p = <0.0001) (figure 11.3). Investigation in to the relationship between geographical area of publication and the compliance with STARD showed no relationship (Kruskal-Wallis 5.05 p=0.65) figure 11.4. Sensitivity analysis showed no significant difference in any of the results.

Table 11.1: Percentage compliance with individual STARD criteria for included test accuracy studies in obstetrics.

STARD item	Description	Percentage compliance obstetric studies (%)
1	*Article is identified as study of diagnostic accuracy*	27.2
2	*States research question/aims*	94.9
3	*Describes study population*	74.4
4	*Describes participant recruitment*	85.1
5	*Describes participant sampling*	36.4
6	*Describes index standard*	59.5
7	*Describes reference standard*	86.7
8a	*Describes technique of index test*	45.1
8b	*Describes technique of reference standard*	0
9a	*Describes cut-off for index test*	96.9
9b	*Describes cut-off for reference standard*	75.9
10a	*Describes persons executing index test*	8.2
10b	*Describes persons executing reference standard*	0
11a	*Were results of index test blinded?*	100
11b	*Were results of reference test blinded?*	8.2
12	*Describes methods for statistics used*	53.3
13	*Describes methods for calculating test reproducibility*	12.3
14	*Reports dates of study*	65.1
15	*Reports characteristics of study population*	67.2
16	*Reports number of eligible patients that did not undergo either test*	69.2
17	*Time interval between tests and any treatment*	11.8
18	*Reports distribution of severity of disease*	86.7
19	*Reports cross tabulation of results*	49.2
20	*Reports adverse events*	0 (100% not applicable)
21	*Reports estimates of diagnostic accuracy*	54.4
22	*Reports how missing results were handled*	63.6
23	*Reports estimates of variability of accuracy*	56.4
24	*Reports estimates of test reproducibility*	12.8
25	*Discuss clinical applicability of findings*	99.5

Figure 11.2: Bar chart showing mean percentage compliance of studies with STARD criteria, line shows trend over time.

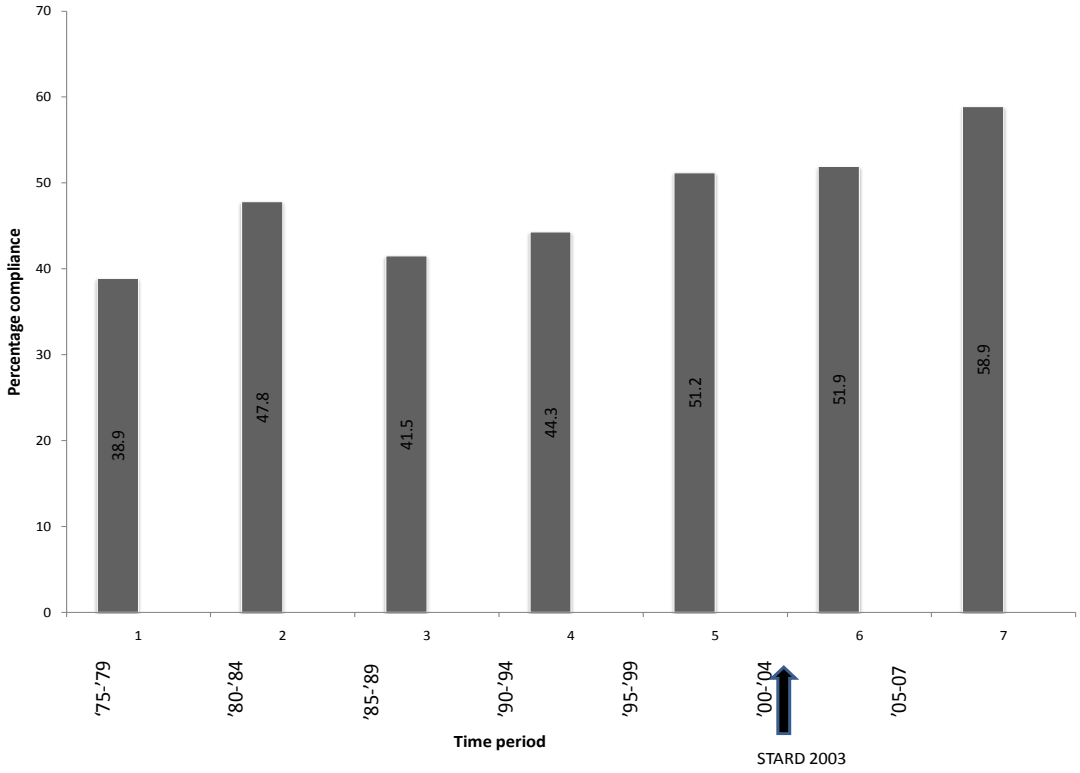


Figure 11.3: Scatter plot showing total compliance with STARD reporting criteria according to sample size.

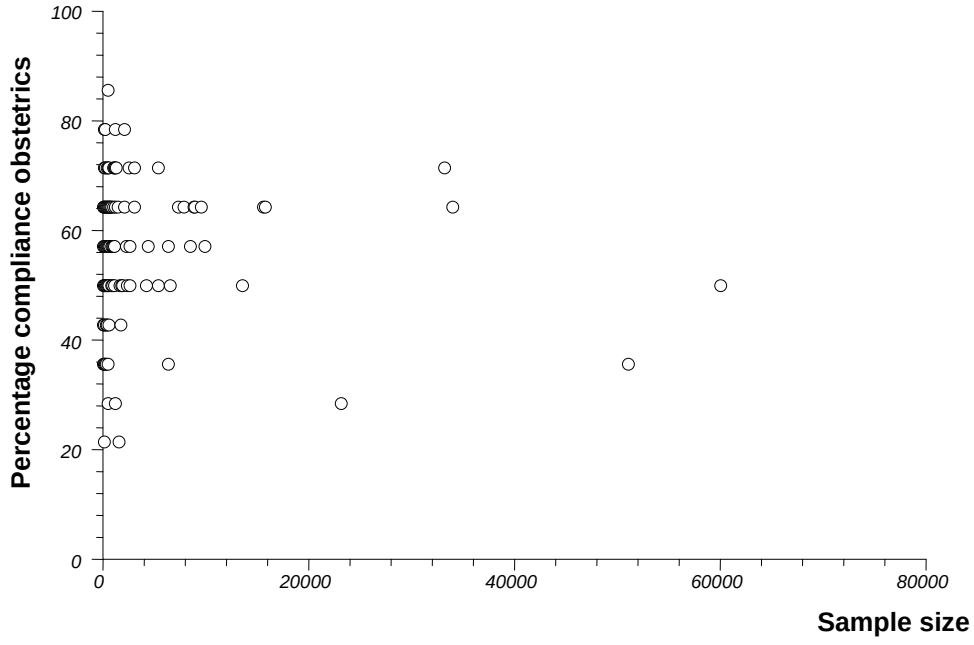
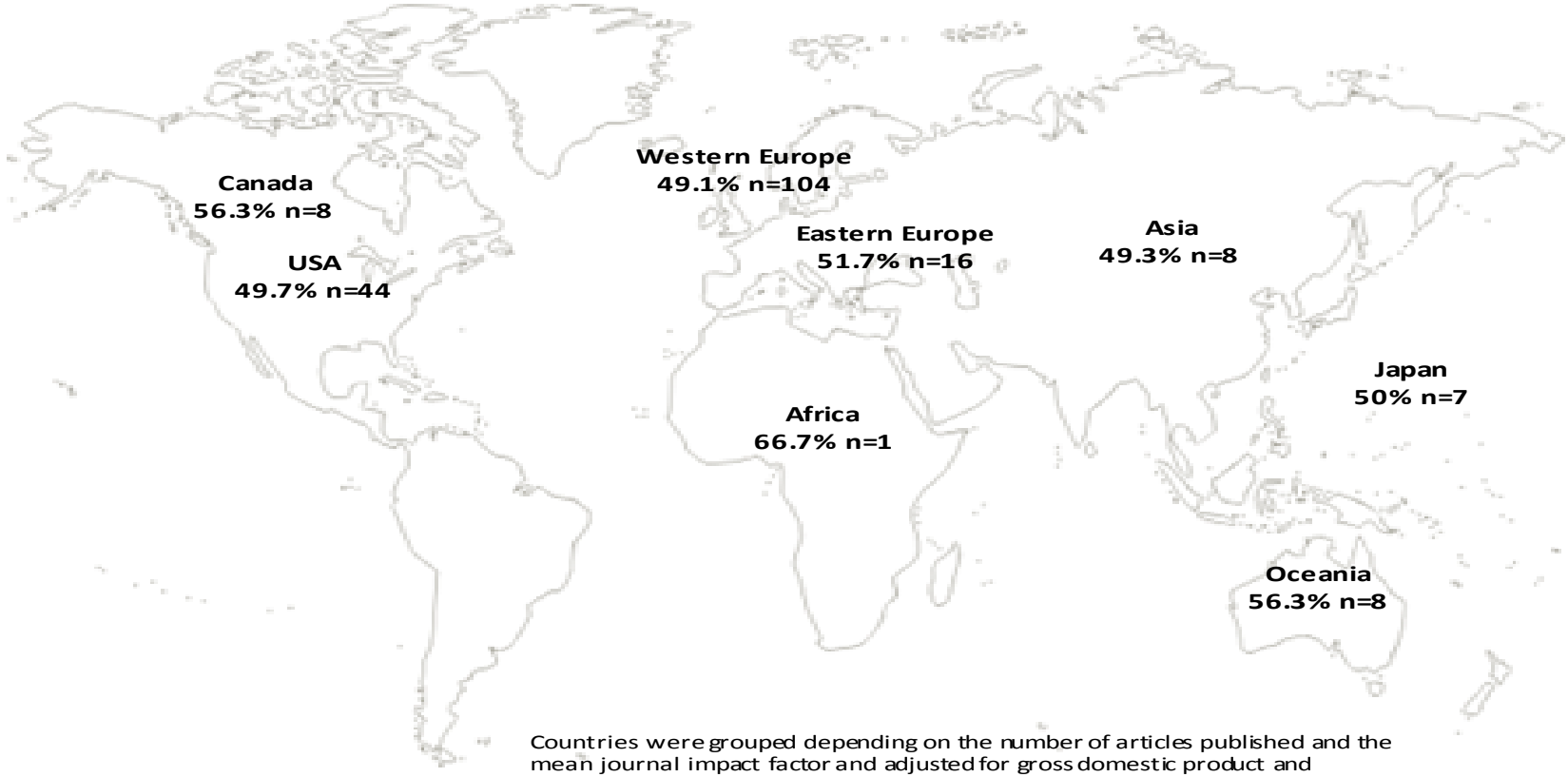


Figure 11.4: World map showing mean percentage compliance of studies with STARD according to geographical area of publication.



11.5 Discussion

The reporting of included studies in this review overall was poor. The geographical origin had no effect on the reporting quality; however the study size showed a positive correlation. There has been a trend in improvement in reporting quality. This may be due to initiatives such as the STARD checklist as well as historical progress in awareness among authors of the need to accurately report studies. There is however still significant room for improvement.

There was poor compliance with STARD in many of the studies in this review, in many studies it was unclear whether the study complied with the reporting item. This lack of clarity could potentially affect the inferences, but in other fields it is well known that unclear reporting is associated with bias[131]. The studies were limited to a subset of conditions within these fields. It is likely that these results can be translated across obstetrics, however care should be taken as to the generalisability of this study.

Poor reporting of a study does not necessarily correlate with bad quality. This is evaluated in chapter 12. Accurate reporting is necessary to allow transparency of a study and to ensure the results are interpreted correctly. The application of the STARD checklist may help prevent the implementation of unnecessary or inaccurate tests which can lead to unnecessary financial expenditure and potentially serious consequences for patients.

11.6 Conclusion

The reporting quality of papers in Obstetrics is improving. This may be due to initiatives such as the STARD checklist as well as historical progress in awareness among authors to accurately report studies. There is however considerable scope for further improvement.

CHAPTER 12: METHODOLOGICAL QUALITY OF TEST ACCURACY STUDIES INCLUDED IN SYSTEMATIC REVIEWS IN OBSTETRICS REVIEWED IN THIS THESIS: SOURCES OF BIAS.

12.1 Abstract

12.1.1 Background

In obstetrics there has been a rapid growth in the development of new tests with research on these presented as test accuracy studies. It is important that the methodology of these studies is such that the potential for bias is minimised. The purpose of this study is to determine the methodological quality of test accuracy studies in obstetrics using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) checklist and to assess sources of bias.

12.1.2 Methods

The included studies of seven systematic reviews, performed as part of this thesis were assessed for methodological quality using the QUADAS checklist. The compliance with each one of the QUADAS criteria was assessed. Using appropriate statistical tests it was investigated whether there was an improvement in study quality since the introduction of the QUADAS checklist, whether a correlation existed between study sample size, country of origin of study and its quality. This study also investigated whether there was

a correlation between reporting and methodological quality and by the use of meta-regression analyses explored for items of quality that were associated with bias.

12.1.3 Results

A total of 195 studies were included. The overall quality of included studies was poor (>50% compliance with 57.1% of quality items). However, the mean compliance with QUADAS showed an improvement post-publication of QUADAS checklist (55.5% versus 59.2%), this did not however reach statistical significance (p=0.1). There was no correlation between study sample size and methodological quality. There was no association with country of origin and methodological quality. Meta-regression analysis showed that no individual quality item had a significant impact on accuracy. There was an association between reporting and methodological quality ($r=0.51$ $p<0.0001$).

12.1.4 Conclusions

A combination of poor methodological quality and poor reporting affects the inferences that can be drawn for test accuracy studies. Further compliance with quality checklists is required to ensure that bias is minimised.

12.1.5 Publications arising from this work

Morris RK, Selman TJ, Zamora J, Khan KS. Methodological quality of test accuracy studies included in systematic reviews in obstetrics and gynaecology: sources of bias. *BMC Women's Health 2010* (In press).

12.2 Introduction

As discussed in chapter 11 obstetrics has seen rapid growth in the development of new tests[3]. For instance, tests designed to detect small for gestational age fetuses have grown in recent years[128;129]. The reporting of the study should allow for the detection of any biases by providing a complete and transparent description of the study participants, methodology and results. Guidelines for the reporting of other study types have widely been accepted e.g. CONSORT[122] for randomised control trials and QUOROM[132] and MOOSE[42] for systematic reviews. When studies of this type are incorporated in systematic reviews, assessment of their methodological quality is necessary to ensure that potential bias is identified and errors in judgement avoided. This allows methodological flaws that can lead to bias and sources of variation that might lead to heterogeneity to be identified. An evidence based methodological quality assessment tool has been developed for such assessments called Quality Assessment of Diagnostic Accuracy Studies (QUADAS)[35]. The need for quality appraisal of included studies in systematic review has been recognised for many years, however how deficiencies in study quality should be addressed in meta-analysis is not as clear[48;133].

The QUADAS initiative provides an assessment tool for the quality of test accuracy studies required when using these in systematic reviews. It combines empirical evidence and expert opinion into a checklist of 14 quality items. As these quality items should be adhered to and then reported in a study, they are directly and indirectly duplicated in the STARD checklist. Chapter 11 assesses the standard of reporting quality in obstetrics using the STARD checklist. Although gaps in reporting of quality item themselves do not necessarily mean that the methodological quality is poor,

interpretation is made difficult. The use of one standard checklist for assessment of study quality in all diagnostic reviews should allow clinicians to make comparable assessment of different studies. Where previous studies have attempted to assess methodological or reporting quality of test accuracy studies, a strong relationship has been found between various quality items and test accuracy results[134]. This study aims to assess the impact of the QUADAS initiative on test accuracy studies in antenatal screening.

12.3 Methods

A prospective protocol was developed to assess the impact of QUADAS on seven systematic reviews performed over the period 2005-2007 as part of this thesis. The included reviews were reviews of Down's syndrome serum screening markers and uterine artery Doppler to predict small for gestational age fetuses in obstetrics[128;129]. These reviews were chosen as they had all been performed by the author who had received training in use of the QUADAS checklist and had reached a consensus *a priori* as to how compliance with the checklist should be assessed for each review. The following questions were addressed: What is the quality of studies in this field? Did the introduction of QUADAS improve quality? Does study size correlate with quality? Is there a geographical pattern to quality? Is there a relationship between compliance with STARD and QUADAS? Which quality items are associated with bias?

The QUADAS checklist was applied to each of the studies included in all the reviews with the reporting item being determined as either present, absent, unclear or not applicable (appendix 4). All studies were assessed in duplicate by the author and by a

second reviewer (Dr Tara Selman) who had received the same training and been involved in the discussions to determine how compliance would be assessed. Where there was disagreement this was resolved by consensus with a third reviewer (Professor Khalid Khan). Results of individual studies were summarized in two by two tables from which the DOR was calculated as a measure of diagnostic accuracy[5]. DOR is the odds of a positive result in a diseased person relative to the odds of a positive result in a non diseased person. In the case of zero entities in the two by two tables 0.5 was added to the cells to enable calculation of DOR[56]. In the event that several tests had been applied to the same patient, the results including the largest number of patients were used in this study or where there was no difference, one index test was selected at random, this ensured patients were only included once.

The percentage compliance of studies with QUADAS items was determined before and after the introduction of QUADAS using the unpaired t test to assess the effect of QUADAS on the methodological quality of studies. With the publication of QUADAS in 2003 the assumption was made that all studies published pre 2005 were published without the benefit of this directorate.

The relationship between sample size and compliance was assessed with QUADAS using Spearman's rank correlation coefficient (Rho). Kruskal Wallis was used to investigate any relationship between geographical distribution and reporting quality. The country of origin of a study was determined by the country of the corresponding author. Where a significant result was found, pairways comparison was made using Conover Inman. Countries were grouped depending on the number of articles published

and the mean journal impact factor and adjusted for gross domestic product and population, based on previous publication[130]. Where there was a large disparity in number of studies per geographical area, some studies were re grouped to avoid large differences in group size and potentially spurious results. The geographical areas used were Oceania, USA, Canada, Asia, Japan, Africa, Eastern Europe and Western.

If the standard of reporting of a study is poor then this can potentially limit the assessment of the quality of study design. To investigate the relationship between reporting and methodological quality, the studies' compliance with STARD and QUADAS was compared using Spearman correlation coefficient.

The final analysis performed was a meta-regression analysis to assess which quality items were associated with bias. Multiple logistic regression models were adjusted to test the effect of individual QUADAS quality items on diagnostic accuracy, measured as DOR. This methodology[135] has been used successfully in demonstrating empirically the effect of bias related to methodological flaws in clinical trials[131;136;137] and in diagnostic studies[75]. The dependent variable in each logistic model was a binary variable representing disease status (diseased verses non diseased) from each meta-analysis. The independent variables included a variable representing test threshold (i.e. the sum of logits of sensitivity and 1-specificity) ; a binary variable for test result (positive verses negative); indicator variables to control for the effect of the primary studies; terms for the "meta-analysis by test result" interaction to control for the different summary diagnostic odds ratios estimating diagnostic accuracy in the included meta-analysis; and the "QUADAS item (dichotomized as Yes verses all other) by test

result" interaction terms to analyze its association with estimates of diagnostic accuracy. The estimated effect of quality characteristic on average diagnostic accuracy is given by the coefficient of the QUADAS test result interaction, which estimates the log of the ratio of diagnostic odds ratios (RDOR) in studies with and without the quality item. Exponentiation of the coefficient yields the RDOR. RDOR greater than 1 are interpreted as follows: those studies which fulfil the quality item overestimate test accuracy compared to those studies not fulfilling that item. RDOR lower than 1 means that those studies without the methodological quality characteristic overestimate diagnostic test performance. This effect is assumed to be constant across meta-analyses. Only meta-analyses that contained studies with and without the characteristic could contribute to this estimate. The RDOR was used as the summary measure of accuracy and dependant variable in the analyses as it is useful as a single indicator of test performance.

In the initial analysis those quality items coded as unclear and not applicable were excluded. For all of the above analysis, due to the uncertainty of whether reporting items coded as unclear represented methodological failure, sensitivity analysis was performed excluding this code and adding it to the not reported group for all comparisons. Similarly sensitivity analysis was also performed to assess the effect of those items assessed as not applicable, with their initially exclusion to the analysis and then addition as if they were reported so as not to penalise studies which had a larger number of not applicable items and would therefore potentially have a seemingly lower compliance with QUADAS.

12.4 Results

A total 195 studies were identified and included in this study (figure 11.1). 85.6% (167/195) of the studies were published prior to the QUADAS initiative. The overall percentage compliance with individual quality items is shown in figure 12.1. The included studies complied adequately >50% of the time for 57.1% (8/14) of the items assessed. Items where quality was uniformly poor (<50%) were an adequate description of the performance of the reference standard, reporting whether the reference test results were interpreted blind to the index test results and whether clinical data were available at the time of test interpretation.

There was an improvement in the mean compliance with quality items after publication of the QUADAS checklist (55.5% versus 59.2%), this did not however which reach statistical significance (p=0.1). Analysis of the correlation between sample size and QUADAS revealed no correlation (Rho=0.14, p=0.06). For these analyses sensitivity analysis as described in the methods section showed no significant difference.

The mean compliance with QUADAS according to country of publication of study is shown in figure 12.2. Investigation in to the relationship between geographical area of publication with QUADAS showed no association between compliance and area (p =0.73). In the meta-regression analysis only QUADAS item 3 (appropriate reference standard) had a marginal impact on diagnostic accuracy (p=0.05) with studies in which an inappropriate reference standards were used overestimated the diagnostic accuracy by 10%. The results are illustrated in figure 12.3.

All included papers were assessed for reporting standard and overall this was poor. The included studies reported adequately >50% of the time for 62.1% (18/ 29) of the items as assessed in this review. There was significant correlation between the percentage compliance of studies with STARD and QUADAS checklists (Rho=0.51, p=<0.0001) which is illustrated in figure 12.4. This figure shows that when studies had a higher standard of reporting it did not necessarily equate to improved quality of methods.

Fig 12.1: Bar chart showing percentage compliance with individual QUADAS criteria for included test accuracy studies in obstetrics.

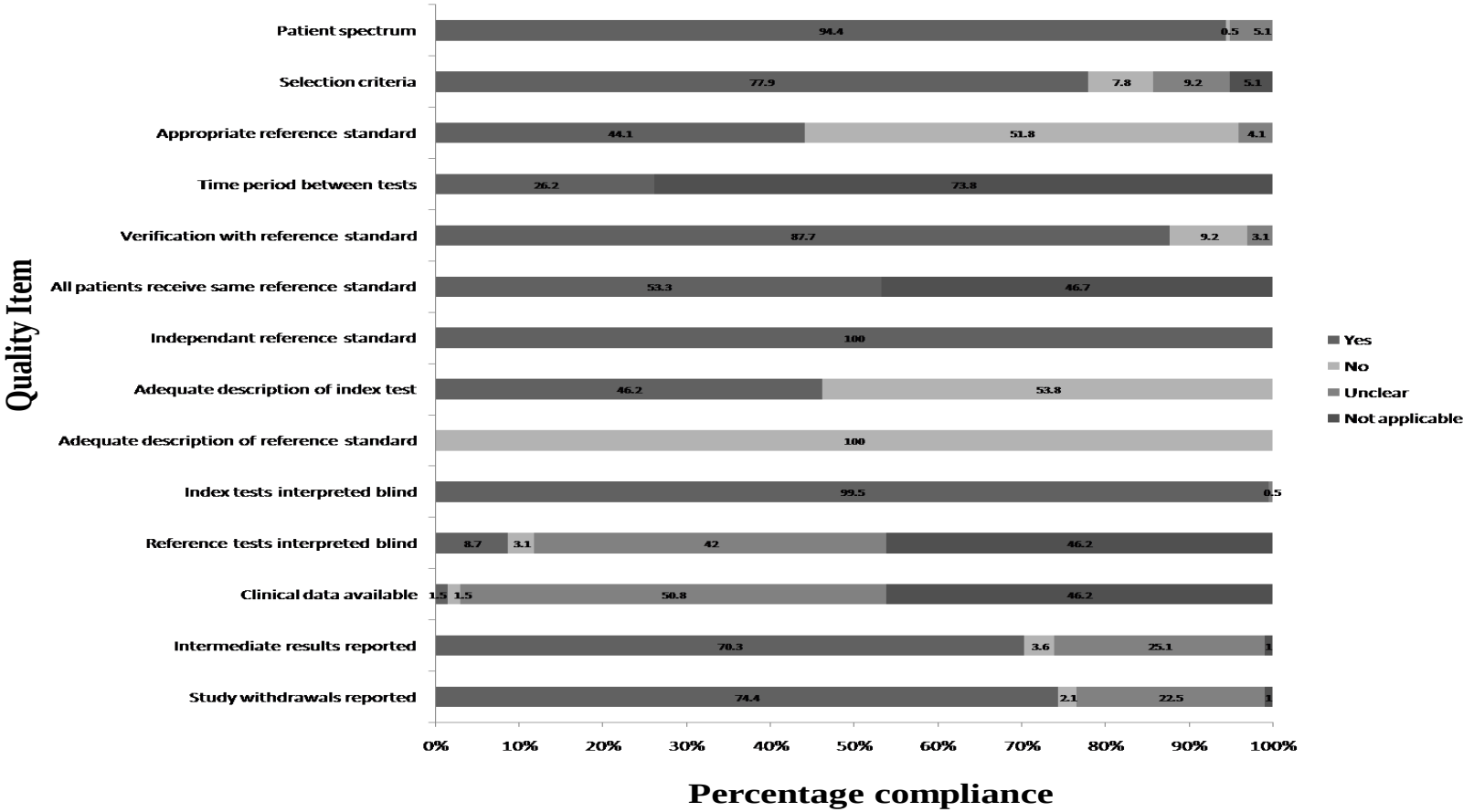
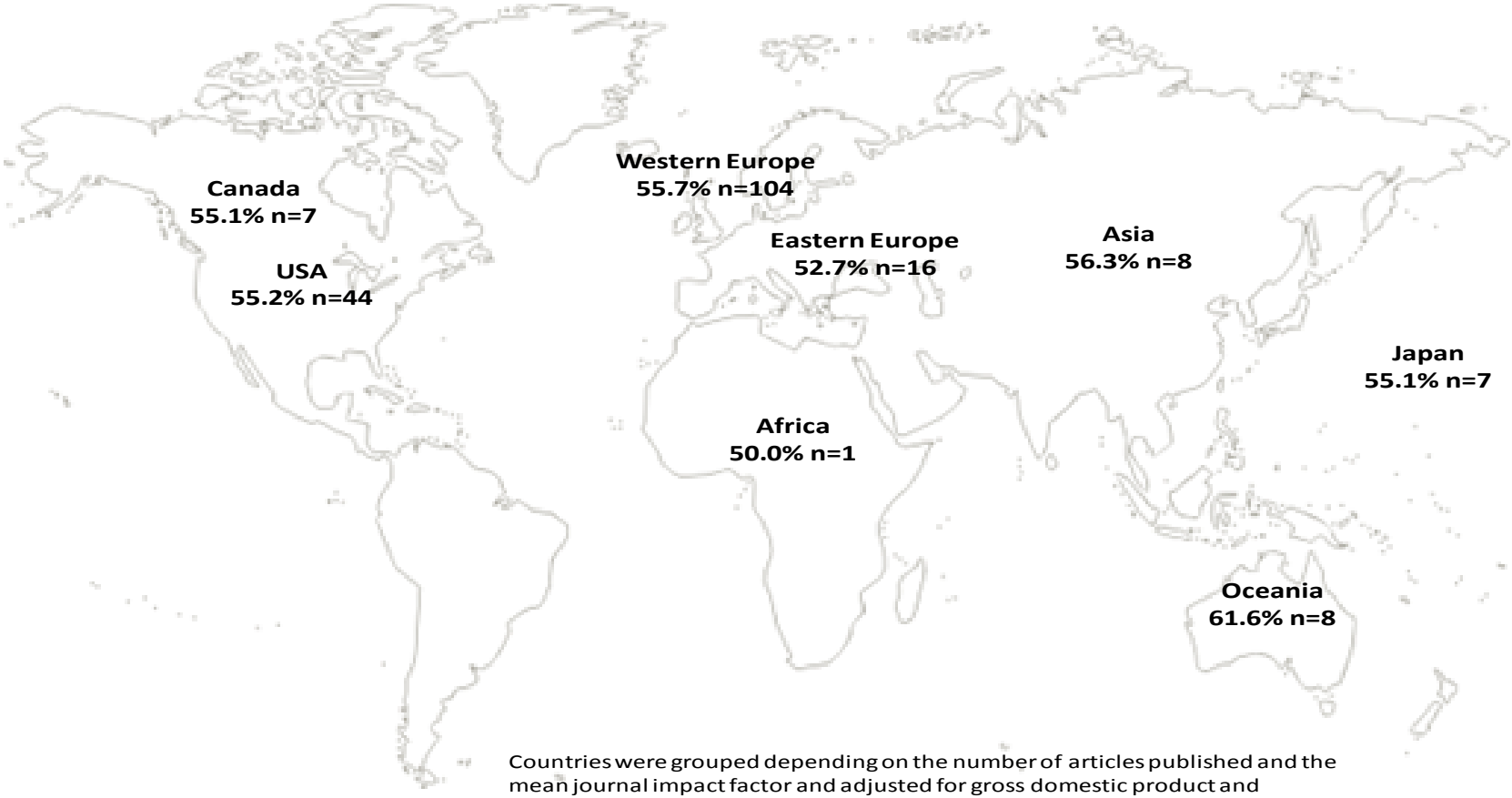
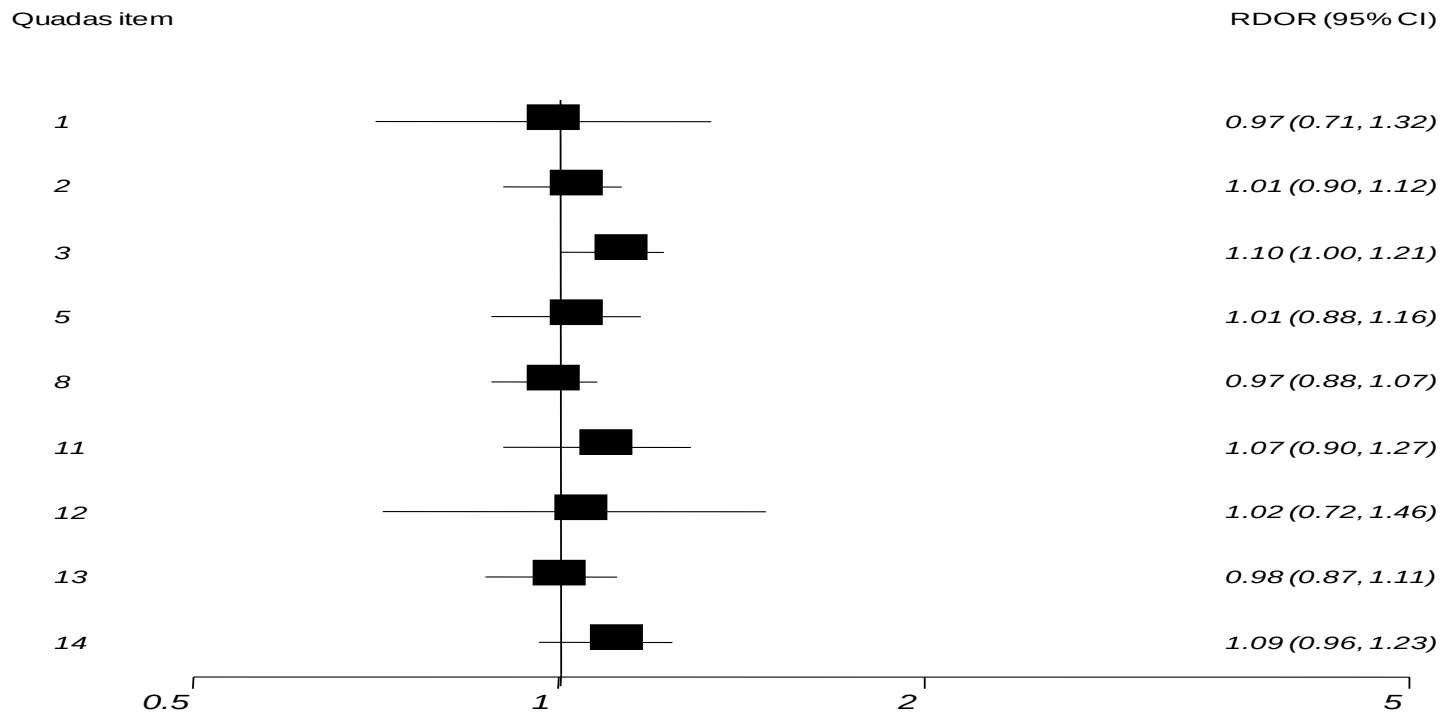


Figure 12.2: World map showing mean percentage compliance of studies with QUADAS according to geographical area of publication.



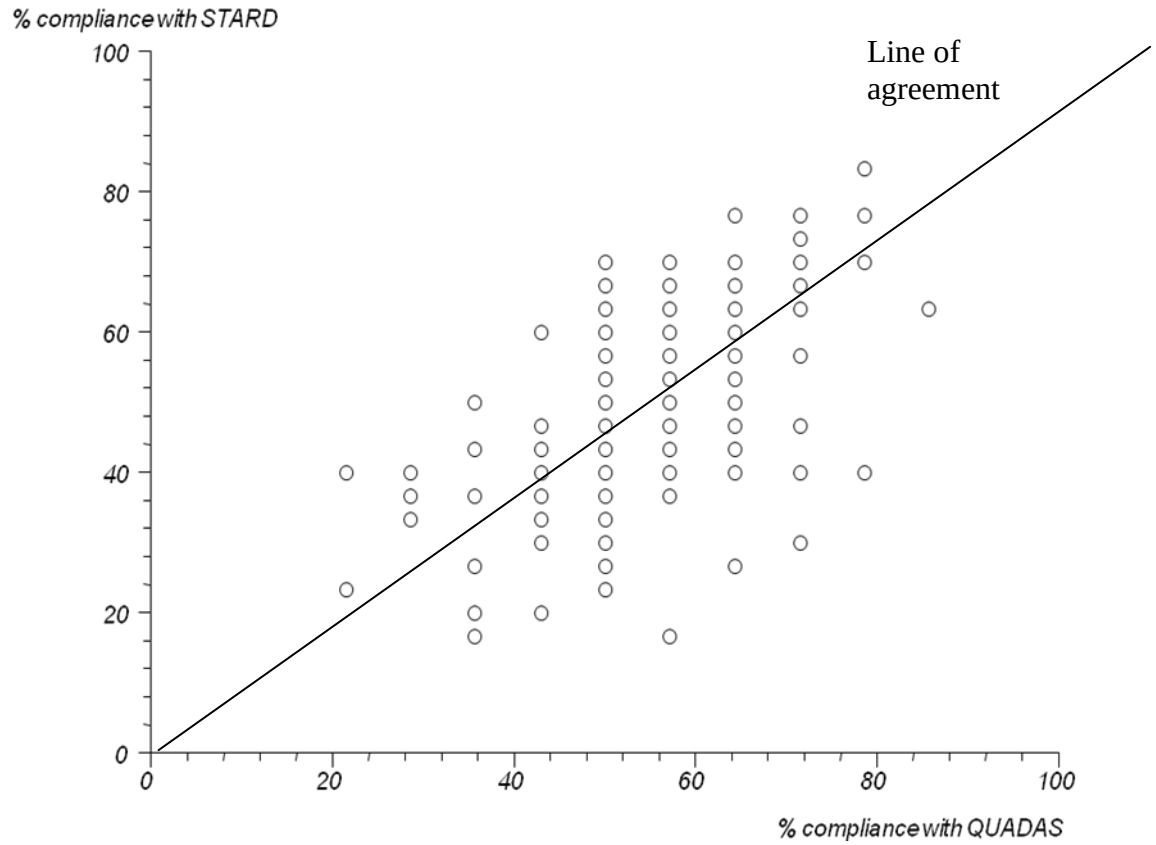
Countries were grouped depending on the number of articles published and the mean journal impact factor and adjusted for gross domestic product and population. Percentage compliance for obstetrics, n=number of studies

Figure 12.3: Effect of compliance with QUADAS quality item on the ratio of the diagnostic odds ratio in studies of test accuracy in obstetrics.



Obstetrics: Q7 and 8 constant compliance; Q4, 6, 10 colinearity
(Colinearity = mathematical problem which impedes adequate estimation of coefficients in model due to two or more variables that are correlated)
RDOR ratio of diagnostic odds ratio
CI confidence interval

Figure 12.4: Scatter plot showing the level of agreement between the percentage compliance of test accuracy studies in obstetrics with the STARD and QUADAS check lists.



12.5 Discussion

This study showed that there was not statistically significant improvement in the methodological quality of test accuracy studies in obstetrics with the introduction of QUADAS. Unsurprisingly, due to the overlap in quality items between the two checklists there was a positive correlation between compliance with STARD and QUADAS. Sample size showed no correlation with compliance. No correlation with

geographical area was seen. Meta regression did not show any significant correlation between compliance with QUADAS item and test accuracy.

The strengths of this study lie in the large number of included studies and meta-analyses, the continuity in assessment using the same two reviewers throughout and the use of tailored checklists to take into account the differences in studies in studies (e.g. the use of the not applicable category). Limitations to this study include the small proportion of included studies that were reported after publication of the QUADAS tool and the overall poor reporting standard of the included papers. As true assessment of a study's methodological quality relies on good reporting thus it must be concluded that the poor methodological quality of the papers in this review may actually reflect a combination of poor study design as well as poor reporting. The investigation into the effect of individual items of study quality on diagnostic accuracy could find no significant relationship between any individual quality item and accuracy. Although it was demonstrated that there was an improvement in methodological quality since the introduction of QUADAS it cannot be concluded that this improvement is due to the QUADAS initiative or due to other factors such as a historical progression in improved methodological techniques.

Recommendations are that all future test accuracy studies adhere to the QUADAS guidelines and that when studies are being included in systematic reviews, reviewers must assess for reporting and methodological quality using the QUADAS items that are relevant to their study area and consider additional items where necessary. As adherence to QUADAS becomes more widespread, the effect of items of methodological quality

on diagnostic accuracy should be reassessed to enable clinicians to interpret the validity and generalisability of results. This type of research will also help to improve test accuracy study design.

12.6 Conclusion

A combination of poor methodological quality and poor reporting affects the inferences that can be drawn from test accuracy studies. Further compliance with quality checklists is required to ensure that bias is minimised.

PART C: REVIEW of SYSTEMATIC REVIEWS OF THE EVIDENCE ON EFFECTIVENESS OF AVAILABLE INTERVENTIONS FOR PREVENTION OF SMALL FOR GESTATIONAL AGE AND COMPROMISE OF FETAL/NEONATAL WELLBEING

CHAPTER 13: REVIEW OF SYSTEMATIC REVIEWS OF THE EXISTING EVIDENCE ON THE EFFECTIVENESS OF AVAILABLE INTERVENTIONS FOR PREVENTION OF SMALL FOR GESTATIONAL AGE FETUSES AND COMPROMISE OF FETAL/NEONATAL WELLBEING.

13.1 Abstract

13.1.1 Background

Previous narrative reviews of effectiveness of interventions for fetal growth restriction have concluded that there are few interventions that are likely to be beneficial and that further high quality research is required. A review of systematic reviews of effectiveness for interventions for fetal growth restriction and compromise of fetal wellbeing was performed to summarise the most up to date evidence and assess the quality of the evidence.

13.1.2 Methods

Electronic searches of the following databases were performed: Medline, Embase, Cochrane Library, DARE (inception to July 2009), hand searching of journal and reference lists, contact with experts. Two reviewers independently selected articles which were systematic reviews of randomised controlled trials reporting on the effectiveness of interventions for prevention of fetal growth restriction and/or compromise of fetal wellbeing. Quality assessment for methodological quality and

reporting quality were assessed for the included reviews. There were no language restrictions applied. Data were extracted on study characteristics, quality and results to construct 2x2 tables. Summary data were presented as relative risks and their 95% confidence intervals for the different interventions.

13.1.3 Results

There were 71 systematic reviews included of which 69 were systematic reviews and one health technology assessment on screening and prevention of pre-term birth which included updated data for two Cochrane reviews. There were 4 non Cochrane systematic reviews. These articles included a total of 733 randomised controlled trials reporting on 42 different interventions. Forty –four reviews included outcomes for fetal growth reporting on 30 different interventions. Sixty one reviews reported on outcomes for adverse perinatal outcome using a total of 15 different outcome measures. For perinatal mortality there were a total of 49 reviews reporting on the effectiveness of 30 different interventions.

13.1.4 Conclusion

After considering the results and the quality of evidence antiplatelets and multiple micronutrient supplements were the interventions that were considered to be effective in preventing the small for gestational age fetus and suitable for use in all pregnant women. For high risk pregnant women the following were considered to be effective: antiplatelets, multiple micronutrient supplements, smoking cessation interventions and progesterone therapy. For prevention/reduction of perinatal mortality antiplatelets and antenatal corticosteroids were the interventions shown to be effective.

13.2 Introduction

Pregnancy and the period of intrauterine growth are a critical and vulnerable time in the life cycle of an individual. Fetuses that are born small for gestational age of low birth weight are known to be at increased risk of intrauterine death, peripartum asphyxia, neonatal morbidity and mortality and even in later life are at increased risk of adult diseases[8;11;12;138]. Low birth weight may be born preterm or at term and within the low birth weight category are those that are appropriately grown or "constitutionally small" and that have failed to achieve their growth potential - the "fetal growth restriction" . When associated with prematurity the outcomes for the growth restricted category are worse[139].

Fetal growth restriction has a broad aetiology and may be classified as being due to fetal, placental or maternal causes. Fetal causes include chromosomal and structural anomalies, inborn errors of metabolism and congenital infections. Maternal factors are those that affect placental transfer e.g. low pre-pregnancy weight, under nutrition, substance abuse, severe anaemia. There are also maternal medical conditions that affect placental implantation and vasculature and hence transfer e.g. pre-eclampsia, autoimmune disease, thrombophilias, renal disease, diabetes and essential hypertension. Of all these factors pre-eclampsia is associated with the most severe impact on fetal growth[140].

The potential for a therapy to be effective thus depends on the nature of the underlying aetiology. A thorough assessment of the fetus and mother must be performed to identify those causes that will not be amenable to therapy i.e. chromosomal anomalies and in

those cases where therapy might be an option determine which may be the most appropriate. The major difficulties in this area are however the lack of accurate predictive and diagnostic tests for the growth restricted fetus and the potential for there to be more than one contributory aetiology e.g. pre-term labour and fetal growth restriction or fetal growth restriction and pre-eclampsia.

Previous narrative reviews of effectiveness of interventions for fetal growth restriction have concluded that there are few interventions that are likely to be beneficial and that further high quality research is required[39]. The same conclusion was reached in the RCOG guideline published in 2002[21]. Since this time there has been further research in this area including the updating of many of the reviews of the Cochrane Pregnancy and Childbirth group in 2009[40]. When evidence is spread across many databases and in the case of fetal growth restriction evidence may be related to other conditions such as pre-eclampsia it can be difficult to access appropriate up to date robust evidence for clinical decision making. Systematic reviews provide a technique to allow individual pieces of research to be collected and if appropriate subjected to meta-analysis[41]. It is essential that these reviews are performed with rigorous methods and include an assessment of study quality if they are to have valid inferences and produce usable summaries to guide medical practice[41]. A review of systematic reviews of effectiveness for interventions for fetal growth restriction and compromise of fetal wellbeing was performed to summarise the most up to date evidence and assess the quality of the evidence.

13.3. Methods

The review was based on a prospective protocol designed following recommendations from the NHS Centre for Reviews and Dissemination[141] and the Cochrane collaboration[41].

13.3.1 Framing the question

A clearly defined question was based on the PICOS criteria:

Population: Pregnant women in any health care setting, at any level of risk. Populations that only included multiple pregnancies were excluded.

Intervention: Any intervention or combination of interventions, applied at any gestation to pregnant women to improve fetal growth or fetal wellbeing. Due to the underlying aetiology and pathophysiology of fetal growth restriction, interventions that were applied to pregnant women for the purposes of reducing their risk of preterm labour or pre-eclampsia were also included.

Comparator: No intervention or placebo or usual care.

Outcomes: Any outcome determined for the baby after birth looking at growth or wellbeing. Percentage deviation from estimated fetal weight at a given gestational age, mean birth weight and birth weight z scores were excluded as these outcomes could not be related to test accuracy data.

Study design: Systematic reviews of randomised controlled trials. Reviews had to be based on a clearly formulated question and use systematic and explicit methods to identify, select and critically appraise the relevant primary research, and to extract and analyse data to be included. Meta-analysis was not a pre-requisite to inclusion.

13.3.2 Identifying the literature

The search was designed using a structured approach with the aim of identifying literature concerning interventions for FGR and compromise of fetal wellbeing using the elements of the framed question (section 13.3.1). Pilot searches were performed to ensure that the search strategies gave an acceptable level of specificity without compromising sensitivity. The search strategy is detailed in appendix 42.

Literature was identified via the following sources:

(a) General bibliographic databases including MEDLINE and EMBASE; from inception to July 2009.

(b) Specialist computer databases – DARE (Database of Abstracts of Reviews of Effectiveness), the Cochrane Library (issue 2009:3) and relevant specialist registers of the Cochrane Collaboration, namely the Pregnancy and Child Birth Group, Health Technology Assessment (HTA) database.

(c) Contact with individual experts and those with an interest in this filed to uncover grey literature and identify any reviews in progress.

(d) Hand- searching of relevant specialist journals in Obstetrics

(e) Checking of reference lists of included articles and narrative review articles

(f) SCISEARCH and Web of Science to identify frequently cited articles and conference abstracts.

All searches were made without language restrictions. A comprehensive database of articles was constructed using Reference Manager 11.0 software.

The titles and abstracts of the citations were assessed for inclusion by two reviewers independently. Potentially relevant citations were obtained and the paper read in full by the two reviewers. Reviews were selected for inclusion by the two reviewers independently using a checklist, the items of this checklist were based on the question elements as stated in section 13.3.1. Disagreements were resolved by consensus with a third reviewer. In the case of duplicate publications or reviews on the same subject published by different groups then a decision was made by the three reviewers as to the most up to date review and/or the review of higher methodological quality for selection. All foreign language papers were translated.

13.3.3 Assessment of the quality of the literature

Included reviews were assessed for methodological quality using a checklist based on the CASP checklist (Critical Appraisal Skills Programme)[142]. The issues assessed when considering the quality of the review were:

1. Did the review ask a clearly structured and focused question?
2. Were selection criteria clearly described?
3. Were all relevant studies identified?
4. Were the included studies synthesized?
5. Was the validity of the included studies assessed?
6. Are sufficient details about the individual included studies presented?

The full checklist with further details on how quality was assessed is shown in appendix 43. Quality assessment was performed independently by two reviewers. Where there were two systematic reviews reporting the same data the review with the highest quality was included only.

13.3.4 Data Extraction

Data were extracted on number of studies included in the review, the methodological quality of the review and the summary results of the review. Data were then extracted from the reviews on the study characteristics, methodological quality and results of the individual included studies. Data were extracted in duplicate by another independent reviewer to ensure accuracy. Disagreements were resolved by consensus.

13.3.5 Description of data

The data extracted from the systematic reviews were presented as tables according to the outcome measure detailing the intervention, population characteristics, comparator and the number of relevant included RCTs. The relative risk (RR) along with 95% confidence intervals and the p value of the z statistic (overall measure of effect) were presented as the summary measure. To enable a meaningful presentation of the data and discussion of the results it was decided to present the data as two tables, one for outcome measures related to fetal growth and the other for perinatal mortality as the main outcome measure for compromise of wellbeing.

13.4 Results

13.4.1 Literature identification and review characteristics

Figure 13.1 summarises the process of literature identification and selection. The references for the included reviews and the details of the individual review characteristics can be found in appendix 44 and 45 respectively. There were 71 systematic reviews included of which 69 were systematic reviews and one health technology assessment on screening and prevention of pre-term birth which included

updated data from two Cochrane reviews. There were 4 non Cochrane systematic reviews. These articles included a total of 733 randomised controlled trials reporting on 42 different interventions. Forty –four reviews included outcomes for fetal growth reporting on 30 different interventions. Small for gestational age (with no threshold noted) was the commonest outcome reported (n=21 studies) followed by birth weight <2500g (n=20 studies). A total of 7 different outcome measures were used across the different reviews. The results for interventions for fetal growth are shown in appendix 46 and figure 13.3 and 13.4.

Sixty one reviews reported on outcomes for adverse perinatal outcome using a total of 15 different outcome measures. For perinatal mortality there were a total of 49 reviews reporting on the effectiveness of 30 different interventions. The results for perinatal mortality are shown in appendix 47 and figure 13.5.

13.4.2 Review Quality

Figure 13.2 summarises the quality of the included reviews. Overall the quality of reviews was good. Fifty-one reviews were assessed as not having asked a structured question based on PICOS (population, intervention, comparator, outcome, study design). In Cochrane reviews, instead of a question, an objective is stated. This was assessed for its compliance with PICOS and in many reviews there was no clear statement of the population or comparator. For a clear statement of selection criteria, 15 reviews were assessed as not being compliant as there was no clear statement of the comparator.

Figure 13.1: Process from initial search to final inclusion for review of systematic reviews of effectiveness for interventions for prevention of fetal growth restriction and compromise of fetal/neonatal wellbeing (up to July 2009). (HTA Health technology assessment; RCT randomised controlled trial).

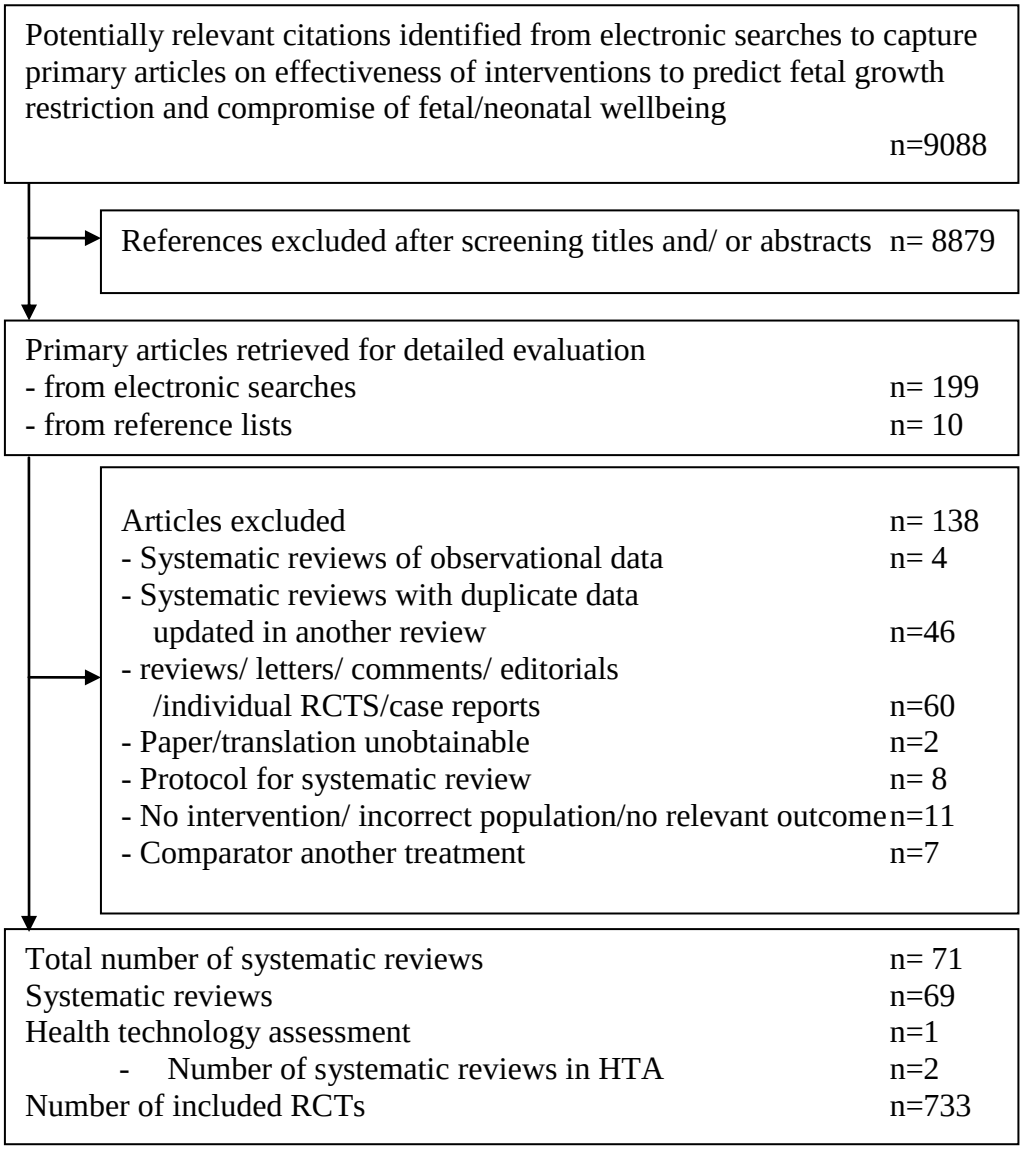
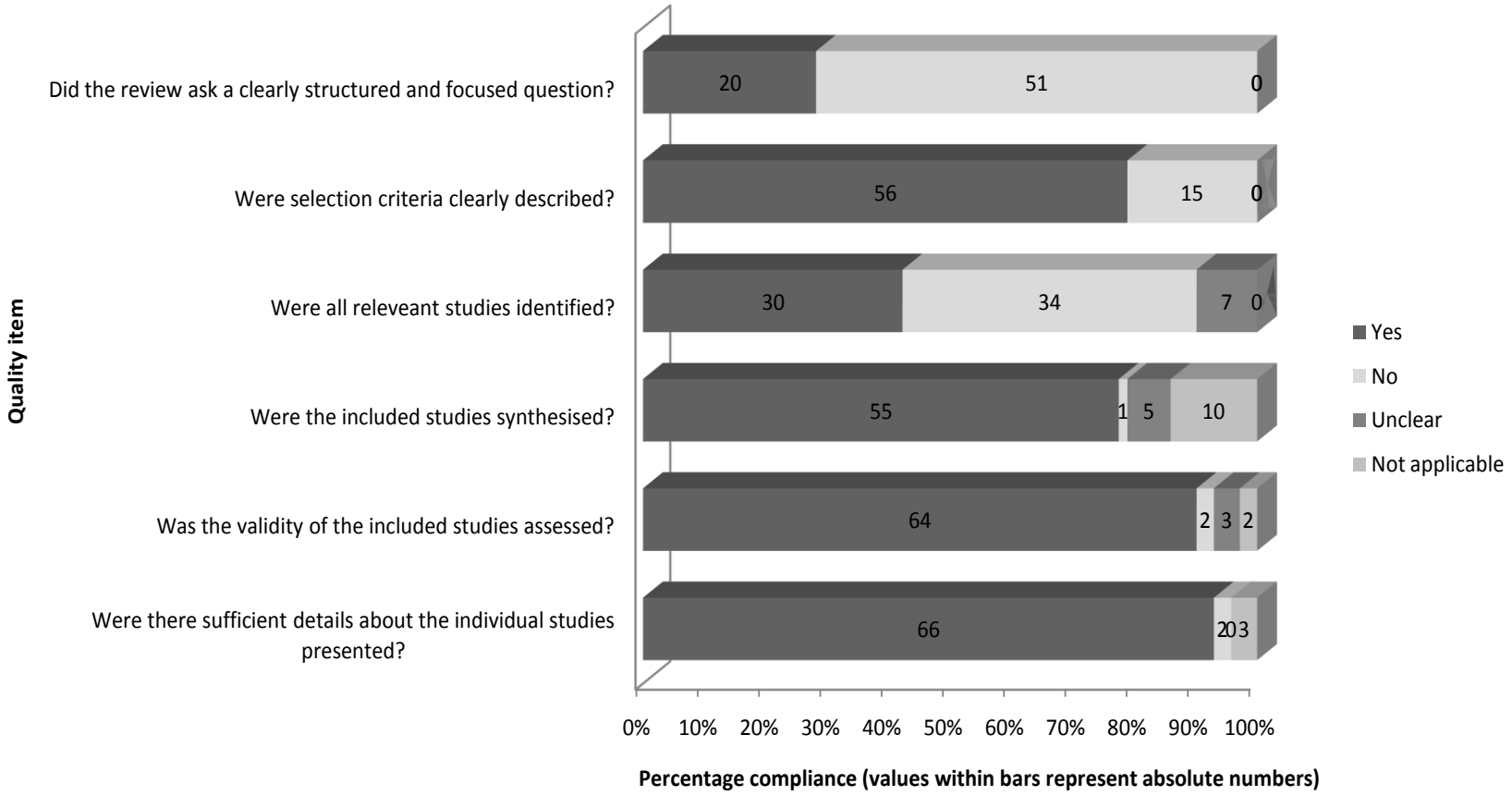


Figure 13.2: Bar chart showing quality assessment of included studies in review of systematic reviews of effectiveness of interventions for fetal growth restriction and compromise of its wellbeing



In 34 reviews it was decided that there was a possibility that not all studies may have been identified as there was no statement regarding additional methods to search grey literature and additional study information e.g. contact with authors and experts within the field. For assessment of appropriateness of data synthesis there were 10 reviews in which there were either no or only one included studies and thus this was assessed as not applicable. In five reviews it was not possible to fully assess the appropriateness of the pooling of data as there had been no statements made about planned sub-group or sensitivity analysis. In 64 of the reviews the validity of the included studies was deemed to have been assessed appropriately i.e. using an appropriate tool, two independent assessors, planned *a priori.* There were two reviews in which there were in-sufficient data presented relating to the individual included studies in the review, neither of these systematic reviews were Cochrane reviews.

13.4.3 Summary of results of effectiveness reviews for prevention of fetal growth restriction

Figure 13.3 summarises the results for effectiveness of interventions for prevention of fetal growth restriction in a general pregnant population and figure 13.4 for a high risk population. The number of included trials in each review ranged from one to 23, with the number if included participants ranging from 10 to 21426. For 11 interventions the reviews included only one RCT.

For a general pregnant population the following interventions had a RR and confidence intervals compatible with an overall positive effect:

- Antiplatelets

- Balanced protein/energy supplementation
- Orally administered magnesium
- Multiple micronutrient supplementation

Antiplatelet agents were extensively reviewed both by a Cochrane review[143] and an individual patient data meta-analysis[144] with the conclusion that they have moderate benefits for prevention of PE and its consequences, with further information required to assess which women are most likely to benefit, when treatment is best started, and at what dose. However, for balanced protein/energy supplementation the authors of the review were concerned regarding the heterogeneity within the included trials and concluded that it was impossible to know whether the benefit was only for women who were undernourished or not[145].

For orally administered magnesium the review authors concluded that the evidence was all of a poor quality and that after excluding one trial from the meta-analysis the effect of benefit was removed[146]. The review on multiple micronutrient supplementation identified nine trials involving 15378 women, there was evidence of an effect on reduction of low birth weight and small for gestational age infants. The review authors however recommended further research in particular to assess adverse effects[147].

Figure 13.3: Summary forest plots of relative risks of various interventions for prevention of fetal growth restriction in a general pregnant population. (Squares represent individual RCTs and diamonds pooled RCTs).

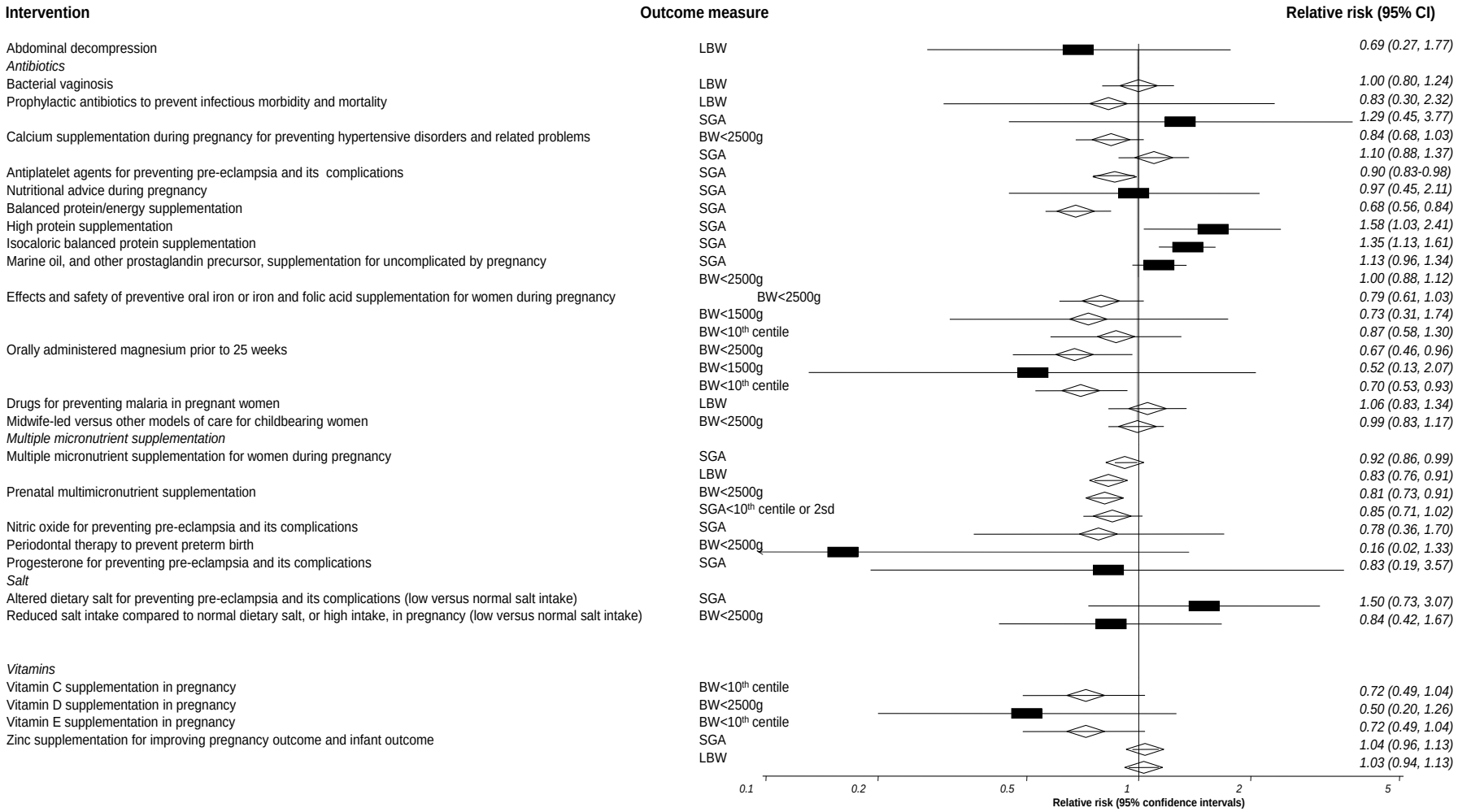
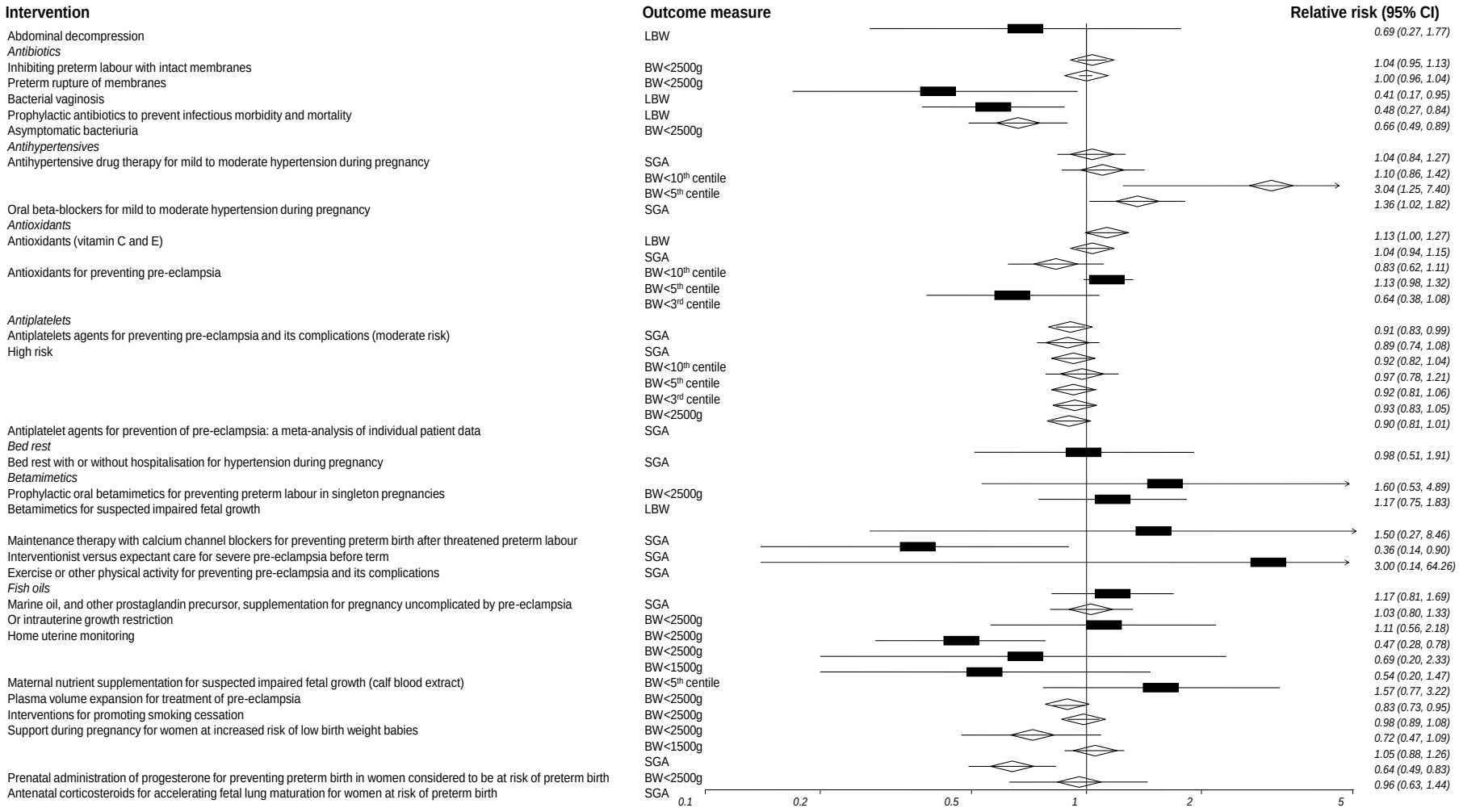


Figure 13.4: Summary forest plots of relative risks of various interventions for prevention of fetal growth restriction in a high risk pregnant population. (Squares represent individual RCTs and diamonds pooled RCTs)



There were two interventions that appeared to have an adverse effect on birth weight:

- Antihypertensives and in particular oral beta-blockers
- High protein and iso-caloric balanced protein supplementation.

For beta-blockers the increase in SGA fetuses was largely due to one small trial of atenolol versus placebo and the conclusion of the review authors was that further research was required[148]. For high protein and iso-caloric balanced protein supplementation the authors advised that there was no potential benefit but a potential for harm[145].

For a high risk population the following interventions had a RR compatible with an overall positive effect:

- Abdominal decompression
- Antibiotics for women with bacterial vaginosis
- Prophylatic antibiotics to prevent infectious morbidity and mortality for women with previous preterm labour
- Antibiotics for women with asymptomatic bacteriuria
- Antiplatelets for women at moderate or high risk of PE
- Interventionist versus expectant care for women with severe PE before term
- Home uterine monitoring for women at risk of preterm labour
- Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth
- Interventions for promoting smoking cessation in pregnancy

In the review on abdominal decompression the authors concluded that all studies had the potential for serious bias and that the therapeutic effect was not clear[149]. For

prophylactic antibiotics the effect was seen in one trial of 253 women who were high risk due to either a previous pre-term labour, low birth weight baby, stillbirth or neonatal death[150]. This trial had a high drop-out rate and thus the results must be interpreted with caution. For home uterine monitoring the authors concluded that the trials were of poor quality and thus the effect was not clear[151].

13.4.4 Summary of results of effectiveness reviews for prevention of perinatal mortality

Figure 13.5 parts A and B summarise the results for effectiveness of interventions for prevention of perinatal mortality. The number of included trials in each review varied from one to 23. The number of included participants varied from 31 to 30672. For 14 interventions the reviews included only one RCT.

The following interventions had an RR and confidence intervals suggesting an overall beneficial effect on perinatal mortality:

- Abdominal decompression in a high risk pregnancy
- Antiplatelets
- Elective caesarean section versus expectant management for delivery of a baby suspected to be small

Figure 13.5: Part A Summary forest plots of relative risks of various interventions for prevention of perinatal mortality (Squares represent individual RCTs and diamonds pooled RCTs)

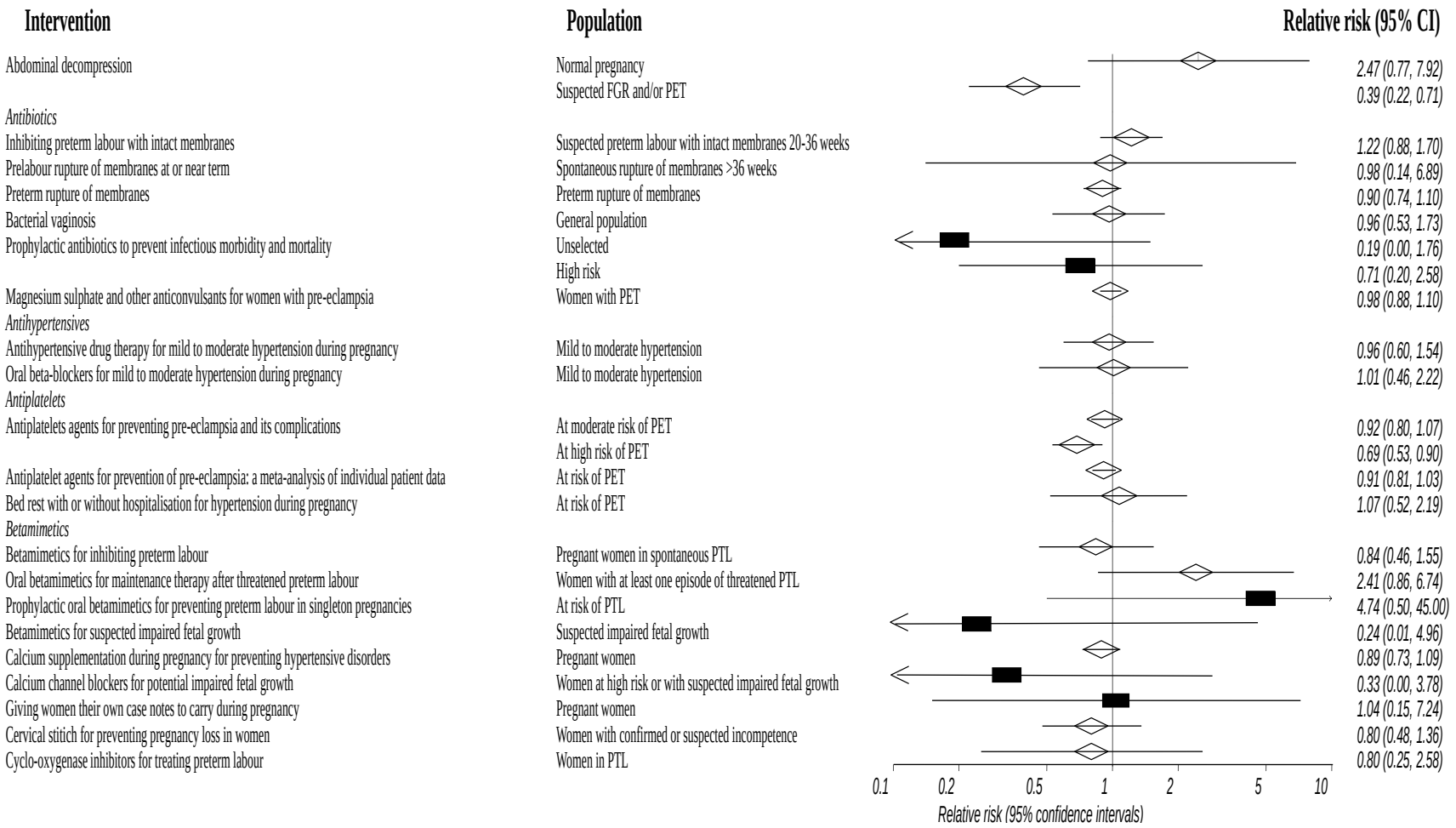
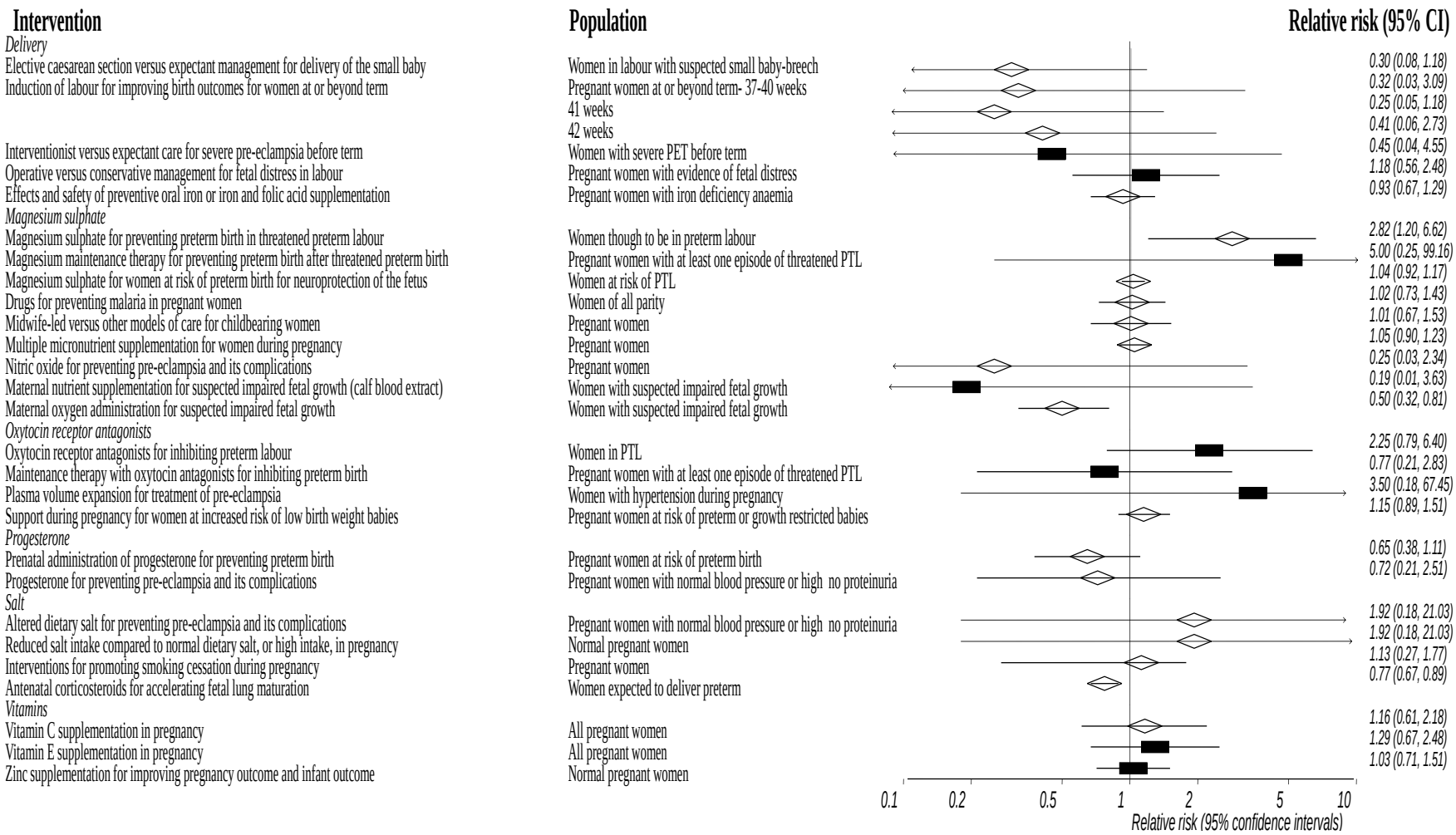


Figure 13.5: Part B Summary forest plots of relative risks of various interventions for prevention of perinatal mortality. (Squares represent individual RCTs and diamonds pooled RCTs).



- Maternal oxygen administration for suspected impaired fetal growth
- Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of pre-term birth.

The limitations of the review on abdominal decompression and the reviews for antiplatelets have already been discussed in section 13.4.3. For elective caesarean section for the suspected small baby the review authors concluded that a policy of elective caesarean section could not be recommended at this time due to the uncertainty around any beneficial effect due to the small numbers of women recruited to the trials[152]. The review on maternal oxygen administration concluded that there were concerns regarding the methods of the included trials, in particular selection bias, and that the trials all had small sample sizes, so further trials with multicentre recruitment were recommended[153].

The evidence from the review on antenatal corticosteroids included 21 studies (3885 women and 4269) and supported the use of single dose of corticosteroids to accelerate fetal lung maturation in women at risk of preterm birth[154].

The following interventions had an RR and confidence intervals suggesting a potential for harm:

- Magnesium sulphate administration for preventing preterm birth in threatened preterm labour.

The review on magnesium sulphate included 23 trials and over 2000 women and the authors concluded that magnesium sulphate was ineffective in delaying or preventing preterm birth, and its use is associated with an increased mortality for the infant[155].

13.5 Discussion

After considering the results and the quality of evidence antiplatelets and multiple micronutrient supplements were the interventions that were considered to be effective in preventing the small for gestational age fetus and suitable for use in all pregnant women. These were thus put forward to the decision analytic model for all pregnant women. For high risk pregnant women the following were put forward antiplatelets, multiple micronutrient supplements, smoking cessation interventions and progesterone therapy. Antibiotics for bacterial vaginosis and asymptomatic bacteriuria and interventionist care for severe PE were considered to represent interventions for very specific sub-groups of the high risk population and thus these interventions were not used in the model. For prevention/reduction of perinatal mortality, antiplatelets and antenatal corticosteroids were the interventions shown to be effective.

Limitations arising from problems with included trials

The assessment of quality of the included reviews was good overall. However, as some of the reviews included only a small number of RCTs this led to relative risks being reported that were based on small sample sizes thus increasing the error rate and decreasing the statistical power. The inferences regarding effectiveness that can be drawn from these reviews are thus limited. It must therefore be concluded that for these interventions there is a lack of evidence on which to determine effectiveness.

For all interventions there was an attempt to determine which pregnant populations (i.e. level of risk) the intervention was most effective or useful for to help determine which interventions should be put forward to the economic model. To determine level of risk the

authors of the included reviews relied upon the description of risk given in the primary RCTs. This description was often inadequate and varied between included trials and across reviews particularly for the high risk population. In order to assess the generalisability of the interventions under review there was careful assessment of the included populations of the individual RCTs and the conclusions of the authors of the systematic reviews was taken into account. There was also significant heterogeneity in the outcome measures used for assessing growth restriction and fetal/neonatal compromise both within included RCTs and across systematic reviews. Caution must thus be exercised when comparing the effectiveness of different interventions to ensure that outcome measures are comparable.

The majority of the included reviews had been updated within the last 5 years. There were however some reviews, particularly those assessing non-contemporary interventions, that, despite an up-date of the search, included non-contemporary data. This introduces concerns regarding the methods of the included RCTs and the applicability of the results to the current pregnant population within the context of modern clinical practice.

Limitations arising from review methods

The review methods employed had strengths as well as weaknesses. The quality of this review is limited by the quality of the included systematic reviews. As 94% of the included systematic reviews were Cochrane reviews they can be considered to have adhered to a rigorous set of methods including a comprehensive search strategy, primary studies restricted to RCTs, peer reviewed and published protocol and have been regularly updated. The fact that they are regularly updated led to the high proportion of

Cochrane reviews within this review. The majority of Cochrane effectiveness reviews within the Pregnancy and Childbirth group were updated in January 2009 thus although a large number of other systematic reviews were identified by the search strategy the vast majority were excluded as the Cochrane reviews were more up to date.

The methods used for this review also allowed a large number of interventions to be assessed as it included any review that assessed an intervention in a pregnant population with a relevant outcome measure for growth restriction and/or fetal/neonatal compromise. This meant that reviews primarily directed at the management of fetal growth restriction were identified and included but also reviews targeted at pre-eclampsia, pre-term labour as well as general antenatal care.

As this was a review of systematic reviews it is a comprehensive summary of the available evidence on interventions that have been systematically reviewed and evaluated by RCTs. It does not however summarise the literature on all available interventions as there are some interventions that may have been assessed by only observational data or not subjected to systematic review.

A weakness of the review methods was that the quality of the individual RCTs was not assessed. This was due to the large number of RCTs being prohibitive within the timescale available. In an attempt to ameliorate this opportunity for bias, the quality of the included trials within the reviews of an intervention determined to be effective was taken into account in decisions regarding whether an intervention should be put forward to the model. This was not however performed for those interventions with a relative

risk suggesting an ineffective intervention and thus there is the potential that the quality of included trials was such that the true effect may be different from that reported. It would be difficult however to assess the probability of this and the conclusion would be that the evidence was not sufficient to confidently determine true effectiveness.

This review only included results for comparison of an intervention against placebo or in a few cases against standard care for antenatal care and delivery interventions. This decision was made due to the large number of interventions under review and the large number of outcome measures used, with the aim of keeping the number of results and hence conclusions manageable and coherent. It is thus possible that some interventions that are effective have not been included in this review as they were compared to other interventions only.

The limitation of this review to systematic reviews of RCTs and exclusion of observational data means that important information that may only truly be revealed by observational studies such as side effect data and longer term outcomes/morbidity is not considered. It must be concluded therefore that this work and the resulting model look only at effectiveness and cost-effectiveness of the interventions and that their use in clinical practice should take into account the side effect profiles and long term outcomes.

Limitations arising from things not done

The decision to include all reviews and interventions with a relevant outcome meant that a considerable number of interventions were considered for this review and hence

for entry to the model. While this is a strength, as the likelihood of an effective intervention being omitted is very small, it does not address the issue of which interventions are important to clinicians and more importantly which are important and acceptable to parents. This could be addressed by a survey of clinicians and parents or consumer groups to identify what is important to them.

Findings in light of these limitations

This review represents the most comprehensive and up to date review of systematic reviews of effectiveness of interventions for fetal growth restriction and compromise of fetal/neonatal wellbeing. It presents good evidence for the effectiveness of a small number of interventions e.g. antioxidants, antiplatelets, antihypertensive, smoking cessation and progesterone. It has also determined that there is considerable uncertainty, either due to concerns regarding the reliability of the evidence or lack of RCTs, for the effectiveness of other interventions e.g. hormones, bed rest, delivery interventions, plasma volume expansion, abdominal decompression.

Considerations for the economic model

As discussed earlier in considering which interventions were to be put forward to the model the following were considered: effectiveness, reliability of evidence and generalisability to the study population.

Recommendations for practice

The 2002 RCOG guidelines recommended the use of antenatal steroids to reduce the incidence of respiratory distress syndrome and delivery in a unit where optimal neonatal

expertise and facilities are available[21]. They concluded that there was insufficient evidence for most interventions and that in the case of aspirin further trials were needed. The evidence in this review shows that antiplatelets and multiple micronutrients in all pregnant women and in addition smoking cessation programmes and progesterone therapy in high risk women reduce the risk of fetal growth restriction compared to placebo. It must be noted however, that of these, only antiplatelets and corticosteroids have been shown to have any effect on perinatal mortality. This information along with information on potential side effects should be discussed with women prior to implementation.

Recommendations and considerations for future research

All interventions discussed here and reported in the literature have looked at prevention of fetal growth restriction and not treatment as this is the ideal. Further research should look at the regulation of fetal growth to allow specific therapies for those cases where growth restriction has developed. Interventions have so far been considered in isolation and it may be that benefit from any single one intervention is always going to be small and that multiple interventions used in combination will be necessary to find a truly effective intervention. This will require careful consideration of side effect profiles and possible adverse interactions of interventions.

Fetal growth restriction and compromise of fetal wellbeing are strongly related to other pregnancy complications such as pre-eclampsia and pre-term labour and women will often have more than one of these complications. It is thus important that future research considers the impact of interventions on other important maternal and neonatal

outcome measures. Future research in this area also needs to be carefully designed with attention paid to population risk and outcome measures to ensure that results are reliable and generalisable.

A difficulty in future research can arise when new interventions are compared in different ways i.e. either randomised against placebo or against an active control. As the number of interventions available increases so does the number of comparisons required to truly assess the effectiveness of the intervention. This can be a considerable burden on resources and can be particularly difficult in the area of obstetric research where recruitment to RCTs can be hampered by the maternal concern regarding the unborn child. A contemporary meta-analytic technique called network meta-analysis[156] or mixed treatment comparisons[157] may need to be considered to allow a unified, coherent analysis of direct and indirect evidence.

13.6 Conclusion

After considering the results and the quality of evidence antiplatelets and multiple micronutrient supplements were the interventions that were considered to be effective in preventing the small for gestational age fetus and suitable for use in all pregnant women. For high risk pregnant women the following were considered to be effective: antiplatelets, multiple micronutrient supplements, smoking cessation interventions and progesterone therapy. For prevention/reduction of perinatal mortality antiplatelets and antenatal corticosteroids were the interventions shown to be effective.

PART D: COST EFFECTIVENESS ANALYSIS WITH ECONOMIC MODELLING

CHAPTER 14: COST EFFECTIVENESS ANALYSIS WITH ECONOMIC MODELLING TO ASSESS TEST AND TREATMENT STRATEGIES FOR THE MANAGEMENT OF SMALL FOR GESTATIONAL AGE FETUSES AND COMPROMISE OF FETAL/NEONATAL WELLBEING.

14.1 Abstract

14.1.1 Background

Identification of the fetus at risk of compromise is crucial to judicious allocation of monitoring resources and use of preventative treatment with the prospect of improving perinatal outcome. To investigate the potential cost-effectiveness of alternative 'test and treat' strategies in the prevention of fetal growth restriction compared to a strategy of no screening in the UK.

14.1.2 Methods

Economic evaluation using a decision tree model based on data from systematic reviews in a population of all pregnant women with sub-group analysis based on population risk. Setting of clinics, General Practices, Health Centres or any setting delivering antenatal care to pregnant women. The main outcome measure was cost-effectiveness based on an outcome of fetal growth restriction avoided.

14.1.3 Results

105 studies were reviewed on the accuracy of 6 different tests; Cochrane reviews and systematic reviews of effectiveness, 44 in total, were used for effectiveness data of a possible 30 interventions. Cost data were based on secondary evidence, supplemented with primary data from local sources. Testing prior to intervention was not shown to be the most cost-effective strategy in the analyses for all pregnant women. Anti-platelet therapy, without prior testing, was highlighted as potentially cost-effective in preventing fetal growth restriction in this population. In high risk women, testing with serum human chorionic gonadotrophin followed by anti-platelet therapy in those that test positive was a potentially cost-effective strategy.

14.1.4 Conclusion

An effective, affordable and safe intervention applied to all mothers without prior testing is likely to be the most cost-effective strategy in the prevention of fetal growth restriction. The results reported in this paper are important for prioritising future research, world-wide.

14.1.5 Publications arising from this work

Morris RK, Malin GL, Tsourapas A, Roberts TE, Khan KS. An economic evaluation of alternative test-intervention strategies to prevent fetal growth restriction in singleton pregnancies. Archive of Disease in Childhood Fetal and Neonatal 2010;95(suppl 1): Fa12.

14.2 Introduction

Restriction of fetal growth and compromise of fetal wellbeing remain significant causes of perinatal death and childhood disability[8;138]. On reaching adulthood, these babies are at greater risk of developing cardiovascular disease, hypertension, and non-insulin dependent diabetes[11;12]. Thus FGR has major direct and indirect costs.

Reliable antenatal identification of FGR is crucial to judicious allocation of monitoring resources and use of preventative treatment[129] with the prospect of improving perinatal outcome. Currently, there is a lack of scientific consensus about the best diagnostic and monitoring strategies for predicting FGR and compromise of fetal wellbeing before birth. Consequently, this has lead to uncertainty regarding the best management of pregnant women with a growth-restricted baby with various strategies proposed. There has however been no formal assessment of the cost-effectiveness of these strategies.

This chapter reports the results of an economic evaluation which used evidence from systematic reviews[128;129] (chapters 5 and 6) on the accuracy of all first and second trimester tests and on the effectiveness of all available interventions (chapter 13) to explore the relative cost-effectiveness of a wide range of potentially available 'test and treat' strategies. The economic evaluation took the form of a cost-effectiveness analysis, using decision analytic modelling based on a primary outcome of cost per case of FGR avoided. The comparator was no screening/testing and no intervention because there is currently no routine UK screening policy for the prevention of FGR. The perspective adopted was that of the NHS.

14.3 Methods

14.3.1 Model structure

An economic evaluation was performed using a decision analytic model to compare test – treatment strategies. The evaluation took the form of a cost-effectiveness analysis based on the outcome cost per case of FGR avoided. The analysis was performed from a health care provider perspective, assessing the financial cost of each strategy; this evaluation did not consider private out-of-pocket costs to patient as there are no data available to estimate this. The model allowed the full range of alternatives, the probabilities and uncertainties of these and the outcomes of each strategy to be considered. It provided a framework for each action to be assigned a cost, with the overall cost of each being the sum of the costs of the consequence weighted by the probability of that consequence. The clinical problem was addressed by constructing two separate decision tree models according to population risk (all pregnant women- model 1 and high risk- model 2). The models were used to synthesise the data on test accuracy and intervention effectiveness in order to highlight the potentially most cost-effective 'test and treat' strategies for the prevention of FGR based on all the available data for each population.

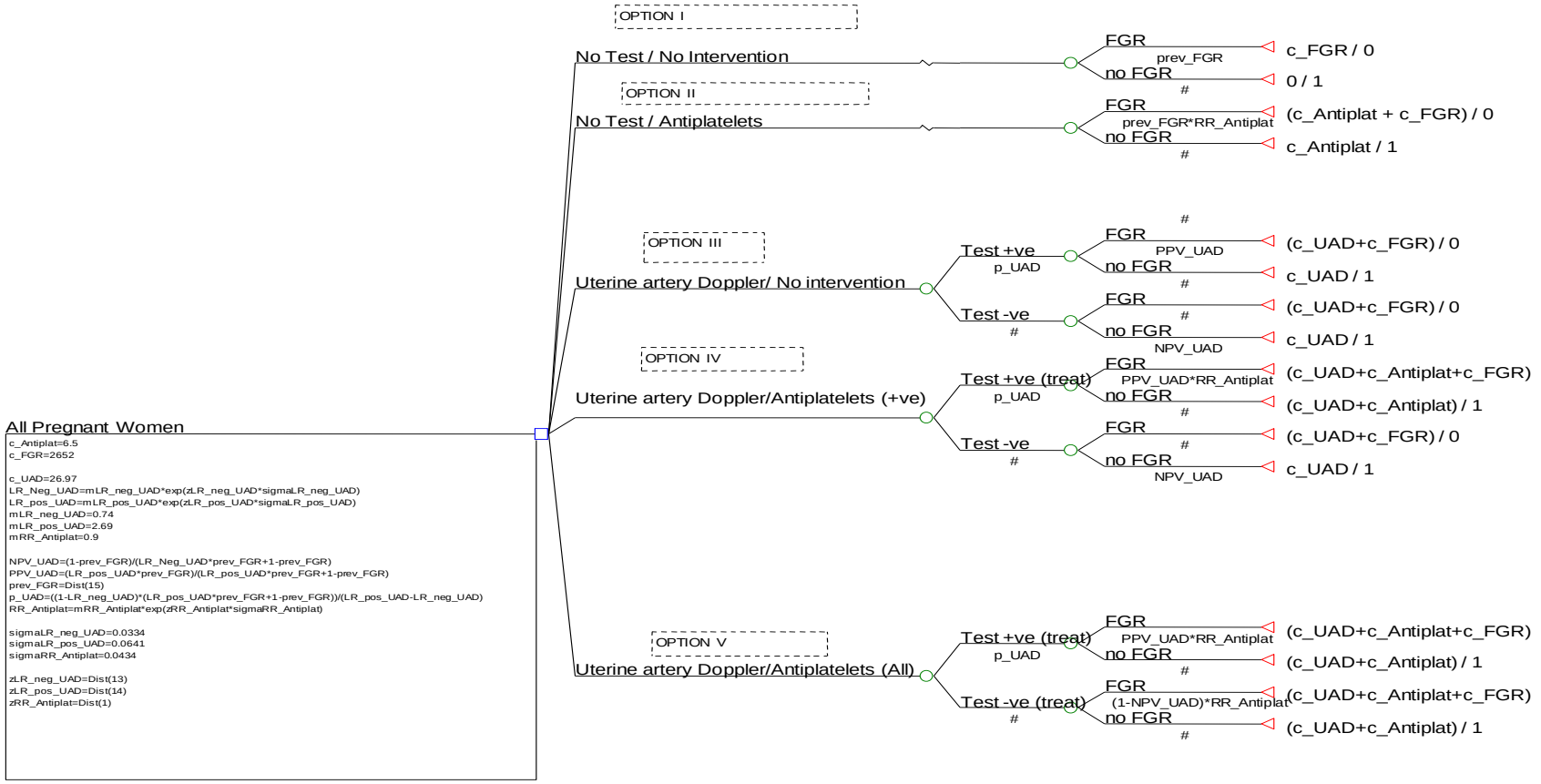
A decision tree was the chosen modelling approach because the time horizons available for both testing and providing the appropriate intervention to women, being within the pregnancy time period, were relatively short and there was no interaction between individuals. The models were constructed in DATA TreeAge Pro Suite 2009[158], an example of one of the models is given in figure 14.1 showing the branches for testing with uterine artery Doppler and treatment with anti-platelets. In this figure each branch to the right of the chance node (round symbol) indicates one way in which the test under

consideration and the treatment can be used together. The diamonds indicate a terminal node or outcome. The box beneath the population under investigation on the far left of the tree, indicates the model parameters being used which will differ depending on the model. Thus for each test and treatment combination, the number of resulting cases of FGR and the associated cost are estimated for the following pathways:

1. No test and no intervention
2. Intervention given to all with no preceding testing
3. Test to all, but no subsequent intervention
4. Test applied to all, followed by the intervention for all those who tested positive
5. Test to all followed by the intervention to all (regardless of test result).

The first pathway represents the comparator for all the other pathways and the common comparator indicating the costs for current clinical practice i.e. no systematic testing and treatment. Pathways 2 and 4 represent the main clinically relevant alternative test treatment strategies. Pathways 3 and 5 are not clinically relevant but are included to give a complete understanding of the relationship between benefits, disbenefits and costs.

Figure 14.1: Example of decision analysis tree showing five pathways for uterine artery Doppler



14.3.2 Inputs to model

Test accuracy

The results from systematic reviews assessing the accuracy of the uterine artery Doppler and Down's syndrome (chapters 5 and 6) serum screening markers were the source of the sensitivity and specificity model parameters[128;129]. These reviews included a total of 105 studies assessing 6 different tests. The actual values used were pooled likelihood ratios generated using the bivariate method of meta-analysis[37]. These values and their associated 95% confidence intervals (CI) are given in table 14.1.

Effectiveness

Systematic reviews of effectiveness as identified by the review of systematic reviews in chapter 13 were the source of data on the effectiveness of interventions. Interventions were eligible for inclusion in the model if the intervention was used in pregnant women with a relevant outcome. The following were also considered: effectiveness, reliability of evidence and generalisability to the population under investigation. The values used were the summary relative risks (RR) along with their 95% CI as shown in table 14.2. The interventions were split into two groups as these are dealt differently by the model. Group 1 includes those treatments in which the 95% CI do not include values>1.0, indicating that the true value of the RR for the treatment is compatible with reducing the number of FGR cases. In group 2, the 95% CI do include values >1.0, i.e. the true value of the RR may be compatible with a worsened outcome.

Table 14.1: Diagnostic test accuracy results for each test provided by systematic reviews of test accuracy – inputs to model

Test (*Subgroup)*	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**	**LR+ve**	**95% CI**	**LR-ve**	**95% CI**
Maternal serum AFP	0.14	(0.10-0.18)	0.94	(0.91-0.96)	2.20	(1.94-2.96)	0.92	(0.90-0.94)
High risk population	*0.31*	*(0.17-0.48)*	*0.94*	*(0.88-0.98)*	*4.54*	*(1.75-11.81)*	*0.70*	*(0.34-1.44)*
Low risk population	*0.13*	*(0.09-0.17)*	*0.94*	*(0.91-0.96)*	*2.16*	*(1.73-2.69)*	*0.93*	*(0.91-0.95)*
Birth weight<10th centile	*0.15*	*(0.10-0.22)*	*0.92*	*(0.87-0.96)*	*1.97*	*(1.48-2.62)*	*0.92*	*(0.88-0.95)*
Birth weight<5th centile	*0.07*	*(0.05-0.10)*	*0.97*	*(0.96-0.98)*	*2.71*	*(1.79-4.12)*	*0.95*	*(0.93-0.98)*
Birth weight<2500g	*0.14*	*(0.09-0.21)*	*0.94*	*(0.89-0.97)*	*2.52*	*(1.74-3.09)*	*0.91*	*(0.88-0.95)*
Maternal serum HCG	0.15	(0.10-0.22)	0.90	(0.86-0.94)	1.57	(1.38-1.78)	0.94	(0.91-0.98)
High risk population	*0.28*	*(0.17-0.41)*	*0.89*	*(0.84-0.93)*	*2.94*	*(1.20-7.17)*	*0.82*	*(0.69-0.99)*
Birth weight<10th centile	*0.15*	*(0.10-0.23)*	*0.91*	*(0.87-0.94)*	*1.76*	*(1.54-2.01)*	*0.93*	*(0.89-0.97)*
Birth weight<5th centile	*0.10*	*(0.06-0.15)*	*0.93*	*(0.90-0.96)*	*1.49*	*(1.13-1.97)*	*0.97*	*(0.94-0.99)*
Birth weight<2500g	*0.19*	*(0.14-0.24)*	*0.85*	*(0.83-0.86)*	*2.70*	*(1.24-5.89)*	*0.89*	*(0.79-1.01)*
Maternal serum unconjugated estriol	0.22	(0.07-0.51)	0.94	(0.86-0.98)	3.88	(1.83-8.22)	0.82	(0.64-1.06)
Birth weight<10th centile	*0.16*	*(0.05-0.40)*	*0.94*	*(0.82-0.98)*	*2.46*	*(1.90-3.17)*	*0.90*	*(0.79-1.03)*
Birth weight<5th centile	*0.34*	*(0.32-0.35)*	*0.98*	*(0.98-0.98)*	*6.54*	*(0.98-43.91)*	*0.59*	*(0.03-13.28)*
Maternal serum PAPPA	0.17	(0.12-0.22)	0.92	(0.88-0.94)	1.98	(1.60-2.45)	0.91	(0.87-0.95)
Birth weight<10th centile	*0.17*	*(0.12-0.24)*	*0.91*	*(0.87-0.94)*	*1.93*	*(1.53-2.44)*	*0.91*	*(0.86-0.95)*
Birth weight<5th centile	*0.18*	*(0.14-0.22)*	*0.92*	*(0.89-0.94)*	*2.17*	*(1.88-2.51)*	*0.90*	*(0.8-,0.92)*
Maternal serum inhibin A	0.11	(0.09-0.12)	0.98	(0.98-0.98)	4.45	(3.92-5.06)	0.92	(0.91-0.93)
Birth weight<10th centile	*0.11*	*(0.09-0.12)*	*0.98*	*(0.98-0.98)*	*4.45*	*(3.92-5.06)*	*0.92*	*(0.91-0.93)*
Birth weight<5th centile	*0.13*	*(0.11-0.15)*	*0.97*	*(0.97-0.98)*	*4.91*	*(4.20-5.73)*	*0.89*	*(0.87-0.91)*
Doppler uterine artery: abnormal waveform	0.34	(0.33,0.36)	0.89	(0.89,0.90)	2.69	(2.37,3.05)	0.74	(0.71,0.79)
High risk population	*0.53*	*(0.44-0.63)*	*0.78*	*(0.71-0.83)*	*2.39*	*(1.90-3.00)*	*0.60*	*(0.50-0.72)*
Low risk population	*0.42*	*(0.30-0.54)*	*0.87*	*(0.81-0.92)*	*3.32*	*(2.39-4.62)*	*0.67*	*(0.56-0.80)*
Birth weight<10th centile	*0.45*	*(0.38-0.52)*	*0.83*	*(0.79-0.87)*	*2.68*	*(2.17-3.30)*	*0.66*	*(0.59-0.74)*
Birth weight<5th centile	*0.51*	*(0.32-0.70)*	*0.85*	*(0.75-0.92)*	*3.43*	*(2.07-5.70)*	*0.57*	*(0.39-0.84)*
Birth weight<3rd centile	*0.41*	*(0.12-0.78)*	*0.86*	*(0.60-0.96)*	*3.00*	*(1.88-4.80)*	*0.69*	*(0.42-1.12)*
Birth weight<2500g	*0.52*	*(0.31-0.73)*	*0.86*	*(0.76-0.92)*	*3.80*	*(2.10-6.87)*	*0.55*	*(0.36-0.86)*

LR likelihood ratio; 95% CI confidence intervals; AFP alpha feto-protein; HCG human chorionic gonadotrophin; PAPPA pregnancy associated plasma protein A

Table 14.2 : Effectiveness data from review of systematic reviews of effectiveness used to inform the model.

	Intervention	Population	Number of RCT	Number of women	Outcome	RR	95% CI
Group 1†	Antiplatelets vs placebo/no intervention	All pregnant women	36	23638	SGA	0.9	(0.83-0.98)
	Interventions for promoting smoking cessation	All pregnant women that smoke	16	9916	BW<2500g	0.83	(0.73-0.95)
	Multiple micronutrient supplementation vs control	All pregnant women	2	2826	SGA	0.92	(0.86-0.99)
	Progesterone vs placebo	Women at increased risk of preterm labour	2	501	BW<2500g	0.64	(0.49-0.83)
Group 2‡	Antiplatelets vs placebo/no intervention	Women at risk of developing PE	13	4239	SGA	0.89	(0.74-1.08)

RCT randomised controlled trial
RR relative risk
CI confidence interval
PE pre-eclampsia
SGA small for gestational age
BW birth weight
† Group 1 are those treatments with an RR whose upper 95% CI is <1.0
‡ Group 2 are those treatments with an RR whose upper 95% CI includes a value compatible with a worsened outcome

Costs

The cost estimates for each test and the outcome are summarised in Table 14.3 and 14.4. All costs were converted to 2009 prices (£ Sterling) using the combined hospital and community index[159]. Since the time horizon of the model was within one year, the discounting of costs and outcomes was not required. The cost data for the tests was retrieved from literature reviews[160;161] and the Birmingham Women's Hospital NHS Foundation Trust (BWH), Birmingham, estimated from the UK Department of Health's latest Health Resource Groups (HRGs) 2009[162]. The cost associated with treatments was estimated based on information on dose and duration described in the included studies in the systematic reviews of effectiveness in the Cochrane library. Where no dose or duration was available, the recommendation in the British National Formulary (BNF) (Volume 57, 2009) was used[13]. Where a dose range was presented the cost of the upper and the lower limit of the dose was used. Costs estimated for hospitalisation associated with treatments were also included. There was no estimate of the cost involved for clinician's time to prescribe the intervention as this was presumed to be constant across all the interventions.

Table 14.3: Estimated costs of diagnostic tests (all models).

Test	Description/Nature of test	Unit cost from Birmingham Women's Hospital (UK£ 2009)	Costs from literature (UK£ 2009)	
			Unit cost (upper and lower estimates)	Source
Maternal serum alpha fetoprotein	Phlebotomoist performs test (5 minutes) - venous blood 2.5ml. Lab technician to analyse (1 hour as part of batch)	14.17	49.11 (42.30-55.91)	Literature[a]
Maternal serum human chorionic gonadotrophin	Phlebotomoist performs test (5 minutes) - venous blood 2.5ml. Lab technician to analyse (1 hour as part of batch)	12.50	49.11 (42.30-55.91)	Proxy based on AFP literature[a]
Maternal serum unconjugated estriol	Phlebotomoist performs test (5 minutes) - venous blood 2.5ml. Lab technician to analyse (1 hour as part of batch)	12.50	49.11 (42.30-55.91)	Proxy based on AFP literature[a]
Maternal serum pregnancy associated plasma protein A	Phlebotomoist performs test (5 minutes) - venous blood 2.5ml. Lab technician to analyse (1 hour as part of batch)	12.50	49.11 (42.30-55.91)	Proxy based on AFP literature[a]
Maternal serum inhibin A	Phlebotomoist performs test (5 minutes) - venous blood 2.5ml. Lab technician to analyse (1 hour as part of batch)	10.00	49.11 (42.30-55.91)	Proxy based on AFP literature[a]
Uterine artery Doppler	Ultrasound scan lasting 10 minutes	26.97	23.07 (20.03-26.13)	Literature[a]

AFP alpha fetoprotein NA not applicable [a]Roberts T, Henderson J, Mugford M, Bricker L, Neilson J, Garcia J. Antenatal ultrasound screening for fetal abnormalities: a systematic review of cost and cost effectiveness studies. *BJOG* 2002;109(1):44-56

Table 14.4: Estimated costs of interventions

Treatment	Nature and dose	Duration	Total Cost UK £ 2009	Bases for cost estimate (comment)	Source of unit
Antiplatelets (principally aspirin)	75-150mg/day	20 weeks	6.23 [4.15 (75mg) - 8.30 (150 mg)]	Aspirin 75mg; 28 tablets @ £0.83 ; 5 packs needed for 140 days. For 150 mg per day, 10 packs required.	BNF[a]
Multiple micronutrients	Multivitamins one capsule per day (ascorbic acid 15mg, nicotinamide 7.5mg, riboflavin 500µg, thiamine hydrochloride 1mg, vitamin A 2500 units, vitamin D 300 units)	Throughout pregnancy, average 30 weeks treatment	2.42	Multivitamin capsules, 20 capsule pack @£0.22- 11 packs required	BNF[a]
Progesterone	300mg IM injection daily to 50 mg IM injection on alternate days	Given for 1 week and then offered again at subsequent visits (therefore cost is 7-14 days of progesterone) stopped if symptoms disappeared or labour started.	130.00 [85.00 (50mg)- 175.00 (300mg)]	50mg /ml in 1 ml ampule = £4.50; 2ml ampule (100mg) = £4.50. Assume 10 days for all. Cost includes the cost of injection at outpatients/clinic for 10 days plus the cost of the drug for 10 days. Cost varies depending on dose. Cost of drug £45.00 (50mg) – £135.00 (300mg) 10 minutes practice nurse time assumed to administer injection. Hourly rate of practice nurse is £24/hour. Thus 10 minutes =£4.00, for 10 days=£40.00	BNF[a] Curtis and en[c]
Smoking cessation	Nicotine patches, 14mg every 24 hours for 4 weeks	Start 1-2 weeks before target stop date, maximum period of treatment 4 weeks	35.80	Nicotine patch, 7 patch pack @ £8.95 (2 packs required)	BNF[a]

[a] British Medical Association and Royal Pharmaceutical Society of Great Britain. *British National Formulary.* March 2009
[b] Holland and Barrett. URL: http://www.hollandandbarrett.com
[c] Curtis L, Netten A. Unit costs of health and social care 2005. Canterbury: PSSRU;2006.
GTN glyceryl trinitrate

A systematic review of the literature was performed to search for relevant cost data related to the outcome. Medline, Embase, British Nursing Index, Cochrane Library (including economic databases) and grey literature databases were searched from inception until July 2009. The search strategy consisted of Medical Subject Heading (MeSH) and keywords related to terms for fetal growth restriction combined with terms related to costs and effectiveness. The full search strategy is shown in appendix 48. The search revealed 3445 citations, 51 of which were selected after scrutiny of the title and abstract by two reviewers (author and Dr Gemma Malin). These papers were obtained in full and one further paper was obtained after checking the reference lists of these papers. To be included in the review papers had to report on the cost of the birth of a baby in a singleton pregnancy at all gestations to the NHS where the baby was either small for gestational age or birth weight <2500g. Costs could be applied over any time period after birth. None of the 52 papers complied with all the inclusion criteria. Cost data were thus calculated from BWH data over a 5 year period, 2004-2008. All babies at BWH with a birth weight <2300g are admitted to either the neonatal intensive care, high dependency care or transitional care. Babies with a birth weight between 2300 and 2500g will be admitted if clinically indicated, those that are well are sent straight to the post-natal ward with no further investigations and thus no extra cost to the NHS above a baby of birth weight >2500g. Over the 5 year time period average length of stay at each level of care was calculated for each baby and then stratified according to year, birth weight and gestation (Table 14.5). There were significant variations in cost according to gestation and birth weight. For the purposes of this analysis it was decided by the author that the most appropriate cost to use was an overall cost for babies born at term ≤ 2500g as this represents the most representative cost for a general pregnant population. The

cost of a normal delivery was removed from the estimate of cost of a baby with FGR as this was presumed to be constant for any outcome and mode of delivery was not being used within any of the comparators of the model.

Prevalence

The prevalence of fetal growth restriction for a general population was obtained from Office of National Statistics data for 2007 (defined as BW<2500g) and the systematic reviews of test accuracy[128;129] giving an overall prevalence of 7.6% (95% CI 5.21-10.5). For the high risk populations the prevalence rate from the systematic reviews of test accuracy were used.

14.3.3 Analysis

The main outcome of each of the models was cost per case of FGR avoided. Quality of life data for this clinical condition were not available in the literature. For each model, a deterministic and probabilistic sensitivity analysis (PSA) was carried out[6]. The deterministic analysis uses the point estimate only, one way sensitivity analysis was performed to examine the impact of changing one variable across its entire plausible range, while keeping all other variables at the point estimate[163]. In PSA, each model parameter is assigned a distribution reflecting the amount and pattern of its variation and cost-effectiveness results are calculated by simultaneously selecting random values from each distribution. This process was repeated 10,000 times in a Monte Carlo simulation of the model to give an indication of how variation in the model parameters leads to variation in the incremental cost effectiveness ratios (ICERs) for a given

combination of a test and treatment pairing. The appropriate distribution to use for both the data on test accuracy (sensitivity and specificity) and for data on intervention effectiveness (RR of developing FGR) was a log normal distribution. The distribution applied for prevalence was a β distribution where overall prevalence ~β(n,r) where n=total number of cases and r=total number of studies size. For costs a gamma distribution was applied.

Assumptions made in the models were that it was appropriate to include only the interventions for which the 95% CI of the relative risk was <1 to avoid including interventions which may be deemed as harmful.

In summary the complete set of analyses performed were:

- *Case 1:* As detailed above, a deterministic sensitivity analysis using data for all the tests and combined with treatments with costs as detailed in tables 14.1-14.5 for model 1, all pregnant women using the appropriate disease prevalence and parameters for an unselected population.

- *Case 2:* As detailed above, a PSA analysis of case 1.

- *Sensitivity analyses:* Sensitivity analyses were performed for case 1 varying the parameters for cost of FGR and using the individual, rather than summary, test accuracy inputs recommended in each of the test accuracy chapters (5-9) e.g. Pulsatility index and notching for uterine artery Doppler in an unselected population.

- *Case 3:* As detailed above, a deterministic sensitivity analysis using data for all the tests and combined with treatments with costs as detailed in tables 14.1-14.5

Table 14.5: Costs for a baby born with a birth weight ≤2500g, according to birth weight and gestation.

	Average cost (£ 2009/2010)										
Year	**All admission ≤2500g**	**Term ≤2500g**	**Term 2000-2500g**	**Term 1500-1999g**	**Term 1000-1499g**	**Preterm ≤2500g**	**Preterm 2000-2500g**	**Preterm 1500-1999g**	**Preterm 1000-1499g**	**Preterm 500-999g**	**Preterm <500g**
2008	12793	2564	2616	1541	5499	14309	3743	8031	15248	33974	51257
2007	12585	1979	2042	1624	NA	13638	4298	7173	13826	32776	NA
2006	13690	3091	2945	3779	NA	14733	4644	8064	21721	35103	NA
2005	14908	2368	2393	2327	NA	16359	4502	9199	20565	48413	23508
2004	16532	3258	3015	4107	NA	17692	4069	8160	20875	42708	35591
Average over 5 years	*14101*	*2652*	*2602*	*2676*	*5499*	*15382*	*4251*	*8125*	*18447*	*38595*	*36785*
Cost of fetal growth restriction per case (input for model)				***£2652 (Incremental cost of health care from birth to discharge from hospital for a baby<2500g)***							

Term ≥ 37 completed weeks of gestation

Preterm < 37 completed weeks of gestation

for model 2, high risk pregnant women using the appropriate disease prevalence and parameters for a high risk population.

- *Case 4:* As detailed above, a PSA analysis of case 3.

- *Sensitivity analyses:* Sensitivity analyses were performed for case 3 varying the parameters for cost of FGR and using the individual, rather than summary, test accuracy inputs recommended in each of the test accuracy chapters (5-9) .

- *Case 5:* A threshold analysis to explore what test accuracy and test cost parameters would be required to optimise cost-effectiveness using the deterministic analysis presented in case 2 as a starting point.

14.4 Results

Main result

The most cost-effective strategy for all pregnant women was "no test/antiplatelets all" and for high risk pregnant women "HCG/antiplatelets +ve" was the most cost-effective strategy.

Case 1: base case for model 1 all pregnant women

Table 14.6 presents the results for the deterministic analysis for all pregnant women. The results are presented incrementally compared to the previous best option. The "no test, antiplatelets to all" strategy dominated throughout as the most cost-effective option at a mean cost of £177 per women. This strategy saves 7 cases of FGR per 1000

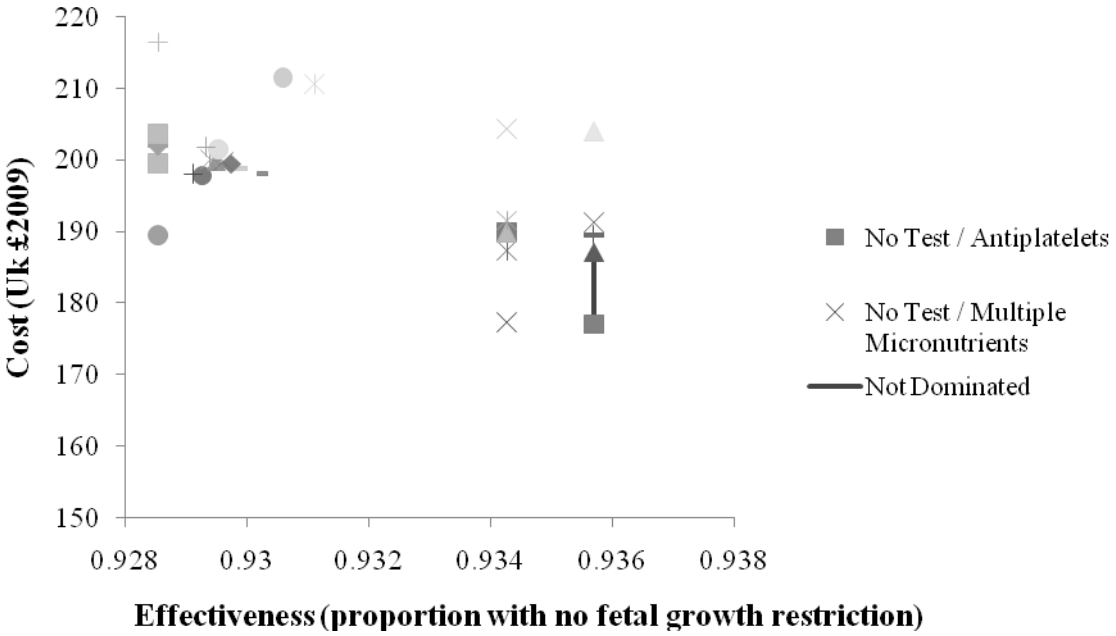
women, a number needed to treat (NNT) to prevent one additional case of FGR of 132. When "no test/antiplatelets" is removed from the model "no test/multiple micronutrients" was dominant. Figure 14.2 demonstrates the results graphically with all the cost-effectiveness estimates for case 1 shown. The nearer an estimate is to the bottom right hand corner of the graph, the greater its effectiveness and the lesser its cost. The strategy "no test/antiplatelets" is seen to dominate all other strategies.

Table 14.6: Case 1, base-case results: costs, effects and ICERs for test/treatment combinations for all pregnant women (model 1).

Test/treatment combination	Mean cost per women (UK £2009)	Difference in costs (UK £2009)	Effectiveness[a]	Absolute risk reduction	Cost effectiveness	ICER[b]
No Test / Antiplatelets	177.05		0.94	0.007[#]	189.22	
No Test / Multiple Micronutrients	177.34	0.29	0.93	0.006	189.82	(Dominated by no test/antiplatelets)
Inhibin A/Antiplatelets (All)	187.05	10.00	0.94	0.007	199.91	(Dominates the rest of the strategies)

[#] Compared to "no test/no treatment"
[a] Effectiveness is defined as the proportion of women with a pregnancy where a fetal growth restricted (FGR) baby is avoided. Therefore, the difference in effectiveness between two strategies is the absolute risk reduction
[b] ICER: incremental cost effectiveness ratio expressed as the additional cost per additional case of FGR avoided.
AFP alpha fetoprotein; HCG human chorionic gonadotrophin; PAPPA pregnancy associated plasma protein A

Figure 14.2: Case 1, base case results: costs, effects and ICERs on cost-effectiveness plane for all combinations of test and treatment pairs in an all pregnant women population



Case 2: probabilistic sensitivity analysis of case 1

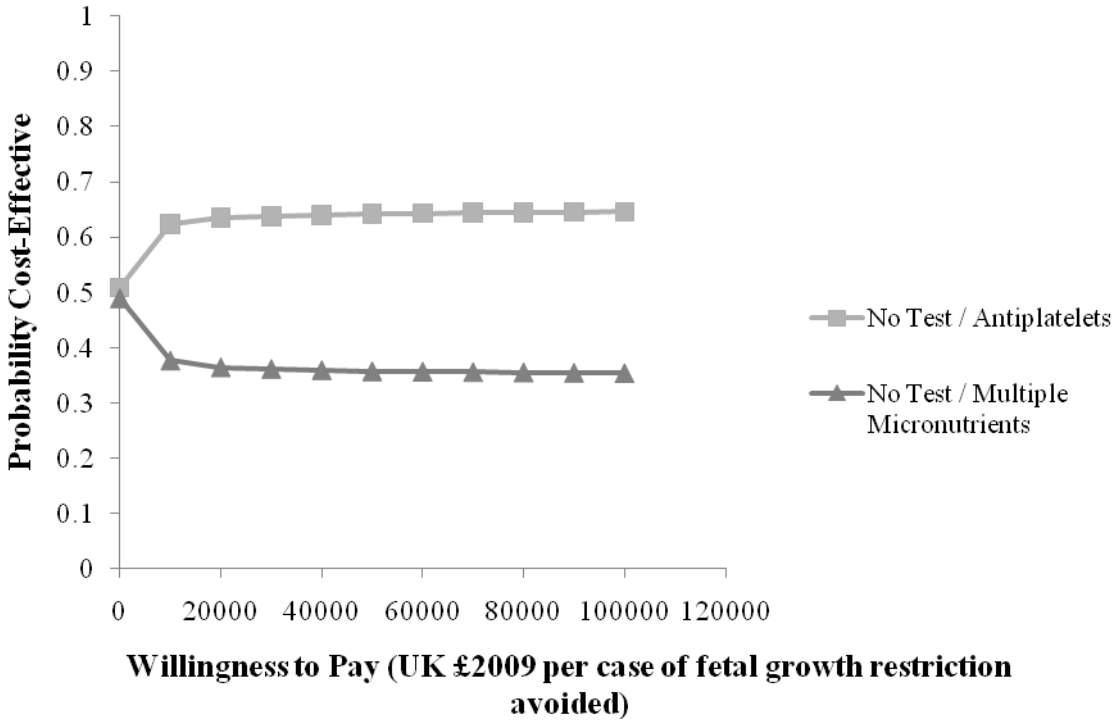
The results of case 2 are presented in table 14.7. The results of the PSA confirm that at all levels of willingness to pay the strategy "no test/antiplatelets" is the dominant option with "no test/multiple micronutrients" is the next dominant. The PSA demonstrates if a policy maker is willing to pay £30,000 per case of FGR avoided, there is a 64% chance that "no test/antiplatelets" is the preferred option with respect to its cost-effectiveness. At the same threshold there is only a 36% chance that "no test/multiple micronutrients" is the preferred option. At a threshold of £100,000 then the results are 65% and 35% respectively. Thus the results are robust for all threshold levels. These results are presented graphically in figure 14.3 as a cost-effectiveness acceptability curve (CEAC) to ensure that this figure is legible the strategies associated with a probability of zero at any threshold level have been removed.

Table 14.7: Case 2, PSA of case 1, results: probability that stated options are the most cost-effective option at different levels of willingness to pay per case of fetal growth restriction avoided.

	Willingness to pay (UK £2009/10)[a]					
	0	**10,000**	**30,000**	**50,000**	**80,000**	**100,000**
Test/Treatment option						
No test/no intervention	0.002	0.0002	0.0002	0	0	0
No test/antiplatelets (All)	0.51	0.623	0.638	0.642	0.644	0.646
No test/multiple micronutrients (All)	0.489	0.377	0.362	0.358	0.356	0.354

[a] Per case of fetal growth restriction avoided
The other test/treatment combinations are not shown as all had a probability of zero for all willingness to pay thresholds

Figure 14.3: Case 2, probabilistic sensitivity analysis of case 1, results for all pregnant women (results with probability of zero across all thresholds removed).



Sensitivity analysis for case 1

As the literature searches had been unable to identify any costs for FGR that could be compared with the calculated costs from BWH data, it was felt appropriate to perform a sensitivity analysis around this cost. Also from the BWH data there were many different costs according to different birth weight thresholds and gestation. Thus the sensitivity analysis used different costs identified from BWH data as shown in table 14.5 as well as analyses to show what would happen to the results if the cost of FGR was reduced to zero.

The results of the sensitivity analysis for case 1 are presented in table 14.8. They demonstrate that at a higher cost level (£14101 and £36785) "no test/antiplatelets" remains the dominant strategy. If the cost of FGR is reduced to £1000 the most cost-effective option becomes "no test/multiple micronutrients". The exact threshold at which the strategies changed was identified as £2450 for "no test/multiple micronutrients" being more effective than "no test/antiplatelets". There was no change in the results when varying the test accuracy inputs.

Case 3: base case for model 2 high risk pregnant women

Table 14.9 presents the results for the deterministic analysis for high risk pregnant women. The results are presented incrementally compared to the previous best option. The "HCG, antiplatelets +ve" was the most cost-effective option at a mean cost of £140 per women. This strategy saves 6 cases of FGR per 1000 women, a number needed to treat (NNT) to prevent one additional case of FGR of 20 compared to 33 for "no test/antiplatelets all". The next most cost-effective options were "AFP/antiplatelets +ve"

Table 14.8: Sensitivity analysis for case 1: deterministic analysis for case 1 when the cost of fetal growth restriction (FGR) is varied from the base case level of £2652

Strategy	Mean cost per woman (UK £2009)	Difference in costs (UK £2009)	Effectiveness	Absolute risk reduction	Cost-effectiveness	ICER[b]
Cost of FGR £36785						
No Test / Antiplatelets[a]	2372.13		0.94		2535.17	
Cost of FGR £14101						
No Test / Antiplatelets[a]	913.33		0.94		976.10	
Cost of FGR £1000						
No Test / Multiple Micronutrients	68.74		0.93		73.58	
No Test / Antiplatelets	70.81	2.07	0.94	0.0014	75.68	1449.09
Cost of FGR £500						
No Test / Multiple Micronutrients	35.87	0.14	0.93	0.0057	38.39	24.80
No Test / Antiplatelets	38.65	2.79	0.94	0.0014	41.31	1949.09
Cost of FGR £50						
No Test / Multiple Micronutrients	5.30	2.80	0.93	0.0057	5.67	489.80
No Test / Antiplatelets	8.75	3.45	0.94	0.0014	9.35	2414.09

[a] Strategy dominates all others

[b] ICER incremental cost-effectiveness ratio

Table 14.9: Case 3, base-case results: costs, effects and ICERs for test/treatment combinations for high risk pregnant women (model 2).

Test/treatment combination	Mean cost per women (UK £2009)	Difference in costs (UK £2009)	Effectiveness[a]	Absolute risk reduction	Cost effectiveness	ICER[b]
HCG/Antiplatelets +ve	140.41		0.79	0.0059[#]	177.31	
AFP/Antiplatelets (+ve)	195.11	54.70	0.79	0.0025	245.63	22118.31423
Uterine artery Doppler/Antiplatelets +ve	299.30	104.18	0.80	0.0042	374.79	24671.15045
No Test/Progesterone	493.22	193.92	0.86	0.0645	571.49	3008.047974
HCG/Progesterone (All)	505.72	12.50	0.86	0.0000	585.97	(This and all subsequent strategies dominated by No test/progesterone)

[#] Compared to "no test/no treatment"

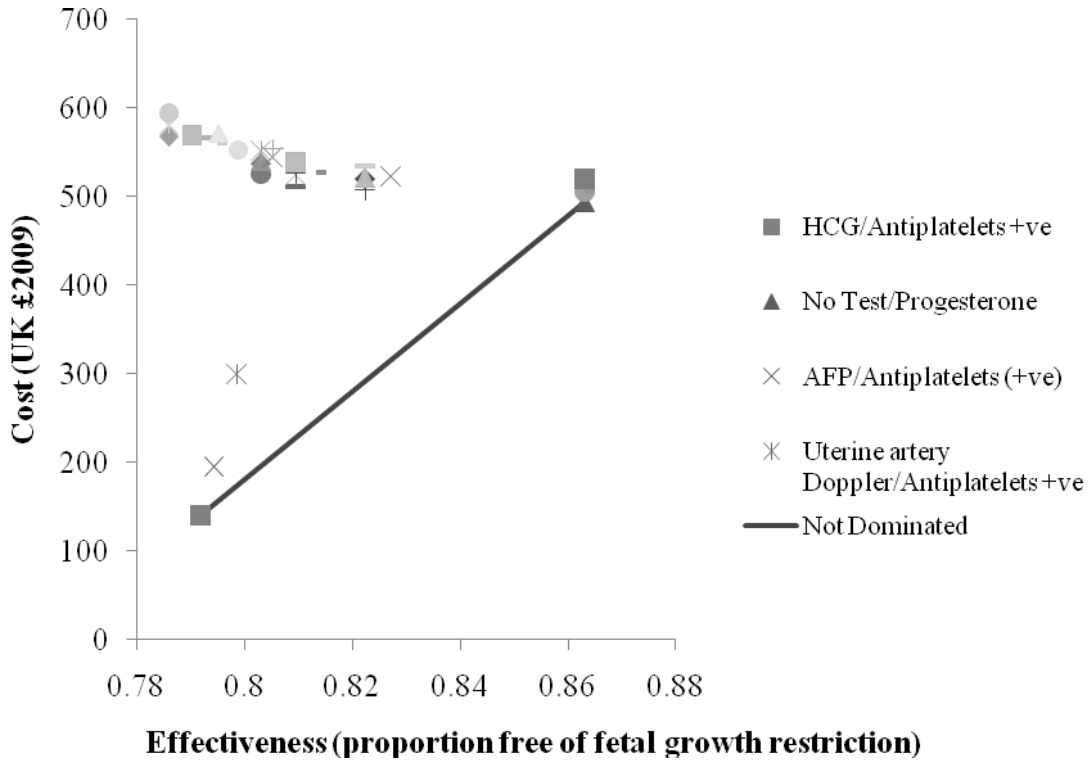
[a] Effectiveness is defined as the proportion of women with a pregnancy where a fetal growth restricted (FGR) baby is avoided. Therefore, the difference in effectiveness between two strategies is the absolute risk reduction

[b] ICER: incremental cost effectiveness ratio expressed as the additional cost per additional case of FGR avoided.

AFP alpha feto-protein; HCG human chorionic gonadotrophin; PAPPA pregnancy associated plasma protein A

with an ICER of 22118 and "uterine artery Doppler/antiplatelets +ve" ICER 24671. Figure 14.4 demonstrates the results graphically with all the cost-effectiveness estimates for case 1 shown. The graph shows that "no test/progesterone all" is more effective but much more costly at a mean cost of £493 per woman compared to £140 for HCG. However, treatment with progesterone can prevent 65 cases per 1000 women of FGR, an NNT of 10.

Figure 14.4: Case 3, base case results: costs, effects and ICERs on cost-effectiveness plane for all combinations of test and treatment pairs in a high risk pregnant population



Case 4: probabilistic sensitivity analysis of case 3

The results of case 4 are presented in table 14.10. The results of the PSA demonstrate that the dominant strategy across all thresholds is "no test/progesterone" with an 87%

Table 14.10: Case 4, PSA of case 3, results: probability that stated options are the most cost-effective option at different levels of willingness to pay per case of fetal growth restriction avoided.

	Willingness to pay (UK £2009/10)[a]					
	0	**10,000**	**30,000**	**50,000**	**80,000**	**100,000**
Test/Treatment option						
No test/no intervention	0	0	0	0	0	0
No test/antiplatelets (All)	0	0.023	0.02	0.017	0.016	0.016
No test/smoking cessation (All)	0	0.056	0.052	0.048	0.045	0.045
No test/multiple micronutrients (All)	0	0	0.001	0.001	0.001	0.001
No test/progesterone (All)	0	0.867	0.928	0.934	0.938	0.939
AFP/Antiplatelets +ve	0.391	0.026	0	0	0	0
HCG/Antiplatelets +ve	0.598	0.029	0	0	0	0
Uterine artery Doppler/antiplatelets +ve	0.011	0	0	0	0	0

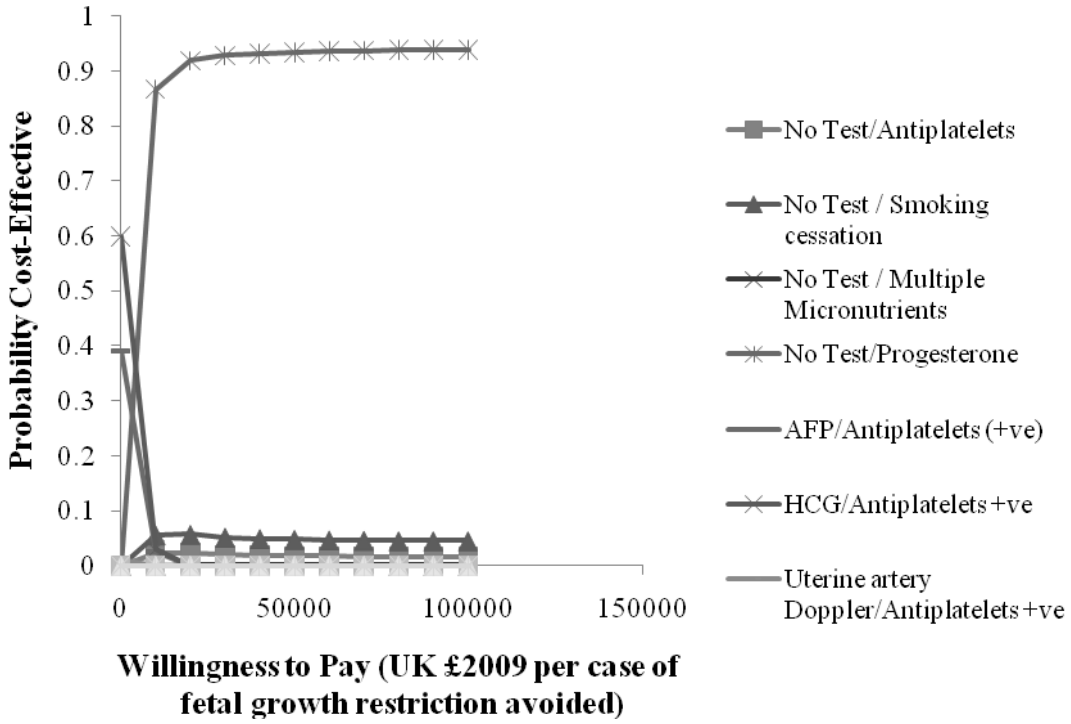
[a] Per case of fetal growth restriction avoided

AFP alpha fetoprotein; HCG human chorionic gonoadotrophin

The other test/treatment combinations are not shown as all had a probability of zero for all willingness to pay thresholds

chance at £10,000 and a 94% chance at £100,000 of this being the most preferred option. These results are presented graphically in figure 14.5 as a CEAC. Further analysis shows that the threshold at which "no test/progesterone" becomes likely to be the dominant option is at £4540. As progesterone is mainly a treatment for high risk women at risk of pre-term labour it was felt appropriate to repeat the PSA with this strategy removed and smoking cessation removed as this can only be used in smokers. The results demonstrated that at a level of £10,000 "HCG/antiplatelets +ve" was likely to be the most cost-effective with a chance of 53% at £30,000 this became "no test/antiplatelets +ve" at 58% and this was likely to be the most cost-effective strategy at all subsequent thresholds.

Figure 14.5: Case 4, probabilistic sensitivity analysis of case 3, results for high risk pregnant women (results with probability of zero across all thresholds removed).



Sensitivity analysis for case 3

This was performed as for case 1 and the results are presented in table 14.11. They demonstrate that at a higher cost level (£14101 and £36785) and at lower cost levels (£1000 and £500) "HCG/antiplatelets +ve" remains the dominant strategy. If the cost of FGR is reduced to £50 the most cost-effective option becomes "no test/multiple micronutrients". The exact threshold at which the strategies changed was identified as £86. There was no change in results when varying the test accuracy inputs.

Table 14.11: Sensitivity analysis for case 3: deterministic analysis for case 3 when the cost of fetal growth restriction (FGR) is varied from the base case level of £2652

Strategy	Mean cost per woman (UK £2009)	Difference in costs (UK £2009)	Effectiveness	Absolute risk reduction	Cost-effectiveness	ICER[a]
Cost of FGR £36785						
HCG/Antiplatelets +ve	1763.20		0.79		2226.61	
AFP/Antiplatelets (+ve)	2500.93	737.73	0.79	0.00	3148.40	298285.31
Uterine artery Doppler/Antiplatelets +ve	3771.33	1270.40	0.80	0.00	4722.59	300838.15
No Test/Progesterone	5168.08	1396.75	0.86	0.06	5988.23	21665.87
Cost of FGR £14101						
HCG/Antiplatelets +ve	684.73		0.79		864.70	
AFP/Antiplatelets (+ve)	968.54	283.81	0.79	0.00	1219.28	114751.13
Uterine artery Doppler/Antiplatelets +ve	1463.90	495.36	0.80	0.00	1833.14	117303.97
No Test/Progesterone	2061.28	597.38	0.86	0.06	2388.39	9266.31
Cost of FGR £1000						
HCG/Antiplatelets +ve	61.87		0.79		78.13	
AFP/Antiplatelets (+ve)	83.51	21.65	0.79	0.00	105.14	8752.13
Uterine artery Doppler/Antiplatelets +ve	131.25	47.74	0.80	0.00	164.36	11304.97
No Test/Antiplatelets	196.96	65.71	0.81	0.01	243.30	5990.88
No Test / Smoking cessation	211.48	2.02	0.82	0.01	256.49	134.84
Uterine artery Doppler/Progesterone (+ve)	237.55	13.57	0.83	0.00	287.19	5167.31
No Test/Progesterone	266.96	29.41	0.86	0.04	309.33	819.45
Cost of FGR £500						
HCG/Antiplatelets +ve	38.10		0.79		48.11	
AFP/Antiplatelets (+ve)	49.74	11.64	0.79	0.00	62.61	4706.68
Uterine artery Doppler/Antiplatelets +ve	80.39	30.66	0.80	0.00	100.67	7259.51
No Test / Multiple Micronutrients	101.44	21.05	0.80	0.00	126.31	4627.97
No Test/Antiplatelets	101.73	0.29	0.81	0.01	125.66	45.17
No Test / Smoking cessation	123.74	9.51	0.82	0.01	150.08	634.84
Uterine artery Doppler/Progesterone (+ve)	151.12	14.88	0.83	0.00	182.70	5667.31
No Test/Progesterone	198.48	47.36	0.86	0.04	229.98	1319.45

Cost of FGR £50						
No Test / Multiple Micronutrients	12.84	2.14	0.80	0.02	15.99	125.23
No Test/Antiplatelets	16.02	3.18	0.81	0.01	19.79	495.17
AFP/ Progesterone (+ve)	41.35	12.83	0.81	0.00	50.85	3389.84
No Test / Smoking cessation	44.77	3.42	0.82	0.01	54.30	305.50
Uterine artery Doppler/Progesterone (+ve)	73.33	16.06	0.83	0.00	88.66	6117.31
No Test/Progesterone	136.85	63.51	0.86	0.04	158.57	1769.45

[a] ICER incremental cost-effectiveness ratio

AFP alpha fetoprotein; HCG human chorionic gonoadotrophin

Case 5: threshold analysis for potentially cost-effective test parameters

This analysis was prompted by the finding that the most cost-effective strategies in all pregnant women involved no prior testing. The model presented in case 2 was used to explore the levels of test accuracy and test cost that would be required to make a "test/treatment" strategy be more effective than a "no test/treat all" strategy. This was performed using two hypothetical tests; test A with a cost of £20 was based on uterine artery Doppler and test B with a cost of £5 based on the costs of first trimester blood tests (e.g. Down's syndrome serum screening). The intervention was chosen as antiplatelets. Thus there were two strategies "hypothetical test A/antiplatelets +ve" and "hypothetical test B/antiplatelets +ve). The model was then run with varying test accuracy parameters. The results are presented in table 14.12 and demonstrate that for a test cost of £20.00 there were no levels of test accuracy that would make a "test/treatment" strategy cost-effective. The cost had to be reduced to £5 with a LR+ve of 20 and LR-ve of 0.02 i.e. an excellent test for the "hypothetical test B/antiplatelets +ve" to be the most cost-effective strategy however; this was only up to a willingness to pay threshold of £1000 (46% chance).

14.5 Discussion

The main finding of the economic evaluation is that in an unselected pregnant population testing using Down's syndrome serum markers or uterine artery Doppler is not cost-effective compared to the strategy of treating all women with antiplatelets in the prevention of FGR. In high risk pregnant women the most cost-effective strategy is to test women with HCG and treat with antiplatelets those with a positive result.

Table 14.12: Case 5, threshold analysis on characteristics of a test that would be cost-effective when combined with the intervention antiplatelets ("hypothetical test/antiplatelets +ve)

Necessary characteristics of test	Test/treatment option	Probability of being most cost-effective option at different levels of willingness to pay for a case of fetal growth restriction avoided				
		0	10,000	30,000	50,000	100,000
Hypothetical test A						
LR+ve 10.0 LR-ve 0.2 Cost=£20.00	Test and treat all positives with antiplatelets	0	0	0	0	0
LR+ve 10.0 LR-ve 0.02 Cost = £20.00	Test and treat all positives with antiplatelets	0	0	0	0	0
LR+ve 20.00 LR-ve 0.2 Cost = £20.00	Test and treat all positives with antiplatelets	0	0	0	0	0
LR+ve 20.00 LR-ve 0.02 Cost = £20.00	Test and treat all positives with antiplatelets	0	0	0	0	0
Hypothetical test B						
LR+ve 10.0 LR-ve 0.2 Cost=£5.00	Test and treat all positives with antiplatelets	0	0	0	0	0
LR+ve 10.0 LR-ve 0.02 Cost = £5.00	Test and treat all positives with antiplatelets	0.326	0.0001	0	0	0
LR+ve 20.00 LR-ve 0.2 Cost = £5.00	Test and treat all positives with antiplatelets	0	0	0	0	0
LR+ve 20.00 LR-ve 0.02 Cost = £5.00	Test and treat all positives with antiplatelets	0.523	0.0008	0.0002	0.0001	0.0001

LR+ve positive likelihood ratio LR-ve negative likelihood ratio

These results were found to be robust in a PSA and sensitivity analysis. Threshold analysis revealed that for a test to be considered as an option prior to treatment in unselected pregnant women it would have to have high levels of accuracy and be relatively cheap (£5). This is likely to be due to the fact that the majority of treatments available are themselves relatively cheap (£2.60 for aspirin) and thus from a cost point of view it will always be preferential to apply treatment to all rather than test first. This has to be interpreted in light of the limitations of the model importantly the lack of inclusion of adverse effects of treatment.

Strengths of the economic evaluation

The model was populated with data acquired through high quality and up to date evidence i.e. the systematic reviews of test accuracy and review of systematic reviews of effectiveness. The model was developed by the author, who received formal training, with the help of an experienced health economist (Angelos Tsourapas, Professor Tracy Roberts) and advice from a modelling expert (Dr Pelham Barton). The model was tested at all stages of development to ensure that it was methodologically correct and clinical advice was taken where necessary regarding the clinical assumptions made (Professor Khalid Khan, Professor Mark Kilby). Interpretation of the results of the model was made by the author following discussion with Angelos Tsourapas.

Limitations of the economic evaluation

There are two main limitations to this work. The first is the constraints of the model and the assumptions made, the second is the limitations in the data used to populate the model.

Model design and assumptions

The model assumes that each pregnant woman will receive multiple tests as is the case in clinical practice but it is assumed that these will be interpreted in isolation i.e. with no reference to risk factors, clinical features or the other test results as would normal happen in clinical practice. The same is true for the interventions. Thus strategies tested consist of a single test+/- single intervention and thus this model does not evaluate combination testing and combination treatments.

The model considers a single outcome "cost per case of FGR avoided". The tests and treatments evaluated are also used in the management of pre-eclampsia and pre-term labour as there is great overlap in the aetiology and management of these conditions. Thus the tests may have a greater value if the outcome used was a combined outcome and the same may be true for the effectiveness of the interventions. However, the converse may be true e.g. in the case of some hypertensives for management of pre-eclampsia there is an increased risk of an FGR baby. The model does also not consider the medium and long term outlooks for these babies e.g. perinatal mortality, neurodevelopmental outcome.

The model does not take into account side effects and adverse events as a result of intervention or indeed testing e.g. increased maternal anxiety. The assumption made was that the side effects of the tests would be negligible as they were not invasive tests and they have been extensively evaluated in clinical practice. For the interventions there is evidence that the side effects of interventions such as aspirin is also negligible[164],

however this is an important consideration particularly when one is advocating a policy of "test none/treat all".

The comparator used in the model was "no test/no intervention". While this is necessary to allow comparison across strategies within the model it is recognised that in clinical practice all pregnant women will have some sort of testing in pregnancy. This may simply be assessment of risk factors or testing may be performed for other reasons but have implications for the outcome of FGR e.g. Down's syndrome serum screening. This is likely to have led to an overestimate of the cost-effectiveness of the strategies investigated compared to that which could be achieved in clinical practice but gain does allow comparison of strategies.

The model cannot take into account any qualitative data e.g. the impact on a woman and her family of testing, the acceptability of the test etc. and thus all interpretations of "cost-effectiveness" are restricted to the perspective of the healthcare payer, in this case the NHS. This will limit the generalisability of the model to other countries and healthcare models however, the information gained from the model regarding the levels of test accuracy and costs required given the available interventions is important for health care researchers and policy makers worldwide.

To the best of the author's knowledge there are no other published economic evaluations in this area with which to make comparisons.

Limitations of primary data within model

Despite the strengths of the methods used to acquire this data it must be recognised that the inputs are limited by the quality of the primary data from which they arise. These limitations have been discussed in chapters 11 and 12. Particular considerations for the model are the wide confidence intervals associated with some of the tests and interventions and the applicability of some of the interventions to subsets of pregnant women. To overcome some of these limitations two models were used with some interventions only being included in the model for high risk women however some intervention e.g. progesterone and smoking cessation might be applied to subgroups only. Attempts to account for this were made in the sensitivity analyses.

Data available within the literature for all costs was sparse despite systematic attempts to find this data. The main conclusions of the model arise due to the low cost of the available treatments thus if these costs were much higher the results may favour testing. The costs for pharmacological treatments were obtained from the BNF and thus can be considered as reasonable estimates and it is highly unlikely that the costs would vary enough to affect the model outputs. It is noted that other costs were not included within the cost of treatments e.g. consultation and prescription time however these costs are considered to be constant across the treatments and would exist for any future interventions. Again due to the very low costs of the treatments it is unlikely that the inclusion of these additional costs would have affected the model outputs.

One of the major limitations is the cost attributed to a case of FGR. As discussed earlier there was no evidence available within the literature to inform this parameter. Cost data were thus obtained from BWH data and this necessitated the use of birth weight<2500g

as the threshold for disease. The limitations of this reference standard have been previously discussed. Attempts were made to make these costs as robust as possible by looking at sufficient years of data to get adequate numbers but to ensure data were contemporary. The costs could only include duration of stay on a neonatal unit or transitional care ward. Thus costs were not captured for mode of delivery, midwifery input, consultant obstetrician input, *in-utero* or *ex-utero* transfer etc. Costs were also determined according to birth weight threshold and gestation to give a range of costs and a very conservative cost of £2652 was used to populate the model. Sensitivity analysis around this cost parameter had no effect on the model outputs. Finally as this data were obtained from BWH data it must be recognised that this represents a regional tertiary referral unit and costs may be different for instance at a district general hospital.

This economic evaluation can only be used to assess the cost-effectiveness of the strategies included within in it. Thus the value of new tests and treatments must be assessed as they are developed and their test accuracy and effectiveness has been properly evaluated e.g. first trimester serum screening markers.

Recommendations for practice

In light of the limitations and assumptions that were made within this health economic evaluation there is insufficient evidence to recommend changes to current clinical practice. However, there are considerable recommendations for future research.

Recommendations for research

The results of this health economic evaluation suggest that research should be directed in three areas. The first is for clinical programmes to assess the use of interventions that are effective and of a low-cost with no prior testing in a pregnant population. One such intervention is aspirin. To implement such a policy would require careful consideration of side effects and patient preferences.

There is further research needed to evaluate test accuracy of low-cost first trimester tests that may have levels of accuracy and costs compatible with a testing strategy and can be employed early enough in gestation to allow preventative treatment to be used.

Finally there is a need for more comprehensive evaluation of currently available tests and treatments and their use in combination within a model, in particular using a comprehensive model that looks at multiple outcomes e.g. pre-eclampsia and FGR.

14.6 Conclusion

The conclusion of this health economic evaluation is that at present there are no tests that are suitable for prior testing in an unselected pregnant population. Considering the current available tests and treatments the most likely cost-effective option will be a low-cost effective treatment with an excellent side-effect profile offered to all pregnant women in the first trimester.

CHAPTER 15: CONCLUSION

15.1 Introduction

This thesis performed an HTA in Obstetrics through evaluation of a range of tests and interventions for SGA fetuses and compromise of fetal/neonatal wellbeing. The thesis achieves the main objectives in that it reports:

1. Summary estimates of accuracy of the following tests for restriction of fetal growth and compromise of fetal wellbeing: five serum screening markers, uterine artery Doppler, umbilical artery Doppler, middle cerebral artery Doppler and ductus venosus Doppler.
2. An evaluation of the relationship between study quality and test accuracy
3. Summary estimates of available treatments for restriction of fetal growth and compromise of its wellbeing.
4. A health economic evaluation and decision analytic model of the combined effects of test and treatments on small for gestational age fetuses.

Each of the previous chapters in this thesis included detailed discussion of the main findings and the conclusions in light of any limitations. This chapter focuses on the main findings of the work undertaken and discusses its strengths and limitations leading to general recommendations for research and practice.

15.2 Summary of main findings

15.2.1 Test accuracy findings

- In total 1,157 papers were read in full with 337 included in the reviews with 472,544 women tested. The median number of women included was 33,292 (interquartile range 13,273-40,637). The median number of studies per test was 60 (interquartile range 31-86). The quality of studies was variable with the overall quality being poor.
- The tests overall for prediction of small gestational age fetuses and adverse perinatal outcome demonstrated low predictive accuracy with no tests having a positive LR>5 and a negative LR<0.5.

15.2.2 Effects of study quality on test accuracy

- A total of 195 studies were included. The overall reporting quality of included studies to the STARD criteria was poor (adequate reporting >50% of the time for 62.1% (18/29) of the items. The overall methodological quality was poor (>50% compliance with 57.1% of quality items).
- There was a positive correlation ($p<0.0001$) between study sample size and reporting quality but not with methodological quality. No correlation with geographical area of publication and compliance with quality criteria could be demonstrated.
- Meta-regression analysis showed that no individual quality item had a significant impact on accuracy. There was an association between reporting and methodological quality ($r=0.51$ $p<0.0001$).

- This work demonstrated that the reporting and methodological quality of papers in Obstetrics is improving but that there is still considerable scope for improvement.

15.2.3 Effectiveness of interventions findings

- There were 71 systematic reviews including a total of 733 RCTs reporting on 42 different interventions.
- After considering the results and the quality of evidence anti platelets and multiple micronutrient supplements were the interventions that were considered to be effective in preventing the small for gestational age fetus and suitable for use in all pregnant women. For high risk pregnant women the following were considered to be effective: anti platelets, multiple micronutrient supplements, smoking cessation interventions and progesterone therapy. For prevention/reduction of prenatal mortality ant platelets and antenatal corticosteroids were the interventions shown to be effective.

15.2.4 Health economic evaluation and decision analytic modelling findings

- Testing prior to intervention was not shown to be the most cost-effective strategy in the analyses for all pregnant women. Anti-platelet therapy, without prior testing, was highlighted as potentially cost-effective in preventing fetal growth restriction in this population.
- In high risk women, testing with serum human chorionic gonadotrophin followed by anti-platelet therapy in those that test positive was a potentially cost-effective strategy.

15.3 Strengths of the thesis

To the best of the author's knowledge there have been no previously reported systematic assessments of test accuracy and effectiveness of interventions in this subject area with decision analytic modelling. This thesis used robust and contemporary methods to achieve the aims and objectives. The evaluation of tests and treatments using a decision analytic model allows the combination of testing with many different treatments and allows them to be systematically assessed and compared. It thus allows a very comprehensive overview of the knowledge to date and by consideration of the strengths and limitations allows important recommendations for future research and clinical practice to be made.

15.4 Limitations of the thesis

15.4.1 Test accuracy limitations

The limitations in the test accuracy systematic reviews were related to:

- The primary data- general poor quality of the included primary studies, the unexplained heterogeneity within and across studies.
- The review methods- the need to assess tests in isolation and thus not assess for diagnostic confounding, the fact that some important tests could not be reviewed in the timescale (section 10.5) and that tests were not assessed in combination.

It is felt however, that the robust methods used within these reviews accounted for some of the limitations and that these reviews still represent the most up to date synthesis of the available evidence for the tests investigated. The work performed in this thesis

looking at effect of quality on results of test accuracy suggests that the impact of poor quality may be minimal. These results are thus still valid despite the limitations.

15.4.2 Effectiveness of interventions limitations

The main limitations in the review of systematic reviews of effectiveness in this subject area were:

- The small number of RCTs for some interventions and the small number of participants in some of the RCTs.
- Heterogeneity in populations and outcome measures.
- The exclusion of observational data and thus possible adverse event data and data on some interventions only assessed by observational studies.
- The lack of assessment of the quality of the individual RCTs.
- The lack of data for combinations of interventions.

15.4.3 Limitations for the economic analysis and decision model

The limitations for the economic analysis and decision model are two fold; the limitations attached to the primary test accuracy and effectiveness data that informed the model and limitations due to model design. The design of the model was limited by:

- Assessment of test-treatment combinations in isolation i.e. with no reference to other clinical data such as risk factors.
- The consideration of a single outcome measure of FGR may have led to an underestimate of the cost-effectiveness of the test-treatment strategies.
- The model does not take into account the acceptability of the strategies to women and clinicians nor does it assess side effects.

- The cost data used to inform the model was of poor quality with little information in the literature available with which to compare and validate the costs.

This economic analysis is still valid despite these limitations due to the robust methods used to determine the test accuracy and effectiveness inputs and the use of sensitivity and PSA analysis to assess the uncertainty around the results.

15.5 Recommendations for practice

Despite the limitations identified it is felt that the methods used ensure that the results are still valid. This thesis has demonstrated that the tests reviewed have a limited use in screening/diagnosis for SGA/compromise of fetal and neonatal wellbeing when used in isolation. The main implications of this work are thus not for recommendations for practice but for future research. An effective, affordable and safe intervention applied to all mothers without prior testing is likely to be the most cost-effective strategy in the prevention of fetal growth restriction. At present aspirin appears to be the most likely intervention however further research needs to be performed particularly looking at interventions in combination and side effects prior to recommending a policy of treating all pregnant women without testing.

15.6 Recommendations for research

Further research in this area needs to consider the use of tests in combination and the role that other diagnostic tools, such as risk factor assessment and clinical features, add to the clinical decision making process. This research needs to be robustly designed, primary test evaluation strategies with reference to the quality criteria of the QUADAS

checklist and include a sample size calculation to ensure that results have sufficient power. There is a particular need for researchers in the area of fetal growth restriction to determine the most appropriate reference standards/outcome measures to be used that truly identify the growth restricted fetus. This will ensure that the primary research is not only directed at the fetuses/pregnancies at risk but will facilitate future systematic reviews and meta-analysis.

To ensure that the results of any future economic analysis and decision model analysis can be translated into recommendations for practice there will be a need for models, and the primary research that informs them, to be able to compare both directly and indirectly all combinations of tests and treatments with consideration of side effects. There will also need to be further primary research to determine accurate costs of the outcomes. This research will also need to be directed to look at the impact of these clinical management strategies on multiple outcomes e.g. pre-eclampsia, pre-term birth and fetal growth restriction to ensure that a truly comprehensive clinical management pathway, that is applicable to a general pregnant population within the NHS can be devised.

PREDICTION AND PREVENTION OF FETAL GROWTH RESTRICTION AND COMPROMISE OF FETAL WELLBEING. SYSTEMATIC REVIEWS AND META-ANALYSES WITH MODEL BASED ECONOMIC EVALUATION

By Rachel Katherine Morris

A thesis submitted to the University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

College of Medical and Dental Sciences

The University of Birmingham

September 2010

Volume II

Volume II

APPENDICES AND REFERENCES

Appendix 1: Table summarising the available literature on tests for prediction of fetal growth restriction and compromise of fetal wellbeing. (* These reviews were for fetal growth restriction and compromise of fetal wellbeing).

Search for tests to predict fetal growth restriction/compromise of fetal wellbeing up to April 2008 – 21,437 citations				
Index Test		**Number of citations identified in Reference Manager 11.0**	**Number of relevant citations**	**Number of existing systematic reviews of test accuracy**
History and examination	Clinical risk scoring	15	6	6
	Abdominal palpation	5	2	0
	Symphyseal fundal height measurement	79	28	1
	Fetal movement counting	Not searched	-	1
Ultrasound Biometry	Abdominal circumference (AC)	16,361	Not screened	1
	Head circumference (HC)	916	Not screened	1
	Biparietal diameter (BPD)	246	Not screened	1
	Femur length (FL)	5171	Not screened	0
	Thoracic diameter	16	2	0
	Abdominal diameter	55	5	0
	Abdominal area	22	5	0
	Chest area	3	0	0
	Liver size	8	0	0
	Thigh circumference	15	3	0
	Subcutaneous fat	34	1	0
	FL/AC	Not searched	-	-
	HC/AC	Not searched	-	-
	FL/HC	Not searched	-	-
	Head area/abdominal area	Not searched	-	-
	FL/thigh circumference	Not searched	-	-
	Estimated fetal weight (EFW)	166	24	1
	Fetal ponderal index	5	4	0
	Total intrauterine volume	13	5	0
	Trunk area x crown rump length	Not searched	-	0
	Growth velocity	56	2	0
	Uterine artery	1366	311	Revised in this thesis

Ultrasound Doppler	Umbilical artery	3896*	393*	Revised in this thesis
	Middle cerebral artery	2004*	75*	Revised in this thesis
	Descending aorta	65	9	0
	Internal carotid artery	5	0	0
	Ductus venosus	637*	46	Revised in this thesis
Ultrasound Other	Amniotic fluid measurements	4869*	310	Revised in this thesis
	Placental grade	3	1	0
	Biophysical profile	4417*	110*	Revised in this thesis
Biochemical and haematological	Ooestriol Alpha fetoprotein Human chorionic gonadotrophin Pregnancy associated plasma protein A Inhibin A	1769	257	Revised in thesis
	Beta-1 gylcoprotein	1	1	0
	Human placental lactogen	109	39	0
	Plasma fibronectin	4	1	0
	Placenta protein 10	3	1	0
	Dehydroepiandrosterone sulphate loading test	11	2	0
	Epidermal growth factor	21	3	0
	Amniotic fluid C amnio peptide	0	0	0
	Serum cystine aminopeptidase	4	2	0
	Schwangerschafts protein 1	0	0	0
	Serum alpha 2-macroglobulin	0	0	0
	Maternal leuokocyte zinc	1	0	0
	Form stability index	0	0	0
Other Tests	Customised growth charts	2	2	1
	Cardiotocography	Not searched	-	1
	Fetal ECG	Not searched	-	0
	Fetal magnetocardiography	Not searched	-	0

Appendix 2: Search strategy for electronic database identification of diagnostic test accuracy studies for prediction of small for gestational age fetuses/fetal growth restriction for reviews of serum markers and uterine artery Doppler.

Host: Ovid

Date of search: April 2006

Years covered by search: 1950-2006

MEDLINE

1. (("Small-for-Gestational Age") OR (Small-for-Gestational Age) OR (lbw) OR (small for gestational age) OR (sgr) OR (small for date*) OR (small for gestation*) OR (fgr) OR (iugr) OR (intrauterine growth retard*) OR (intrauterine growth restrict*) OR (fetal growth retard*) OR (fetal growth restrict*) OR (growth restrict*) OR (growth retard*) OR ("Placental Insufficiency"[MeSH]) OR ("Fetal Growth Retardation"[MeSH]) OR ("Infant, Low Birth Weight"[MeSH])) OR (low birth weight)
2. ("Pregnant Women"[MeSH] OR "Pregnancy"[MeSH] OR "Pregnancy Outcome"[MeSH]) OR (pregnan*)
3. Sensitivity and Specificity[MeSH] OR predict* OR diagnose* OR diagnosi* OR diagnost* OR accura*
4. 1 AND 2 AND 3
5. (((((("cohort studies"[mh] OR "case-control studies"[MeSH Terms]) OR "risk"[mh]) OR "epidemiologic factors"[MeSH Terms]) OR ("odds"[tw] AND "ratio*"[tw])) OR ("relative"[tw] AND "risk"[tw])) OR ("case"[tw] AND "control*"[tw])
6. 1 AND 2 AND 5

EMBASE

1. exp Fetus Growth/
2. low birth weight.mp. or exp Low Birth Weight/
3. exp Intrauterine Growth Retardation/
4. Intrauterine Growth Retard$.mp.
5. Growth Retard$.mp.

6. Fetal Growth Retard$.mp.
7. intrauterine growth restrict$.mp.
8. fetal growth restrict$.mp.
9. growth restrict$.mp.
10. exp Small for Date Infant/
11. Small for gestational age.mp.
12. Small for date$.mp.
13. Small for gestation$.mp.
14. fgr.mp.
15. iugr.mp.
16. sga.mp.
17. or/1-16
18. exp pregnancy/
19. exp Pregnant Woman/
20. pregnancy outcome.mp.
21. pregnan$.mp.
22. pregnant wom$.mp.
23. exp Placenta Insufficiency/
24. or/18-23
25. (sensitiv$ or detect$ or accura$ or specific$ or reliab$ or positive or negative or diagnos$).mp. or di.fs.
26. 17 and 24 and 25
27. cohort analysis/
28. exp risk/
29. (odds$ adj ratio$).mp.
30. (relative adj risk).mp.
31. case control study/
32. (case$ adj control$).mp.
33. (causa$ or predispos$).mp.

34. or/27-33

35. 17 and 24 and 34

COCHRANE LIBRARY

1. **small for gestational age** in All Fields in all products
2. **sga** in All Fields in all products
3. **small for date** in All Fields in all products
4. **fgr** in All Fields in all products
5. **lbw** in All Fields in all products
6. **iugr** in All Fields in all products
7. **intrauterine growth retard*** in All Fields in all products
8. **fetal growth retardation** in All Fields in all products
9. **fetal growth retard*** in All Fields in all products
10. **growth restrict*** in All Fields in all products
11. **growth retard*** in All Fields in all products
12. **low birth weight** in All Fields in all products
13. MeSH descriptor **Placental Insufficiency** explode all trees in MeSH products
14. **placental insufficiency** in All Fields in all products
15. (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14)
16. **pregnancy** in All Fields in all products
17. MeSH descriptor **Pregnant Women** explode all trees in MeSH products
18. MeSH descriptor **Pregnancy** explode all trees in MeSH products
19. MeSH descriptor **Pregnancy Outcome** explode all trees in MeSH products
20. (#16 OR #17 OR #18 OR #19)
21. MeSH descriptor **Sensitivity and Specificity** explode all trees in MeSH products
22. **predict* OR diagnose* OR diagnosi* OR diagnost*** in All Fields in all products
23. (#21 OR #22)
24. (#15 AND #20 AND #23)

25. MeSH descriptor **Cohort Studies** explode all trees in MeSH products
26. MeSH descriptor **Case-Control** Studies explode all trees in MeSH products
27. MeSH descriptor **Risk** explode all trees in MeSH products
28. MeSH descriptor **Epidemiologic Factors** explode all trees in MeSH products
29. (odds AND ratio) OR (relative AND risk) OR (case AND control) in All Fields in all products
30. (#25 OR #26 OR #27 OR #28 OR #29)
31. (#15 AND #20 AND #30)

Appendix 3: The Standards of Reporting in Diagnostic Accuracy (STARD) checklist34.

Section and Topic	Item		Code 1 2 3 4
TITLE, ABSTRACT AND KEYWORDS			
	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading " sensitivity and specificity")	**Yes □ No □ Unclear □ N/A □**
INTRODUCTION			
	2	State the research questions or aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups	**Yes □ No □ Unclear □ N/A □**
METHODS			
Participants	**3**	Describe the **study population**: the inclusion and exclusion criteria and the settings and locations where the data were collected.	**Yes □ No □ Unclear □ N/A □**
	4	Describe **participant recruitment**: was this based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	**Yes □ No □ Unclear □ N/A □**
	5	Describe **participant sampling**: was this a consecutive series of participants defined by selection criteria in items 3 and 4? If not, specify how participants were further selected.	**1= consecutive 2=random 3= unclear 4=N/a**
	6	Describe data collection: was data collection planned before the index tests and reference standard were performed (prospective study) or after (retrospective study)?	**1= prospective 2= retrospective 3= unclear 4= N/a**
Test Methods	**7**	Describe the reference standards and its rationale.	**Yes □ No □ Unclear □ N/A □**
	8	Describe technical specifications of material and methods involved, including how and when measurements were taken, or cite references for a) index test or b) reference test	**Yes □ No □ Unclear □ N/A □**
	9	Describe definition of and rationale for the units, cut-off points, or categories of the results of the a) index test and b) reference standard.	**Yes □ No □ Unclear □ N/A □**

	10	Describe the number, training and expertise of the persons executing and reading the a) index tests and b) reference standards.	**Yes** □ **No** □ **Unclear** □ **N/A** □
	11	Were the readers of the a) index test and b) reference standards blind (masked) to the results of the other test? Describe any other clinical information available to the readers.	**Yes** □ **No** □ **Unclear** □ **N/A** □
Statistical Methods	**12**	Describe methods for calculating or comparing methods of a) diagnostic accuracy and the statistical methods used to b) quantify uncertainty (e.g. 95% CI)	**Yes** □ **No** □ **Unclear** □ **N/A** □
	13	Describe methods for calculating test reproducibility, if done.	**Yes** □ **No** □ **Unclear** □ **N/A** □
RESULTS			
Participants	**14**	Report when study was done, including beginning and ending dates of recruitment	**Yes** □ **No** □ **Unclear** □ **N/A** □
	15	Report clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, co morbidity, current treatments, recruitment centres)	**Yes** □ **No** □ **Unclear** □ **N/A** □
	16	Report the number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test.	**Yes** □ **No** □ **Unclear** □ **N/A** □
Test results	**17**	Report time interval from the index tests to the reference standard, and any treatment administered between.	**Yes** □ **No** □ **Unclear** □ **N/A** □
	18	Report distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	**Yes** □ **No** □ **Unclear** □ **N/A** □
	19	Report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	**Yes** □ **No** □ **Unclear** □ **N/A** □
	20	Report any adverse events form performing the index tests or the reference standard.	**Yes** □ **No** □ **Unclear** □ **N/A** □
Estimates	**21**	Report estimates of a) diagnostic accuracy and b) measures of statistical uncertainty (e.g. 95% CI)	**Yes** □ **No** □ **Unclear** □ **N/A** □
	22	Report how indeterminate results, missing responses and outliers of the index tests were handled.	**Yes** □ **No** □ **Unclear** □ **N/A** □
	23	Report estimates of variability of diagnostic accuracy between	**Yes** □ **No** □ **Unclear** □ **N/A** □

		subgroups of participants, readers or centres, if done.	
	24	Report estimates of test reproducibility, if done.	**Yes** □ **No** □ **Unclear** □ **N/A** □
DISCUSSION			
	25	Discuss the clinical applicability of the study findings.	**Yes** □ **No** □ **Unclear** □ **N/A** □

Appendix 4: The Quality Assessment of Diagnostic Accuracy Studies (QUADAS) checklist35.

Description	QUADAS Item	Code 1	2	3
Was the spectrum of patients representative of the patients who will receive the test in practice?	**1**	Yes □	No □	Unclear □
Were selection criteria clearly described?	**2**	Yes □	No □	Unclear □
Is the reference standard likely to correctly classify the target condition?	**3**	Yes □	No □	Unclear □
Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	**4**	Yes □	No □	Unclear □
Did the whole study population or a random selection of the sample, receive verification using a reference standard for diagnosis?	**5**	Yes □	No □	Unclear □
Did patients receive the same reference standard regardless of the index test result?	**6**	Yes □	No □	Unclear □
Was the reference standard independent of the index test?	**7**	Yes □	No □	Unclear □
Was the execution of the index test described in sufficient detail to permit replication of the test?	**8**	Yes □	No □	Unclear □
Was the execution of the reference standard described in sufficient detail to permit its replication?	**9**	Yes □	No □	Unclear □
Were the index test results interpreted without the knowledge of the results of the reference standard?	**10**	Yes □	No □	Unclear □
Were the reference standard results interpreted without knowledge of the index test results?	**11**	Yes □	No □	Unclear □
Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	**12**	Yes □	No □	Unclear □
Were uninterpretable / intermediate test results reported?	**13**	Yes □	No □	Unclear □
Were withdrawals from the study explained?	**14**	Yes □	No □	Unclear □

Appendix 5: Data extraction form for review of Down's syndrome markers to predict small for gestational age fetuses.

Section A: Study Information

1)Ref ID:		**4)Publication year:**	
2)Rev name:		**5)First Author:**	
3)Country:		**6)Language:**	

Section B: Data Retrieval for Down's screening study

Population

7) Healthcare Centre:

Primary care ☐$_1$ Secondary care ☐$_2$ Mixed ☐$_3$ Other ☐$_4$ Unreported ☐$_5$

8) Setting:

In-patient ☐$_1$ Out-patient ☐$_2$ Mixed ☐$_3$ Unreported ☐$_4$ Other ☐$_5$ _____

9) Number of participating centres: _____

10) Gestation at time of index test:

<20 weeks ☐$_1$ 20-24 weeks ☐$_2$ 24-28 weeks ☐$_3$ 28-34 weeks ☐$_4$ 34-37 weeks ☐$_5$ 37-40 weeks ☐$_6$ > 40 weeks ☐$_7$ Unreported ☐$_8$ Other _____

10.i) Mean (range)_____ Unreported ☐$_3$

10.ii) Median (range) _____ Unreported ☐$_3$

11) Pregnancy:

Low Risk ☐$_1$ High Risk ☐$_2$ Unselected ☐$_3$ Unreported ☐$_4$

11.i) State high risk conditions: Unreported ☐$_3$

12) Were patients with the following conditions excluded/not included?

12.i) Previous IUGR: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.ii) Insulin dependent diabetes mellitus: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.iii) Chronic renal disease: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.iv) Systemic lupus erythematosus: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.v) Antiphospholipid syndrome: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$ 12.vi) Chronic hypertension: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$ 12.vii) Pre-eclampsia: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.viii) Foetal chromosomal/structural anomalies: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

13) Did all patients have singleton pregnancies?:

Yes ☐$_1$ No ☐$_2$ ----------/---------- Unreported ☐$_3$

14) Were all patients primigravid?:

Yes ☐$_1$ No ☐$_2$ ----------/---------- Unreported ☐$_3$

15) List other eligibility/ in-/exclusion criteria:

Not applicable ☐$_3$

16) Study population: (describe age (mean +/- SD or median/range), ethnicity, smoking, BMI etc.)

Unreported ☐$_3$

17) Start of patient inclusion (year) :

Unreported ☐3

18) End of patient inclusion (year) :

Unreported ☐3

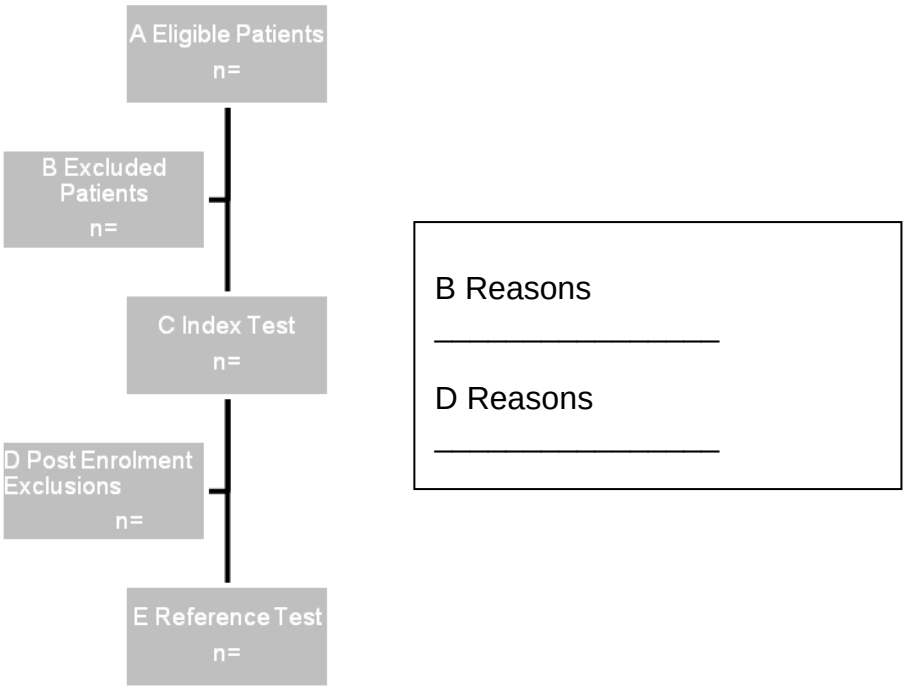
19) Study Design:

cohort ☐1 case control ☐2 RCT/CCT ☐3 cross sectional ☐4 before and after ☐5 case series ☐6 (no _____) other ☐7

19.i) Data collection: prospective ☐1 retrospective ☐2 unreported ☐3 other ☐4

19.ii) Enrolment: consecutive ☐1 arbitary (random) ☐2 unreported ☐3 other ☐4

20) Numbers:



21) Completeness of Verification:

(= E / C x 100 = %)> 90% ☐1 81-90% ☐2 < 81% ☐3

Index Test

22) Description of technique:

Adequate ☐1 Inadequate ☐2

23) Timing of measurement (from delivery):

< 7days ☐1 7-14 days ☐2 14 -28 days ☐3 > 28 days ☐4 Mixture ☐5

Unreported ☐6

23.i) Median gestational age at delivery _____

unreported ☐3

24) Measurement AFP:

Method of sample analysis:

24.i) Test/Analysis method:

RIA ☐1 EIA ☐2 FEIA ☐3 FIA ☐4 MEIA ☐5 Unreported ☐6

24.ii) Laboratory/Machine used_____

unreported ☐3

24.iii) Software used for calculating Mom:

_____Unreported ☐3

24.iv) Cut-off used (and data-set if reported)**:**_____Unreported ☐3

24.v) Any corrections made: (e.g. weight, height, race, IDDM etc)

25) Measurement Beta HCG:

Method of sample analysis:

25.i) Test/Analysis method:

RIA ☐$_1$ EIA ☐$_2$ FEIA ☐$_3$ FIA ☐$_4$ MEIA ☐$_5$ Unreported ☐$_6$

25.ii) Laboratory/Machine used_____

unreported ☐$_3$

25.iii) Software used for calculating Mom:

_____Unreported ☐$_3$

25.iv) Cut-off used (and data –set if

25.v) Any corrections made: (e.g. weight, height, race, IDDM etc)

26) Measurement Ooestriol:

Method of sample analysis:

26.i) Test/Analysis method:

RIA ☐$_1$ EIA ☐$_2$ FEIA ☐$_3$ FIA ☐$_4$ MEIA ☐$_5$ Unreported ☐$_6$

26.ii) Laboratory/Machine used_____unreported ☐$_3$

26.iii) Software used for calculating Mom: _____ __unreported ☐$_3$

26.iv) Cut-off used (and data-set if reported)**:** _____unreported ☐$_3$

26.v) Any corrections made: (e.g. weight, height, race, IDDM etc)

7) Measurement PAPP-A:

Method of sample analysis:

27.i) Test/Analysis method:

RIA ☐$_1$ EIA ☐$_2$ FEIA ☐$_3$ FIA ☐$_4$ MEIA ☐$_5$ Unreported ☐$_6$

27.ii) Laboratory/Machine used_____ unreported ☐3

27.iii) Software used for calculating Mom: _____Unreported ☐3

27.iv) Cut-off used (and data-set if reported)**:**

27.v) Any corrections made: (e.g. weight, height, race, IDDM etc)

28) Measurement Inhibin:

Method of sample analysis:

28.i) Test/Analysis method:

RIA ☐1 EIA ☐2 FEIA ☐3 FIA ☐4 MEIA ☐5 Unreported ☐6

28.ii) Laboratory/Machine used_____ unreported ☐3

28.iii) Software used for calculating Mom: _____Unreported ☐3

28.iv) Cut-off used (and data-set if reported)**:**

28.v) Any corrections made: (e.g. weight, height, race, IDDM etc)

Reference Standard / Outcome

25) Measured blind form diagnostic test: Yes ☐1 No ☐2 Unclear ☐3

26) Measurement for FGR: Birthweight ☐1 Neonatal ponderal index ☐2

Skin fold thickness ☐3 MAC / OFC ☐4 Other ☐5 _____

27) Threshold: < 3rd centile ☐1 < 5th centile ☐2 < 10th centile ☐3 < 25th centile ☐4 > 2SD ☐5 Other ☐6 _____

Unclear ☐7

28) What data set was used to define threshold?

_____

unreported ☐3

29)Timing of measurement: At delivery ☐1 Within 24 hrs ☐2 > 24 hrs ☐3

Mixture ☐4 Unreported ☐5

Results

		Reference Test: Threshold:		
Index test, Measurement:		Positive	Negative	**Total**
	Positive	**TP**	**FP**	
Threshold:	Negative	**FN**	**TN**	
	Total			

Appendix 6: Guide to quality assessment of included studies in review of Down's syndrome markers to predict small for gestational age fetuses.

Feature	Quadas Number	Applicability and criteria fulfilled when
Population spectrum	1	Refers to severity of underlying target condition, demographic features and presence of differential diagnoses and/or co-morbidity. For study to be classified as adequate: Appropriate spectrum – pregnant women, either unselected or selected (high or low risk) in any health care setting. Ideally there was prospective, consecutive recruitment.
Selection Criteria	2	Refers to inclusion/exclusion criteria. For an unselected population this would not be applicable. For a selected population high risk conditions must be explicitly documented. If the inclusion criteria for the categories were not explicitly described then the category was unclear.
Appropriate Reference standard	3	SGA: birth weight < 10^{th} centile adjusted for gestational age and based on local population values and absolute birth weight threshold < 2500g. Severe SGA: birth weight < 5^{th} or < 3^{rd} centile or < 1750g. Neonatal ponderal index < 10^{th} centile, skin fold thickness, and mid-arm circumference/head circumference were also assessed.
Time period between tests	4	Time period needs to be short enough to ensure that target condition does not change. For this review this was always graded as N/A.
Verification Bias	5	If >90% of patients or a random selection of patients received verification with reference standard then answer was yes, even if the reference standard was not the same for all patients. If the number was <90% or a non-random selection then the answer was no. Unclear was utilised when the percentage could not be calculated or no information was given.
Number of reference standards used	6	This is N/A to this review: no invasive reference test.
Independent reference standard	7	The results of the index test are not incorporated in the definition of small for gestational age/fetal growth restriction. For this review the answer will always be yes.
Adequate description of index test	8	To be graded as adequate the description must include: cut-off used, assay used and manufacturer of assay/machine used.
Adequate description of reference standard	9	Birth weight: timing of measurement, scales used, whether baby clothed or not. Neonatal ponderal index: description of birth weight and length measurement as above. Skin fold thickness: description of site of measurement, instrument used and timing of measurement. Mid-arm circumference/ head circumference: see skin fold thickness. If this information was not provided this was classified as unclear.

Blinding of index test	10	For this review this answer will always be yes, as the reference standards can only be performed after delivery. In the case of retrospective analysis of blood samples this will also be yes as fully automated.
Blinding of reference standard	11	To confirm that blinding was present a statement in the text to the effect of "clinicians were blinded/unaware of the results of the --- test". If there was a statement to the contrary the answer was no. If no statement existed the answer was unclear. If test were entirely objective (or an independent laboratory was used) then this was N/A.
Availability of clinical data	12	Clinical data refers to any information relating to the patient obtained by direct observation (e.g. age, sex, symptoms, BMI). If clinical data will be available when the test is interpreted in practice then this should be available when the test is evaluated. In this review the test was fully automated and thus is N/A.
Intermediate results	13	If uninterpretable, failed or intermediate results are documented or no such events occurred then the answer is yes. If it was apparent that such results have occurred but are not reported then the answer was no. If not clear whether all results were reported then answer was unclear.
Withdrawals from study	14	If clear what happened to all patients within the study e.g. flow diagram then answer was yes. If some did not receive both index and reference standard then answer was no.
Intervention	A	If after receiving the index test patients received any medical or surgical intervention then the answer was yes, and the type of intervention recorded. If a statement existed that no intervention was given the answer was no. If no statement existed and no interventions were given then the answer was unclear.

Appendix 7: References for studies included in review of Down's serum screening markers.

Akinbiyi AA. Unexplained elevated maternal serum alpha-fetoprotein in singleton pregnancies as a predictor of fetal risk. *Int J Gynaecol Obstet.* 1996;53(1):17-21.

Audibert F, Benchimol Y, Benattar C, Champagne C, Frydman R. Prediction of preeclampsia or intrauterine growth restriction by second trimester serum screening and uterine Doppler velocimetry. *Fetal Diagn Ther.* 2005;20(1):48-53.

Benn PA, Horne D, Briganti S, Rodis JF, Clive JM. Elevated second-trimester maternal serum hCG alone or in combination with elevated alpha-fetoprotein. *Obstet Gynecol.* 1996;87(2):217-222.

Bernstein IM, Barth RA, Miller R, Capeless EL. Elevated maternal serum alpha-fetoprotein: association with placental sonolucencies, fetomaternal hemorrhage, vaginal bleeding, and pregnancy outcome in the absence of fetal anomalies. *Obstet Gynecol.* 1992;79(1):71-74.

Bewley S, Chard T, Grudzinskas G, Cooper D, Campbell S. Early prediction of uteroplacental complications of pregnancy using Doppler ultrasound, placental function tests and combination testing. *Ultrasound Obstet Gynecol.* 1992;2(5):333-337.

Bloxam DL, Williams NR, Waskett RJ, Stewart SG. Disturbed zinc metabolism and reduced birthweight related to raised maternal serum alpha-fetoprotein in normal human pregnancies. *Acta Obstet Gynecol Scand.* 1994;73(10):758-764.

Brajenovic-Milic B, Tislaric D, Zuvic-Butorac M, Back J, PErovic O, Ristic S, Mimica M, Kapovic M. Elevated second-trimester free beta-hCG as an isolated finding and pregnancy outcomes. *Fetal Diagn Ther.* 2004;19(6):483-487.

Brazerol WF, Grover S, Donnenfeld AE. Unexplained elevated maternal serum alpha-fetoprotein levels and perinatal outcome in an urban clinic population. *Am J Obstet Gynecol.* 1994;171(4):1030-1035.

Bremme K, Eneroth P, Nilsson B, Marsk L, Hagenfeldt L. Outcome of pregnancy in relation to maternal serum alpha-fetoprotein levels in the second trimester. An evaluation of a screening program and a longitudinal follow-up. *Gynecol Obstet Invest.* 1988;26(3):191-205.

Brock DJ, Barron L, Raab GM. The potential of mid-trimester maternal plasma alpha-fetoprotein measurement in predicting of low birth weight. *Br J Obstet Gynaecol.* 1980;87(7):582-585.

Buckland CM, Thom H, Campbell AG. Maternal serum alpha-fetoprotein levels in low birth weight singleton pregnancies. *J Perinat Med.* 1984;12(3):127-132.

Burton BK. Outcome of pregnancy in patients with unexplained elevated or low levels of maternal serum alpha-fetoprotein. *Obstetrics & Gynecology.* 1988;72(5):709-713.

Capeless EL, Kelleher PC, Walters CP. Elevated maternal serum alpha-fetoprotein levels and maternal risk factors: Their association with pregnancy complications. *J Reprod Med.* 1992;37(3):257-260.

Chapman SJ, Brumfield CG, Wenstrom KD, Dubard MB. Pregnancy outcomes following false-positive multiple marker screening tests. *Am J Perinat.* 1997;14(8):475-478.

Chard T, Rice A, Kitau MJ, Hird V, Grudzinskas JG, Nysenbaum AM. Mid-trimester levels of alphafetoprotein in the screening of low birthweight. *Br J Obstet Gynaecol.* 1986;93(1):36-38.

Chitayat D, Farrell SA, Huang T, Meier C, Wyatt PR, Summers AM. Double-positive maternal serum screening results for Down's syndrome and open neural tube defects: An indicator for fetal structural or chromosomal abnormalities and adverse obstetric outcomes. *Am J Obstet Gynecol.* 2002;187(3):758-763.

Cho S, Durfee KK, Keel BA, Parks LH. Perinatal outcomes in a prospective matched pair study of pregnancy and unexplained elevated or low AFP screening. *J Perinat Med.* 1997;25(6):476-483.

Cox MU, Kingdom J, Whittle M, McNay M, Fleming J, Bowman A, Crossley J, Aitken D. [Unexplained increase in maternal plasma in alpha-fetoproteins in the second trimester of pregnancy: perinatal results]. *Rev Chil Obstet Ginecol.* 1995;60(3):174-180.

Cusick W, Rodis JF, Vintzileos AM, Albini SM, McMahon M, Campbell WA. Predicting pregnancy outcome from the degree of maternal serum alpha-fetoprotein elevation. *J Reprod Med.* 1996;41(5):327-332.

Di MM, Piazzi G, Ferrari A, Lotzniker M, Gerola O, Spinillo A. A simultaneous evaluation of second trimester serum AFP, hCG and unconjugated ooestriol as predictors of small for gestational age births. *Italian Journal of Gynaecology & Obstetrics.* 1998;10(4):155-159.

Doran TA, Valentine GH, Wong PY, Wielgosz G, Benzie RJ, Soltan HC, Jenner MR, Morland PA, Montogomery RJ, Allen LC. Maternal serum alpha-fetoprotein screening: report of a Canadian pilot project. *CMAJ.* 1987;137(4):285-293.

Dugoff L, Hobbins JC, Malone FD, Porter TF, Luthy D, Comstock CH, Hankins G, Berkowitz RL, Merkatz I, Craigo SD, Timor-Tritsch IE, Carr SR, Wolfe HM, Vidavar J, D'Alton ME. First-trimester maternal serum PAPP-A and free-beta subunit human chorionic gonadotropin concentrations and nuchal translucency are associated with obstetric complications: a population-based screening study (the FASTER Trial). *Am J Obstet Gynecol.* 2004;191(4):1446-1451.

Dugoff L, Hobbins JC, Malone FD, Vidaver J, Sullivan L, Canick JA, Lambert-Messerlian GM, Porter TF, Luthy DA, Comstock CH, Saade G, Eddleman K, Merkatz IR, Craigo SD, Timor-Tritsch IE, Carr SR, Wolfe HM, D'Alton ME, FASTER Trial Research Consortium. Quad screen as a predictor of adverse pregnancy outcome. *Obstet Gynecol.* 2005;106(2):260-267.

Dungan JS, Shulman LP, Phillips OP, Simpson JL, Meyer NL, Grevengood C, Elias S. Positive serum screening for fetal Down's syndrome does not predict adverse pregnancy outcome in absence of fetal aneuploidy. *J Soc Gynecol Investig.* 1994;1(1):55-58.

Duric K, Skrablin S, Lesin J, Kalafatic D, Kuvacic I, Suchanek E. Second trimester total human chorionic gonadotropin, alpha-fetoprotein and unconjugated oestriol in predicting pregnancy complications other than fetal aneuploidy. *Eur J Obstet Gynecol Reprod Biol.* 2003;110(1):12-15.

Endres LK, Krotz S, Grobman WA. Isolated low second-trimester maternal serum beta-human chorionic gonadotropin is not associated with adverse pregnancy outcome. *Am J Obstet Gynecol.* 2003;189(3):755-757.

Evans J, Stokes IM. Outcome of pregnancies associated with raised serum and normal amniotic fluid alpha fetoprotein concentrations. *Br Med J (Clin Res Ed).* 1984;288(6429):1494.

Ghosh A, Tang MH, Tai D, Nie G, Ma HK. Justification of maternal serum alphafetoprotein screening in a population with low incidence of neural tube defects. *Prenat Diagn.* 1986;6(2):83-87.

Gonen R, Perez R, David M, Dar H, Merksamer R, Sharf M. The association between unexplained second-trimester maternal serum hCG elevation and pregnancy complications. *Obstet Gynecol.* 1992;80(1):83-86.

Gordon YB, Lewis JD, Leighton M, Kitau MJ, Clarke PC, Chard T. Maternal serum alpha-fetoprotein levels as an index of fetal risk. *Am J Obstet Gynecol.* 1979;133(4):422-424.

Haddad B, Abirached F, Louis-Sylvestre C, Le Blond J, Paniel BJ, Zorn JR. Predictive value of early human chorionic gonadotrophin serum profiles for fetal growth retardation. *Hum Reprod.* 1999;14(11):2872-2875Haddow JE, Kloza EM, Smith DE, Knight GJ. Data from an alpha-fetoprotein pilot screening program in Maine. *Obstet Gynecol.* 1983;62(5):556-60.

Haddow JE, Knight GJ, Kloza EM, Palomaki GE. Alpha-fetoprotein, vaginal bleeding and pregnancy risk. *Br J Obstet Gynaecol.* 1986;93(6):589-593.

Haddow JE, Palomaki GE, Knight GJ. Can low birth weight after elevated maternal serum alpha-fetoprotein be explained by maternal weight? *Obstet Gynecol.* 1987;70(1):26-28.

Hamilton MP, Abdalla HI, Whitfield CR. Significance of raised maternal serum alpha-fetoprotein in singleton pregnancies with normally formed fetuses. *Obstet Gynecol.* 1985;65(4):465-470.

Hayashi M, Kohzu H. [Clinical studies on maternal serum alpha-fetoprotein (MSAFP) of 532 cases--the median value, pregnancy outcomes and clinical evaluation]. *Nippon Sanka Fujinka Gakkai Zasshi.* 1992;44(10):1211-1218.

Heikkila A, Makkonen N, Heinonen S, Kirkinen P. Elevated maternal serum hCG in the second trimester increases prematurity rate and need for neonatal intensive care in primiparous preeclamptic pregnancies. *Hypertens Pregnancy.* 2001;20(1):99-106.

Heinonen S, Ryynanen M, Kirkinen P, Saarikoski S. Elevated midtrimester maternal serum hCG in chromosomally normal pregnancies is associated with preeclampsia and velamentous umbilical cord insertion. *Am J Perinatol.* 1996;13(7):437-441.

Heinonen S, Ryynanen M, Kirkinen P. The effects on fetal development of high alpha-fetoprotein and maternal smoking. *Am J Public Health.* 1999;89(4):561-563.

Hershkovitz R, Erez O, Sheiner E, Landau D, Mankuta D, Mazor M. Elevated maternal mid-trimester chorionic gonadotropin > or =4 MoM is associated with fetal cerebral blood flow redistribution. *Acta Obstet Gynecol Scand.* 2003;8(1)2:22-27.

Hershkovitz R, De SM, Kingdom J. Mid-trimester placentation assessment in high-risk pregnancies using maternal serum screening and uterine artery Doppler. *Hypertens Pregnancy.* 2005;24(3):273-280.

Jauniaux E, Gulbis B, Tunkel S, Ramsay B, Campbell S, Meuris S. Maternal serum testing for alpha-fetoprotein and human chorionic gonadotropin in high-risk pregnancies. *Prenat Diagn.* 1996;16(12):1129-1135.

Kavak ZN, Basgul A, Elter K, Uygur M, Gokaslan H. The efficacy of first-trimester PAPP-A and free betahCG levels for predicting adverse pregnancy outcome. *J Perinat Med.* 2006;34(2):145-148.

Kim SY, Kim SK, Lee JS, Kim IK, Lee K. The prediction of adverse pregnancy outcome using low unconjugated oestriol in the second trimester of pregnancy without risk of Down's syndrome. *Yonsei Med J.* 2000;41(2):226-229.

Kiran TS, Bethel J, Bhal PS. Correlation of abnormal second trimester maternal serum alpha-fetoprotein (MSAFP) levels and adverse pregnancy outcome. *J Obstet Gynaecol.* 2005;25(3):253-256.

Krantz D, Goetzl L, Simpson JL, Thom E, Zachery J, Hallahan TW, Silver R, Pergament E, Platt LD, Filkins K, Johnson A, Mahoney M, Hogge WA, Wilson RD, Mohide P, Hershey D, Wapner R, First Trimester Maternal Serum Biochemistry and Fetal Nuchal Translucency Screening (BUN) Study Group. Association of extreme first-trimester free human chorionic gonadotropin-beta, pregnancy-associated plasma protein

A, and nuchal translucency with intrauterine growth restriction and other adverse pregnancy outcomes. *Am J Obstet Gynecol.* 2004;191(4):1452-1458.

Kuo PL, Lin CC, Lin YH, Guo HR. Placental sonolucency and pregnancy outcome in women with elevated second trimester serum alpha-fetoprotein levels. *J Formos Med Assoc.* 2003;102(5):319-325.

Kwik M, Morris J. Association between first trimester maternal serum pregnancy associated plasma protein-A and adverse pregnancy outcome. *Aust N Z J Obstet Gynaecol.* 2003;43(6):438-442.

Legge M, Duff GB, Pike L, Aickin DR. Second trimester maternal serum alpha-fetoprotein as an indicator of fetal risk. *Aust N Z J Obstet Gynaecol.* 1985;25(4):266-268.

Lepage N, Chitayat D, Kingdom J, Huang T. Association between second-trimester isolated high maternal serum maternal serum human chorionic gonadotropin levels and obstetric complications in singleton and twin pregnancies. *Am J Obstet Gynecol.* 2003;188(5):1354-1359.

Lieppman RE, Williams MA, Cheng EY, Resta R, Zingheim R, Hickok DE, Luthy DA. An association between elevated levels of human chorionic gonadotropin in the midtrimester and adverse pregnancy outcome. *Am J Obstet Gynecol.* 1993;168(6 Pt 1):1852-1856.

Markestad T, Bergsjo P, Aakvaag A, Lie RT, Jacobsen G, Hoffman HJ, Bakketeig LS. Prediction of fetal growth based on maternal serum concentrations of human chorionic gonadotropin, human placental lactogen and oestriol. *Acta Obstet Gynecol Scand - Supplt.* 1997;165:50-55.

Milunsky A, Jick SS, Bruell CL, Maclaughlin DS, Tsung YK, Jick H, Rothman KJ, Willett W. Predictive values, relative risks, and overall benefits of high and low maternal serum alpha-fetoprotein screening in singleton pregnancies: new epidemiologic data. *Am J Obstet Gynecol.* 1989;161(2):291-297.

Milunsky A, Nebiolo L. Maternal serum triple analyte screening and adverse pregnancy outcome. *Fetal Diagn Ther.* 1996;11(4):249-253.

Miyakoshi K, Tanaka M, Gabionza D, Ishimoto H, Miyazaki T, Yoshimura Y. Prediction of smallness for gestational age by maternal serum human chorionic gonadotropin levels and by uterine artery Doppler study. *Fetal Diagn Ther.* 2001;16(1):42-46.

Morssink LP, Kornman LH, Beekhuis JR, De Wolf BT, Mantingh A. Abnormal levels of maternal serum human chorionic gonadotropin and alpha-fetoprotein in the second trimester: relation to fetal weight and preterm delivery. *Prenat Diagn.* 1995;15(11):1041-1046.

Mwambingu FT. Association between unaccountably high maternal alpha feto-protein and increased fetal risk. *Acta Obstet Gynecol Scand* 1985;64(3):275-276.

Naylor CS, Porto M, Cohen B, Garite TJ. Pregnancy outcome in Hispanic patients with unexplained positive triple marker screening for Down's syndrome. *J Matern Fetal Med.* 2001;10(1):20-22.

Odibo AO, Sehdev HM, Stamilio DM, Macones GA. Evaluating the thresholds of abnormal second trimester multiple marker screening tests associated with intra-uterine growth restriction. *Am J Perinatol.* 2006;23(6):363-7.

Ogle R, Jauniaux E, Pahal GS, Dell E, Sheldrake A, Rodeck C. Serum screening for Down's syndrome and adverse pregnancy outcomes: a case-controlled study. *Prenat Diagn.* 2000;20(2):96-99.

Onderoglu LS, Kabukcu A. Elevated second trimester human chorionic gonadotropin level associated with adverse pregnancy outcome. *Int J Gynaecol Obstet.* 1997;56(3):245-249.

Ong CY, Liao AW, Spencer K, Munim S, Nicolaides KH. First trimester maternal serum free beta human chorionic gonadotrophin and pregnancy associated plasma protein A as predictors of pregnancy complications. *BJOG.* 2000;107(10):1265-1270.

Pergament E, Stein AK, Fiddler M, Cho NH, Kupferminc MJ. Adverse pregnancy outcome after a false-positive screen for Down's syndrome using multiple markers. *Obstet Gynecol.* 1995;86(2):255-258.

Pilalis A, Souka AP, Antsaklis P, Daskalakis G, Papantoniou N, Mesogitis S, Antsaklis A. Screening for pre-eclampsia and fetal growth restriction by uterine artery Doppler and PAPP-A at 11-14 weeks' gestation. *Ultrasound Obstet Gynecol.* 2007;29(2):135-140.

Roop AP, Boughman JA, Blitzer MG. Study of the relationship between elevated maternal serum alpha-fetoprotein and adverse pregnancy outcome. *Maryland Medical Journal.* 1991;40(9):779-784.

Secher NJ, Eriksen PS, Hansen PK, Lenstrup C, Thisted J. High maternal serum alpha-fetoprotein and low birthweight. A study concerning possible correlation between maternal weight, smoking, and serum alpha-fetoprotein and subsequent birthweight and gestational age. *Am J Perinatol.* 1985;2(2):78-81.

Simpson JL, Palomaki GE, Mercer B, Haddow JE, Andersen R, Sibai B, Elias S. Associations between adverse perinatal outcome and serially obtained second- and third-trimester maternal serum alpha-fetoprotein measurements. *Am J Obstet Gynecol.* 1995;173(6):1742-1748.

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. *J Clin Endocrinol Metab.* 2002;8(4)7:1762-1767.

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

Sritippayawan S, Vachirasrisoontra C. Adverse pregnancy outcomes after a false-positive second trimester serum screen for Down's syndrome in Thai pregnant women. *J Med Assoc Thai.* 2005;88(4):449-454.

Summers AM, Huang T, Meier C, Wyatt PR. The implications of a false positive second-trimester serum screen for Down's syndrome. *Obstet Gynecol.* 2003;101(6):1301-1306.

Tanaka M, Natori M, Ishimoto, Gohda N, Kiyokawa K, Yamauchi J, Miyazaki T, Kpbayashi T, Nozawa S. [Screening for adverse perinatal outcome by biomarkers in maternal serum; a comparison of elevated maternal serum alpha-fetoprotein versus human chorionic gonadotropin]. *Nippon Sanka Fujinka Gakkai Zasshi.* 1994;46(11):1229-1233.

Towner D, Gandhi S, El KD. Obstetric outcomes in women with elevated maternal serum human chorionic gonadotropin. *Am J Obstet Gynecol.* 2006;194(6):1676-81..

Tul N, Pusenjak S, Osredkar J, Spencer K, Novak-Antolic Z. Predicting complications of pregnancy with first-trimester maternal serum free-betahCG, PAPP-A and inhibin-A. *Prenat Diagn.* 2003;23(12):990-996.

Wald N, Cuckle H, Stirrat GM, Bennett MJ, Turnbull AC. Maternal serum-alpha-fetoprotein and low birth-weight. *Lancet.* 1977;2(8032):268-270.

Wald NJ, Cuckle HS, Boreham J, Turnbull AC. Maternal serum alpha-fetoprotein and birth weight. *Br J Obstet Gynaecol.* 1980;87(10):860-863.

Waller DK, Lustig LS, Cunningham GC, Feuchtbaum LB, Hook EB. The association between maternal serum alpha-fetoprotein and preterm birth, small for gestational age , preeclampsia, and placental complications. *Obstet Gynecol.* 1996;88(5):816-822.

Weiner CP, Grant SS, Williamson RA. Relationship between second trimester maternal serum alpha-fetoprotein and umbilical artery Doppler velocimetry and their association with preterm delivery. *Am J Perinatol.* 1991;8(4):263-268.

Wenstrom KD, Sipes SL, Williamson RA, Grant SS, Trawick DC, Estle LC. Prediction of pregnancy outcome with single versus serial maternal serum alpha-fetoprotein tests. *Am J Obstet Gynecol.* 1992;167(6):1529-1533.

Wenstrom KD, Owen J, Davis RO, Brumfield CG. Prognostic significance of unexplained elevated amniotic fluid alpha-fetoprotein. *Obstet Gynecol.* 1996;87(2):213-216

Westergaard JG, Teisner B, Hau J, Grudzinskas JG. Placental protein measurements in complicated pregnancies. I. Intrauterine growth retardation. *Br J Obstet Gynaecol.* 1984;91(12):1216-1223.

Williams MA, Hickok DE, Zingheim RW, Luthy DA, Kimelman J, Nyberg DA, Mahon BS. Elevated maternal serum alpha-fetoprotein levels and midtrimester placental abnormalities in relation to subsequent adverse pregnancy outcomes. *Am J Obstet Gynecol.* 1992;167(4 Pt 1):1032-1037.

Yaron Y, Cherry M, Kramer RL, O'Brien JE, Hallak M, Johnson MP, Evans MI. Second-trimester maternal serum marker screening: maternal serum alpha-fetoprotein, beta-human chorionic gonadotropin, oestriol, and their various combinations as predictors of pregnancy outcome. *Am J Obstet Gynecol.* 1999;181(4):968-974.

Yaron Y, Ochshorn Y, Heifetz S, Lehavi O, Sapir Y, Orr-Urtreger A. First trimester maternal serum free human chorionic gonadotropin as a predictor of adverse pregnancy outcome. *Fetal Diagn Ther.* 2002;17(6):352-356.

Zarzour SJ, Gabert HA, Diket AL, St AM, Miller JM, Jr. Abnormal maternal serum alpha fetoprotein and pregnancy outcome. *J Matern Fetal Med.* 1998;7(6):304-307.

Appendix 8: Study characteristics of included studies for maternal serum biochemical (Down's syndrome) screening to predict small for gestational age fetuses.

First Author (year)	Population Age (country/study design)	No of women analysed	Gestational age at test (weeks)	Incidence of SGA (%)	Reference standard SGA	Details of Index test
Akinbiyi (1996)	INC: singletons, 12% primips, EXC: structural and chromosomal anomalies 18-47 years (UK) (case control, matched, index test)	300	16-18	7.33	BW <2500g	AFP,RIA >2.0 MoM
Audibert (2005)	INC: double test and uterine artery Doppler 18-26 weeks, singleton, primips 48.5% EXC: structural and chromosomal abnormalities, multiple pregnancies, increased NT, delivery<24 weeks, 8 lost to follow up. Mean age 30.9+/-4.5 years (France) (cohort)	2615	14-18	8.70	BW<10th centile (local)	Method not reported AFP>1.5MoM HCG>1.5 MoM, >2.0MoM
Benn (1996)	INC: singletons EXC: structural and chromosomal anomalies, IDDM Age not reported (USA) (case control, matched test)	1079	15-21.9	3.06	BW<10th centile	HCG, Method not reported >3 MoM

Bernstein (1992)	INC: singletons EXC: structural and chromosomal abnormalities 26.3+/-4.3 years (USA) (cohort, prospective)	234	17-21	10.7	SGA (no threshold)	AFP, Method not reported >2.0 MoM
Bewley (1992)	INC: singletons Age not reported (UK) (Cohort, prospective)	172	16-24	14.5	BW<10th centile (local)	AFP and HCG,RIA >90th centile PAPP-A, RIA >90th and <10th centile
Bloxam (1994)	INC: singletons EXC: structural and chromosomal abnormalities Age not reported (UK)(cohort)	147	16-18	14.3	BW≤10th centile	AFP, Method not reported >1.7MoM
Brajenovic-Milic (2004)	INC: singletons, primips 58.5% EXC: IDDM, structural and chromosomal anomalies, false positive NTD screen (MSAFP≥2.o MoM), screen positive Down's test ≥1:250, smokers Mean age 27.9+/-4.3 years (Croatia) (cohort)	1507	15-20	4.45	BW<10th centile	Free ßHCG Method not reported ≥2.0 MoM
Brazerol (1994)	INC: singletons EXC: structural and chromosomal abnormalities Mean 23.8 years	774	15-20	4.52	IUGR (no threshold)	AFP, Method not reported

	(USA)(cohort, prospective)					≥2.0MoM
Bremme (1988)	INC: singletons EXC: structural and chromosomal abnormalities Age not reported (Sweden)(case control)	222	16-17	7.66	SGA (no threshold)	AFP, RIA (Behringwerke) >83.3µg/l
Brock (1980)	INC: singletons Age not reported (UK)(Case control, retrospective, outcome)	226	15-22	50.0	BW<2500g	AFP, RIA ≥1.0,1.5,2.0,2.5,3.0 MoM
Buckland (1984)	INC: singletons Age not reported (UK)(Case control, retrospective, outcome)	325	16-20	62.5	BW<10th centile (sex, parity, local)	AFP,RIA ≥2.0 MoM
Burton (1988)	INC: screening programme EXC: oligohydramnios Age not reported (USA)(Case control, nested cohort)	15512	16-18	7.33	BW<2500g	AFP, RIA/EIA (Amersham/Abbott) >2.5, <0.5 MoM
Capeless (1992)	INC: screening programme Age not reported (USA) (cohort, prospective)	358	16-20	3.91	BW<10th centile	AFP, Method not reported >2.0 MoM
Chapman	INC: maternal age ≥30, amniocentesis EXC: structural and	1135	15-20	3.44	BW<10th	Triple test, RIA

(1997)	chromosomal anomalies (USA) (cohort, retrospective)				centile (local)	DS>1:190
Chard (1986)	INC: singletons EXC: delivery < 28 weeks Age not reported (UK) (cohort, prospective)	476	15-18	10.3 17.6	BW≤2500g BW<10th centile (local)	AFP, Method not reported ≥90th centile
Chitayat (2002)	INC: singletons Age not reported (Canada) (case control, test)	1134	Second trimester	3.35	SGA no threshold	DS≥1:385 and AFP≥2.2MoM Method not reported
Cho (1997)	INC: singletons EXC: structural and chromosomal abnormalities Mean 25.8 +/- 5.8 years (USA) (case control, prospective, matched index test)	255	14-20	8.23	BW<10th centile	AFP, RIA (Kallastaad) ≤0.5, ≥2.5, 4.0 MoM
Cox (1995)	INC: singletons EXC: structural and chromosomal anomalies, invasive procedures, birth < 24 weeks Age not reported (Scotland) (cohort)	15705	16-20	2.29	BW<5th centile (local)	AFP, RIA >2.0 MoM
Cusick (1996)	INC: singletons, 47% primips, no fetoplacental abnormality EXC: structural and chromosomal anomalies, 2 placental abruption, TOP	333	15-20	10.8	BW<10th centile	AFP, Method not reported

	Mean age 27.1 (15-42) years (USA) (cohort, retrospective)				(local, sex)	≥2.5 ≥3.0 MoM
Di Mario (1998)	INC: singletons EXC: structural and chromosomal anomalies, previous PE, IDDM, delivery < 26 weeks Mean 30.6 +/- 3.6 years (Italy) (cohort)	547	16-18	11.6	BW<10th centile (local)	RIA (Johnson and Johnson) AFP and HCG ≥2.0 MoM UE3 ≤0.7MoM
Doran (1987)	INC: patients at low genetic risk, singleton Age not reported (Canada) (cohort, prospective)	7307	16-18	1.35	BW<10th centile	AFP, RIA (WHO Behring) >2.0 MoM
Dugoff (2004)	INC: singletons, 45.1% primips EXC: IDDM, structural and chromosomal anomalies Mean age 30.1+/-5.77 years (16-53) (USA) (cohort, prospective)	33995	10+3-13+6	8.80 3.82	BW<10th centile (local) BW<5th centile (local)	PAPP-A ELISA (Diagnostics, Texas) ≤10th, <5th, ≤ 1st centile
Dugoff (2005)	EXC: structural and chromosomal anomalies Mean age 30.2 +/- 5.71 (16-53) years (USA) (cohort, prospective)	33145	15-18+6	8.90 3.90	BW<10th centile (local) BW≤5th	AFP and HCG, Chemiluminescent immunoassay (Diagnostics)

					centile	≥2.0 MoM UE3 RIA (Diagnostics) ≤0.5MoM Inhibin A, ELISA (Serotec) ≥2.0MoM
Dungan (1994)	INC: singleton EXC: structural and chromosomal abnormalities, maternal age >35 years (USA) (case control, matched test)	198	15-20	5.66	BW<5th centile	DS≥1:270
Duric (2003)	INC: singletons EXC: structural and chromosomal anomalies Age <35 years (Croatia) (cohort, retrospective)	673	15-22	5.20	BW<10th (local)	RIA AFP≥2.0 MoM Total HCG ≥2.02MoM UE3 ≤0.74MoM
Endres (2003)	INC: AFP>0.5 but <2.0 MoM, HCG≤0.5 MoM and oestriol >0.6 and <2.0 MoM Mean age 30+/-6 years (USA) (case control matched)	438	15-20	7.30	BW<2500g	ßHCG Method not reported ≤0.5 MoM
Evans (1984)	INC: screening programme Age not reported	220	16-18	7.73 3.18	BW<2500g BW<10th	AFP, RIA (Amersham)

	(UK) (case control, unmatched, index test)				centile	>95th centile
Ghosh (1986)	EXC: structural and chromosomal anomalies Age not reported (Hong Kong) (cohort, prospective)	9838	15-20	3.03	SGA (no threshold)	AFP, RIA >2.0,≥2.8,3.0,4.0 MoM
Gonen (1992)	INC: HCG>2.5 MoM, singleton, USS dating EXC: structural and chromosomal anomalies, AFP>2.5 MoM Age not reported (Israel) (cohort)	493	16-20	7.91	BW<10th centile	HCG, method not reported, (Delfia), >2.5MoM
Gordon (1979)	INC: singletons EXC: structural and chromosomal anomalies, delivery < 28 weeks Age not reported (UK) (cohort, prospective)	828	16-22	4.35	BW<2500g	AFP, RIA >95th centile
Haddad (1999)	INC: singleton, IVF, primips 86% Mean age 33.6+/-4.2 (France) (cohort, retrospective)	180	13-35 days	10.60	BW<10th (local)	HCG, Method not reported <10th, >90th centile
Haddow (1983)	INC: singletons EXC: structural and chromosomal anomalies Age not reported (USA) (cohort, prospective)	2984	15-20	4.50	BW<2500g	AFP, RIA (Oxford) ≥2.0,3.0 MoM

Haddow (1986)	INC: singletons EXC: neural tube defects Age not reported (USA) (cohort)	6531	15-20	3.94	BW<2500g	AFP, RIA ≥2.0MoM
Haddow (1987)	INC: singletons EXC: structural and chromosomal anomalies Age not reported (USA) (cohort, prospective)	9507	15-20	4.08	BW<2500g	AFP, RIA (Maine) ≥2.0 MoM
Hamilton (1985)	INC: singletons EXC: structural and chromosomal anomalies Age not reported (Scotland) (case control, prospective, matched, index test)	372	16-20	15.90 17.70 4.30 9.95	BW<2500g BW<10th centile (sex, local, parity) BW<1500g BW<5th centile (sex, local, parity)	AFP, Method not reported >2.5 MoM
Hayashi (1992)	INC: screening programme Age not reported (Japan) (cohort, prospective)	532	12-19	1.50	SGA (no threshold)	AFP, Method not reported ≥2.5 MoM

Heikkila (2001)	INC: singletons, primips, pre-eclampsia EXC: structural and chromosomal anomalies Mean age 26.8+/-5.1 years (Finland) (cohort, prospective)	487 471	15-16	2.22 2.95	BW<10th centile BW<2500g	HCG, Immunoassay (Abbott) ≥2.5MoM
Heinonen (1996)	INC: singletons EXC: structural and chromosomal anomalies, pregnancy loss < 24 weeks Mean age not reported (Finland) (case control, matched, test)	5290	15	4.63	BW<2500g	Total ßHCG (IMX Abbott) ≥2.0 MoM, >4 MoM
Heinonen (1999)	INC: singletons EXC: structural and chromosomal anomalies, women that stopped smoking during study Mean age 27.4 years (Finland) (cohort)	1421	15-18	12.30 19.60	BW<2500g BW<10th centile (sex)	AFP< RIA (Clinical chemistry) >2.5 MoM
Hershkovitz (2003)	EXC: structural and chromosomal anomalies Mean age (AFP≥4.0MoM) 29.9+/-10.1 years (Canada) (cohort, prospective)	121	15-18	3.31	BW<10th centile	HCG, Method not reported ≥4.0MoM
Hershkovitz (2005)	INC: chronic hypertension, previous PE, thrombophilia Median age 29 (21-40) (Canada) (cohort)	88	15-18	26.1	BW<10th centile (sex)	Method not reported AFP >2.0 MoM HCG ≥3.0MoM
Jauniaux	INC: singletons, abnormal uterine artery Doppler EXC: structural	41	20-24	39.00	BW<10th	AFP, FEIA

(1996)	and chromosomal anomalies Age not reported (UK) (Cohort)				centile	(Hybritech) ≥2.5 MoM IRMA (Biomeriuex) Total ßHCG>2.5MoM Free ßHCG>2.5MoM
Kavak (2006)	INC: singletons, 50-4% primips EXC: IDDM, chronic hypertension, fetal abnormalities Mean age 30.4+/-5 years (Turkey) (cohort)	476	First trimester	7.35	BW<10th centile	PAPP-A, Random access immunoassay (Kryptor) <0.69MoM, <0.4 MoM(roc determined)
Kiran (2005)	INC: singletons, low risk EXC: structural and chromosomal anomalies Mean age not reported. (UK) (cohort)	6297	Second trimester	4.10	BW<2500g	AFP, Method not reported >2.0 MoM
Kowalczyk (1998)	INC: singletons, 31.7% primips, AFP and HCG>2.0 MoM EXC: structural and chromosomal anomalies Age <35 years (USA) (cohort)	309	15-21	8.74	BW<10th	UE3, RIA ≤0.75 MoM

Krantz (2004)	INC: first trimester screening EXC: chromosomal and structural anomalies Age not reported (USA) (cohort, retrospective)	6276	10+4 – 13+6	6.26	BW<10th centile (ga, local, sex)	Free ßHCG <1st and <5th centile, >90th and 99th centile Papp-a <1st and 5th centile, >90th and 99th centile
Kuo (2003)	INC: singletons EXC: structural and chromosomal anomalies, abnormal HCG or Down's risk, IDDM. Mean age 28.3 +/-0.3 years (Taiwan) (case control, unmatched, index test)	247	15-20	4.70 8.80	BW<2500g BW<10th centile	AFP, Method not reported >2.0 MoM
Kwik (2003)	INC: singletons EXC: structural and chromosomal anomalies Mean age 32.7 (15-42) (Australia) (Cohort retrospective)	827	77-97 days	6.65	BW<10th centile (local)	PAPP-A ELISA (diagnostics) <0.3,<0.5 MoM
Legge (1985)	NC: singletons EXC: structural and chromosomal anomalies Age not reported (New Zealand) (cohort)	507	10-24	8.68	BW<10th centile	AFP, RIA (Biodata) ≥2.0 MoM
Lepage (2003)	INC: MSAFP<2.0 MoM EXC: structural and chromosomal anomalies, IDDM Age not reported	2256	Second trimester	2.34	BW<10th centile	HCG, Method not reported ≥ 4.0 MoM

	(Canada) (case control, matched, test)					
Lieppmann (1993)	INC: singleton, Down's risk >1:195 EXC: structural and chromosomal anomalies, women with normal HCG but raised AFP or oestriol Mean age not reported. (USA) (cohort, prospective)	60	15-18	10.20 5.87	BW<10th centile (local) BW<2500g	HCG, RIA (MAIAClone Serono) ≥ 2.0 MoM
Markestaad (1997)	INC: multips Mean age if SGA 28.8+/-0.4 years, non-SGA 30.2+/-0.4 years (USA) (cohort, prospective)	216	<20	47.22	BW<15th centile (sex, parity, local)	HCG, Immunoreactive (Seano) <10th centile UE3, RIA (Amersham) <10th centile
Milunsky (1989)	INC: singletons 20-34 years (90%) (USA) (cohort)	13486	15-20	2.28	BW<5.5 pounds	AFP, RIA (Clinical assays) ≥2.0 MoM, ≤0.4MoM
Milunsky (1996)	INC: singletons EXC:IDDM, structural and chromosomal anomalies Mean age 30.3 years	78	15-24	10.20	BW<2500g	DS≥1:270

	(USA) (case control matched test)					
Miyakoshi (2001)	INC: singletons, primips 70% EXC: structural and chromosomal anomalies Mean age 38.9+/-1.8 (Japan) (cohort, retrospective)	359	15-18	11.42	BW<10th centile	HCG, Method not reported >2.0 MoM
Morssink (1995)	INC: singletons EXC: structural and chromosomal anomalies, IDDM, delivery < 28 weeks Age not reported (Netherlands) (cohort)	8892	15-20	10.10	BW<10th centile (local)	AFP and HCG, EIA (Abbott) >2.5 MoM
Mwambingu (1985)	INC: singletons EXC: structural and chromosomal anomalies Age not reported (Scotland) (cohort)	282	16-18	13.48	SGA (no threshold)	AFP, Method not reported >2.5 MoM or > 97th centile
Naylor (2001)	INC: singletons EXC: IDDM, hypertension, maternal illnesses associated with adverse pregnancy outcome Mean age 27.8+/-7.7 years (USA) (case control, prospective)	150	15-24	5.33	BW<10th centile	DS>1:190
Odibo (2006)	INC: singletons EXC: structural and chromosomal anomalies Mean age 25.8 +/-7.0	2040	Second trimester	12.50	BW<5th centile	Method not reported AFP>2.0 MoM

	(USA) (case control, retrospective, outcome)				(local)	HCG >2.5 MoM UE3≤0.9MoM
Ogle (2000)	INC: singletons, 53% primips EXC: structural and chromosomal anomalies Mean age not reported (UK) (case control index test, nested cohort)	544	15-18	3.31	IUGR (threshold not reported)	DS>1:270 AFP MEIA Free βHCG ELISA
Onderoglu (1997)	INC: singletons EXC: IDDM, structural and chromosomal anomalies, MSAFP >2.0MoM, raised AFP and HCG Mean age 30.1+/-5.2 years (Turkey) (case control, nested)	562	15-20	3.56	BW<10th centile	HCG, Dunzen method >2.0 MoM
Ong (2000)	INC: singletons,32.0% primips Mean age 29.2 (15-45) (UK) (cohort, retrospective)	5297	10-14	7.46 3.23	BW<10th centile BW<5th centile	Free ßHCG and PAPP-A, random access immunoassay (Kryptor) <5th, <10th centile and <median
Pergament (1995)	INC: singletons, age <35 years, amniocentesis EXC: structural and chromosomal anomalies Mean age cases 30.0+/-3.8, controls 30.0+/-3.7 years (USA) (cohort, retrospective)	174	15-20	1.72	BW<10th centile	DS>1:250 All RIA

Pilalis (2007)	INC: singletons, Papp-a, TVS uterine artery Doppler, known outcome EXC: 4 miscarriages, 11 terminations Mean age 29 (15-45) (Greece) (cohort, prospective)	878	11-14	10.7 4.00	BW<10th centile BW<5th centile	PAPP-A, immunoassay (Kryptor) ≤5th and <10th centile
Roop (1991)	EXC: lost to follow up Mean age 27.7 years (USA) (Cohort)	1703	15-20	3.23	IUGR (no threshold)	AFP, RIA (Clinical assays) >2.3 MoM
Secher (1985)	INC: singletons, primips, birth > 28 weeks EXC: neural tube defects Age not reported (Denmark) (cohort)	1739	16-18	10.60 5.29	BW<10th centile (local) BW<5th centile (local)	AFP, RIA >1.0, 1.5, 2.0 MoM
Simpson (1995)	INC: singletons EXC: structural and chromosomal anomalies Age not reported (USA) (cohort)	650	15-20	10.50 2.77	BW<2500g BW<10th	AFP, EIA (Hybritech Tandem ERA, Abbott) ≥2.0 MoM
Smith	INC: singletons, primips 44.4%, EXC: structural and	8839	8-14	4.18	BW<5th	Free ßHCG and

(2002)	chromosomal anomalies Median age 30.7 years (UK) (cohort, prospective)				centile	PAPP-A, random access immunoassay (Kryptor) <5th centile
Smith (2006)	INC: singletons, screening programme, birth ≥24 weeks Median age 29 (25-33) (UK) (cohort, prospective)	8483	15-21	4.16	SGA (no threshold)	AFP, Method not reported ≥1.7 MoM (97th centile) PAPP-A, method not reported <5th centile
Sritippayawan (2005)	INC: singletons EXC: uninterpretable results, delivery at another hospital, bad obstetric or past medical or family history, structural or chromosomal anomalies Mean age cases 34.5(6.4) , controls 33.7 (5.2) years (Thailand) (case control, matched test)	330	14-21	1.21 5.15	BW<10th centile BW<2500g	DS>1:270 AFP – RIA HCG - EIA
Summers (2003)	INC: singletons EXC: IDDM, structural and chromosomal anomalies, positive NTD screen Median age 34 years (Canada) (case control, retrospective, nested cohort, index test)	23098	115 days median	1.68	SGA (no threshold)	DS>1:385

Tanaka (1994)	INC: singletons Age not reported (Japan) (cohort)	1097	15-18	10.00	BW<2500g	AFP RIA, HCG TRFIA ≥2.0 MoM
Towner (2006)	INC: screened, singleton EXC: pregnancy loss <20 weeks Mean age 26.7 +/-6 (USA) (case control, matched, test)	618	<20 weeks	8.41	BW<10th centile (ga, local, sex)	HCG, method not reported ≥2.0 MoM
Tul (2003)	INC: singletons, 51% primips Mean age 30.4 (18-44) years (Slovenia) (cohort, retrospective)	1004	10-14	5.07	BW<10th centile (local)	PAPP-A, Random access immunoassay (Kryptor) ≤0.5MoM
Wald (1977)	INC: singletons EXC: structural and chromosomal anomalies Mean age not reported (UK) (case control, prospective, matched, index test)	188	4-22	9.04	BW<2500g	AFP, RIA ≥ 3.0 MoM
Wald (1980)	INC: singletons EXC: structural and chromosomal anomalies Mean age not reported (UK) (cohort)	4198	16-18	5.40	BW≤2500g	AFP, Method not reported > 2.0 MoM
Waller (1996)	INC: singletons EXC: structural and chromosomal anomalies Mean age 27 years (USA) (cohort, retrospective)	51008	15-19	5.16	BW<5th centile	AFP, EIA (Abbott) >1.0,2.0,2.5 MoM <0.44 MoM (1st

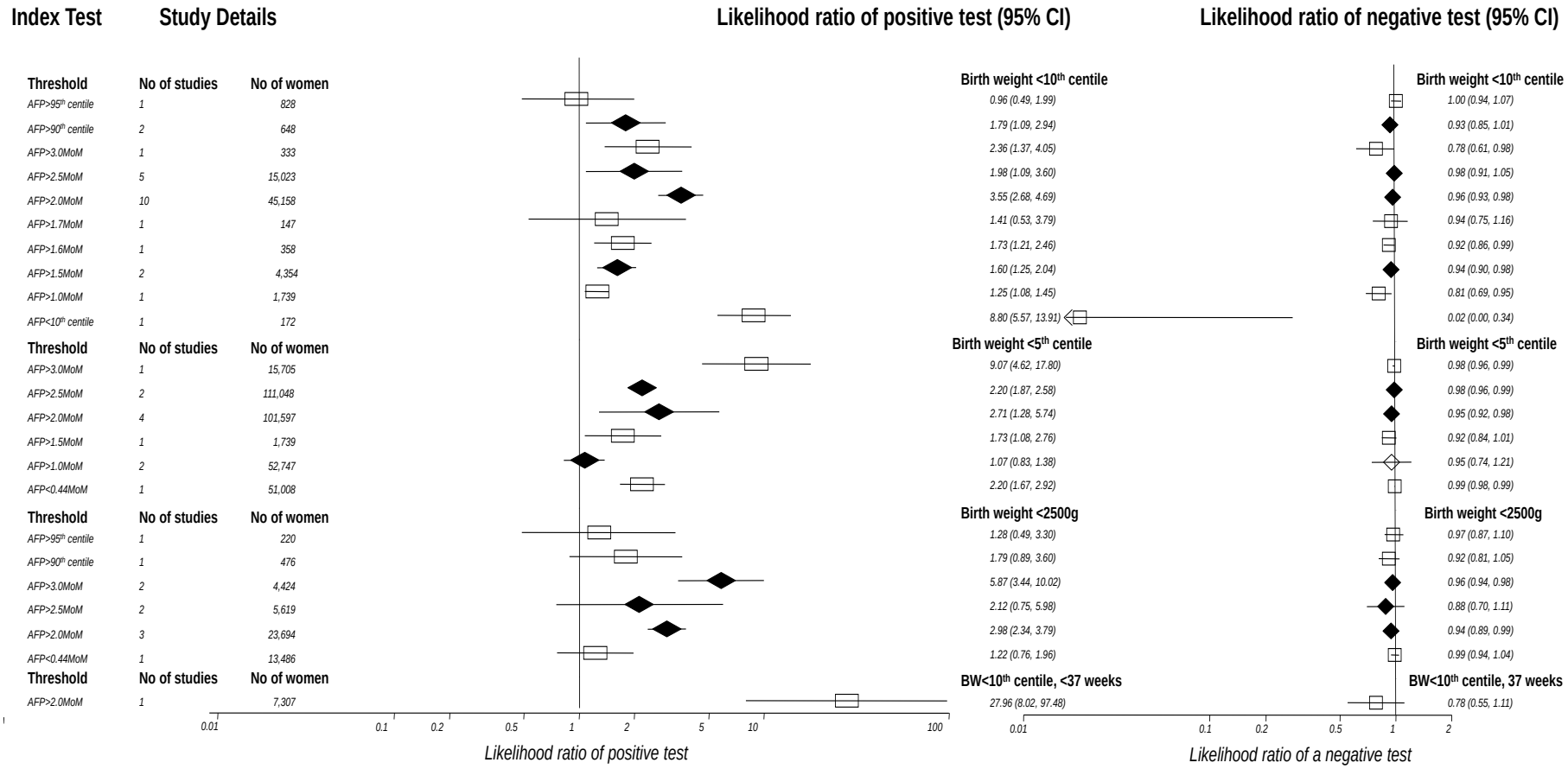
						centile)
Weiner (1991)	INC: singletons, referred for amniocentesis EXC: structural and chromosomal anomalies, TOP, miscarriage Mean age 32.7 +/-6.0 (USA) (cohort)	144	Second trimester	6.94	BW<10th centile	AFP, RIA (Amersham) >2.0 MoM
Wenstrom (1992)	INC: singletons, screened with raised AFP and repeat sample EXC: structural and chromosomal anomalies Age not reported (USA) (Cohort)	440	15-20	12.0	SGA no threshold	AFP, method not reported ≥2.5 MoM
Wenstrom (1996)	INC: singletons EXC: raised AFP or acetylcholinesterase in amniotic fluid, blood contamination of amniotic fluid, structural or chromosomal anomalies Age not reported (USA) (cohort, retrospective)	4336(FGR) 4614 (PE)	Second trimester	3.48	BW<10th (local)	AFP< RIA (Sanofi Pasteur) ≥2.5 MoM
Westergaard (1984)	INC: singletons Age not reported (Denmark) (cohort prospective)	208	18-22	15.9	BW<10th centile (local) and phenotypic signs of FGR	PAPP-A, EIA, <10th centile

Williams (1992)	INC: singletons EXC: structural and chromosomal anomalies Mean age 28.8 +/- 4.5 (USA) (case control, prospective, unmatched test)	412	Second trimester	18.70 14.3	BW<2500g BW<10th centile (local)	AFP, EIA (Hybritech) ≥2.0 MoM
Yaron (1999)	EXC: structural and chromosomal anomalies Age not reported (USA) (cohort)	60040 45565 24504 20907	14-22	2.47 2.32 4.93 1.76	BW<5th centile	AFP, RIA (Sanofi) >2.5 MoM ßHCG IRMA (Biodata) >2.5 MoM UE3 ComPEitive immunoasay <0.5 MoM Triple test
Yaron (2002)	INC: singletons EXC: structural and chromosomal anomalies Mean age 30.4+/-4.3 years (USA) (cohort)	1622	10-13	3.02	BW<5th centile	Free ßHCG FIA (Delfia Wallace) >5.0,4.0,3.0,2.0,1.0 MoM
Yuong Kim (2000)	INC: singletons EXC: IDDM, HCG>2MoM, AFP>2MoM Mean age 29.0+/-2.6 years (Korea) (cohort)	1096	15-21	3.65	BW<10th centile	UE3, Method not reported ≤0.75 MoM

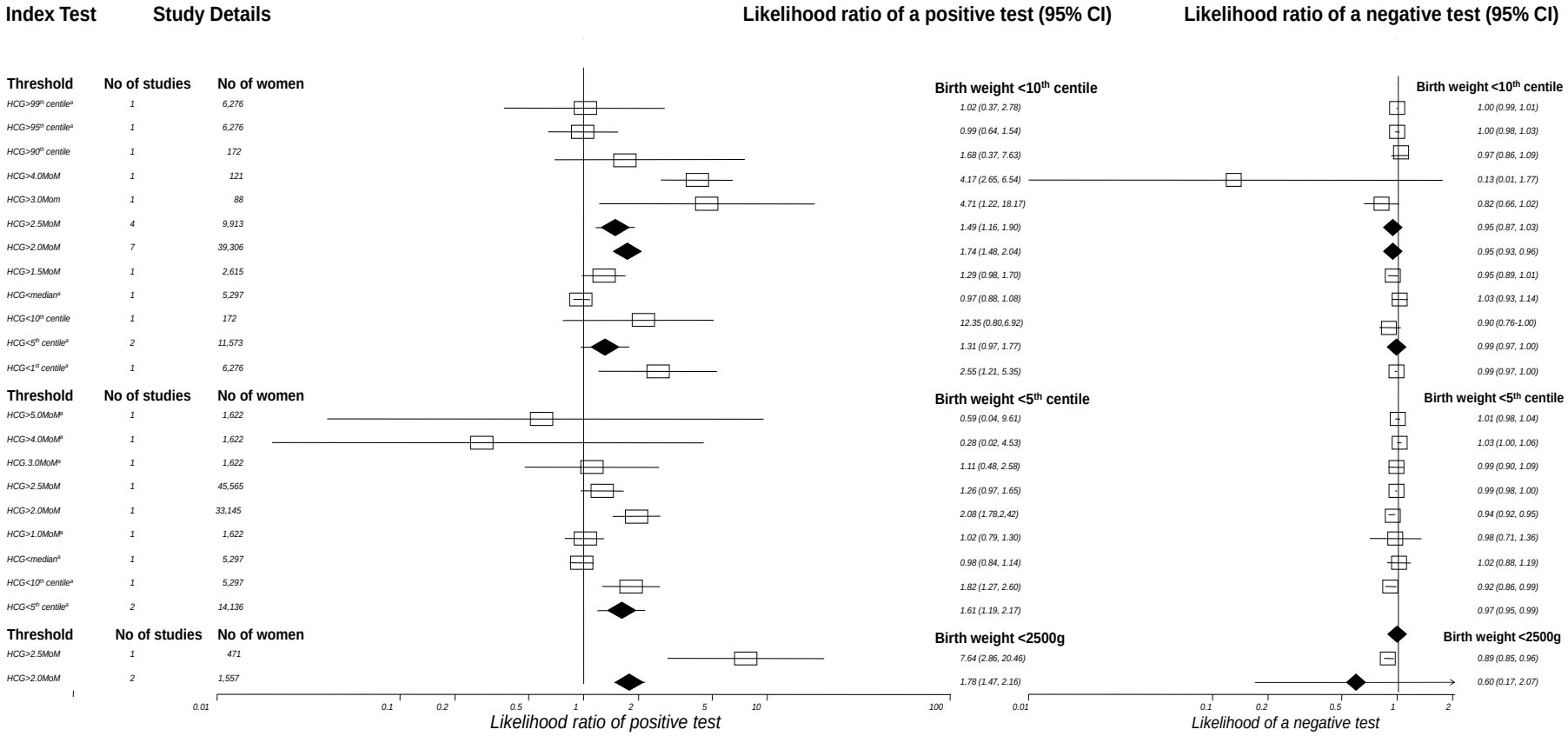
Zarzour (1998)	INC: amniocentesis EXC: abdominal wall defects Age not reported (USA) (cohort, retrospective)	1904	14-20	8.25	BW<2500g	AFP, RIA >2.0 MoM

FPR false positive rate; hrs hour; INC inclusion; EXC exclusion; AFP alpha-fetoprotein; HCG human chorionic gonadotrophin;UE3 unconjugated oestriol; PAPP-A pregnancy associated plasma protein A; PE preeclampsia; PIH pregnancy induced hypertension; FGR fetal growth restriction; IDDM diabetes mellitus; AID auto immune disease; APS antiphospholipid syndrome; SLE systemic lupus erythematodes; MoM multiples of the median. TP true positives; BW birth weight; TVS transvaginal; UK United Kingdom; USA United States of America; NA not applicable; RIA random access immunoassay; ELISA enzyme linked immunoabsorbent assay; EIA enzyme immunoassay; FEIA fluroenzyme immunoassay; IRMA immunoradiometric assay; MEIA microparticle enzyme immunoassay; IFMA immunoflurometric assay; TRFIA time resolved flurometricimmuno assay; mg milligrams; mmmHg millimetres of mercury; µg/l mircrograms per litre; pg/ml pictograms per millilitre; g grams; NT nuchal translucency; ROC receiver operating characteristic curve; SGA small for gestational age; NTD neural tube defects; MSAFP maternal serum alpha feto-protein; BP blood pressure; IUGR intrauterine growth restriction; DS Down's syndrome; TOP termination of pregnancy; USS ultrasound scan; IVF in-vitro fertilisation; ga gestational age; sd standard deviation, % percent

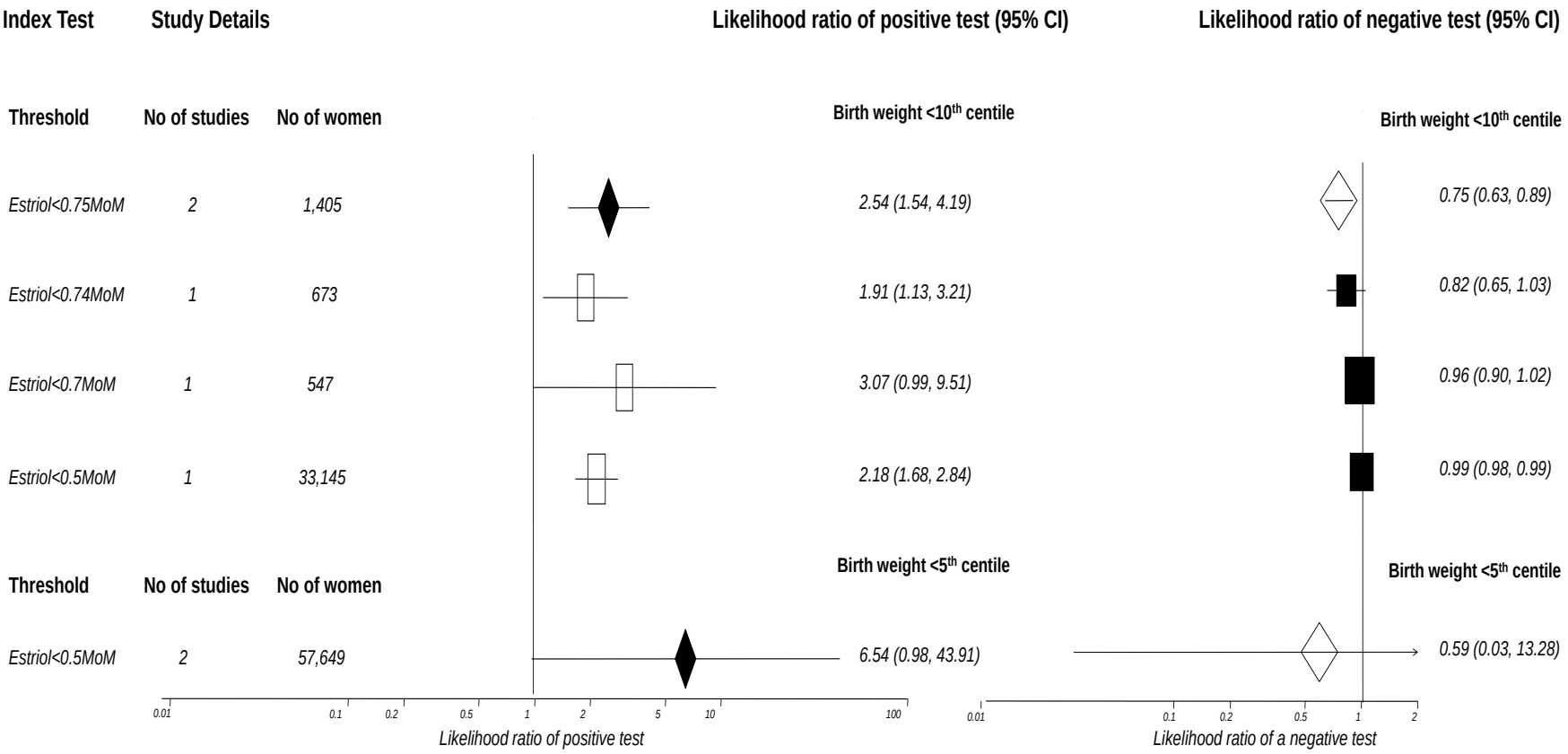
Appendix 9: Forest plot of likelihood ratios and 95% confidence intervals for alpha feto-protein (AFP) to predict small for gestational age fetuses. (squares represent individual studies and diamonds pooled results).



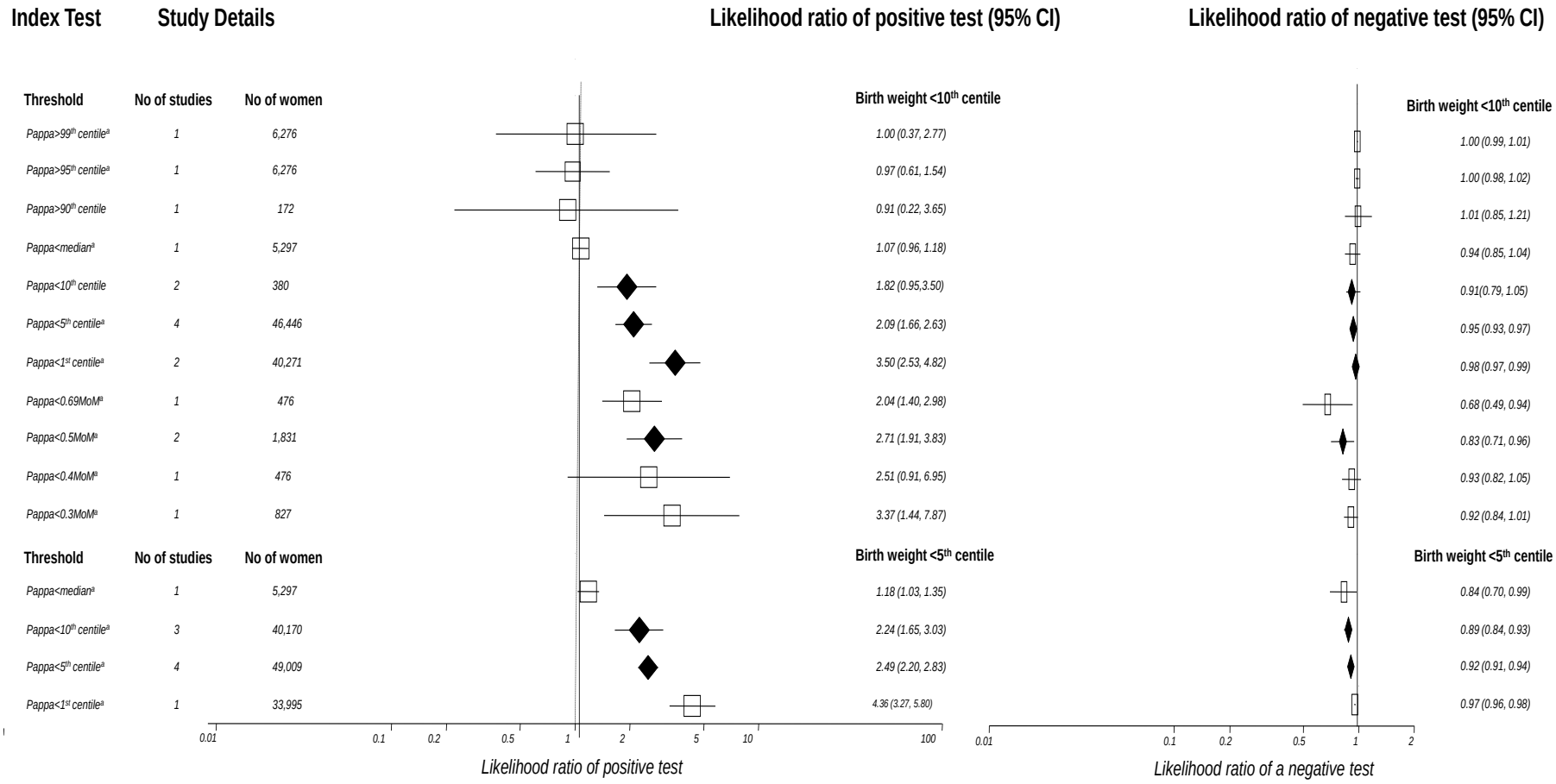
Appendix 10: Forest plot of likelihood ratios and 95% confidence intervals for human chorionic gonadotrophin (HCG) to predict small for gestational age fetuses. (squares represent individual results and diamonds pooled results).



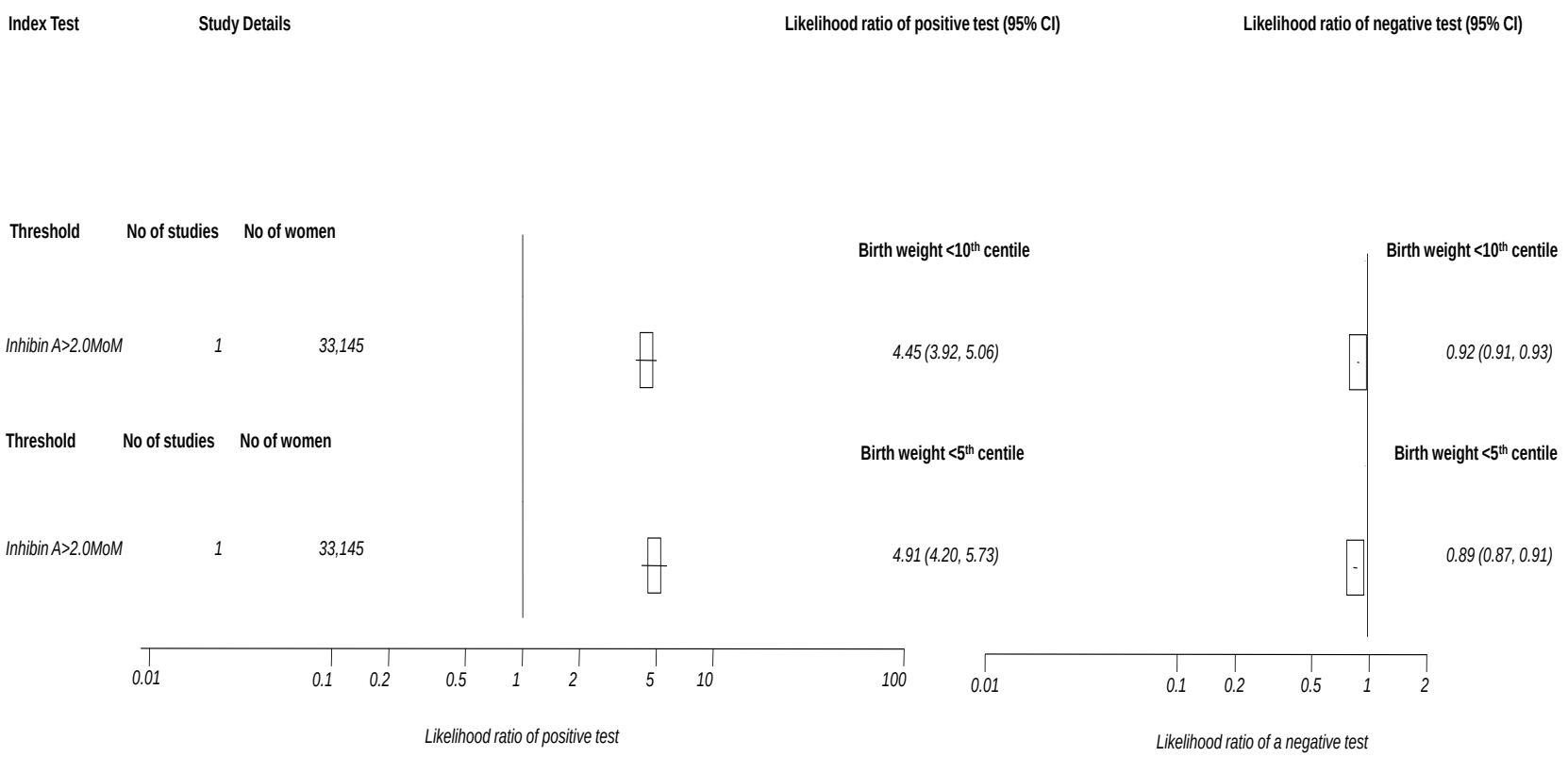
Appendix 11: Forest plot of likelihood ratios and 95% confidence intervals for unconjugated oestriol to predict small for gestational age fetuses. (squares represent individual studies and diamonds pooled results).



Appendix 12: Forest plot of likelihood ratios and 95% confidence intervals for pregnancy associated plasma protein A (PAPPA) to predict small for gestational age fetuses. (squares represent individual studies and diamonds pooled results).



Appendix 13: Forest plot of likelihood ratios and 95% confidence intervals for inhibin A to predict small for gestational age fetuses. (squares represent individual studies and diamonds pooled results).



Appendix 14: Analyses according to gestation of testing of accuracy of biochemical screening to predict small for gestational age fetuses. (CI confidence intervals, HCG human chorionic gonadotrophin, BW birth weight, MoM multiple of median, PAPPA pregnancy associated plasma protein A)

		Small for gestational age		
Analyte *Subgroup*	***Positive Likelihood Ratio (95% CI)***	***Negative Likelihood Ratio (95% CI)***	***Sensitivity (95% CI)***	***Specificity (95% CI)***
HCG>90th centile (BW<10th centile)				
Trimester				
First	1.48 (0.57-3.81)	0.92 (0.72-1.17)	0.21 (0.06-0.46)	0.86 (0.79-0.91)
Second	1.68 (0.37-7.63)	0.97 (0.86-1.09)	0.08 (0.01-0.26)	0.95 (0.90-0.98)
HCG<10th centile (BW<10th centile)				
Trimester				
First	1.29 (0.05-33.56)	1.14 (0.53-2.43)	0.13 (0.10-0.16)	0.60 (0.57-0.63)
Second	2.35 (0.80-6.92)	0.90 (0.76-1.08)	0.16 (0.05-0.36)	0.93 (0.88-0.97)
HCG>2.0MoM (BW<5th centile)				
Trimester				
First	0.96 (0.55-1.68)	1.01 (0.88-1.17)	0.20 (0.10-0.34)	0.79 (0.77-0.81)
Second	2.08 (1.78-2.42)	0.94 (0.92-0.95)	0.12 (0.10-0.14)	0.94 (0.94-0.95)
PAPPA<10th centile (BW<10th centile)				
Trimester				
First	1.68 (1.25-2.27)	0.93 (0.88-0.98)	0.17 (0.16-0.19)	0.90 (0.89-0.90)
Second	1.82 (0.95-3.50)	0.91 (0.75-1.05)	0.20 (0.10-0.33)	0.89 (0.85-0.92)

Appendix 15: Subgroup analyses of accuracy of biochemical screening to predict small for gestational age fetuses. (CI confidence intervals, AFP alpha feto-protein, HCG human chorionic gonadotrophin, BW birth weight, MoM multiple of median)

	Small for gestational age			
Analyte *Subgroup*	***Positive Likelihood Ratio (95% CI)***	***Negative Likelihood Ratio (95% CI)***	***Sensitivity (95% CI)***	***Specificity (95% CI)***
AFP>2.0MoM (BW<10th centile)				
Incidence				
>10%	2.69 (1.36-5.31)	0.98 (0.96-1.00)	0.04 (0.02-0.08)	0.98 (0.98-0.99)
≤10%	3.71 (2.66-5.16)	0.93 (0.88-0.97)	0.06 (0.05-0.07)	0.98 (0.98-0.98)
HCG>2.0MoM(BW<10th centile)				
Incidence				
>10%	1.53 (1.1-2.12)	0.89 (0.77-1.04)	0.29 (0.22-0.37)	0.79 (0.77-0.82)
≤10%	1.92 (1.72-2.13)	0.95 (0.94-0.96)	0.11 (0.1-0.12)	0.94 (0.94-0.95)

Appendix 16: Data extraction form for review of uterine artery Doppler to predict small for gestational age fetuses.

Section A: Study Information

1)Ref ID:		**4)Publication year:**	
2)Rev name:		**5)First Author:**	
3)Country:		**6)Language:**	

Section B: Data Retrieval for Uterine Artery Doppler Study

Population

7) Healthcare Centre:

Primary care ☐$_1$ Secondary care ☐$_2$ Mixed ☐$_3$ Other ☐$_4$ Unreported ☐$_5$

8) Setting:

In-patient ☐$_1$ Out-patient ☐$_2$ Mixed ☐$_3$ Unreported ☐$_4$ Other ☐$_5$ _____

9) Number of participating centres: _____

10) Gestation at time of index test:

<20 weeks ☐$_1$ 20-24 weeks ☐$_2$ 24-28 weeks ☐$_3$ 28-34 weeks ☐$_4$ 34-37 weeks ☐$_5$ 37-40 weeks ☐$_6$ > 40 weeks ☐$_7$ Unreported ☐$_8$ Other _____

10.i) Mean (range)_____ Unreported ☐$_3$

10.ii) Median (range) _____ Unreported ☐$_3$

11) Pregnancy:

Low Risk ☐$_1$ High Risk ☐$_2$ Unselected ☐$_3$ Unreported ☐$_4$

11.i) State high risk conditions: Unreported ☐$_3$

12) Were patients with the following conditions excluded/not included?

12.i) Previous IUGR: Yes ☐$_1$ No ☐$_2$

Unreported ☐$_3$

12.ii) Insulin dependant diabetes mellitus: Yes ☐$_1$ No ☐$_2$

Unreported ☐$_3$

12.iii) Chronic renal disease: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.iv) Systemic lupus erythematosus: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.v) Antiphospholipid syndrome: Yes ☐$_1$ No ☐$_2$

Unreported ☐$_3$ 12.vi) Chronic hypertension: Yes ☐$_1$ No ☐$_2$

Unreported ☐$_3$ 12.vii) Pre-eclampsia: Yes ☐$_1$ No ☐$_2$

Unreported ☐$_3$

12.viii) Foetal chromosomal/structural anomalies: Yes ☐$_1$ No ☐$_2$

Unreported ☐$_3$

13) Did all patients have singleton pregnancies?:

Yes ☐$_1$ No ☐$_2$ ----------/---------- Unreported ☐$_3$

14) Were all patients primigravid?:

Yes ☐$_1$ No ☐$_2$ ----------/---------- Unreported ☐$_3$

15) List other eligibility/ in-/exclusion criteria:

Not applicable ☐$_3$

16) Study population: (describe age (mean +/- SD or median/range), ethnicity, smoking, BMI etc.)

Unreported ☐$_3$

17) Start of patient inclusion (year) :

Unreported ☐3

18) End of patient inclusion (year) :

Unreported ☐3

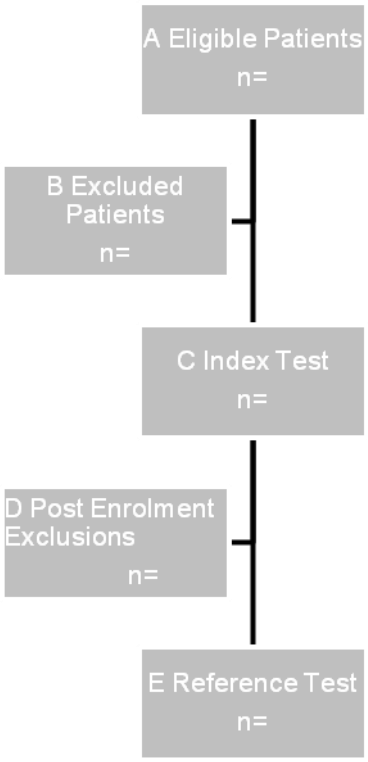
19) Study Design:

cohort ☐1 case control ☐2 RCT/CCT ☐3 cross sectional ☐4 before and after ☐5 case series ☐6 (no _____) other ☐7

19.i) Data collection: prospective ☐1 retrospective ☐2 unreported ☐3 other ☐4

19.ii) Enrolment: consecutive ☐1 arbitary (random) ☐2 unreported ☐3 other ☐4

20) Numbers:



B Reasons _____

D Reasons _____

21) Completeness of Verification:

(= E / C x 100 = %)

> 90% ☐$_1$ 81-90% ☐$_2$ < 81% ☐$_3$

Index Test

22) Description of technique:

Adequate ☐$_1$ Inadequate ☐$_2$

23) Timing of measurement (from delivery):

< 7days ☐$_1$ 7-14 days ☐$_2$ 14 -28 days ☐$_3$ > 28 days ☐$_4$ Mixture ☐$_5$

Unreported ☐$_6$

23.i) Median gestational age at delivery _____

unreported ☐$_3$

24) Measurement:

SCANNING:

24.i) Operator:

Single ☐$_1$ Multiple ☐$_2$ Unreported ☐$_3$

24.ii) Operator experience_____

unreported ☐$_3$

24.iii) Scanning Route: Transabdominal ☐$_1$ Transvaginal ☐$_2$ Unreported ☐$_3$

DOPPLER:

24.iv) Method: Continuous wave Doppler ☐$_1$ Pulsed wave Doppler ☐$_2$ Colour mapping ☐$_3$

Unreported ☐$_4$

24.v) Measurement parameter: Resistance index (RI) ☐$_1$ Systolic / diastolic ratio ☐$_2$ Diastolic / systolic ratio ☐$_3$ Unilateral Diastolic notch ☐$_4$ Bilateral diastolic notch ☐$_5$ Pulsatility index (PI) ☐$_6$ Time averaged velocity (TAV) ☐$_7$ Time averaged maximum velocity (TAMXV) ☐$_8$ Minimum velocity ☐$_9$ Unreported ☐$_{10}$

24.vi) Cut-off level for waveform ratio: > 2 SD ☐$_1$ > 95th centile ☐$_2$ > 90th centile ☐$_3$

> 80th centile ☐$_4$ > 50th centile ☐$_5$ < 10th centile ☐$_6$ < 5th centile ☐$_7$ Unreported/NA ☐$_8$

Other/Threshold data set:

_____

24.vii) Machine: _____ unreported ☐$_3$

24.viii) Probe: _____

unreported ☐3

24.ix) High pass filter: _____

unreported ☐3

24.x) Pulse rePEition frequency: _____

_____

unreported ☐3

24.xi) Size of sampling gate: _____

unreported ☐3

24.xii) Site : _____

unreported ☐3

24.xiii) Angel of insonation: _____

unreported ☐3

24.xiv) Number of consecutive waveforms: _____

unreported ☐3

24.xv) Were both sides measured: Yes ☐1 No ☐2 Unreported ☐3

24.xvi) Other information:

_____

Reference Standard / Outcome

25) Measured blind form diagnostic test: Yes ☐1 No ☐2 Unclear ☐3

26) Measurement for FGR: Birthweight ☐1 Neonatal ponderal index ☐2

Skin fold thickness ☐3 MAC / OFC ☐4 Other ☐5 _____

27) Threshold: < 3rd centile ☐1 < 5th centile ☐2 < 10th centile ☐3 < 25th centile ☐4

> 2SD ☐$_5$ Other ☐$_6$ _____ Unclear ☐$_7$

28) What data set was used to define threshold?

_____

unreported ☐$_3$

29)**Timing of measurement:** At delivery ☐$_1$ Within 24 hrs ☐$_2$ > 24 hrs ☐$_3$ Mixture ☐$_4$ Unreported ☐$_5$

Results

	Reference Test: Threshold:			
Index test, Measurement:		Positive	Negative	**Total**
	Positive	**TP**	**FP**	
Threshold:	Negative	**FN**	**TN**	
	Total			

Appendix 17: Guide to quality assessment of included studies in review of uterine artery Doppler to predict small for gestational age fetuses.

QUADAS question	Applicability and criteria fulfilled when
1. Representative spectrum of patients? (spectrum bias)	Pregnant women, consecutively recruited and prospective design.
2. Clearly described patient selection criteria? (selection bias)	Information on chronic hypertension, diabetes mellitus, parity, singleton/multiple pregnancies, previous preeclampsia/ fetal growth restriction available.
3. Reference standard correctly classifies target condition?	SGA: birth weight < 10^{th} centile adjusted for gestational age and based on local population values and absolute birth weight threshold < 2500g. Severe FGR: birth weight < 5^{th} or < 3^{rd} centile or < 1750g. Neonatal ponderal index < 10^{th} centile, skin fold thickness, and mid-arm circumference/head circumference were also assessed.
4. Time between tests short enough to be sure that target condition did not change? (disease progression bias)	Not applicable
5. Whole or random selection of study population received verification using a reference test? (partial verification bias)	All patients or a random selection received verification with reference standard (even if reference standard not the same for all patients).
6. Did patients receive the same reference test regardless of index test result? (differential verification bias)	All patients received same reference test (this is likely because the index test is non-invasive).
7. Reference test independent of index test? (incorporation bias)	The results of the index test are not incorporated in the definition of SGA.

8. Execution of index test described in sufficient detail?	Type of Doppler (e.g. color wave, pulsed, etc), site of measurement, measurement parameter and cut off level used, transvaginal or transabdominal route.
9. Execution of reference test described in sufficient detail?	Birthweight: timing of measurement, scales used, whether baby clothed or not. Neonatal ponderal index: description of birth weight and length measurement as above. Skin fold thickness: description of site of measurement, instrument used and timing of measurement. Mid-arm circumference/ head circumference: see skin fold thickness.
10 Blind interpretation of index test results (review bias)	Always fulfilled, reference test results not yet available when index test (Doppler) is performed (prediction).
10 Blind interpretation of reference test results (review bias)	Statement in text, such as "assessors were blind to Doppler results".
11. Same clinical data available when tests results were interpreted as would be available when test used in practice?	Any information to the patient obtained by direct observation (age, symptoms, BMI) normally available when test is interpreted in practice and similar data were available when interpreting the test in the study or if data not available when interpreted and not available in practice.
12. Uninterpretable/ intermediate test results reported?	All test results including uninterpretable/ intermediate are reported.
13. Were withdrawals from the study explained?	Clear what happened to all patients in study e.g. flow diagram (follow-up).
Additional	
14. Was there any preventative intervention?	Patients after having uterine artery Doppler received any of the following: aspirin, low molecular weight heparin, vitamin C or E, antihypertensive medication, saline infusion, oxygen

Appendix 18: References for studies included in review of uterine artery Doppler.

Aardema MW, Lander M, Oosterhof H, De Wolf BT, Aarnoudse JG. Doppler ultrasound screening predicts recurrence of poor pregnancy outcome in subsequent pregnancies, but not the recurrence of PIH or preeclampsia. Hypertens.Pregnancy. 2000;19:281-88.

Albaiges G, Missfelder-Lobos H, Lees C, Parra M, Nicolaides KH. One-stage screening for pregnancy complications by color Doppler assessment of the uterine arteries at 23 weeks' gestation. Obstet.Gynecol. 2000;96:559-64.

Alkazaleh F, Chaddha V, Viero S, Malik A, Anastasiades C, Sroka H et al. Second-trimester prediction of severe placental complications in women with combined elevations in alpha-fetoprotein and human chorionic gonadotrophin. Am.J.Obstet.Gynecol. 2006;194:821-27.

Arenas J, Fernandez I, Rodriguez-Mon C, Dupla B, Diez E, Gonzalez-Garcia A. Doppler screening of the uterine arteries to predict complications during pregnancy. [Spanish]. Clinica e Investigacion en Ginecologia y Obstetricia 2003;30:178-84.

Audibert F, Benchimol Y, Benattar C, Champagne C, Frydman R. Prediction of preeclampsia or intrauterine growth restriction by second trimester serum screening and uterine Doppler velocimetry. Fetal Diagn.Ther. 2005;20:48-53.

Axt-Fliedner R, Schwarze A, Nelles I, Altgassen C, Friedrich M, Schmidt W et al. The value of uterine artery Doppler ultrasound in the prediction of severe complications in a risk population. Arch.Gynecol.Obstet. 2004;271:53-58.

Bassim S, Lange M, Salzer H. Uterine artery doppler and pregnancy complications in unselected women. Geburtshilfe und Frauenheilkunde 2006;66:59-62.

Bower S, Schuchter K, Campbell S. Doppler ultrasound screening as part of routine antenatal scanning: prediction of pre-eclampsia and intrauterine growth retardation. BJOG. 1993;100:989-94.

Caforio L, Testa AC, Mastromarino C, Carducci B, Ciampelli M, Mansueto D et al. Predictive value of uterine artery velocimetry at midgestation in low- and high-risk populations: a new perspective. Fetal Diagn.Ther. 1999;14:201-05.

Carbillon L, Uzan M, Largilliere C, Perrot N, Tigaizin A, Paries J et al. Prospective evaluation of uterine artery flow velocity waveforms at 12-14 and 22-24 weeks of gestation in relation to pregnancy outcome and birth weight. Fetal Diagn.Ther. 2004;19:381-84.

Caruso A, De Carolis S, Ferrazzani S, Valesini G, Caforio L, Mancuso S. Pregnancy outcome in relation to uterine artery flow velocity waveforms and clinical characteristics in women with antiphospholipid syndrome. Obstet.Gynecol. 1993;82:970-77.

Caruso A, Caforio L, Testa AC, Ferrazzani S, Mastromarino C, Mancuso S. Chronic hypertension in pregnancy: color Doppler investigation of uterine arteries as a predictive test for superimposed preeclampsia and adverse perinatal outcome. J.Perinat.Med. 1996;24:141-53.

Coleman MA, McCowan LM, North RA. Mid-trimester uterine artery Doppler screening as a predictor of adverse pregnancy outcome in high-risk women. Ultrasound Obstet.Gynecol. 2000;15:7-12.

Degani S, Leibovich Z, Shapiro I, Gonen R, Ohel G. Early second-trimester low umbilical coiling index predicts small-for-gestational-age fetuses. J.Ultrasound Med. 2001;20:1183-88.

Driul L, Springolo F, Pezzani I, Casarsa S, Plaino L, Ianni A et al. [Pathological monolateral Doppler velocimetry of the uterine artery and materno-fetal outcome]. Minerva Ginecol. 2002;54:397-402.

Dugoff L, Lynch AM, Cioffi-Ragan D, Hobbins JC, Schultz LK, Malone FD et al. First trimester uterine artery Doppler abnormalities predict subsequent intrauterine growth restriction. Am.J.Obstet.Gynecol. 2005;193:1208-12.

Ferrier C, North RA, Becker G, Cincotta R, Fairley K, Kincaid-Smith P. Uterine artery waveform as a predictor of pregnancy outcome in women with underlying renal disease. Clin.Nephrol. 1994;42:362-68.

Frusca T, Soregaroli M, Danti L, Guandalini F, Lojacono A, Scalvi L et al. Uterine artery velocimetry as a screening test in patients with previous preeclampsia. Int.J.Gynaecol.Obstet. 1996;8:94-98.

Frusca T, Soregaroli M, Valcamonico A, Guandalini F, Danti L. Doppler velocimetry of the uterine arteries in nulliparous women. Early Hum.Dev. 1997;48:177-85.

Geipel A, Ludwig M, Germer U, Katalinic A, Diedrich K, Gembruch U. Uterine artery Doppler velocimetry and the outcome of pregnancies resulting from ICSI. Hum.Reprod. 2001;16:1397-402.

Geipel A, Berg C, Germer U, Katalinic A, Krapp M, Smrcek J et al. Doppler assessment of the uterine circulation in the second trimester in twin pregnancies: prediction of pre-eclampsia, fetal growth restriction and birth weight discordance. Ultrasound Obstet.Gynecol. 2002;20:541-45.

Gomez O, Martinez JM, Figueras F, Del RM, Borobio V, Puerto B et al. Uterine artery Doppler at 11-14 weeks of gestation to screen for hypertensive disorders and associated complications in an unselected population. Ultrasound Obstet.Gynecol. 2005;26:490-94.

Haddad B, Uzan M, Breart G, Uzan S. Uterine Doppler wave form and the prediction of the recurrence of pre-eclampsia and intra-uterine growth retardation in patients treated with low-dose aspirin. Eur.J.Obstet.Gynecol.Reprod.Biol. 1995;62:179-83.

Hafner E, Metzenbauer M, Hofinger D, Stonek F, Schuchter K, Waldhor T et al. Comparison between three-dimensional placental volume at 12 weeks and uterine artery impedance/notching at 22 weeks in screening for pregnancy-induced hypertension, pre-eclampsia and fetal growth restriction in a low-risk population. Ultrasound Obstet.Gynecol. 2006;27:652-57.

Harrington K, Carpenter RG, Goldfrad C, Campbell S. Transvaginal Doppler ultrasound of the uteroplacental circulation in the early prediction of pre-eclampsia and intrauterine growth retardation. BJOG. 1997;104:674-81.

Harrington K, Cooper D, Lees C, Hecher K, Campbell S. Doppler ultrasound of the uterine arteries: the importance of bilateral notching in the prediction of pre-eclampsia, placental abruption or delivery of a small-for-gestational-age baby. Ultrasound Obstet.Gynecol. 1996;7:182-88.

Harrington K, Fayyad A, Thakur V, Aquilina J. The value of uterine artery Doppler in the prediction of uteroplacental complications in multiparous women. Ultrasound Obstet.Gynecol. 2004;23:50-55.

Hershkovitz R, De SM, Kingdom J. Mid-trimester placentation assessment in high-risk pregnancies using maternal serum screening and uterine artery Doppler. Hypertens.Pregnancy. 2005;24:273-80.

Jorn H, Dinkloh C, Ritter S, Fendel H, Rath W. Clinical value of mid-pregnancy uterine artery Doppler velocimetry in the prediction of pregnancy complications. [German]. Geburtshilfe und Frauenheilkunde. 2003;63:49-55.

Konchak PS, Bernstein IM, Capeless EL. Uterine artery Doppler velocimetry in the detection of adverse obstetric outcomes in women with unexplained elevated maternal serum alpha-fetoprotein levels. Am.J.Obstet.Gynecol. 1995;173:1115-19.

Kurdi W, Campbell S, Aquilina J, England P, Harrington K. The role of color Doppler imaging of the uterine arteries at 20 weeks' gestation in stratifying antenatal care. Ultrasound Obstet.Gynecol. 1998;12:339-45.

Kurdi W, Fayyad A, Thakur V, Harrington K. Delayed normalization of uterine artery Doppler waveforms is not a benign phenomenon. Eur.J.Obstet.Gynecol.Reprod.Biol. 2004;117:20-23.

Le Thi HD, Wechsler B, Vauthier-Brouzes D, Duhaut P, Costedoat N, Andreu MR et al. The second trimester Doppler ultrasound examination is the best predictor of late pregnancy outcome in systemic lupus erythematosus and/or the antiphospholipid syndrome. Rheumatology.(Oxford) 2006;45:332-38.

Liberati M, Rotmensch S, Zannolli P, Perrino S, Celentano C, Tiboni GM et al. Uterine artery Doppler velocimetry in pregnant women with lateral placentas. J.Perinat.Med. 1997;25:133-38.

Marchesoni D, Pezzani I, Springolo F, Ianni A, Casarsa S, Zavarise D et al. The use of uterine artery Doppler as a screening test for pre-eclampsia. Int.J.Gynaecol.Obstet. 2003;15:15-20.

Martin AM, Bindra R, Curcio P, Cicero S, Nicolaides KH. Screening for pre-eclampsia and fetal growth restriction by uterine artery Doppler at 11-14 weeks of gestation. Ultrasound Obstet.Gynecol. 2001;18:583-86.

Miyakoshi K, Tanaka M, Gabionza D, Ishimoto H, Miyazaki T, Yoshimura Y. Prediction of smallness for gestational age by maternal serum human chorionic gonadotropin levels and by uterine artery Doppler study. Fetal Diagn.Ther. 2001;16:42-46.

Morris JM, Fay RA, Ellwood DA, Cook C-M, Devonald KJ. A randomized controlled trial of aspirin in patients with abnormal uterine artery blood flow. Obstet.Gynecol. 1996;87:74-78.

Nagtegaal MJ, van RS, McGavin S, Dekker G. Use of uterine Doppler in an Australian level II maternity hospital. Aust.N.Z.J.Obstet.Gynaecol. 2005;45:424-29.

North RA, Ferrier C, Long D, Townend K, Kincaid-Smith P. Uterine artery Doppler flow velocity waveforms in the second trimester for the prediction of preeclampsia and fetal growth retardation. Obstet.Gynecol. 1994;83:378-86.

Ohkuchi A, Minakami H, Sato I, Mori H, Nakano T, Tateno M. Predicting the risk of pre-eclampsia and a small-for-gestational-age infant by quantitative assessment of the diastolic notch in uterine artery flow velocity waveforms in unselected women. Ultrasound Obstet.Gynecol. 2000;16:171-78.

Onalan R, Onalan G, Gunenc Z, Karabulut E. Combining 2nd-Trimester Maternal Serum Homocysteine Levels and Uterine Artery Doppler for Prediction of Preeclampsia and Isolated Intrauterine Growth Restriction. Gynecol.Obstet.Invest 2005;61:142-48.

Papageorghiou AT, Yu CK, Bindra R, Pandis G, Nicolaides KH. Multicenter screening for pre-eclampsia and fetal growth restriction by transvaginal uterine artery Doppler at 23 weeks of gestation. Ultrasound Obstet.Gynecol. 2001;18:441-49.

Park YW, Lim JC, Kim YH, Kwon HS. Uterine artery Doppler velocimetry during mid-second trimester to predict complications of pregnancy based on unilateral or bilateral abnormalities. Yonsei Med.J. 2005;46:652-57.

Phupong V, Dejthevaporn T, Tanawattanacharoen S, Manotaya S, Tannirandorn Y, Charoenvidhya D. Predicting the risk of preeclampsia and small for gestational age by uterine artery Doppler in low-risk women. Arch.Gynecol.Obstet. 2003;268:158-61.

Prefumo F, Guven M, Ganapathy R, Thilaganathan B. The longitudinal variation in uterine artery blood flow pattern in relation to birth weight. Obstet.Gynecol. 2004;103:764-68.

Schwarze A, Nelles I, Krapp M, Friedrich M, Schmidt W, Diedrich K et al. Doppler ultrasound of the uterine artery in the prediction of severe complications during low-risk pregnancies. Arch.Gynecol.Obstet. 2005;271:46-52.

Sekizuka N, Murakoshi T, Yoshizawa H. The uterine and spiral artery flow velocity waveforms in early pregnancy: A transvaginal color and pulsed Doppler study. J.Matern.Fetal.Invest. 1994;4:229-32.

Soregaroli M, Valcamonico A, Scalvi L, Danti L, Frusca T. Late normalisation of uterine artery velocimetry in high risk pregnancy. Eur.J.Obstet.Gynecol.Reprod.Biol. 2001;95:42-45.

Soutif C, Prevost A, Andre M. [Value of the systematic uterine Doppler in the primiparous woman. A series of 315 cases]. J.Gynecol.Obstet.Biol.Reprod.(Paris) 1996;25:819-23.

Subtil D, Goeusse P, Houfflin-Debarge V, Puech F, Lequien P, Breart G et al. Randomised comparison of uterine artery Doppler and aspirin (100 mg) with placebo in nulliparous women: the Essai Regional Aspirine Mere-Enfant study (Part 2). BJOG. 2003;110:485-91.

Todros T, Ferrazzi E, Arduini D, Bastonero S, Bezzeccheri V, Biolcati M et al. Performance of Doppler ultrasonography as a screening test in low risk pregnancies: results of a multicentric study. J.Ultrasound Med. 1995;14:343-48.

Uludag S, Madazli R, Benian A, Ocak V, Erez S. The relationship between maternal serum a-fetoprotein and uterine artery Doppler findings at 20-24 weeks' gestation for prediction of preeclampsia and intrauterine growth retardation. Marmara.Med.J. 2002;15:167-74.

Vainio M, Kujansuu E, Koivisto AM, Maenpaa J. Bilateral notching of uterine arteries at 12--14 weeks of gestation for prediction of hypertensive disorders of pregnancy. Acta Obstet.Gynecol.Scand. 2005;84:1062-67.

Valensise H, Romanini C. Second-trimester uterine artery flow velocity waveform and oral glucose tolerance test as a means of predicting intrauterine growth retardation. Ultrasound Obstet.Gynecol. 1993;3:412-16.

Valensise H, Bezzeccheri V, Rizzo G, Tranquilli AL, Garzetti GG, Romanini C. Doppler velocimetry of the uterine artery as a screening test for gestational hypertension. Ultrasound Obstet.Gynecol. 1993;3:18-22.

Valensise H, Romanini C. Uterine Doppler in the identification of patients at risk for hypertension and IUGR. J.Perinat.Med. 1994;22 Suppl 1:69-72.

Van den Elzen HJ, Cohen-Overbeek TE, Grobbee DE, Quartero RW, Wladimiroff JW. Early uterine artery Doppler velocimetry and the outcome of pregnancy in women aged 35 years and older. Ultrasound Obstet.Gynecol. 1995;5:328-33.

Venkat-Raman N, Backos M, Teoh TG, Lo WT, Regan L. Uterine artery Doppler in predicting pregnancy outcome in women with antiphospholipid syndrome. Obstet.Gynecol. 2001;98:235-42.

Zimmermann P, Eirio V, Koskinen J, Niemi K, Nyman R, Kujansuu E et al. Effect of low-dose aspirin treatment on vascular resistance in the uterine, uteroplacental, renal and umbilical arteries - A prospective longitudinal study on a high risk population with persistent notch in the uterine arteries. Eur.J.Ultrasound 1997;5:17-30.

Yu CK, Papageorghiou AT, Boli A, Cacho AM, Nicolaides KH. Screening for pre-eclampsia and fetal growth restriction in twin pregnancies at 23 weeks of gestation by transvaginal uterine artery Doppler. Ultrasound Obstet.Gynecol. 2002;20:535-40.

Appendix 19: Study characteristics of included studies on uterine artery Doppler and fetal growth restriction: low risk and unselected populations.

Author (year) Country	Gestation (weeks)	N (%SGA)	Details index test	Index test (cut-off, centiles or absolute thresholds)	Reference test (centiles or absolute thresholds (g))	Results (TP; FP; FN; TN)	Population (study design)*
Albaiges (2000) UK	22-25	1757 (8.9) for BW<10th 1757 (2.9) for BW<3rd	Unreported route, color + PW, crossover	1) BilateraI notching or mean PI > 1.45 (95th) 2) Bilateral notching only 3) Mean PI > 1.45 only 4) Mean PI > 1.45 + bilateral notching	a) BW < 10th b) BW < 3rd (local and gestational values)	1a) 32;96;111;1518 1b) 16;112;36;1593 2a) 19;58;124;1556 2b) 12;68;40;1637 3a) 30;60;113;1554 3b) 35;84;17;1621 4a) 17;21;126;1593 4b) 12;27;40;1678	IN: singleton pregnancies, routine antenatal care
Arenas	20	319 (8.2)	TA, color + PW,	1) Mean RI ≥ 0.59 (75th)	BW < 10th (local and	1) 12;77;14;216	IN: unselected women,

(2003) Spain			crossover	2) Mean RI ≥ 0.52 (50^{th}) 3) Mean RI ≥ 0.65 (90^{th}) 4) Mean RI ≥ 0.71 (95^{th})	gestational values)	2) 16;159;10;134 3) 5;32;21;261 4) 3;16;23;277	EX: multiple pregnancies, congenital defects
Audibert (2005) France	18-26	2615 (8.8)	Unreported route, type and site	1) Bilateral notching 2) Unilateral notching	BW < 10^{th} (local values)	1) 30;85;200;2300 2) 52;267;178;2118	IN: AFP and hCG testing at 14-18 weeks and ultrasound screening (USS). EX: women without USS 10-14 weeks for dating, women with raised NT, no Doppler at 18-26 weeks delivery< 24 weeks
Bassim (2006)	20-24	490 (5.3)	TA, unreported type, crossover	1) Bilateral notching	BW < 10th (local values)	1) 8;28;18;436	IN: routine screening
Bower (1993) UK	18-22	2058 (3.5) for BW<3^{rd} 2058 (5.2) for BW<5^{th} 2058	TA, CW, crossover	1) RI > 95^{th} centile either side +/- any notching	a) BW < 10^{th} b) BW < 5^{th} c) BW < 3^{rd} (local and gestational values)	1a) 84;245;141;1588 1b) 49;280;57;1672 1c) 34;295;39;1690	IN: unselected women. EX: multiple pregnancies, outside gestational age, fetal anomalies

		(10.9) for BW<10th					
Caforio (1999) Italy	i) 18-20 ii) 22-24	a) 530 (5.1) b) 530 (10.9)	Unreported route, color + pulsed, crossover	1) RI > 90th	a) BW < 1750g b) BW < 2500g	1ai) 16;147;11;355 1aii) 14;127;11;378 1bi) 33;134;25;338 1bii) 31;98;32;369	IN: healthy nulliparae. EX: congenital defects, chromosomal abnormalities, multiple pregnancies, infections, Rhesus isoimmunisation, non immune hydrops, PPROM, IUD, delivery < 26 weeks;
Carbillon (2004) France	12-14 and 22-24	243 (9.4)	TA, unreported type, ascending branch	1) No notching at 12-14 weeks v uni- or bilateral notching 2) Bilateral notching at 22-24 weeks	FGR no threshold	1) 19;120;4;100 2) 6;15;17;205	IN: routine ultrasound screening
Driul (2002) Italy	24	830 (1.8)	Unreported route, color, crossover	1) RI ≥ 0.6 or unilateral notching	FGR unreported threshold	1) 8;103;7;722	Not reported
Dugoff (2005) UK	10-14	1008 (1.2)	TA, color + PW, crossover	1) Mean RI ≥ 0.81(95th) 2) Mean RI ≥ 0.78 (90th) 3) Mean RI ≥ 0.70 (75th)	BW < 10th (local and gestational values)	1) 2;49;10;947 2) 4;102;8;894 3) 8;248;4;748	EX: structural or chromosomal anomalies, fetal genetic syndrome, IUD < 24 weeks,

							congenital uterine malformation
Frusca (1997) Italy	24	a) 419 (7.1) b) 419 (2.6)	TA, color + PW, crossover	1) RI > 0.58	a) BW < 10th (local and gestational values) b) BW < 3rd	1a) 13;23;17;366 1b) 6;30;5;378	IN: nulliparae without risk factors, EX: CH, DM, AID
Geipel (2001) Germany	18-24	114 (8.8)	Unreported route, color, crossover	1) Bilateral notching + mean RI > 0.55 or unilateral notching + mean RI > 0.65 or no notching + mean RI > 0.7	BW < 10th (local values)	1) 5;19;5;85	IN: singleton pregnancies (control group of ICSI)
Gomez (2005) Spain	11-14	999 (3.7)	TVS, color + pulsed, cervicocorporeal junction	1) Mean PI > 95th	a) BW < 5th (local and gestational values)	1a) 9;44;28;918	EX: fetal anomalies, women treated with aspirin, heparin or antihypertensive medication before enrolment
Hafner (2006) Germany	21-23	2489 (6.8)	TA, color + PW, crossover	1) Bilateral notching 2) Mean PI ≥ 90th	BW < 10th (unreported dataset)	1) 37;189;131;2132 2) 34;218;134;2103	IN: all singleton pregnancies
Harrington (1997) UK	12-16	623 (19.9)	TVS, unreported type and site	1) Bilateral vs unilateral or no notching 2) Bilateral or unilateral vs	BW < 10th (local and gestational values)	1) 42;163;19;399 2) 53;294;8;268	IN: unselected singleton pregnancies

				no notching			
Harrington (1996) UK	24	1204 (10.8)	TA, color + PW, crossover	1) Any notching or mean RI > 95th 2) Unilateral or bilateral notching 3) Unilateral notching 4) Bilateral notching	BW < 10th (local and gestational values)	1) 42;68;89;1005 2) 42;68;89;1005 3) 18;44;113;1029 4) 24;24;107;1049	IN: unselected women. EX: multiple pregnancies, fetal anomalies, PE or FGR ≤ 24 weeks
Harrington (2004) UK	19-21	458 (5.2)	TA, color + pulsed, crossover	1) Bilateral notching + mean RI ≥ 0.55 (50th) or unilateral notching and mean RI ≥ 0.65 (80th)	BW < 5th (local values)	1) 8;33;15;402	IN: unselected multiparae with singleton pregnancies EX: fetal anomalies
Jorn (2003) Germany	18-24	602 (4.8)	TA, color, ascending branch	1) Mean RI > 0.61	BW < 5th (local and gestational values)	1) 22;46;7;527	Not reported
Kurdi (1998) UK	19-21	946 (16.5) for BW<10th 946 (6.0) for BW<5th	TA, color + PW, crossover	1) Bilateral notching + mean RI > 0.55(50th) or unilateral notching + mean RI > 0.65 (90th), or no notching + mean RI > 0.70 (90th) 2) Bilateral notching + mean RI > 0.55 (50th)	a) BW < 10th b) BW < 5th (local and gestational values)	1a) 70;146;86;644 1b) 27;189;30;700 2b) 21;96;36;793	IN: unselected women EX: multiple pregnancies, fetal anomalies, women already on low dose aspirin
Kurdi (2004) UK	19-21 and 24-26	779 (13.2) for BW<10th	TA, color + PW, crossover	1) Bilateral notching + mean RI > 0.55 (50th) or unilateral notching + mean RI > 0.65	a) BW < 10th b) BW < 3rd (local and gestational values)	1 or 2a) 16;33;87;643 1 or 2 b)	sub-group analysis of cohort in Kurdi (1998) with normalization of

				(90^{th}), or no notching + mean RI > 0.70 (90^{th}) 2) Bilateral notching + mean RI >f0.55 (50^{th})		29;45;4;701	Doppler at 24-26 weeks
		779 (4.2) for BW<3^{rd}					
Liberati (1997) Israel	22-24	481 (8.5)	TA, color + PW, crossover	1) Mean RI ≥ 90^{th} 2) Unilateral notching or mean RI ≥ 90^{th}	BW < 10^{th} (local values)	1) 11;18;30;422 2) 16;34;25;406	EX: preterm delivery, multiple pregnancies, major fetal anomalies
Marchesoni (2003) Italy	24	900 (1.7)	Unreported route and type, crossover	1) Bilateral vs unilateral or no notching 2) Bilateral or uni vs no notching	BW < 3^{rd}	1) 0;60;15;825 2) 8;153;7;722	IN: unselected women
Martin (2001)UK	11-14	3045 (9.5)	TA, color + PW, ascending branch	1) Mean PI > 2.35 (95^{th})	BW < 10^{th} (local and gestational values)	1) 34;121;256;2639	IN: routine antenatal care
Miyakoshi (2001) Japan	21-24	359 (11.4)	Unreported route, color + PW, crossover	1) Mean PI > 95^{th} +/- unilateral notching	BW < 10^{th} (local and gestational values)	1) 10;18;31;300	Not reported
Morris (1996) Australia	18	a) 768 (12.7) b) 768 (4.6) c) 679 (10.9)	TA, color + PW, crossover	1) S/D > 3.0 (90^{th}) + any notching or S/D > 3.3 (2SD)	a) BW < 10^{th} (local values) b) BW < 3^{rd} c) PI < 10^{th}	1a) 21;54;77;616 1b) 9;66;26;667 1c) 15;46;59;559	IN: all nulliparae **(RCT)**
Nort	19-24	457 (6.6)	TA, color + PW,	1) RI > 90^{th}	BW < 10^{th} (local and	1) 15;41;15;386	IN: healthy nulliparae,

(1994) Australia			crossover	2) S/D > 90th 3) RI > 0.53 4) RI > 0.54 5) RI > 0.55 6) RI > 0.56 7) RI > 0.57	gestational values)	2) 14;47;16;380 3) 17;86;13;342 4) 16;74;14;354 5) 16;60;14;368 6) 16;45;14;383 7) 14;38;16;390	EX: renal disease, DM
Ohkuchi (2000) Japan	16-23	288 (6.3)	TA, color + PW, crossover	1) RI > 70th 2) S/D > 78th 3) NDI > 0.14 4) Any notching 5) Bilateral notching	BW < 10th (local and gestational values)	1) 10;102;8;168 2) 9;69;9;201 3) 6;21;12;249 4) 6;53;12;217 5) 3;25;15;245	IN: unselected healthy women with singleton pregnancies
Onala (2005) Turkey	19-21	406 (10.1)	Unreported route, type and site	1) Bilateral notching + mean RI > 0.55 or unilateral notching + mean RI > 0.65 or no notching + mean RI > 0.7	BW < 5th (local values)	1) 15;19;26;346	IN: fasting serum tHcy levels. EX: multiple pregnancies, history of PE, hypertension < 20 weeks, altered renal function, DM, chronic disease, fetal anomalies, folic acid use > 12 wks, special folate diet, treatment with antifolate drugs, age >40 yrs
Papageorghiou (2001) UK	20-24	7851 (9.4)	TVS, color + PW, level of	Mean PI > 1.63 (95th)	BW < 10th (local and gestational values)	1) 121;280;619;6831	IN: singleton pregnancies attending

			internal cervical os				for routine antenatal care
Park (2005) Korea	20-24	1090 (9.5)	TA, color + PW, previously reported	1) S/D > 2SD +/- any notching 2) S/D > 2SD +/- bilateral notching only	BW < 10th (local and gestational values)	1) 38;227;66;759 2) 18;51;86;935	EX: multiple pregnancies, fetal anomalies, cardiovascular and renal diseases, DM, FGR, PIH, preterm labour before performance of Doppler.
Phupong (2003) Thailand	22-28 (24.9 ± 1.9)	324 (1.9)	TA, color + pulsed, crossover	1) S/D + 2SD a/o bilateral notching	BW < 10^{th} (gestational values)	1) 4;56;2;262	IN: healthy nullparae and multiparae. EX: multiple pregnancies, renal and cardiovascular disease, DM, fetal anomalies
Prefumo (2004) UK	i) 11-14 ii) 18-23	662 (9.8)	Unreported route, color, ascending branch	1) Bilateral notching	BW < 10^{th} (local, gestational and sex values)	1i) 37;214;28;383 1ii) 4;25;61;572	EX:multiparae, fetal abnormalities, concurrent maternal disease and gestational diabetes
Schwarze (2005) Germany	19-26	346 (10.1)	TA, color, crossover	1) Any RI > 0.58 2) Both RI > 0.58 3) Any RI > 0.7	BW < 10^{th}	1) 18;116;17;195 2) 4;47;31;264 3) 7;41;28;270	EX: essential hypertension, DM, AID, history of PE, FGR,

				4) Both RI > 0.7 5) Any notching 6) Bilateral notching		4) 12;65;23;246 5) 20;124;15;187 6) 10;97;25;214	IUD, placental abruption in previous pregnancies, multiple pregnancies, fetal abnormalities
Sekizuka (1994) Japan	6-13	135 (7.4)	TVS, color + pulsed, internal os	1) Bilateral notching	BW < 10th (local and gestational values)	1) 9;115;1;10	EX: threatened abortions and fibroids
Soutif (1996) France	21	315 (10.4)	Unreported route and site, color + PW	1) S/D > 2.6 +/- any notching	BW < 10th (local values)	1) 10;40;23;242	EX: nephropathy, CH, DM, systemic disorders, multiple pregnancies
Subtil (2003) France/ Belgium	22-24	a) 1186 (1.9) b) 1186 (10.3)	Unreported route and type, crossover	1) RI ≥ 0.61 or any notching	a) BW ≤ 3rd (local and gestational values) b) BW ≤ 10th	1a) 11;228;12;935 1b) 45;194;78;869	Routine Doppler examination followed by a prescription for aspirin in case of abnormal Doppler findings versus placebo **(RCT)**
Todros (1995) Italy	a) 19-24	916 (4.6)	Unreported route and site, CW + pulsed	1a) S/D > 2.7	BW < 10th (local values)	1a) 5;54;37;820	IN: singleton pregnancies, no prepregnancy pathology, no obstetric risk, no chromosomal or structural anomalies
Uludag (2002) Turkey	18-20	80 (11.3)	TA, color + PW, unreported site	1) Bilateral notching	BW < 10th (unreported data set)	1) 6;11;3;60	IN: non-smokers, EX: DM, fetal anomalies,

							multiple pregnancies
Valensise (1993) Italy	24	192 (14.6)	Unreported route and site, color	1) RI > 0.58 (mean)	BW < 10th (gestational values	1) 24;15;4;149	IN: low risk (n=104) primiparae, no current or previous relevant medical history. High risk (n=88) history of PIH, FGR, IUD. EX: FGR on ultrasound screening or oligohydramnios
Valensise (1993) Italy	24	272 (7.7)	TA, color + PW, crossover	1) RI > 0.58	BW < 10th (local values)	1) 14;12;7;239	EX: history of hypertension, DM, SLE, pharmacological induction of ovulation, fetal or chromosomal abnormalities
Zimmerman (1997) Finland	20-24	55 (7.2)	Unreported route, PW, crossover	1) Bilateral notching	BW < 10th (local values)	1) 3;27;1;24	IN: low risk (n=29) or high risk (n=26; family or personal history of PE, CH or FGR or IUD) (RCT)

* Studies are cohort studies unless otherwise stated (randomised controlled trial (RCT)).TP true positives; FP false positives; FN false negatives; TN true negatives; FPR false positive rate; BW birthweight;, PI pulsatility index; RI resistance index; a/o and/or; IN inclusion; EX exclusion; AFP alpha-fetoprotein; hCG human chorionic gonadotrophin; PE preeclampsia; PIH pregnancy induced hypertension; SGA small for gestational age; DM diabetes mellitus; AID auto immune disease;

APLS antiphospholipid syndrome; SLE systemic lupus erythematodes; IUD intra uterine demise; TTTS twin transfusion syndrome; MoM multiples of the median; PW pulsed waved; CW continuous waved, TA transabdominal, TVS transvaginal.

Appendix 20: Study characteristics of included studies on uterine artery Doppler and fetal growth restriction: high risk populations.

Author (year) Country	Gestation (weeks)	N (% SGA)	Details index test	Index test (cut-off, centiles or absolute thresholds)	Reference test (centiles or absolute thresholds (g))	Results (TP; FP; FN; TN)	Population (study design)*
Aardema (2000) Netherlands	21-22	94 (10.6)	Unreported route, color + PW, crossover	1) PI ≥ 1.3 2) Any notching	BW < 10th centile (local, gestational and sex values)	1) 8;26;2;58 2) 5;20;5;64	IN: multiparae with history of hypertensive disorders in previous pregnancy, but no current pathology, singleton pregnancies
Alkazaleh (2006) Canada	19-23	50 (52.0)	Unreported route, color + PW, crossover	1) Mean PI > 1.45 or bilateral notching	BW < 10th (sex and gestational values)	1) 21;7;5;17	IN: AFP > 2.0 MoM and hCG > 2.5 MoM
Axt-Fliedner (2005) Germany	19-26	52 (13.5)	TA, color, crossover	1) Any RI > 0.58 2) Both RI > 0.58 3) Any RI > 0.7 4) Both RI > 0.7 5) Any notching 6) Bilateral notching	BW < 10th	1) 7;24;0;21 2) 5;9;2;36 3) 3;4;4;41 4) 2;1;5;44 5) 5;22;2;23 6) 5;10;2;36	IN: high risk singleton pregnancies: history of PE, FGR, IUD, abruption
Caforio (1999) Italy	a) 18-20 b) 22-24	i) 335 (15.2) ii) 335	Unreported route, color + PW,	1) RI > 90th centile	i) BW < 1750g ii) BW < 2500g	1ai) 36;88;15;196 1aii) 68;58;44;165 1bi) 44;75;13;202	IN: CH, DM, AID, SLE, renal disease; history of stillbirths, FGR, PE,

		(33.4)	crossover			1bii) 73;49;47;166	habitual abortion
Caruso (1993) Italy	18-24	a) 28 (10.7) b) 28 (39.0)	TA, color + PW, crossover	1) RI > 90th	a) BW < 10th (local values) b) BW < 1750g	1a) 2;6;1;19 1b) 9;0;2;17	IN: APLS
Caruso (1996) Italy	23-24	a) 42 (9.5) b) 42 (42.0)	TA, color crossover	1) Mean RI > 90th	a) BW < 10th (loca values) b) BW < 2500g	1a) 4;11;0;27 1b) 14;1;4;23	IN: CH, singleton pregnancies. EX: AID, fetal anomalies, Rhesus isoimmunisation
Coleman (2000) New Zealand	22-24	116 (26.7)	TA, color, crossover	1) Any RI > 0.58 2) Both RI > 0.58 3) Any RI ≥ 0.7 4) Both RI ≥ 0.7 5) Any notching 6) Bilateral notching 7) Any RI > 0.58 and any notching 8) Any RI > 0.7 and any notching 9) Both RI > 0.7 and any notching	BW < 10th (local values)	1) 26;52;5;33 2) 14;18;17;67 3) 17;23;14;62 4) 8;4;23;81 5) 19;26;12;59 6) 11;9;20;76 7) 19;26;12;59 8) 15;16;16;69 9) 8;3;23;82	IN: essential and secondary hypertension, renal disease, SLE, APLS, previous PE or placental abruption. EX: multiple pregnancies, fetal abnormalities

Degani (2001) Israel	15	124 (31.5)	TA, color + PW, unreported site	1) Mean PI > 1.12	BW < 10th (gestationa values)	1) 9;17;30;68	IN: singleton pregnancies, accurately dated, previous SGA infant, no chromosomal or structural anomalies
Ferrier (1994) Australia	19-24	51 (11.8)	TA, color, crossover	1) RI > 90th +/- notching on placental side or highest value if midline	BW < 10th (local and gestational values)	1) 5;9;1;36	IN: renal disease other than diabetic nephropathy
Frusca (1996) Italy	24	56 (23.2)	TA, color + pulsed, crossover	1) RI > 0.58 2) Any notching 3) Bilateral notching and RI > 0.58	BW < 10th	1) 11;13;2;30 2) 11;10;2;33 3) 6;2;7;41	IN: previous history of PE, normal blood pressure after that pregnancy
Geipel (2002) Germany	18-24	256 (17.6)	Unreported route and type, crossover	1) RI > 95th (singleton ref) 2) RI > 95th (twin ref) 3) RI > 95th + any notching (twin ref) 4) Any notching 5) Bilateral notching	BW < 10th centile (twin reference ranges local population)	1) 4;4;41;207 2) 12;24;33;187 3) 11;13;34;198 4) 16;27;29;184 5) 4;10;41;201	IN: dichorionic twins EX: fetal malformation, PPROM, unclear chorionicity, unavailable outcome
Geipel (2001) Germany	18-24	114 (14.0)	Unreported route, color, crossover	1) Bilateral notching + mean RI > 0.55 or unilateral notching + mean RI > 0.65 or no notching + mean RI > 0.7	BW < 10th (local values)	1) 7;20;9;78	IN: ICSI patients, singleton pregnancies
Geipel (2001)	18-24	32 (18.8)	Unreported route, color,	1) Bilateral notching only	BW < 10th (local values)	1) 4;4;2;22	IN: ICSI patients, twin pregnancies

Germany			crossover				
Geipel (2001) Germany	18-24	32 (21.8)	Unreported route, color, crossover	1) Bilateral notching only	BW < 10^{th} (local values)	1) 3;5;4;20	IN: twin pregnancies, (control group of (ICSI)
Haddad (1995) France	First at mean 23.8	48 (18.7)	Unreported route and site, CW	1) D/S < 10^{th} unilateral 2) D/S < 10^{th} unilateral +/- diastolic notching 3) Unilateral diastolic notching	BW < 10^{th} (local and gestational values)	1) 7;15;2;24 2) 8;18;1;21 3) 7;11;2;28	IN: aspirin treatment because of poor previous outcome, PE, eclampsia, HELLP, abruption, IUGR, IUD
Harringto (2004) UK	19-21	170 (10.2)	TA, color + pw, crossover	1) Bilateral notching + mean RI ≥ 0.55 (50^{th} centile) or unilateral notching + mean RI ≥ 0.65 (80^{th} centile)	BW < 5^{th} (local values)	1) 11;29;15;115	IN: CH, previous PE, GH, FGR, preterm labour, abruption, IUD, DM, renal disease, other medical diseases, EX: fetal anomalies
Hershkovitz (2005) UK	24	88 (26.1)	Unreported route, color + PW, crossover	1) PI (mean) > 95^{th} +/- any notching 2) PI (mean) > 95^{th} +/- bilateral notching	BW < 10^{th} centile (local, gestational and sex)	1) 17;16;6;49 2) 12;10;11;55	IN: CH, history of severe PE, thrombophilia
Konchak (1995) USA	17-22	a) 103 (17.4) b) 103 (13.5)	TA, color + PW, crossover	1) Unilateral notching 2) RI ≥ 0.7 (95^{th} centile)	a) BW < 10^{th} centile (local values) b) BW < 2500g	1a) 3;6;15;79 1b) 4;5;10;84 2a) 3;8;15;77 2b) 5;6;9;83	IN: AFP > 2.0 MoM twice or >2.5 MoM once. Singleton pregnancies, no fetal anomalies, normal amniotic fluid volume

Le Thi Huong (2006) France	2nd trimester	100 (18.0)	Unreported route and type, crossover	Any notching	BW < 10th (gestational values)	1) 3;15;12;70	IN: SLE, APLS
Nagtegaal (2005) Australia	18-2[illegible]	a) 182 (19.8) b) 182 (12.1)	Unreported route and site color + PW	1) Mean RI ≥ 0.58 + any/no notching or mean RI < 0.58 + bilateral notching	a) BW < 10th (local and gestational values) b) BW < 5th	1a) 23;104;13;42 1b) 14;113;8;47	IN: previous PE, FGR, placental abruption, recurrent miscarriages, unexplained stillbirth, CH, IDDM, thrombophilia, positive family history of PE
Soregaroli (2001) Italy	24	282 (18.0)	TA, color, crossover	1) RI > 0.6 +/- bilateral notching	BW < 10th (local and gestational values)	1) 40;45;10;187	IN: high risk pregnancy: history of GH, PE, FGR, IUD; CH, AID, renal diseases EX: multiple pregnancies, fetal chromosomal anomalies, pregnancy complications <24 wks
Vainio (2005) Finland	12-14	72 (5.6)	TVS, color, uterocervical junction	1) Bilateral notching	BW < 10th (local values)	1) 3;40;1;28	IN: high risk for PE, EX: GA <12 or >14 wks, asthma, allergy aspirin, peptic ulcer, prostaglandin inhibitors < 10 days of

							investigation
Valensise (1994) Italy	22 and 24	16 (43.7)	Unreported route, CW, crossover	1) RI > 0.58 (+2SD) + notching	BW < 10th (local and gestational)	1) 7;1;0;8	IN: CH
Van den Elzen (1995) Netherlands	a) 7-11 b) 12-13 c) 23-27	a) 320 (10.0) b) 341 (9.4) c) 351 (9.6)	Unreported route, color, crossover	1) PI > 25th a) 1.52; b)1.24; c) 0.96 2) PI > 50th a) 1.78; b)1.41; c)1.09 3) PI > 75th a) 2.07; b) 1.66; c)1.23	BW < 10th (local, sex and gestational values)	1a) 22;218;10;70 1b) 26;229;6;80 1c) 26;238;9;78 2a) 17;144;15;144 2b) 18;153;14;156 2c) 18;159;17;157 3a) 13;67;19;221 3b) 13;73;19;236 3c) 13;75;22;241	IN: age > 35 yrs at 20 weeks, DBP < 85mmHg, no history of hypertension, cardiovascular disease or diabetes, viable singleton pregnancy < 11 weeks
Venkat-raman (2001) UK	a) 16-18 b) 22-24	a) 164 (12.8) b) 163 (13.5)	Unreported route, color crossover	1) Any notching 2) Bilateral notching	BW < 10th (local, sex and gestational values)	1a) 16;62;5;81 1b) 9;21;13;120 2a) 12;25;9;118 2b) 5;8;17;133	IN: recurrent miscarriage and positive APL antibodies (no SLE or tromboembolic disease)
Yu (2002) UK	22-24	a) 351 (8.8) b) 351 (23.6) c) 351 (6.3) d) 351	TVS, color + PW, level of internal os	1) Mean PI > 1.5 (95th) 2) Bilateral notching	a) BW < 5th both twins (local, gestational singleton reference) b) BW < 5th one twin c) BW < 3rd both twins d) BW < 3rd one twin	1a) 3;15;28;305 1b) 5;13;78;255 1c) 3;15;19;314 1d) 4;14;59;274 2a) 1;11;30;309 2b) 5;7;78;261 2c) 1;11;21;318	IN: twin pregnancies, 2 live fetuses, no fetal abnormality, no TTTS

	(17.9)		2d) 5;7;58;281

* Studies are cohort studies unless otherwise stated (randomised controlled trial (RCT)). TP true positives; FP false positives; FN false negatives; TN true negatives; FPR false positive rate; BW birthweight; DBP diastolic blood pressure, PI pulsatility index; RI resistance index; a/o and/or; IN inclusion; EX exclusion; AFP alpha-fetoprotein; hCG human chorionic gonadotrophin; PE preeclampsia; PIH pregnancy induced hypertension; SGA small for gestational age;; DM diabetes mellitus; AID auto immune disease; APLS antiphospholipid syndrome; SLE systemic lupus erythematodes; IUD intra uterine demise; TTTS twin transfusion syndrome; MoM multiples of the median; PW pulsed waved; CW continuous waved, TA transabdominal, TVS transvaginal.

Appendix 21: Pooled and single estimates for uterine artery Doppler predicting small for gestational age fetuses: exclusion of studies that applied preventative treatment (sensitivity analysis).

Doppler index	No of studies	No of women	Sensitivity % (95% CI)	Specificity % (95% CI)	LR positive (95% CI)	LR negative (95% CI)
Low risk/ unselected: birth weight < 10th centile or < 2500g, 2nd trimester Doppler testing						
RI (0.58 or 90th)	9	3304	53 (42-64)	87 (79-94)	4.0 (1.6-6.3)	0.54 (0.41-0.68)
RI (0.7 or 95th)	2	665	16 (10-23)	91 (86-97)	1.9 (0.5-3.3)	0.92 (0.81-1.0)
PI	3	12097	18 (16-19)	95 (92-97)	3.4 (1.7-5.1)	0.87 (0.84-0.90)
Bilateral notching	10	10174	21 (13-28)	93 (90-96)	3.0 (1.8-4.2)	0.85 (0.78-0.92)
Unilateral notching	2	3819	17 (16-19)	93 (91-95)	2.4 (2.0-2.9)	0.89 (0.89-0.89)
Any notching	4	2162	44 (32-57)	82 (72-92)	2.5 (1.4-3.5)	0.68 (0.56-0.80)
RI or notching	4	3857	36 (33-40)	90 (83-97)	3.5 (2.0-4.9)	0.71 (0.66-0.76)
RI and notching	1	946	45 (37-53)	82 (79-84)	2.4 (1.9-3.0)	0.68 0.58-0.77)
PI or notching	2	2116	23 (19-27)	94 (93-95)	3.9 (3.0-4.7)	0.82 (0.77-0.87)
PI and notching	1	1757	12 (7-18)	99 (98-99)	9.1 (5.0-16.7)	0.89 (0.85-0.93)
S/D ratio	3	1661	34 (10-57)	88 (79-96)	2.7 (1.6-3.9)	0.76 (0.54-0.97)
Notch index	1	288	33 (13-59)	92 (88-95)	4.3 (1.9-8.4)	0.72 (0.49-0.91)
S/D or notching	1	1090	37 (27-47)	92 (88-95)	1.6 (1.2-2.1)	0.82 (0.70-0.94)
High risk: birth weight < 10th centile or < 2500g, 2nd trimester Doppler testing						
RI (0.58 or 90^{th})	3	643	68 (58-78)	73 (71-75)	2.5 (2.4-2.6)	0.44 (0.31-0.56)
RI (0.7 or 95^{th})	3	411	26 (18-33)	89 (97-92)	2.4 (1.6-3.3)	0.83 (0.75-0.91)
PI	2	445	58 (25-91)	75 (72-78)	2.3 (1.0-3.6)	0.56 (0.12-1.0)
Bilateral notching	2	279	17 (10-25)	92 (90-95)	2.2 (1.1-3.4)	0.90 (0.81-0.98)

Unilateral notching	1	103	17 (4-41)	93 (85-97)	2.4 (0.67-7.8)	0.90 (0.72-1.0)
Any notching	5	522	51 (44-58)	77 (68-86)	2.2 (1.3-3.2)	0.63 (0.52-0.75)
RI or notching	2	296	58 (48-67)	56 (17-95)	1.3 (0.14-2.5)	0.76 (0.20-1.3)
RI and notching	2	272	35 (25-44)	94 (91-96)	5.4 (3.0-7.9)	0.70 (0.60-0.80)
PI or notching	1	88	52 (31-73)	85 (74-92)	3.4 (1.7-6.4)	0.57 (0.38-0.81)
D/S ratio	1	48	78 (40-97)	62 (45-77)	2.0 (1.1-2.7)	0.36 (0.10-0.94)
D/S or notching	1	48	89 (52-100)	54 (37-70)	1.9 (1.1-2.2)	0.44 (0.31-0.56)

Appendix 22: Pooled and single estimates for uterine artery Doppler predicting small for gestational age fetuses in low risk populations: high quality studies (sensitivity analysis)

Doppler index	No of studies	No of women	Sensitivity % (95% CI)	Specificity % (95% CI)	LR positive (95% CI)	LR negative (95% CI)
Birth weight < 10th centile or < 2500g/ second trimester Doppler testing						
RI (0.58 or 90th)	2	634	53 (43-62)	62 (60-65)	1.4 (1.1-1.7)	0.76 (0.60-0.91)
RI (0.7 or 95th)	1	346	20 (8-37)	87 (83-90)	1.5 (0.72-2.9)	0.92 (0.75-1.0)
Bilateral notching	4	932	28 (21-34)	80 (63-97)	1.4 (0.13-2.7)	0.90 (0.69-1.1)
Any notching	3	958	51 (42-60)	75 (65-86)	2.1 (1.1-3.0)	0.65 (0.50-0.80)
RI or notching	1	2058	37 (31-44)	87 (85-88)	2.8 (2.3-3.4)	0.72 (0.65-0.79)
RI and notching	1	946	45 (37-53)	82 (79-84)	2.4 (1.9-3.0)	0.68 (0.58-0.77)
S/D ratio	1	288	50 (26-74)	74 (69-80)	2.0 (1.1-2.9)	0.67 (0.39-0.96)
Notch index	1	288	33 (13-59)	92 (88-95)	4.3 (1.9-8.4)	0.72 (0.49-0.91)
S/D or notching	1	768	21 (14-31)	92 (90-94)	2.7 (1.7-4.1)	0.86 (0.77-0.93)
Birth weight < 5th centile, < 3rd centile or < 1750g/ second trimester Doppler testing						
RI or notching	1	2058	37 (31-44)	87 (85-88)	2.8 (2.3-3.4)	0.72 (0.65-0.79)
RI and notching	2	1404	44 (36-51)	87 (76-96)	3.4 (0.86-5.8)	0.65 (0.53-0.76)
S/D or notching	1	768	26 (12-43)	91 (89-93)	2.9 (1.5-4.9)	0.82 (0.64-0.95)
Birth weight < 10th centile or < 2500g/ first trimester Doppler testing						
RI (>0.70 or 95th)	1	1008	67 (35-90)	75 (72-78)	2.7 (1.6-3.5)	0.44 (0.18-0.81)
PI	1	3045	44 (36-51)	87 (76-96)	3.4 (0.86-5.8)	0.65 (0.53-0.76)
Bilateral notching	1	135	90 (56-100)	8 (4-14)	0.98 (0.68-1.1)	1.3 (0.21-6.0)

Appendix 23: Pooled and single estimates for uterine artery Doppler predicting small for gestational age fetuses in high risk populations: high quality studies (sensitivity analysis)

Doppler index	No of studies	No of women	Sensitivity % (95% CI)	Specificity % (95% CI)	LR positive (95% CI)	LR negative (95% CI)
Birth weight < 10th centile or < 2500g/ second trimester Doppler testing						
RI (0.58 or 90th)	4	242	84 (77-91)	64 (49-79)	2.3 (1.4-3.3)	0.24 (0.12-0.37)
RI (0.7 or 95th)	1	116	55 (36-73)	73 (62-82)	2.0 (1.2-3.1)	0.62 (0.41-0.88)
Bilateral notching	2	279	30 (21-39)	92 (89-96)	4.0 (1.96-6.0)	0.76 (0.66-0.85)
Any notching	4	435	52 (26-78)	79 (73-85)	2.5 (1.5-3.5)	0.61 (0.31-0.92)
RI or notching	2	333	80 (73-88)	81 (77-84)	4.1 (3.3-4.9)	0.24 (0.15-0.34)
RI and notching	2	172	57 (46-67)	78 (73-83)	2.6 (1.8-3.4)	0.55 (0.42-0.69)
PI or notching	1	50	81 (61-93)	71 (49-87)	2.8 (1.6-4.6)	0.27 (0.13-0.56)
Birth weight < 5th centile, < 3rd centile or < 1750g/ second trimester Doppler testing						
PI	1	351	6 (2-14)	95 (92-97)	1.2 (0.47-3.2)	0.99 (0.92-1.0)
Bilateral notching	1	351	6 (2-14)	97 (95-99)	2.3 (0.79-6.7)	0.97 (0.91-1.0)
RI and notching	1	170	42 (23-63)	80 (72-86)	2.1 (1.2-3.4)	0.72 (0.50-0.95)
Birth weight < 10th centile or < 2500g/ first trimester Doppler testing						
Bilateral notching	1	72	75 (18-99)	41 (29-54)	1.3 (0.50-1.7)	0.61 (0.11-1.8)

Appendix 24: Search strategy for systematic review of umbilical artery Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing.

Host: Ovid

Date of search: March 2009

Years covered by search: 1950-2009

1. exp Pregnant Women/

2. exp Pregnancy/

3. pregnan$.mp.

4. exp Prenatal Diagnosis/

5. exp Ultrasonography, Prenatal/

6. exp Ultrasonography, Doppler/

7. arterial doppler.mp.

8. doppler velocimetry.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

9. doppler ultrason$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

10. umbilical arter$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

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

Appendix 25: Data extraction form for systematic review of umbilical artery Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing.

Section A: Study Information

1)Ref ID:		**4)Publication year:**	
2)Rev name:		**5)First Author:**	
3)Country:		**6)Language:**	

Section B: Data Retrieval for Umbilical Artery Doppler Study

Population

7) Healthcare Centre:

Primary care ☐$_1$ Secondary care ☐$_2$ Mixed ☐$_3$ Other ☐$_4$ Unreported ☐$_5$

8) Setting:

In-patient ☐$_1$ Out-patient ☐$_2$ Mixed ☐$_3$ Unreported ☐$_4$ Other ☐$_5$

9) Number of participating centres: _____

10) Gestation at time of index test:

<20 weeks ☐$_1$ 20-24 weeks ☐$_2$ 24-28 weeks ☐$_3$ 28-34 weeks ☐$_4$ 34-37 weeks ☐$_5$ 37-40 weeks ☐$_6$ > 40 weeks ☐$_7$ Unreported ☐$_8$ Other _____

10.i) Mean (range)_____ Unreported ☐$_3$

10.ii) Median (range) _____ Unreported ☐$_3$

11) Pregnancy:

Low Risk ☐$_1$ High Risk ☐$_2$ Unselected ☐$_3$ Unreported ☐$_4$

11.i) State high risk conditions: Unreported ☐$_3$

12) Were patients with the following conditions excluded/not included?

12.i) Previous IUGR: Yes ☐1 No ☐2

Unreported ☐3

12.ii) Insulin dependent diabetes mellitus: Yes ☐1 No ☐2

Unreported ☐3

12.iii) Chronic renal disease: Yes ☐1 No ☐2 Unreported ☐3

12.iv) Systemic lupus erythematosus: Yes ☐1 No ☐2 Unreported ☐3

12.v) Antiphospholipid syndrome: Yes ☐1 No ☐2 Unreported ☐3 12.vi) Chronic hypertension: Yes ☐1 No ☐2 Unreported ☐3

12.vii) Pre-eclampsia: Yes ☐1 No ☐2 Unreported ☐3

12.viii) Foetal chromosomal/structural anomalies: Yes ☐1 No ☐2 Unreported ☐3

13) Did all patients have singleton pregnancies?:

Yes ☐1 No ☐2 ----------/---------- Unreported ☐3

14) Were all patients primigravid?:

Yes ☐1 No ☐2 ----------/---------- Unreported ☐3

15) List other eligibility/ in-/exclusion criteria:

Not applicable ☐3

16) Study population: (describe age (mean +/- SD or median/range), ethnicity, smoking, BMI etc.)

Unreported ☐3

17) Start of patient inclusion (year) :

Unreported ☐3

18) End of patient inclusion (year) :

Unreported ☐3

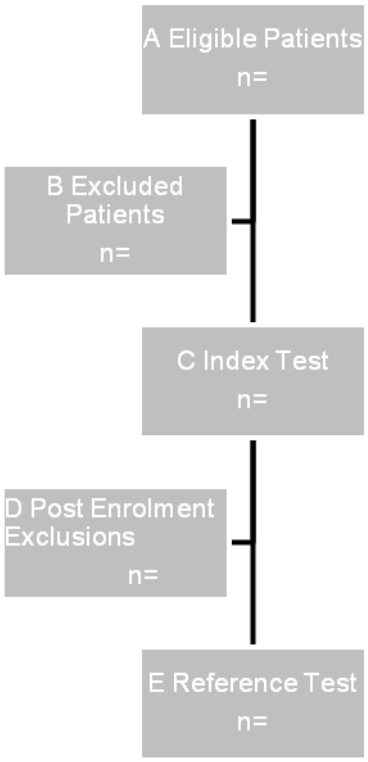
19) Study Design:

cohort ☐1 case control ☐2 RCT/CCT ☐3 cross sectional ☐4 before and after ☐5 case series ☐6 (no _____) other ☐7

19.i) Data collection: prospective ☐1 retrospective ☐2 unreported ☐3 other ☐4

19.ii) Enrolment: consecutive ☐1 arbitary (random) ☐2 unreported ☐3 other ☐4

20) Numbers:



B Reasons _____

D Reasons _____

21) Completeness of Verification:

(= E / C x 100 = %)

> 90% ☐$_1$ 81-90% ☐$_2$ < 81% ☐$_3$

Index Test

22) Description of technique:

Adequate ☐$_1$ Inadequate ☐$_2$

23) Timing of measurement (from delivery):

< 7days ☐$_1$ 7-14 days ☐$_2$ 14 -28 days ☐$_3$ > 28 days ☐$_4$ Mixture ☐$_5$

Unreported ☐$_6$

23.i) Median gestational age at delivery _____

unreported ☐$_3$

24) Measurement:

SCANNING:

24.i) Operator:

Single ☐$_1$ Multiple ☐$_2$ Unreported ☐$_3$

24.ii) Operator experience_____

unreported ☐3

24.iii) Scanning Route: Transabdominal ☐1 Transvaginal ☐2 Unreported ☐3

DOPPLER:

24.iv) Method: Continuous wave Doppler ☐1 Pulsed wave Doppler ☐2 Colour mapping ☐3 Unreported ☐4

24.v) Measurement parameter: Resistance index (RI) ☐1 Systolic / diastolic ratio ☐2 Diastolic / systolic ratio ☐3 Unilateral Diastolic notch ☐4 Bilateral diastolic notch ☐5 Pulsatility index (PI) ☐6 Reduced EDF ☐7 Absent EDF ☐8 Reversed EDF ☐9 Unreported ☐10

24.vi) Cut-off level for waveform ratio: > 2 SD ☐1 > 95th centile ☐2 > 90th centile ☐3

> 80th centile ☐4 > 50th centile ☐5 < 10th centile ☐6 < 5th centile ☐7 Unreported/NA ☐8

Other/Threshold data set:

_____

24.vii) Machine: _____ unreported ☐3

24.viii) Probe:

_____unreported ☐3

24.ix) High pass filter: _____ unreported ☐3

24.x) Pulse rePEition frequency: _____ unreported ☐3

24.xi) Size of sampling gate:

_____unreported ☐3

24.xii) Site :

_____unreported ☐3

24.xiii) Angel of insonation: _____ unreported ☐3

24.xiv) Number of consecutive waveforms: _____ unreported ☐3

24.xv) Other information:

_____

Reference Standard / Outcome

25) Measured blind form diagnostic test: Yes ☐1 No ☐2 Unclear ☐3

26) Measurement for FGR: Birthweight ☐1 Neonatal ponderal index ☐2

Skin fold thickness ☐3 MAC / OFC ☐4 Other ☐5

_____

27) Threshold: < 3rd centile ☐1 < 5th centile ☐2 < 10th centile ☐3 < 25th centile ☐4

> 2SD ☐5 Other ☐6 _____ Unclear ☐7

28) What data set was used to define threshold?

_____ unreported ☐3

29)Timing of measurement: At delivery ☐$_1$ Within 24 hrs ☐$_2$ > 24 hrs ☐$_3$

Mixture ☐$_4$ Unreported ☐$_5$

30) Marker of wellbeing e.g. Apgar score, perinatal mortality

31) Threshold and data set (if applicable):

32) Measured blind form diagnostic test: Yes ☐$_1$ No ☐$_2$ Unclear ☐$_3$

Results

	Reference Test: Threshold:			
Index test, Measurement:		Positive	Negative	**Total**
	Positive	**TP**	**FP**	
Threshold:	Negative	**FN**	**TN**	
	Total			

Appendix 26: Guide to quality assessment of included studies in systematic review of umbilical artery Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing.

Feature	Item	Applicability and criteria fulfilled when
Population spectrum	1	Refers to severity of underlying target condition, demographic features and presence of differential diagnoses and/or co-morbidity. For study to be classified as adequate: Appropriate spectrum – pregnant women, either unselected or selected (high or low risk) in any health care setting. Ideally there was prospective, consecutive recruitment.
Selection Criteria	2	Refers to inclusion/exclusion criteria. For an unselected population this would not be applicable. For a selected population high risk conditions must be explicitly documented. If the inclusion criteria for the categories were not explicitly described then the category was unclear.
Appropriate Reference standard	3	SGA: birth weight < 10^{th} centile adjusted for gestational age and based on local population values and absolute birth weight threshold < 2500g. Severe SGA: birth weight < 5^{th} or < 3^{rd} centile or < 1750g. Neonatal ponderal index < 10^{th} centile, skin fold thickness, and mid-arm circumference/head circumference were also assessed. For reference standards for wellbeing: any test performed after birth e.g. cord pH, Apgar scores, perinatal death, admission to NICU, cerebral palsy
Time period between tests	4	Time period needs to be short enough to ensure that target condition does not change. For this review this was always graded as N/A.
Verification Bias	5	If >90% of patients or a random selection of patients received verification with reference standard then answer was yes, even if the reference standard was not the same for all patients. If the number was <90% or a non-random selection then the answer was no. Unclear was utilised when the percentage could not be calculated or no information was given.
Number of reference standards used	6	This is N/A to this review: no invasive reference test.

Independent reference standard	7	The results of the index test are not incorporated in the definition of fetal growth restriction/fetal wellbeing. For this review the answer will always be yes.
Adequate description of index test	8	Type of Doppler (e.g. color wave, pulsed, etc), site of measurement, measurement parameter and cut off level used, transvaginal or transabdominal route.
Adequate description of reference standard	9	Birth weight: timing of measurement, scales used, whether baby clothed or not. Neonatal ponderal index: description of birth weight and length measurement as above. Skin fold thickness: description of site of measurement, instrument used and timing of measurement. Mid-arm circumference/ head circumference: see skin fold thickness. Wellbeing measurements: timing of measurement, threshold used, adequate description of how measurement performed or details of outcome.If this information was not provided this was classified as unclear.
Blinding of index test	10	For this review this answer will always be yes, as the reference standards can only be performed after delivery.
Blinding of reference standard	11	To confirm that blinding was present a statement in the text to the effect of "clinicians were blinded/unaware of the results of the Doppler test". If there was a statement to the contrary the answer was no. If no statement existed the answer was unclear.
Availability of clinical data	12	Clinical data refers to any information relating to the patient obtained by direct observation (e.g. age, sex, symptoms, BMI). If clinical data will be available when the test is interpreted in practice then this should be available when the test is evaluated.
Intermediate results	13	If uninterpretable, failed or intermediate results are documented or no such events occurred then the answer is yes. If it was apparent that such results have occurred but are not reported then the answer was no. If not clear whether all results were reported then answer was unclear.
Withdrawals from study	14	If clear what happened to all patients within the study e.g. flow diagram then answer was yes. If some did not receive both index and reference standard then answer was no.
Intervention	A	If after receiving the index test patients received any medical or surgical intervention then the answer was yes, and the type of intervention recorded. If a statement existed that no intervention was given the answer was no. If no statement existed and no interventions were given then the answer was unclear.

Appendix 27: Reference list of included studies in systematic review of accuracy of umbilical artery Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing.

Anyaegbunam A, Brustman L, Langer O. A longitudinal evaluation of the efficacy of umbilical Doppler velocimetry in the diagnosis of intrauterine growth retardation. *International Journal of Gynaecology & Obstetrics 34(2):121-5.* 1991.

Arauz JF, Leon JC, Velasquez PR, Jimenez GA, Perez CJ. Umbilical artery Doppler velocimetry and adverse perinatal outcome in severe pre-eclampsia. *Ginecologia y Obstetricia de Mexico.* 2008;76:440-449.

Arduini D, Rizzo G, Romanini C, Mancuso S. Fetal blood flow velocity waveforms as predictors of growth retardation. *Obstetrics and Gynecology* 1987;70:7-10.

Arduini D, Rizzo G, Soliani A, Romanini C. Doppler velocimetry versus nonstress test in the antepartum monitoring of low-risk pregnancies. *Journal of Ultrasound in Medicine* 1991;10(6):331-335.

Arduini D, Rizzo G. Prediction of fetal outcome in small for gestational age fetuses: comparison of Doppler measurements obtained from different fetal vessels. *Journal of Perinatal Medicine.* 1992;29-38.

Atkinson MW, Maher JE, Owen J, Hauth JC, Goldenberg RL, Copper RL. The predictive value of umbilical artery Doppler studies for preeclampsia or fetal growth

retardation in a preeclampsia prevention trial. *Obstetrics & Gynecology 83(4):609-12.* 1994.

Baschat AA, Weiner CP. Umbilical artery doppler screening for detection of the small fetus in need of antepartum surveillance. *American Journal of Obstetrics & Gynecology 182(1 Pt 1):154-8.* 2000.

Beattie RB, Dornan JC. Antenatal screening for intrauterine growth retardation with umbilical artery Doppler ultrasonography. *BMJ 298(6674):631-5.* 1989.

Bekedam DJ, Visser GH, van der Zee AG, Snijders RJ, Poelmann-Weesjes G. Abnormal velocity waveforms of the umbilical artery in growth retarded fetuses: relationship to antepartum late heart rate decelerations and outcome. *Early Human Development 24(1):79-89.* 1990.

Berkowitz GS, Chitkara U, Rosenberg J et al. Sonographic estimation of fetal weight and Doppler analysis of umbilical artery velocimetry in the prediction of intrauterine growth retardation: A prospective study. *American Journal of Obstetrics and Gynecology* 1988;158(5):1149-1153.

Berkowitz GS, Mehalek KE, Chitkara U, Rosenberg J, Cogswell C, Berkowitz RL. Doppler umbilical velocimetry in the prediction of adverse outcome in pregnancies at risk for intrauterine growth retardation. *Obstetrics and Gynecology* 1988;71(5)742-746.

Bilar M, Plonka T, Engel-Pietrzak K, Ronin-Walknowska E. The prediction of the condition of newborns on the basis of antepartum CTG monitoring and antepartum

Doppler flow velocity in umbilical arteries in women delivered by caesarean section. *Polski Merkuriusz Lekarski* 2005;18(104)200-204.

Bo HY, Hee CS, Syng WK. The efficacy of Doppler umbilical artery velocimetry in identifying fetal acidosis: A comparison with fetal biophysical profile. *Journal of Ultrasound in Medicine* 1992;11(1)1-6.

Bo HY, Ig HO, Pyl RL, Woo JK, Hee CS, Syng WK. Is an abnormal Doppler umbilical artery waveform ratio a risk factor for poor perinatal outcome in the non-small for gestational age fetus? *American Journal of Perinatology* 1993;10(3)245-249.

Bo Hyun Y, Lee CM, Kim SW. An abnormal umbilical artery waveform: a strong and independent predictor of adverse perinatal outcome in patients with preeclampsia. *American Journal of Obstetrics & Gynecology* 1994 171(3):713-21.

Bracero LA, Beneck D, Kirshenbaum N, Peiffer M, Stalter P, Schulman H. Doppler velocimetry and placental disease. *American Journal of Obstetrics and Gynecology* 1989;161(2)388-393.

Bracero LA, Figueroa R, Byrne DW, Han HJ. Comparison of umbilical Doppler velocimetry, nonstress testing, and biophysical profile in pregnancies complicated by diabetes. *Journal of Ultrasound in Medicine* 1996 15(4):301-8.

Brar HS, Medearis AL, DeVore GR, Platt LD. A comparative study of fetal umbilical velocimetry with continuous- and pulsed-wave Doppler ultrasonography in high-risk

pregnancies: Relationship to outcome. *American Journal of Obstetrics and Gynecology* 1989;160(2): 375-378.

Brar HS, Medearis AL, DeVore GR, Platt LD. Maternal and fetal blood flow velocity waveforms in patients with preterm labor: Prediction of successful tocolysis. *American Journal of Obstetrics and Gynecology* 1988;159(4):947-950.

Bruinse HW, Sijmons EA, Reuwer PJ. Clinical value of screening for fetal growth retardation by Doppler ultrasound. *Journal of Ultrasound in Medicine* 1989 8(4):207-9.

Bruner JB, Levy DW, Arger PH. Doppler ultrasonography of the umbilical cord in complicated pregnancies. *Southern Medical Journal* 1993 86(4):418-22.

Burke G, Stuart B, Crowley P, Scanaill SN, Drumm J. Is intrauterine growth retardation with normal umbilical artery blood flow a benign condition? BMJ 1990 300(6731):1044-5.

Carroll BC, Bruner JP. Umbilical artery doppler velocimetry in pregnancies complicated by oligohydramnios. *Journal of Reproductive Medicine for the Obstetrician and Gynecologist* 2000;45(7):562-566.

Chambers SE, Hoskins PR, Haddad NG, Johnstone FD, McDicken WN, Muir BB. A comparison of fetal abdominal circumference measurements and Doppler ultrasound in the prediction of small-for-dates babies and fetal compromise. *British Journal of Obstetrics & Gynaecology 96(7):803-8.* 1989.

Chan FY, Pun TC, Lam P, Lam C, Lee CP, Lam YH. Fetal cerebral Doppler studies as a predictor of perinatal outcome and subsequent neurologic handicap. *Obstetrics & Gynecology* 1996 87(6):981-8.

Chang TC, Robson SC, Spencer JA, Gallivan S. Identification of fetal growth retardation: comparison of Doppler waveform indices and serial ultrasound measurements of abdominal circumference and fetal weight. *Obstetrics &* Gynecology1993 82(2):230-6.

Chang TC, Robson SC, Spencer JA, Gallivan S. Prediction of perinatal morbidity at term in small fetuses: comparison of fetal growth and Doppler ultrasound. *British Journal of Obstetrics & Gynaecology* 1994 101(5):422-7.

Chanprapaph P, Tongsong T, Siriaree S. Validity of antenatal diagnosis of intrauterine growth restriction by umbilical Doppler waveform index. *Journal of the Medical Association of Thailand 87(5):492-6.* 2004.

Chua S, Arulkumaran S, Kurup A, Anandakumar C, Selemat N, Ratnam SS. Search for the most predictive tests of fetal well-being in early labor. *Journal of Perinatal Medicine 24(3):199-206.* 1996.

Cosmi E, Ambrosini G, D'Antona D, Saccardi C, Mari G. Doppler, cardiotocography, and biophysical profile changes in growth-restricted fetuses. *Obstetrics & Gynecology 106(6):1240-5.* 2005.

Craigo SD, Beach ML, Harvey-Wilkes KB, D'Alton ME. Ultrasound predictors of neonatal outcome in intrauterine growth restriction. *American Journal of Perinatology 13(8):465-71.* 1996.

de Rochembeau, Rudigoz RC, Le MG, Gaucherand P, Chollat L, Dargent D. The contribution of umbilical Doppler velocimetry in suspected echographic IUGR. *Journal de Gynecologie, Obstetrique et Biologie de la Reproduction 17(8):1021-30.* 1988.

de Rocehmbeau, Jabbour N, Mellier G. Umbilical Doppler velocimetry in prolonged pregnancies. *Revue Francaise de Gynecologie et d Obstetrique 87(5):289-94.* 1992.

Degani S, Paltiely Y, Lewinsky R, Shapiro I, Sharf M. Fetal blood flow velocity waveforms in pregnancies complicated by intrauterine growth retardation. *Israel Journal of Medical Sciences 26(5):250-4.* 1990.

Degani S, Gonen R, Shapiro I, Paltiely Y, Sharf M. Doppler flow velocity waveforms in fetal surveillance of twins: a prospective longitudinal study. *Journal of Ultrasound in Medicine 11(10):537-41.* 1992.

Dempster J, Mires GJ, Patel N, Taylor DJ. Umbilical artery velocity waveforms: Poor association with small-for-gestational-age babies. *British Journal of Obstetrics and Gynaecology* 1989;96(6)692-696.

Divon MY, Guidetti DA, Braverman JJ, Oberlander E, Langer O, Merkatz IR. Intrauterine growth retardation--a prospective study of the diagnostic value of real-time

sonography combined with umbilical artery flow velocimetry. *Obstetrics & Gynecology 72(4):611-4.* 1988.

Divon MY, Girz BA, Sklar A, Guidetti DA, Langer O. Discordant twins--a prospective study of the diagnostic value of real-time ultrasonography combined with umbilical artery velocimetry. *American Journal of Obstetrics & Gynecology 161(3):757-60.* 1989.

Dubinsky T, Lau M, Powell F et al. Predicting poor neonatal outcome: a comparative study of noninvasive antenatal testing methods. *AJR American Journal of Roentgenology 168(3):827-31.* 1997.

Eronen M, Kari A, Pesonen E, Kaaja R, Wallgren EI, Hallman M. Value of absent or retrograde end-diastolic flow in fetal aorta and umbilical artery as a predictor of perinatal outcome in pregnancy-induced hypertension. *Acta Paediatrica 82(11):919-24.* 1993.

Ezra Y, Jones J, Farine D. Umbilical artery waveforms in triplet and quadruplet pregnancies. *Gynecologic & Obstetric Investigation 47(4):239-43.* 1999.

Faber R, Robel R, Viehweg B. Predictive value of uteroplacental and fetal perfusion in premature labour as measured by Doppler ultrasound *Geburtshilfe und Frauenheilkunde* 1996;56(12):660-664.

Farine D, Granovsky-Grisaru S, Ryan G et al. Umbilical artery blood flow velocity in pregnancies complicated by systemic lupus erythematosus. *Journal of Clinical Ultrasound 26(8):379-82.* 1998.

Farmakides G, Schulman H, Winter D, Ducey J, Guzman E, Penny B. Prenatal surveillance using nonstress testing and Doppler velocimetry. *Obstetrics and Gynecology* 1988;71(2):184-187.

Ferchiou-Cherif M, Zhioua F, Hafsia S et al. Evaluation of neonatal prognosis using Doppler velocimeter in cases of a high risk fetus. *Pediatrie 48(11):828-31.* 1993.

Figueras RF, Bennasar M, Eixarch E et al. Arterial, venous and intracardiac parameters in growth-restricted fetuses: Associations with adverse perinatal outcome. *Ultrasound Review of Obstetrics and Gynecology* 2004;4(3):179-185.

Figueras F, Eixarch E, Gratacos E, Gardosi J. Predictiveness of antenatal umbilical artery Doppler for adverse pregnancy outcome in small-for-gestational-age babies according to customised birthweight centiles: population-based study. *BJOG: An International Journal of Obstetrics & Gynaecology.* 2008;115:590-594.

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

Fong KW, Ohlsson A, Hannah ME et al. Prediction of perinatal outcome in fetuses suspected to have intrauterine growth restriction: Doppler US study of fetal cerebral, renal, and umbilical arteries. *Radiology 213(3):681-9.* 1999.

Forouzan I, Cohen AW. Can umbilical and arcuate artery Doppler velocimetry predict fetal distress among prolonged pregnancies? *Journal of Ultrasound in Medicine 10(1):15-7.* 1991.

Gaziano E, Knox GE, Wager GP, Bendel RP, Boyce DJ, Olson J. The predictability of the small-for-gestational-age infant by real-time ultrasound-derived measurements combined with pulsed Doppler umbilical artery velocimetry. *American Journal of Obstetrics & Gynecology 158(6 Pt 1):1431-9.* 1988.

Gaziano EP, Knox GE, Bendel RP, Calvin S, Brandt D. Is pulsed Doppler velocimetry useful in the management of multiple-gestation pregnancies? *American Journal of Obstetrics & Gynecology 164(6 Pt 1):1426-31; discussion 1431-3.* 1991.

Gaziano EP, Knox H, Ferrera B, Brandt DG, Calvin SE, Knox GE. Is it time to reassess the risk for the growth-retarded fetus with normal Doppler velocimetry of the umbilical artery? *American Journal of Obstetrics & Gynecology 170(6):1734-41; discussion 1741-3.* 1994.

Giles WB, Trudinger BJ, Cook CM, Connelly A. Umbilical artery flow velocity waveforms and twin pregnancy outcome. *Obstetrics and Gynecology* 1988;72(6):94-897.

Ghosh GS, Gudmundsson S. Uterine and umbilical artery Doppler are comparable in predicting perinatal outcome of growth-restricted fetuses. *BJOG: An International Journal of Obstetrics and Gynaecology* 2009;116(3):424-430.

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, & Reproductive Biology 71(1):11-9.* 1997.

Gonzalez P, Robert JA, Espinoza R, Gormaz G, Tzunekawa H. Blood flow in umbilical artery in fetus with intrauterine growth retardation. *Revista Chilena de Obstetricia y Ginecologia 60(1):23-7.* 1995.

Gonzalez JM, Stamilio DM, Ural S, Macones GA, Odibo AO. Relationship between abnormal fetal testing and adverse perinatal outcomes in intrauterine growth restriction. *American Journal of Obstetrics & Gynecology.* 2007;e48-e51.

Gramellini D, Piantelli G, Verrotti C, Fieni S, Chiaie LD, Kaihura C. Doppler velocimetry and non stress test in severe fetal growth restriction. *Clinical & Experimental Obstetrics & Gynecology 28(1):33-9.* 2001.

Gudmundsson S, Marsal K. Ultrasound Doppler evaluation of uteroplacental and fetoplacental circulation in pre-eclampsia. *Archives of Gynecology and Obstetrics* 1988;243(4):199-206.

Gudmundsson S, Marsal K. Blood velocity waveforms in the fetal aorta and umbilical artery as predictors of fetal outcome: a comparison. *American Journal of Perinatology 8(1):1-6.* 1991.

Guzman E, Schulman H, Bracero L, Rochelson B, Farmakides G, Coury A. Uterine-umbilical artery Doppler velocimetry in pregnant women with systemic lupus erythematosus. *Journal of Ultrasound in Medicine 11(6):275-81.* 1992.

Hack KEA, Kaandorp JJ, Derks JB, Elias SG, Pistorius L, Visser GHA. The value of umbilical artery Doppler velocimetry in the antenatal surveillance of uncomplicated monochorionic twin pregnancies. *Ultrasound in Obstetrics and Gynecology* 2008;31(6):662-668.

Haddad NG, Johnstone FD, Chambers SE, Hoskins PR, McDicken WN. Umbilical artery doppler flow velocity wave form analysis and the outcome of hypertensive pregancies. *Journal of Obstetrics and Gynaecology* 1988;9(1):9-13.

Hastie SJ, Danskin F, Neilson JP, Whittle MJ. Prediction of the small for gestational age twin fetus by Doppler umbilical artery waveform analysis. *Obstetrics & Gynecology 74(5):730-3.* 1989.

Hastie SJ, Brown MF, Whittle MJ. Predictive values of umbilical artery waveforms and repeat cardiotocography in pregnancies complicated by nonreactive cardiotocography. *European Journal of Obstetrics Gynecology and Reproductive Biology* 1990;34(1-2):67-72.

Hecher K, Ertl U, Spernol R, Haselbach H, Szalay S. Comparison of the clinical value of 2 frequently used indices of fetal blood flow analysis (pulsatility index PI and resistance index RI). *Geburtshilfe und Frauenheilkunde 48(11):809-13.* 1988.

Hitschold T, Weiss E, Berle P. Umbilical artery flow velocity waveforms, mode of delivery and fetal outcome in post-term pregnancy. *Zeitschrift fur Geburtshilfe und Perinatologie* 1988;192(5):197-202.

Hutter W, Grab D, Schneider D, Terinde R, Wolf A. Continuous-wave Doppler investigation of uteroplacental vessels in high-risk pregnancies as predictor of fetal growth retardation and pregnancy-induced hypertension. *Gynecologic & Obstetric Investigation 38(2):90-5.* 1994.

Joern H, Schroeder W, Sassen R, Rath W. Predictive value of a single CTG, ultrasound and Doppler examination to diagnose acute and chronic placental insufficiency in multiple pregnancies. *Journal of Perinatal Medicine* 1997;25(4):325-332.

Joern H, Rath W. Correlation of Doppler velocimetry findings in twin pregnancies including course of pregnancy and fetal outcome. *Fetal Diagnosis and Therapy* 2000;15(3):160-164.

Jorn H, Funk A, Fendel H. [Doppler ultrasound diagnosis in post-term pregnancy. *Geburtshilfe und Frauenheilkunde 53(9):603-8.* 1993.

Jorn H, Funk A, Kuhlwein H, Schmidt A. Doppler velocity waveform analysis of the fetal venous precardiac blood flow in normal and complicated pregnancies. *Zeitschrift fur Geburtshilfe und Perinatologie* 1994;198(2):56-61.

Karsdorp VHM, Van Vugt JMG, Van Geijn HP et al. Clinical significance of absent or reversed end diastolic velocity waveforms in umbilical artery. *Lancet* 1994;344(8938):1664-1668.

Kay HH, Carroll BB, Dahmus M, Killam AP. Sonographic measurements with umbilical and uterine artery Doppler analysis in suspected intrauterine growth retardation. *Journal of Reproductive Medicine 36(1):65-8.* 1991.

Kofinas AD, Penry M, Nelson LH, Meis PJ, Swain M. Uterine and umbilical artery flow velocity waveform analysis in pregnancies complicated by chronic hypertension or preeclampsia. *Southern Medical Journal* 1990;83(2):150-155.

Lakhkar BN, Rajagopal KV, Gourisankar PT. Doppler prediction of adverse perinatal outcome in PIH and IUGR. *Indian Journal of Radiology & Imaging* 2006;16(1):109-116.

Le Thi HD, Wechsler B, Vauthier-Brouzes D et al. The second trimester Doppler ultrasound examination is the best predictor of late pregnancy outcome in systemic lupus erythematosus and/or the antiphospholipid syndrome. *Rheumatology 45(3):332-8.* 2006.

Lombardi SJ, Rosemond R, Ball R, Entman SS, Boehm FH. Umbilical artery velocimetry as a predictor of adverse outcome in pregnancies complicated by oligohydramnios. *Obstetrics & Gynecology 74(3 Pt 1):338-41.* 1989.

Lowery CL, Jr., Goldenberg RL, Baker RC et al. Screening tests for intrauterine growth retardation: a comparison of umbilical artery Doppler to real-time ultrasound. *Echocardiography 7(2):159-64.* 1990.

Maria FG, Cherchi PL, D'Antona D et al. Umbilical artery pulsatility index in pregnancies complicated by insulin-dependent diabetes mellitus without hypertension. *Gynecologic and Obstetric Investigation* 2001;51(3):173-177.

Maulik D, Yarlagadda P, Youngblood JP, Ciston P. The diagnostic efficacy of the umbilical arterial systolic/diastolic ratio as a screening tool: a prospective blinded study. *American Journal of Obstetrics and Gynecology* 1990;162(6):1518-1525.

Maunu J, Ekholm E, Parkkola R et al. Antenatal Doppler measurements and early brain injury in very low birth weight . *Journal of Pediatrics 150(1):51-56 e1.* 2007.

McCowan LM, Duggan PM. Abnormal internal carotid and umbilical artery Doppler in the small for gestational age fetus predicts an adverse outcome. *Early Human Development 30(3):249-59.* 1992.

McCowan LME, Naden RP. Doppler ultrasound in pregnancies with hypertension. *Australian and New Zealand Journal of Obstetrics and Gynaecology* 1992;32(3):225-230.

McCowan LM, Harding JE, Stewart AW. Umbilical artery Doppler studies in small for gestational age babies reflect disease severity. *BJOG: An International Journal of Obstetrics & Gynaecology 107(7):916-25.* 2000.

Miller JM, Jr., Gabert HA. Comparison of dynamic image and pulsed Doppler ultrasonography for the diagnosis of the small-for-gestational-age fetus. *American Journal of Obstetrics & Gynecology 166(6 Pt 1):1820-3; discussion 1823-6.* 1992.

Miyashita S, Chiba Y. Doppler studies can predict long-term outcome of growth-restricted fetuses. *Journal of Medical Ultrasound* 2002;10(2):86-93.

Moon AE. Using S/D ratios to predict fetal outcome. *Journal of Diagnostic Medical Sonography* 1999;15(2):59-64.

Mulders LGM, Wijn PFF, Jongsma HW, Hein PR. A comparative study of three indices of umbilical blood flow in relation to prediction of growth retardation. *Journal of Perinatal Medicine* 1987;15(1):3-12.

Mulders LGM, Jongsma HW, Hein PR. Uterine and umbilical artery blood flow velocity waveforms and their validity in the prediction of fetal compromise. *European Journal of Obstetrics Gynecology and Reproductive Biology* 1989;31(2):143-154.

Newnham JP, Patterson LL, James IR, Diepeveen DA, Reid SE. An evaluation of the efficacy of Doppler flow velocity waveform analysis as a screening test in pregnancy. *American Journal of Obstetrics and Gynecology* 1990;162(2):403-410.

Niknafs P, Sibbald J. Accuracy of single ultrasound parameters in detection of fetal growth restriction. *American Journal of Perinatology 18(6):325-34.* 2001.

Nordstrom UL, Patel NB, Taylor DJ. Umbilical artery waveform analysis and biophysical profile. A comparison of two methods to identify compromised fetuses.

European Journal of Obstetrics Gynecology and Reproductive Biology 1989;30(3):245-251.

Odendaal HJ, Theron AM, Carstens MH, de JM, Grove D. Intrauterine deaths in high-risk pregnancies with normal and borderline umbilicial artery Doppler flow velocity waveforms. *South African Journal of Obstetrics and Gynaecology* 2008;14(1):33-37.

Ogunyemi D, Stanley R, Lynch C, Edwards D, Fukushima T. Umbilical artery velocimetry in predicting perinatal outcome with intrapartum fetal distress. *Obstetrics and Gynecology* 1992;80(3):377-380.

Ott WJ. Comparison of dynamic image and pulsed Doppler ultrasonography for the diagnosis of intrauterine growth retardation. *Journal of Clinical Ultrasound 18(1):3-7.* 1990.

Ott WJ. Intrauterine growth restriction and Doppler ultrasonography. *Journal of Ultrasound in Medicine.* 2000;quiz.

Owen P, Khan KS, Howie P. Single and serial estimates of amniotic fluid volume and umbilical artery resistance in the prediction of intrauterine growth restriction. *Ultrasound in Obstetrics & Gynecology 13(6):415-9.* 1999.

Ozcan T, Sbracia M, d'Ancona RL, Copel JA, Mari G. Arterial and venous Doppler velocimetry in the severely growth-restricted fetus and associations with adverse perinatal outcome. *Ultrasound in Obstetrics & Gynecology 12(1):39-44.* 1998.

Ozden S, Ficicioglu C, Guner R, Arioglu P, Oral O. Comparison of the intrapartum analysis of Doppler blood flow velocity waveform of the umbilical artery and fetal heart rate tracing for the prediction of perinatal outcome. *Journal of Obstetrics and Gynaecology* 1998;18(5):445-450.

Pattinson RC, Kriegler E, Odendaal HJ, Muller LM, Kirsten G. Increased placental resistance and late decelerations associated with severe proteinuric hypertension predicts poor fetal outcome. *South African Medical Journal Suid-Afrikaanse Tydskrif Vir Geneeskunde 75(5):211-4.* 1989.

Pattinson RC, Norman K, Odendaal HJ. The use of Doppler velocimetry of the umbilical artery before 24 weeks' gestation to screen for high-risk pregnancies. *South African Medical Journal* 1993;83(10):734-736.

Pattinson RC, Odendaal HJ, Kirsten G. The relationship between absent end-diastolic velocities of the umbilical artery and perinatal mortality and morbidity. *Early Human Development* 1993;33(1):61-69.

Poulain P, Palaric JC, Paris-Liado J, Jacquemart F. Fetal umbilical Doppler in a population of 541 high-risk pregnancies: prediction of perinatal mortality and morbidity. Doppler Study Group. *European Journal of Obstetrics, Gynecology, & Reproductive Biology 54(3):191-6.* 1994.

Puzey MS, Lindow SW. The value of umbilical artery Doppler analysis in the prediction of fetal acidosis in labour. *South African Medical Journal 81(5):251-3.* 1992.

Rocca MM, Said MS, Khamis MY, Ghanem IA, Karkour TA. The value of Doppler study of the umbilical artery in predicting perinatal outcome in pre-eclamptic patients. *Journal of Obstetrics & Gynaecology 21(5):427-31.* 1995.

Rochelson BL, Schulman H, Fleischer A. The clinical significance of Doppler umbilical artery velocimetry in the small for gestational age fetus. *American Journal of Obstetrics and Gynecology* 1987;156(5):1223-1226.

Rochelson B, Bracero LA, Porte J, Farmakides G. Diagnosis of intrauterine growth retardation as a two-step process with morphometric ultrasound and Doppler umbilical artery velocimetry. *Journal of Reproductive Medicine 37(11):925-9.* 1992.

Rognerud Jensen OH, Guimaraes MS. Prediction of fetal outcome by Doppler examination and by the non-stress test. *Acta Obstetricia et Gynecologica Scandinavica 70(4-5):271-4.* 1991.

Rognerud Jensen OH. Doppler velocimetry in twin pregnancy. *European Journal of Obstetrics Gynecology and Reproductive Biology* 1992;45(1):9-12.

Rudigoz RC, Thome Saint-Paul M, Gindre T, de RB, Gaucherand P, Wong A. Umbilical and cerebral Doppler velocimetry hypertension in pregnancy. *Journal de Gynecologie, Obstetrique et Biologie de la Reproduction.* 1991;436-442.

Sarno J, Ahn MO, Brar HS, Phelan JP, Platt LD. Intrapartum Doppler velocimetry, amniotic fluid volume, and fetal heart rate as predictors of subsequent fetal distress. I.

An initial report. *American Journal of Obstetrics and Gynecology* 1989;161(6):1508-1514.

Schulman H, Winter D, Farmakides G et al. Pregnancy surveillance with Doppler velocimetry of uterine and umbilical arteries. *American Journal of Obstetrics & Gynecology 160(1):192-6.* 1989.

Sezik M, Tuncay G, Yapar EG. Prediction of Adverse Neonatal Outcomes in Preeclampsia by Absent or Reversed End-Diastolic Flow Velocity in the Umbilical Artery. *Gynecologic and Obstetric Investigation* 2004;57(2):109-113.

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

Skodler WD, Philipp K, Pateisky N, Reinold E. Practical use of blood flow measurement with pulsed ultrasound in the 3d trimester of an at-risk pregnancy. *Ultraschall in der Medizin 10(4):230-4.* 1989.

Soothill PW, Ajayi RA, Campbell S, Nicolaides KH. Prediction of morbidity in small and normally grown fetuses by fetal heart rate variability, biophysical profile score and umbilical artery Doppler studies. *British Journal of Obstetrics & Gynaecology 100(8):742-5.* 1993.

Soregaroli M, Bonera R, Danti L et al. Prognostic role of umbilical artery Doppler velocimetry in growth-restricted fetuses. *Journal of Maternal-Fetal & Neonatal Medicine 11(3):199-203.* 2002.

Spinillo A, Montanari L, Bergante C, Gaia G, Chiara A, Fazzi E. Prognostic value of umbilical artery Doppler studies in unselected preterm deliveries. *Obstetrics & Gynecology 105(3):613-20.* 2005.

Strigini FA, De LG, Lencioni G, Scida P, Giusti G, Genazzani AR. Middle cerebral artery velocimetry: different clinical relevance depending on umbilical velocimetry. *Obstetrics & Gynecology 90(6):953-7.* 1997.

Szalay S, Hecher K, Spernol R, Wimmer-Hebein D. Doppler ultrasound of the umbilical artery of the dystrophic child. *Ultraschall in der Medizin 12(3):124-6.* 1991.

Tchirikov M, Strohner M, Forster D, Huneke B. A combination of umbilical artery PI and normalized blood flow volume in the umbilical vein: Venous-arterial index for the prediction of fetal outcome. *European Journal of Obstetrics Gynecology and Reproductive Biology* 2009;142(2):129-133.

Thiebaugeorges O, Ancel P-Y, Goffinet F, Breart G. A population-based study of 518 very preterm neonates from high-risk pregnancies: Prognostic value of umbilical and cerebral artery Doppler velocimetry for mortality before discharge and severe neurological morbidity. *European Journal of Obstetrics Gynecology and Reproductive Biology* 2006;128(1-2):69-76.

To WW, Chan AM, Mok KM. Use of umbilical-cerebral Doppler ratios in predicting fetal growth restriction in near-term fetuses. *Australian & New Zealand Journal of Obstetrics & Gynaecology 45(2):130-6.* 2005.

Todros T, Ferrazzi E, Arduini D et al. Performance of Doppler ultrasonography as a screening test in low risk pregnancies: results of a multicentric study. *Journal of Ultrasound in Medicine 14(5):343-8.* 1995.

Todros T, Ronco G, Fianchino O et al. Accuracy of the umbilical arteries Doppler flow velocity waveforms in detecting adverse perinatal outcomes in a high-risk population. *Acta Obstetricia et Gynecologica Scandinavica 75(2):113-9.* 1996.

Torres PJ, Gratacos E, Alonso PL. Umbilical artery Doppler ultrasound predicts low birth weight and fetal death in hypertensive pregnancies. *Acta Obstetricia et Gynecologica Scandinavica 74(5):352-5.* 1995.

Trudinger BJ, Cook CM, Giles WB et al. Fetal umbilical artery velocity waveforms and subsequent neonatal outcome. *British Journal of Obstetrics & Gynaecology 98(4):378-84.* 1991.

Tyrrell S, Obaid AH, Lilford RJ. Umbilical artery Doppler velocimetry as a predictor of fetal hypoxia and acidosis at birth. *Obstetrics & Gynecology 74(3 Pt 1):332-7.* 1989.

van Asselt K, Gudmundsson S, Lindqvist P, Marsal K. Uterine and umbilical artery velocimetry in pre-eclampsia. *Acta Obstetricia et Gynecologica Scandinavica 77(6):614-9.* 1998.

Vergani P, Andreotti C, Roncaglia N et al. Doppler predictors of adverse neonatal outcome in the growth restricted fetus at 34 weeks' gestation or beyond. *American Journal of Obstetrics and Gynecology* 2003;189(4):1007-1011.

Vintzileos AM, Campbell WA, Rodis JF, McLean DA, Fleming AD, Scorza WE. The relationship between fetal biophysical assessment, umbilical artery velocimetry, and fetal acidosis. *Obstetrics & Gynecology 77(4):622-6.* 1991.

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

Weiner Z, Farmakides G, Schulman H, Lopresti S, Schneider E. Surveillance of growth-retarded fetuses with computerized fetal heart rate monitoring combined with Doppler velocimetry of the umbilical and uterine arteries. *Journal of Reproductive Medicine 41(2):112-8.* 1996.

Wong SF, Chan FY, Cincotta RB, McIntyre DH, Stone M. Use of umbilical artery Doppler velocimetry in the monitoring of pregnancy in women with pre-existing diabetes. *Australian & New Zealand Journal of Obstetrics & Gynaecology 43(4):302-6.* 2003.

Worrell JA, Fleischer AC, Drolshagen LF, Durmon GR, Shah DM. Duplex Doppler sonography of the umbilical arteries: Predictive value in IUGR and correlation with birth weight. *Ultrasound in Medicine and Biology* 1991;17(3):207-210.

Yildirim G, Turhan E, Aslan H et al. Perinatal and neonatal outcomes of growth restricted fetuses with positive end diastolic and absent or reversed umbilical artery doppler waveforms. *Saudi Medical Journal.* 2008;29:403-408.

Zhou R, Li W, Liu S, Tang L. Pulsed-wave Doppler ultrasound evaluation of flow velocity waveforms of fetal umbilical artery and uterine artery as predictors of neonatal outcome. *Hua-Hsi i Ko Ta Hsueh Hsueh Pao [Journal of West China University of Medical Sciences] 22(4):424-7.* 1991.

Zimmermann P, Alback T, Koskinen J, Vaalamo P, Tuimala R, Ranta T. Doppler flow velocimetry of the umbilical artery, uteroplacental arteries and fetal middle cerebral artery in prolonged pregnancy. *Ultrasound in Obstetrics & Gynecology 5(3):189-97.* 1995.

Appendix 28: Table of study characteristics of included studies for systematic review of umbilical artery Doppler to predict small for gestational age fetuses / compromise of fetal wellbeing.

First Author (year)	Population Maternal age (yrs) (country/study design)	No of fetuses analysed	Gestational age at test (weeks)	Reference Standard SGA	Incidence of SGA (%)	Reference standard Fetal compromise	Details of Index test
Anyaegbunam (1990)	High risk. INC: Patients with hypertension or suspected SGA Hypertension and SGA mean age 29+/-10. (USA) (Cohort, prospective)	149	Third trimester (32-40)	BW<10th centile	19.4%	NA	TA, pulsed, site not reported, SD≥3
Arauz (2008)	High risk INC: women with severe pre-eclampsia (one of systolic BP>160mmHg or diastolic>110mmHg 6 hrs apart, proteinuria>2grms in 24 hrs or 3+ on dipstick twice 6 hrs apart without UTI, altered vision, epigastric pain, oliguria, pulmonary oedema, thromobocytopenia, abnormal hepatic function) EXC: multiple pregnancy, essential hypertension,	43	27-33 weeks (test to delivery interval 7 days)	BW<5th centile	32.6%	Admission to NICU, RDS, IVH, NEC, perinatal mortality	Route not reported, pulsed and color, middle portion of cord, PI>95th centile or AREDF

	diabetes mellitus, autoimmune conditions, kidney disease, chromosomal and structural malformations. Mean maternal age 30 years (sd+/-5.34) normal Doppler 30+/-5.40 abnormal Doppler (Mexico) (Cohort, prospective)						
Arduini (1987)	High risk. INC: Singleton, suspected SGA with confirmed EDD, at risk of hypertension EXC: Patients receiving tocolytics Mean age not reported (Italy) (Cohort, prospective)	75	26-28 weeks Test to delivery interval mean 8.3 weeks +/- 2.1 (5-12)	BW<10th centile (adjusted for ga/local/height /weight/parity/ sex)	30.7%	NA	TA, pulsed, luminal centre, PI mean+1sd
Arduini (1991)	Low risk. INC: 1000 patients with low risk pregnancies, singleton, certain gestational age. Mean age 28.71+/-4.76 (Italy) (Cohort)	1000	Mean 38.29+/- 1.57 weeks. Test to delivery interval mean	BW<10th centile	6.4%	Adverse perinatal outcome (one or more of cs for fetal distress, Apgar score <7 at 5 mins, admission to NICU for	Route not reported, continuous wave, site not reported, SD>3.

			13.21+/-7.93 days.			asphyxia for >48hrs).	
Arduini (1992)	High risk. INC: Singleton, accurate EDD, AC<5th centile or EFW<10th centile, successful Doppler. EXC: No chromosomal or structural anomalies. Mean age 29.4+/-4.3 (18-36) (Italy) (Cohort)	120	Mean 32.2+/-3.0 (24-36) weeks. First investigation after diagnosis used for analysis.	NA	NA	Adverse perinatal outcome (one or more of perinatal death, cs due to abnormal FHR, Apgar score <7 at 5 mins, admission to NICU for asphyxia >48hrs).	Route not reported, color+pulsed, site not reported, PI>95th centile.
Atkinson (1994)	Low risk. INC: Singleton, low risk nullips enrolled on double blind trial of low dose (60mg) aspirin for PE prevention. EXC: Renal disease, collagen vascular disease, diabetes mellitus, multiple gestations, chronic hypertension. Mean age 19.9+/-2.7. (USA) (RCT, prospective, consecutive)	490	20-42 weeks.	BW<10th centile (ga/local)	6.73%	NA	TA, continuous wave, site not reported, SD>90th centile for ga.

Baschat (2000)	High risk. INC: Ultrasonographic biometric results suggestive of IUGR, delivery >23+6 weeks. Mean age not reported. (USA) (Cohort, consecutive)	302	Third trimester. Mean test to delivery interval 2 weeks (1day-9 weeks).	BW<10th centile (local/ga)	35.4%	Admission to NICU, NEC, acidaemia (umbilical cord artery and vein pH<10th centile).	TA, method and site not reported, SD mean +2sd.
Beattie (1989)	Low risk. INC: 2097 ultrasonically dated singleton pregnancies attending hospital within 7 days of their 28th gestational week. Mean age 26.3+/-5.5. (UK) (Cohort)	2097	28, 34 and 38 weeks.	BW<5th centile	4.14%	NA	TA, continuous wave, characteristic waveform, PI/RI/SD>90th centile.
Bekedam (1990)	High risk. INC: Patients admitted for suspected IUGR who developed late heart rate decelerations, delivery by elective cs, BW<10th centile, accurate gestation EXC: fetal chromosomal and structural anomalies. Mean age normal Doppler 34; abnormal Doppler	70	Within 72 hrs of delivery. Mean age at delivery normal	NA	NA	Intra-uterine death, neonatal death, intubation>7days , IVH, NEC.	Route not reported, continuous wave, site not reported, AEDF or raised PI (cut-

	31 (Netherlands) (Cohort, retrospective, consecutive)		Doppler 33 weeks; abnormal 33.2 weeks.				off not reported.)
Berkowitz (1988)	High risk. INC: Singleton, known risk factors or clinical suspicion of IUGR Mean age 26 +/-6.0 (USA) (Cohort, prospective)	168	30-42 weeks, mean 31.6.	BW<10th centile	25.0%	NA	TA, continuous wave, site not reported, mean SD≥3
Berkowit (1988)	High risk. INC: Singletons, known risk factors or EFW<10th centile, accurate gestation. Mean age normal Doppler 26.5+/-6.1; abnormal 27.1+/-5.5 (USA)(Cohort, consecutive)	129	30-42 weeks	NA	NA	Apgar at 1 and 5 mins <7, neonatal resuscitation, admission to NICU, RDS, neonatal death, perinatal morbidity.	TA, continuous wave, site not reported, mean SD≥3
Bilar	High risk.	213	Third	NA	NA	Poor outcome	TA, method

(2005)	INC: Pregnant women delivered by cs (for preventive and urgent indications before labour) who had umbilical artery Doppler 7 days prior) Age not reported (Poland) (Cohort)		trimester.			(Apgar at 5 mins<4 and or umbilical artery pH≤7.20/BE≤11 mmol/l	and site not reported, PI/SD mean+2sd, AREDF.
Bo Hyun Yoon (1992)	High risk. INC: Singletons, delivery by cs within 16 hrs of ultrasound, intact membranes, not in labour EXC: structural and chromosomal abnormalities Mean age abnormal Doppler 28.9+/-3.6 years, normal Doppler 28.4+/- 3.5 years. (Korea)(Cohort)	105	Within 16 hours of delivery	NA	NA	Umbilical artery pH<7.20	TA, pulsed, site not reported, SD>2sd above mean for ga/local population.
Bo Hyun Yoon (1993)	Low risk. INC: Singletons, known gestational age EXC: Doppler to delivery interval>7days, delivery of SGA infant, Doppler obtained during labour or therapy with ritodrine, multiple pregnancy, delivery outside institution, unavailability of follow-up, gestational age <26 weeks at delivery Mean age 29.6+/-3.2 (Korea)(Cohort, prospective)	328	Third trimester (test to delivery interval ≤7 days).	NA	NA	Apgar 1 and 5 mins<7, admission to NICU, perinatal death.	TA, pulsed, free floating portion of cord, SD>2sd above mean ga/local population.

Bo Hyun Yoon (1994)	High risk. INC: 72 consecutive patients with preeclampsia, singletons EXC: multiple pregnancy, congenital malformations Mean age normal Doppler 29.2+/-3.3; abnormal 29.1+/-4.0 (Korea) (Cohort, consecutive)	72	Within 7 days of delivery. Mean ga at delivery normal Doppler 38.1+/-2.7; abnormal 32.3+/-3.9 weeks.	NA	NA	Adverse perinatal outcome (fetal distress requiring cs, Apgar <7 at 5 mins, significant neonatal morbidity or perinatal death)	TA, pulsed, free loop of cord, PI>2sd above mean for ga/local population.
Bracero (1989)	High risk. INC: Women who had umbilical artery Doppler and quantitative placental examinations Mean age not reported (USA)(Cohort)	47	Last test before delivery used for analysis. Mean ga at delivery	NA	NA	Admission to NICU	Route not reported, continuous wave, site not reported, SD≥3.

			normal Doppler 32+/-4.0, abnormal 36+/-5.0				
Bracero (1996)	High risk. INC: Singletons, availability of mid-trimester glycosylated haemoglobin, test within 1 week of delivery. EXC: chromosomal and structural anomalies. Mean age 29.3+/-.4 (USA) (Cohort, retrospective)	207	Within 1 week of delivery	BW mean<2sd for ga	1.93%	Hypocalcaemia, hypoglycaemia, hyperbilirubinaemia, RDS.	Route not reported cw+ pulsed, free loop of cord, SD≥3.
Brar (1989)	High risk. INC: High risk pregnancies (chronic hypertension, PIH, IUGR, SLE, post dates, diabetes, decreased fetal movements) Age not reported (USA) (Cohort)	200	Third trimester. Test within 7 days of delivery.	BW<10th centile, local values.	10.5%	Apgar score at 5 mins <7.	TA,pulsed, free loop of cord, SD>3.

Brar (1989)	High risk. INC: Premature uterine contractions ≥2 in 10 mins at <36 completed weeks, cervical change and cervical dilatation<4cm. EXC: Ruptured membranes, known IUGR, medical complications of pregnancy or mature lung profile. Mean age 24.5+/-2.6 (18-32) (USA) (Cohort, consecutive)	92	Mean 32.7+/-1.8 weeks (27-36)	BW<10th centile (local/ga)	3.26%	Apgar at 1 and 5 mins <7, neonatal death.	TA, continuous wave, site not reported, SD>3.5.
Bruinse (1989)	Unselected population. INC: Singleton, unselected women chosen at random. Age not reported (Netherlands) (Cohort, prospective, random)	393	28th and 34th week.	BW or PI <10th centile (ga/local)	22.6%	NA	Route not reported, pulsed, site not reported, PI>95th centile (local).
Bruner (1993)	High risk. INC: Women with various pregnancy complications Mean age abnormal Doppler 24.8+/-6.6 (16-40); normal Doppler 26.1+/-7.0 (14-41) (USA) (Cohort)	92	16 weeks, last test before delivery used in analysis.	BW<10th cenitle (ga/local)	25.0%	NA	TA, continuous wave, characteristic waveform, SD>95th cenitle (ga/local).

Burke (1990)	High risk. INC: Singletons, suspected IUGR (AC<5th centile) Age not reported. (Ireland) (Cohort, prospective, consecutive)	166	Not reported	BW<5th centile (ga/local)	57.8%	NA	Route not reported, continuous wave, site not reported, SD mean>2sd or AREDF.
Carroll (2000)	High risk. INC: Singletons, intact membranes, oligohydramnios (AFI<5th centile) EXC: congenital anomalies Mean age normal Doppler 26.1+/-7.0 (14-41); abnormal Doppler 24.8+/-6.6 (16-40) (USA) (Cohort, retrospective)	86	Mean normal Doppler 35.0 weeks, abnormal 31.4 weeks	BW<10th centile	36.8%	Perinatal morbidity (SGA, preterm delivery, hyperbilirubinaemia, blood transfusion, cardiovascular or pulmonary complications) or admission to NICU.	TA, pulsed, free floating portion of cord, SD>95th centile (ga/local).
Chambers (1989)	High risk. INC: 145 patients in third trimester, high risk (suspected SGA, hypertensive disorder) Age not reported.	145	28-39 weeks, mean 35+/-3.2	BW<10th centile (sex/parity)	58.6%	NA	Route not reported, continuous wave, site not

	(UK) (Cohort, prospective)						reported, RI>2sd.
Chan (1996)	High risk. INC: Severe IUGR (EFW<2sd), severe PE (diastolic BP>100mmHg), proteinuria (>300mg/24hrs or 2+ on dipstix), major congenital anomalies, clinically evident placental abruption. Mean age 28.7+/-5.0 (Hong Kong) (Cohort, prospective)	71	Mean 31.9+/-5.1 weeks, mean test to delivery interval 1.2+/-1.6 weeks	BW ratio<0.75	47.9%	NEC, major perinatal morbidity or mortality	TA, method not reported, site not reported, SD>90th or 97th centile.
Chang (1993)	High risk. INC: AC<10th centile, anomaly scan performed, accurate dates, delivery>36 weeks. Mean age 28.6+/-4.9 (UK) (Cohort, prospective)	133	Last scan median ga 220 days (182-270), median interval last scan to delivery	PI, subscapular thickness, MAC/HC<2sd	14.4% (PI)	NA	TA, color+pulsed, lumen away from insertion, PI mean>1.5sd (ROC analysis)

			5 days (0-14)				
Chang (1994)	High risk. INC: AC<10th centile, anomaly scan performed, accurate dates, delivery>36 weeks. Mean age 28.6+/-4.9 (UK) (Cohort, prospective)	104	Last scan median ga 220 days (182-270), median interval last scan to delivery 5 days (0-14)	NA	NA	Adverse perinatal outcome (one or more of acidaemia at birth (pH<10th centile), fetal distress in labour, NICU admission)	TA, color+pulsed, lumen away from insertion, PI mean>1.5sd (ROC analysis)
Chanprapaph (2004)	High risk. INC: Singleton, clinical suspicion of IUGR (SFH<3cm expected height), accurate dates, Doppler within 14 days of delivery Mean age 28.24+/-6.36 (16-45) (Thailand) (Cohort)	212	30-42 weeks, test within 14 days of delivery, mean ga at	BW<10th centile (local/ga)	50.9%	NA	TA, color, free floating portion of cord, SD≥3.

			delivery 37.66+/-1.81 weeks				
Chua (1996)	Unselected population INC: Singleton, live fetus, cephalic, intact membranes, >37 weeks, admitted to labour ward. Age not reported. (Singapore) (Cohort, prospective)	1092	>37 weeks	NA	NA	Apgar at 1 or 5 mins <7, need for assisted ventilation, admission to NICU.	TA, continuous wave, site not reported, PI>1.2.
Cosmi (2005)	High risk INC: Singleton, ga established before 20 weeks, absence of maternal pathology, delivery before 32 weeks, forward umbilical diastole, normal AFI≥5.0cm, absence of pulsation in umbilical vein, forward Ductus venosus flow, last Doppler within 24 hours of delivery EXC: structural anomalies Mean maternal age abnormal Doppler 32 (27-39), normal Doppler 31 (24-37) (Italy and USA) (Cohort, prospective)	145	24-30.4 weeks. Test for analysis performed within 24 hours of delivery.	NA	NA	Neonatal death	Route not reported, color+pusled, site abdominal origin of umbilical vein, REDF.

Craigo (1996)	High risk. INC: Singleton, prenatal diagnosis of IUGR (EFW<10th centile) EXC: Fetal abnormalities Age 11-42 years (USA) (Cohort, retrospective)	59	Not reported	NA	NA	Neonatal death, NEC, BPD, IVH	Route, method and site not reported, elevated SD ratio, cut-off not reported.
De Rochambeau (1988)	High risk INC: Singleton, Ac<10th centile. Age not reported. (France) (Cohort)	117	20-42 weeks	BW<10th centile BW<3rd centile	69.2% 49.6%	Perinatal mortality	Route not reported, continuous +pulsed wave, site not reported, RI>99th centile (local).
De Rochambeau (1992)	High risk INC: Singleton, post dates, accurate dating prior to 17 weeks Age not reported. (France) (Cohort)	80	>40+3 weeks, test performed every 2 days and last before	NA	NA	Umbilical artery pH<7.20	Route not reported, continuous wave+pulsed, site not reported, RI≥0.54

			delivery used for analysis				
Degani (1990)	High risk INC: Singleton, accurate gestation, EFW<10th centile Age not reported. (Israel) (Cohort)	49	29-40 weeks	BW<10th centile (ga/parity/sex) BW mean<2sd	77.6% 59.2%	NA	Route not reported color+pw, site not reported, PI>2sd.
Degani (1992)	High risk INC: Twin pregnancies >24 weeks Age not reported. (Israel) (Cohort, consecutive)	242	>24 weeks	BW<10th centile (ga/parity/sex) Either twin	24.8%	NA	Route not reported pulsed wave, site not reported, PI>2sd.
Dempster (1989)	High risk INC: High risk singletons (suspected SGA, hypertension, APH, diabetes, preterm labour) Age not reported. (UK) (Cohort)	205	Within 7 days of delivery. Mean ga at delivery abnormal	BW<10th centile (ga/parity/height/weight/sex)	40.0%	NA	TA, continuous wave, site not reported, SD>95th centile (local).

			Doppler 36.7 (30-40), normal 38.2 (32-41)				
Divon (1988)	High risk INC: Singleton, suspected IUGR, accurate dates, intact membranes, delivery within 2 weeks of ultrasound Age not reported. (USA) (Cohort, prospective)	127	Within 2 weeks of delivery. Mean ga at delivery 39.2+/-2.4 in IUGR, 38.5+/-2.5 non-IUGR	BW<10th centile (ga/local)	35.4%	NA	Route not reported continuous wave, site not reported, SD>3.
Divon (1989)	High risk INC: Twins, third trimester, accurate gestation, intact membranes, delivery within 2 weeks of Doppler	58	Within 2 weeks of delivery. Mean ga	BW discordancy> 15%	31.0%	NA	Route not reported continuous wave, site not

	Age not reported. (USA) (Cohort, consecutive)		at delivery 37.4+/-1.2 weeks.				reported, SD discordancy>15%
Dubinsky (1997)	High risk INC: Suspected SGA EXC: two foetuses with intrapartum complications resulting in poor outcomes Age not reported. (USA) (Cohort, prospective, consecutive)	97	Third trimester	NA	NA	Poor neonatal outcome (CS for fetal distress, fetal death, IUD, IVH, cerebral infarction, admission to NICU>10 days, admission to NICU at term, preterm delivery)	Route, method and site not reported, SD>4.0 (ROC analysis)
Eronen (1993)	High risk INC: PIH (BP>140/90 on more than 2 occasions >6 hrs apart), delivery prior to 34 weeks EXC: Fetal malformations Mean age abnormal Doppler 29.2+/-4.3; normal Doppler 28.2+/-4.3	41	24-34 weeks	NA	NA	RDS, NEC, BPD	Route not reported, continuous and pulsed wave, free loops of cord, AREDF

	(Sweden) (Cohort)						(at three separate sites).
Ezra (1997)	High risk INC: Triplet and quadruplet pregnancies Age not reported (Canada) (Cohort, consecutive)	73	Within 2 weeks of delivery. Mean ga at delivery 33+/-2.8 weeks (24-37)	BW<5th centile (triplet growth curves)	8.21%	Neonatal death, stillbirth, perinatal death, admission to NICU	TA, pulsed wave, site not reported, AEDF.
Faber (1996)	High risk INC: Threatened preterm labour (>3 contractions in 30 mins, maximum cervical dilatation 3cm, no other obstetric complications, Doppler performed within 48 hrs of admission and after 2 weeks of treatment with tocolysis+/-antibioitcs Age not reported (Germany) (Cohort, prospective)	114	24-34 weeks, mean 30.1 weeks	BW<10th centile	70.2%	Apgar at 5 mins ≤7, umbilical artery pH<7.20	Route, method and site not reported, PI>90th centile.
Farine (1998)	High risk INC: Pregnant women with SLE, 45% treated	56	24-35 weeks	BW<10th centile	28.6%	Admission to NICU	Route not reported colour

	with aspirin, prednisolone, azathioprine. Median age abnormal Doppler 29 yrs (IQR 25-29); normal Doppler 31 yrs (IQR 25-29) (Canada) (Cohort, retrospective)			(ga/local)			and pulsed wave, site not reported, AREDF.
Farmakides (1988)	High risk INC: Women referred for non stress test (hypertension, diabetes, post dates, congenital anomaly, suspected IUGR) Age not reported (USA) (Cohort)	140	Third trimester	NA	NA	Admission to NICU, assisted ventilation	Route not reported, continuous wave, site not reported, SD>3.
Ferchiou-Cherif (1993)	High risk INC: Singleton, high risk (hypertension, diabetes, history of IUD, hydramnios, oligohydramnios, suspected IUGR, placenta praevia) Age not reported (Tunisia) (Cohort)	52	30-41 weeks, test for analysis within 5 weeks of delivery	BW<3rd centile	30.8%	Neonatal morbidity	Route not reported, pulsed wave, site not reported, RI no cut-off reported.
Figueras (2004)	High risk INC: Singletons, feta size<5th centile EXC: Neonates with a birth weight>10th centile, congenital and structural anomalies	108	>26 weeks. Test for analysis	NA	NA	Adverse perinatal outcome [umbilical artery pH<7.10 or	Route and method not reported, free floating loop of

	Median age 30.45 (3.25) (Spain) (Cohort, prospective)		within 3 days of delivery, median ga at delivery 31.1 (2.346)			neonatal morbidity (severe IVH, HIE, retinopathy, seizures, NEC, RDS requiring ventilation, intubation, admission to NICU), perinatal mortality]	cord. AREDF.
Figueras (2008)	High risk INC: 369 singleton fetuses identified as SGA on customised charts antenatally had umbilical artert Doppler performed EXC: multiple pregnancies, congenital anomalies, insufficient data for customised birth weight percentile Mean maternal age 30.3+/-5.3 years (Spain) (Cohort, retrospective, consecutive)	365	>30 weeks (test to delivery interval within 2 weeks)	NA	NA	Arterial cord pH<7.10, Apgar at 5 mins<7, admission to NICU, neonatal morbidity and perinatal death	Route not reported, method and site not reported, PI>95th centile

Fischer (1991)	High risk INC: Women ≥287 days, accurate gestation, singleton pregnancy EXC: Hypertension, diabetes mellitus, renal disease, multiple gestations, suspected IUGR, fetal anomaly, substance abuse Mean age 24.3 (15-40) (USA) (Cohort)	75	>41 weeks, mean test to delivery interval 2 days (all within 8 days)	BW<10th centile PI<10th centile (local/ga)	5.33% 13.9%	Abnormal perinatal outcome (one of operative delivery due to non-reassuring FHR, umbilical artery pH<7.15 and vein<7.2, 5 minute Apgar<7, meconium below the cords, admission to NICU, BW<10th centile)	TA, continuous wave, site not reported, SD≥2.40 (ROC analysis).
Fong (1999)	High risk INC: Singletons, >24 weeks, confirmed gestation, ultrasound EFW or AC <10th centile EXC: Major congenital or chromosomal anomalies Mean age 30.3+/-5.6 (Canada) (Cohort)	293	At study entry and 36 weeks, mean 32.6+/-3.7. Mean test to delivery	NA	NA	Adverse perinatal outcome – major (perinatal death, HIE, major IVH, PVLM, NEC), minor (CS for fetal distress, arterial cord	TA, pulsed wave, middle of cord, PI>95th centile

			interval 2.4+/-2.6 weeks.			pH<7.1, Apgar at 5 mins<7)	
Forouzan (1991)	High risk INC: Prolonged pregnancies (≥41 weeks), no medical or obstetric problem, normal amniotic fluid volume and NST or CST EXC: Clinical or ultrasonic evidence of IUGR, oligohydramnios (MPD<2cm), positive NST or CST, abnormal BPS. Age not reported (USA) (Cohort)	30	>41 weeks, test for analysis within 72 hrs of delivery	NA	NA	Poor outcome (abnormal fetal monitoring during labour, umbilical artery pH≤7.2, scalp pH≤7.2, intrapartum hypoxia requiring NICU admission)	TA, color, site not reported, SD≥mean +1sd
Gaziano (1988)	High risk INC: Previous abnormal ultrasound, multiple gestations, suspected IUGR, abnormal MSAFP, history of suspected anomalies, other abnormalities Age range 15-45 years (USA) (Cohort)	230	15-44 weeks, last test before delivery used for analysis	BW≤10th centile	10.4%	NA	Route not reported, pulsed wave, characteristic waveform, SD≥4/5.

Gaziano (1991)	High risk INC: Multiple pregnancies Age not reported (USA) (Cohort)	207	Abnormal Doppler mean 27.7, normal 28.6, last before delivery used for analysis	BW<1500g	23.7%	Apgar at 5 mins ≤6, stillbirths.	Route not reported, pulsed wave, midsegment of cord, SD no cut-off reported.
Gaziano (1994)	High risk INC: Multiple and singleton pregnancies, suspected SGA (EFW<10th centile). EXC: Major chromosomal and structural anomalies Age not reported (USA) (Cohort, consecutive)	90	Test to delivery interval mean 5.2 days; mean ga at delivery normal Doppler 33.3+/-2.9;abnormal	BW<10th centile	37.8%	BW<10th centile and need for admission to NICU	Route not reported, pulsed wave, midsegment of cord, SD≥mean+2sd

			32.4+/-3.7				
Giles (1988)	High risk INC: Twin pregnancies Age not reported (Australia) (Cohort)	165	28-32 weeks	BW<10th centile (either twin)	49.1%	Admission to NICU or need for ventilation either twin.	Route not reported, continuous wave, site not reported, SD cut-off not reported.
Ghosh (2008)	High risk INC: pregnancies suspected of FGR diagnosed by fetal biometry (EFW<2sd or decline of more than 1 sd USS 2 weeks apart) EXC: multiple pregnancies, congenital malformations and chromosomal abnormalites, IUFD (n=6) were excluded Age not reported (Sweden) (Cohort, prospective)	353	Mean 34.6+/-3.2 weeks (test to delivery interval mean 19.6+/-18 days)	BW<2sd	56.8%	Admission to NICU	Route not reported, method and site not reported, PI>2sd or AREDF
Goffinet (1997)	Low risk INC: Singletons, routine consultation before 28 weeks	1903		BW<10th centile BW<3rd centile	8.09% 2.68% 4.26%	Apgar at 1 and 5 mins <7, resuscitation	

	EXC: Hypertension, diabetes, previous IUD/SGA/PIH/PE, fetal biometry<10th centile, multiple pregnancy, abnormalities Mean age normal Doppler 27.9+/-5.2, abnormal mean 28.0+/-5.5 (France) (Cohort)		28-34 weeks	(local/ga) BW<2500g		required	Route, method and site not reported, RI>90th centile ga.
Gonzalez (1995)	High risk INC: Suspected IUGR (EFW<10th centile) Age range 18-37 years (Chile) (Cohort)	74	26-40 weeks, test for analysis within 24 hours of delivery	NA	NA	Umbilical vein pH<7.16 and pO2<18%	Route, method and site not reported, AREDF.
Gonzalez (2007)	High risk INC: Singleton, EFW<5th centile EXC: Chromosomal and structural anomalies Mean age adverse outcome 26.7+/-7.0, normal outcome 25.9+/-6.8 (USA) (Cohort, retrospective)	151	Last before delivery, mean ga at delivery abnormal	NA	NA	Umbilical artery pH<7, RDS, PVL, IVH, perinatal mortality, NEC. Adverse outcome (one or more of	Route, method and site not reported, AREDF.

			outcome 30.8+/-3.6, normal outcome 37.0+/-2.5			above)	
Gramellini (2001)	High risk INC: Singletons, ultrasound dating prior to 20th week, AC<2.5th centile, Doppler within 2 weeks of birth. EXC: Chromosomal and structural anomalies Age not reported (Italy) (Cohort)	53	24-35 weeks, test for analysis within 2 weeks of delivery	BW<10th centile	96.2%	Admission to NICU, Apgar at 5 mins<7, NEC, IVH, RDS, perinatal mortality	Route not reported, color, intermediate section of cord, PI>95th centile
Gudmundsson (1988)	High risk INC: Pregnancy complicated by preeclampsia Mean age 28 +/-5.5 (Sweden) (Cohort)	58	Mean ga at delivery 258+/-19 days, test to delivery interval	BW≤mean -2sd (local/ga)	29.3%	NA	Route not reported, pulsed wave, characteristic waveform, PI mean>2sd for ga.

			mean 6 days (0-19)				
Gudmundsson (1991)	High risk INC: Singleton pregnancy, EFW≥15% below expected Mean age 27 years (17-42) (Sweden) (Cohort)	139	>32 weeks, mean ga at delivery 265+/-18 days (203-291); test to delivery interval mean 6 days (0-21)	BW≤mean -2sd (local/ga)	51.8%	Apgar at 1 and 5 mins<7, umbilical artery pH≤ 7.10	TA, pulsed wave, site not reported, PI mean >2sd for ga/
Guzman (1992)	High risk INC: Women with SLE Mean age normal Doppler 27.9+/-4.4; abnormal 27.2+/-5.4	27	Not reported	BW≤10th centile (local/ga)	25.9%	Admission to NICU for >24 hours, need for positive pressure	Route not reported, continuous wave, site not

	(USA) (Cohort, retrospective)					ventilation, perinatal death.	reported, SD mean>2sd <30 weeks and >3.0 after 30 weeks
Hack (2008)	Multiple pregnancies INC: 67 women with monochorionic twin pregnancies EXC: monoamniotic twin pregnancies, pregnancies complicated by TTTS, 2 TOP due to HELLP and trisomy 21. Mean maternal age 31 years (Netherlands) (Cohort, retrospective)	134	Median 35.2 (20-39+5), test to delivery interval median 3 days (0-17)	BW<2000g	32.1%	Stillbirth, neonatal morbidity and mortality	Route, site and method not reported, PI>2sd+/-AREDF either twin.
Haddad (1988)	High risk INC: Patients with hypertension (PIH, PE or chronic) Mean age PIH 24.8+/-4.2; PE 22.2+/-4.0; chronic 30.7+/-3.1 (UK) (Cohort)	101	Once weekly until delivery, first result used for analysis	BW<10th centile (ga/sex/parity)	25.7%	NA	Route not reported, continuous wave, characteristic waveform, RI>95th centile
Hastie	High risk	56	36-39	BW≤5th	21.4%	NA	TA, continuous

(1989)	INC: Consecutive unselected twin pregnancies Age not reported (UK) (Cohort, consecutive)		weeks, test performed monthly	centile either twin (ga/local)			wave, characteristic waveform, SD>90th centile for ga.
Hastie (1990)	High risk INC: 50 pregnancies with non-reactive CTG, >28 weeks EXC: Major congenital anomalies, rhesus isoimmunisation, premature rupture of membranes Age not reported (UK) (Cohort)	35	>28 weeks, mean 37 (28-42)	NA	NA	Adverse perinatal outcome [perinatal death, SGA (BW<10th centile), obstetric intervention for fetal distress]	TA, method not reported, characteristic waveform, SD>90th centile ga
Hecher (1988)	Population risk not reported INC: not reported EXC: not reported Age not reported (Germany) (Cohort, prospective and retrospective)	188	30-41 weeks	NA	NA	Adverse perinatal outcome [SGA (BW<10th centile) and/or operative delivery for suspected fetal hypoxia, admission to	Route not reported, color and pulsed wave, site not reported, RI/PI mean+2sd (local values for third trimester)

						NICU]	
Hitschold (1988)	High risk INC: Ultrasound before 20 weeks, accurate dates, post dates pregnancies + 1-17 days post term. Age not reported (Germany) (Cohort)	130	281-297 days, test for analysis within 10 days of delivery	NA	NA	Umbilical artery pH<7.20 or <7.10	Route not reported, pulsed wave, site not reported, SD>2.3 or RI>95th centile
Hutter (1994)	High risk INC: Not reported EXC: Not reported Age not reported (Germany) (Cohort)	559	Within 6 days of delivery; 41% delivered before 37 weeks, 49% before 33 weeks.	BW<10th centile (ga/local)	17.9%	NA	Route not reported, continuous wave, site not reported, RI>90th centile.
Joern (1997)	High risk INC: 130 multiple pregnancies (122 twins, 8 triplets)	261	Mean 34 weeks (26-42).	BW<10th centile (local/ga)	31.0%	Adverse perinatal outcome (umbilical artery	Route not reported, color and pulsed

	EXC: Chromosomal and structural anomalies Mean age 29 yrs (19-40) (Germany) (Cohort)		Mean test to delivery interval 19 days.	Any fetus		pH<7.20, 5 min Apgar<8 or transfer to NICU)	wave, site not reported, SD>3.
Joern (2000)	High risk INC: Twin pregnancies (68 monochorionic, 128 dichorionic) Median age 30 (20-41) (Germany) (Cohort, prospective)	412	Third trimester, median 35 weeks. Median test to delivery interval 9 days (0-15)	BW<10th centile (local/ga) Either twin BW<1500g	12.6%	Umbilical artery pH≤7.15, admission to NICU, Apgar at 5 mins<8, perinatal mortality.	TA, method and site not reported, PI>95th centile.
Jorn (1993)	High risk INC: Post-term, accurate gestation Age not reported (Germany) (Cohort)	165	>40 weeks	BW<10th centile (local/ga)	7.23%	NA	TA, pulsed wave and color, site not reported, SD>3
Jorn (1994)	Unselected Age not reported	120	Third trimester	BW<10th centile	21.7%	NA	Doppler no details, no

	(Germany) (Cohort)						threshold reported
Karsdorp (1994)	High risk INC: Singletons, confirmed ga, hypertension or suspected IUGR (EFW<5th centile) EXC: Chromosomal or structural anomalies Mean maternal age normal Doppler 30.6 (95% CI 29.8-31.4); abnormal Doppler 30.1 (29.2-31.0) (Netherlands, UK, Germany, Italy) (Cohort, prospective, consecutive)	459	Mean ga normal Doppler 29.8 (95% CI 29.1-30.4), abnormal Doppler 30.1 (29.2-31.0)	NA	NA	Perinatal mortality	TA, method not reported, free loop of cord, AREDF.
Kay (1991)	High risk INC: Singletons, clinically suspected IUGR, accurate ga Age not reported (USA) (Cohort, retrospective)	48	Nearest to delivery, delivery range 26-41 weeks	BW<10th centile (local/ga)	37.5%	NA	Route not reported, pulsed wave, free loop of cord, SD (average of standard values)

Kofinas (1990)	High risk INC: 36 patients with chronic hypertension, 7 chronic and PE, 25 PE Age not reported (USA) (Cohort)	68	Late second or third trimester, nearest to delivery used for analysis; mean ga at delivery abnormal Doppler 33.5+/-1.03, normal 38.2+/-0.24	BW≤10th centile	30.9%	Apgar at 1 and 5 mins <7, admission to NICU	TA, continuous wave, characteristic waveform, SD>95th centile (local values)
Lakhkar (2006)	High risk INC: 58 singleton pregnancies>30 weeks with severe PE (standard criteria) and or suspected IUGR (EFW<10th centile)	58	Within 10 days of delivery	NA	NA	Major adverse perinatal outcome (perinatal deaths,	TA, pulsed wave, free loop of cord, SD/RI/PI>2sd

	EXC: Multiple gestations and congenital anomalies Mean age 27.3 years (India) (Cohort, prospective)					HIE, IVH, PVL, pulmonary haemorrhage, NEC). Minor adverse perinatal outcome (cs for fetal distress, Apgar at 5 mins<7, admission to NICU)	
Le Thi Huong (2006)	High risk INC: Women diagnosed with SLE and/or APS Age not reported (France) (Cohort, prospective, consecutive)	100	Second trimester	NA	NA	Fetal or neonatal death; adverse perinatal outcome (fetal or neonatal death, PE, eclampsia or HELLP, premature birth, IUGR)	Route and method not reported, placental end of cord, AREDF.
Lombardi (1989)	High risk INC: Women with subjective oligohydramnios,	22	Normal Doppler	BW<10th centile	45.4%	NA	Route not reported,

	intact membranes Mean age normal Doppler 25.6+/-5.2; abnormal 26.8+/-5.7 (USA) (Cohort, prospective)		35.9+/-2.4; abnormal 34.0+/-2.2; last before delivery used for analysis	(ga/local)			continuous wave, characteristic waveform, SD>95th centile.
Lowery (1990)	High risk INC: Singletons, women with risk factors for IUGR (smoking, short stature, low prepregnancy weight, previous low birth weight) EXC: Multiple pregnancies, fetal anomalies, failure to complete both studies Age not reported (USA) (Cohort, prospective)	271	24-40 weeks (subgroup analysis based on test to delivery interval of 2 weeks)	BW<10th centile (local/ga)	9.23%	NA	TA, continuous wave, characteristic waveform, SD>2sd
Maria Fadda (2001)	High risk INC: Pregnancies complicated by IDDM, normotensive EXC: PIH	67	From second trimester; last	BW<10th centile (local/ga)	44.8%	RDS, admission to NICU for >2 days	TA, color and pulsed wave, site not reported,

	Age (Italy) (Cohort, prospective)		before delivery used for analysis; mean ga at delivery normal Doppler 38+/-1.9; abnormal Doppler 36+/-1.2				PI>95th centile
Maulik (1990)	High risk INC: Singletons, women at high risk for adverse outcome Age not reported (USA) (Cohort, prospective, consecutive)	350	34-36 weeks	BW<10th centile	12.3%	Adverse perinatal outcome (one or more of BW,10th centile, Apgar at 5 mins<7, umbilical artery pH at birth <7.20, thick meconium, fetal distress in labour,	Route not reported, continuous wave, site not reported, SD>3.

						neonatal complications needed admission to NICU)	
Maunu (2006)	High risk. INC: Preterm birth (<37 weeks), VLBW<1500g. Mean age 30,2 (+/-5.2) years (Finland) (Cohort, prospective)	67	Mean 28+2 weeks (24-36) (last Doppler performed within 7 days of delivery)	NA	NA	MRI at term abnormal (e.g. IVH, ventriculomegaly, ischaemic lesions)	Route, method, site not reported, PI>95th centile
McCowan (1992)	High risk INC: Singleton, 29 women admitted to the antenatal ward with suspected SGA (AC<5th centile) and at birth BW≤2sd below the mean, accurate gestation Age not reported (New Zealand) (Cohort)	29	Within 7 days of delivery	NA	NA	Adverse perinatal outcome [fetal distress and or acidosis at birth (umbilical artery pH<7.15 and base deficit>7) or	TA, pulsed wave, mid section of cord, PI mean +2sd

						perinatal death]	
McCowan (1992)	High risk INC: 29 women with hypertension Mean age 28 years (23-42) (New Zealand) (Cohort)	29	24 weeks or less	BW mean -2sd	31.0%	NA	Route not reported, pulsed wave, mid section of cord, PI mean +2sd
McCowan (2000)	High risk INC: Singletons, suspected SGA (AC<10th centile), pregnancies taking part in one of two RCTs (aspirin study and fetal surveillance study for women with abnormal umbilical artery Dopplers), women included for anlaysis were those that gave birth to an SGA baby (BW<10th centile) EXC: Fetal chromosomal or structural anomalies Mean age abnormal Doppler 28.2+/-6.0; normal 25.6+/-5.4 (New Zealand) (Cohort, prospective)	186	24-36 weeks	PI<10th centile	43.7%	Admission to NICU>48 hrs, perinatal death, acidosis (cord srterial pH<7.15 and BE>8mmol/l)	TA, method not reported, mid section of cord, RI>95th centile
Miller (1991)	High risk INC: Singletons, 136 women at high risk for fetal	136	SGA 254.7+/-	BW<10th centile	33.8%	NA	Route not reported,

	growth abnormalities, delivery within 3 weeks of ultrasound, accurate gestation, intact membranes EXC: Fetal chromosomal and structural anomalies Age not reported (USA) (Cohort, prospective)		23.2 days; non-SGA 267.4+/-14.0. Test to delivery interval SGA 6.5+/-6.3 days; non SGA 7.0+/-6.0 days.	(local/ga)			pulsed wave, free floating portion of cord, SD≥3.0
Miyashita (2002)	High risk INC: Singletons, 119 fetuses suspected IUGR (EFW<1,5sd) EXC: Structural and chromosomal anomalies Age not reported (Japan) (Cohort, prospective)	119	24-36 weeks, test for analysis within 10 days of delivery	NA	NA	Adverse perinatal outcome (neoanatal death, infantile death, cerebral palsy and or developmental retardation)	TA, pulsed wave, site not reported, RI≥1.0
Moon	Low risk	96	>30	BW<10th	25.0%	NA	Route not

(1999)	INC: Singletons, >30 weeks, accurate gestation EXC: Multiple gestations Mean age 27 (14-42) (USA) (Cohort, retrospective)		weeks	centile (local/ga)			reported, pulsed wave, free floating portion of cord, SD>3.
Mulders (1987)	Mixed risk INC: Singletons, 30 patients admitted with suspected IUGR or hypertension, 18 women with uncomplicated pregnancy, all accurate dates Age not reported (Netherlands) (Cohort)	48	Mean 34+/- 2 weeks	BW<10th centile (local/ga)	31.3%	NA	TA, continuous wave, site not reported, PI≥1.1 (ROC analysis)
Mulders (1989)	Mixed risk INC: Singletons, 99 gravid women (30 admitted due to complications) Age not reported (Netherlands) (Cohort)	99	Third trimester Last before delivery (1-28 days)	BW<10th centile (local/ga)	34.3%	Apgar at 1 and 5 mins<7, umbilical artery pH≤7.15	TA, continuous wave, site not reported, PI≥mean +1.64 sd (95th centile)
Newnham (1990)	Medium risk INC: Singleton, pregnant women attending antenatal clinic before 18 weeks	516	18,24,28 and 34 weeks	BW<10th centile (local/ga/hei	9.88%	Hypoxia (operative delivery for CTG	TA, color and pulsed wave, site not

	EXC: Chromosomal and structural anomalies Mean age 26.0 +/-5.1 (Australia) (Cohort, prospective)			ght/parity/se x)		abnormalities, uterine artery pH<7.20, 5 min Apgar<7)	reported, SD>95th centile
Niknafs (2001)	Low risk INC: 219 women seen in routine antenatal clinic, accurate dates Mean age normal Doppler 25.8+/-7.0; abnormal 25.9+/-6.4 (Australia) (Cohort)	219	Not reported	PI≤20th centile PI≤10th centile	20.1% 9.13%	NA	Route, method and site not reported, SD mean>3sd
Nordstrom (1989)	High risk INC: Singletons, 69 women at high risk (suspected IUGR, PIH,PE, APH, abdominal pain, IDDM or GDM, polyhydramnios, placenta praevia, threatened preterm labour, unstable lie, UTI, adenexal mass, maternal collagen disease), confirmed gestation Age not reported (UK) (Cohort)	69	28-42 weeks, test for analysis last before delivery, median interval 2 days	NA	NA	Severe and moderate compromise (BW<10th centile, fetal distress during labour, Apgar at 5 mins<7)	TA, pulsed wave, site not reported, SD>95th centile
Odendaal	High risk		Not	BW<10th	26.3%	Intrauterine death	Route not

(2008)	INC: Pregnant women with suspected poor fetal growth Mean age 29 years (13-46) (South Africa) (Cohort, retrospective, consecutive)		reported	centile (local/ga)			reported, continuous wave, site not reported, RI>95th centile/AREDF
Ogunyemi (1992)	High risk INC: Patients in labour with a presumptive diagnosis of fetal distress based on abnormal fetal heart rate patterns, accurate gestation. Mean age normal Doppler 25.5+/-0.72; abnormal Doppler 25.6+/-1.4 (USA) (Cohort, prospective, consecutive)	102	In labour, mean ga normal Doppler 39.7+/-0.24;abnormal 37.9+/-0.76; test for analysis within 10 hrs of delivery (median 1 hr)	BW<10th centile (local/ga)	5.88%	Poor perinatal outcome [SGA, Apgar <7 at 1 min, umbilical artery pH<7.12, presence of meconium below the cords, neonatal hospital stay>3 days, NICU admission, neonatal morbidity (RDS, hypoglycaemia, sepsis)]	TA, continuous wave, characteristic waveform, SD≥3 after 30 weeks

Ott (1990)	High risk INC: Fetuses with suspected IUGR, completed studies, delivery within 21 days of last examination (11 twin and 93 singleton) Age not reported (USA) (Cohort)	104	Within 21 days of delivery	BW<10th centile (local/ga)	27.9%	NA	Route not reported, pulsed wave, free floating portion of cord, SD≥3.
Ott (2000)	High risk INC: Singletons, delivered within 2 weeks of last examination, indications for ultrasound (wellbeing, PIH, preterm labour, IUGR, oligohydramnios, chronic hypertension, IDDM, bleeding, collagen vascular disease, polyhydramnios) Age not reported (USA) (Cohort, prospective)	578	Within 2 weeks of delivery	NA	NA	Neonatal morbidity (IVH, HIE, retinopathy, seizures, NEC, sepsis, preterm delivery)	Route and method not reported, free floating portion of cord, SD>90th centile
Owen (1999)	Low risk INC: Singletons, 313 women attending antenatal clinic, gestational age<85days, no risk factors for accelerated or restricted growth Age not reported (UK) (Cohort, prospective)	257	Last and penultimate but one examination (4 weeks separation	Skinfold thickness <10th centile, PI<25th centile MAC:OFC<-1sd	13.4% 15.2% 7.2%	NA	Route and method not reported, free floating portion of cord, PI (ROC analysis)

			)				
Ozcan (1998)	High risk INC: Singletons, EFW<5th centile, Doppler within 2 weeks of delivery EXC: Chromosomal and structural anomalies, BW<5th centile Age (USA) (Cohort)	19	Mean 28.2 (27-31.4); test for analysis within 2 weeks of delivery median 2 days (0-14)	NA	NA	Perinatal death (fetal demise or neonatal death within first 30 days), Apgar at 5 mins<7, NEC.	TA, color and pulsed, insertion at fetal abdomen, AREDF.
Ozden (1998)	Unselected INC: Singletons, 99 randomly selected term pregnant women in labour Age not reported (Turkey) (RCT, prospective, random)	99	In labour	NA	NA	Apgar at 1 and 5 mins<7, umbilical artery pH<7.10, neonatal death	Route not reported, pulsed wave, free floating portion of cord, SD>3 and or AREDF.
Pattinson (1989)	High risk INC: Patients with severe proteinuric	46	Within 3 days of	BW<10th centile	36.9%	Apgar at 1 and 5 mins<6, NEC,	Route not reported,

	hypertension (BP 160/110 mmHg and ≥2+ protein on dipstix) Median age normal Doppler 24 (17-41); abnormal Doppler 25 (20-41) (South Africa) (Cohort)		delivery	(local/ga)		RDS, perinatal mortality	pulsed wave, site not reported, SD≥6.
Pattinson (1993)	Low risk INC: Umbilical artery Doppler performed in women presenting for routine dating ultrasound Age not reported (South Africa) (Cohort)	481	16-24 weeks	BW<10th centile	8.11%	Perinatal death	TA, continuous wave, characteristic waveform, AEDF
Pattinson (1993)	High risk INC: High risk pregnancies at risk of placental insufficiency Age not reported (South Africa) (Cohort)	348	>24 weeks	BW<10th centile (local/ga)	55.5%	Perinatal mortality (death of a baby with a birth weight>500g or >24 weeks gestation occurring within 28 days of delivery)	Route not reported, continuous wave, site not reported, RI>95th centile
Poulain	High risk	541	28-34	BW<10th	20.1%	Apgar at 1 and 5	Route, method

(1994)	INC: High risk pregnancy requiring ultrasound (IUGR, hypertension, history of obstetric hypertension, previous IUD or previous IUGR, beta-mimetic treatment or maternal diabetes) Age not reported (France) (Cohort, prospective)		weeks, first Doppler performed used for analysis mean test to delivery interval 6.6+/-3.2 days	centile (local/ga) BW<3rd centile	10.1%	mins<7, IUD.	and site not reported, AEDF or RI>90th centile.
Puzey (1992)	High risk INC: Patients in labour, no evidence of fetal distress before labour, >35 weeks, umbilical artery Doppler performed if decelerations on CTG. Age not reported (South Africa) (Cohort)	42	Doppler performed in labour within 30 mins of umbilical cord vein sample	NA	NA	Umbilical vein pH<7.2	Route not reported, continuous wave, site not reported, PI>2sd above mean (>1.15)
Rocca	High risk	113	Third	NA	NA	Apgar at 5	TA, pulsed

(1995)	INC: Singleton, 113 patients with PE Mean age 26.3+/-2.3 (Egypt) (Cohort)		trimester, weekly from 28 weeks until delivery, last test performed used for analysis			mins<7, perinatal death	wave, site not reported, SD≥3
Rochelson (1987)	High risk INC: All women who delivered an SGA (BW<10th centile) infant and had antenatal Doppler, accurate gestational age Age not reported (USA) (Cohort)	54	Third trimester	NA	NA	Admission to NICU, positive pressure ventilation, perinatal mortality	TA, continuous wave, site not reported, SD>3 in third trimester, <30 weeks SD mean>2sd
Rochelson (1992)	High risk INC: 40 women with an ultrasound diagnosis of IUGR (poor growth by serial BPD/AC or EFW<10th centile) Age not reported	40	27-42 weeks; test for analysis performe	BW<10th centile (local/ga)	65.0%	NA	TA, continuous wave, characteristic waveform, SD≥3

	(USA) (Cohort)		d within 3 weeks of delivery				
Rognerud Jensen (1991)	High risk INC: Confirmed ga, umbilical artery Doppler within 7 days of birth. High risk conditions include PE/hypertension, diabetes mellitus, poor obstetric history, suspected IUGR, abnormal fetal heart rate, decreased fetal movements, imminent preterm delivery) Age not reported (Norway) (Cohort)	94	30-42 weeks, test for analysis within 7 days of delivery, median ga at delivery normal Doppler 38 (30-42), abnormal 35 (30-40)	BW<5th centile (local/ga/sex)	29.8%	Apgar at 1 and 5 mins ≤7, admission to NICU, perinatal mortality.	TA, pulsed wave, site not reported, SD>3 after 30 weeks.
Rognerud Jensen	High risk INC: 50 women with twin pregnancies	100	28 weeks onwards;	BW<5th centile	25.0%	Apgar at 1 min and 5 mins<7	Route not reported,

(1992)	Mean age 29.5 (18-39) (Norway) (Cohort, consecutive)		test for analysis performed within 7 days of delivery; mean ga at delivery 37 weeks (30-40)	(singleton values)			pulsed wave, site not reported, RI≥80%
Rudigoz (1991)	High risk INC: 26 pregnancies with maternal hypertension (BP>130/90) Age not reported (France) (Cohort)	28	Not reported	BW<10th centile (ga/local)	46.4%	NA	Route, method, site not reported, RI no threshold
Sarno (1989)	Unselected INC: Singletons, ≥36 weeks, vertex presentation and latent phase of labour Mean age 26.0+/-5.3 (USA) (Cohort)	109	≥36 completed week; mean 40.2+/-2.0	BW<10th centile	1.83%	Perinatal mortality, Apgar at 1 and 5 mins<7	TA, continuous wave, characteristic waveform, SD>3.

Schulman (1989)	Low risk INC: 255 women with routine prenatal care, monthly Doppler of uterine and umbilical arteries Mean age 28.4 years (USA) (Cohort)	255	>20 weeks	BW<15th centile (local/ga)	9.02%	NA	Route not reported, continuous wave, characteristic waveform, SD>3 after 30 weeks
Sezik (2004)	High risk INC: Singleton, 270 pre-eclamptic women EXC: Multiple gestations, glucose intolerance, preexisiting diabetes, major congenital malformations Mean age abnormal Doppler 29.8+/-1.2; normal Doppler 28.1+/-0.4 (Turkey) (Cohort)	270	Abnormal Doppler 31.5+/-0.4; normal Doppler 35.0+/-0.2	BW<10th centile BW<2500g BW<1500g	44.1% 24.8% 24.8%	Apgar at 1 and 5 mins <4, RDS, ICH, seizures, neonatal mortality	TA, continuous wave, free floating portion of cord, AREDF
Sijmons (1989)	Unselected INC: Singletons, random selection of hospital populations, confirmed ga Age not reported (Netherlands) (Cohort, prospective, consecutive)	394	28weeks+/-6 days and 34 weeks+/-6 days	BW<2.3rd centile BW<10th centile PI<3rd centile	3.3% 22.6% 4.26% 10.2%	NA	TA, pulsed wave, characteristic waveform, PI>95th centile

				PI<10th centile (ga/local)			
Skodler (1989)	High risk INC: Not reported EXC: Not reported Age not reported (Germany) (Cohort)	163	Third trimester	BW<10th centile (local/ga)	12.9%	NA	No details
Soothill (1993)	High risk INC: Singleton pregnancies booked for antenatal care and referred for fetal surveillance (suspected SGA, hypertension, post dates, reduced fetal movements, APH, pain) seen within one week of delivery, no clinical suspicion of ruptured membranes, delivery>32 weeks EXC: Fetal chromosomal and structural anomalies Mean age not reported (UK) (Cohort, prospective, consecutive)	191	Within one week of delivery	NA	NA	Neonatal morbidity (one or more of fetal heart rate abnormalities in labour leading to cs, 5 min Apgar<7, umbilical venous pH<7.15, admission to NICU)	Method not reported, continuous wave, site not reported, PI>97.5th centile
Soregaroli	High risk	578	Mean	NA	NA	Apgar at 5	Route not

(2002)	INC: Singletons with a diagnosis of IUGR (AC<2sd), first ultrasound within 20 weeks and Doppler performed Age not reported (Italy) (Cohort, retrospective)		31+/-4 (19-38); test for analysis performed within 48 hrs of delivery			mins<7, admission to NICU, RDS, IVH, NEC, ROP, IUD, neonatal death, perinatal mortality (IUD and deaths up to 28 days)	reported, color, site not reported, PI>2sd or AREDF.
Spinillo (2005)	High risk INC: All pregnant women delivered at department of singleton fetus between 24 and 35 weeks with a umbilical artery Doppler prior to delivery EXC: Congenital malformations and uncertain dates Mean age normal growth 30.5+/-5.4; abnormal growth 30.2+/-4.3 (Italy) (Cohort, retrospective)	316	24-25 weeks; test for analysis performed within 7days of delivery	NA	NA	Neonatal death, ICH.	Route not reported, pulsed wave and color, site not reported, SD≥95th centile (local/ga)
Strigini (1997)	High risk INC: Singletons, suspected FGR, poor obstetric history, preterm labour, hypertension, diabetes,	576	25-41 weeks; mean	BW<10th centile	17.9%	Adverse perinatal outcome (fetal death or death	TA, pulsed wave and color, free loop

	reduced fetal movements, APH) EXC: Multiple pregnancies, chromosomal and structural anomalies Mean age not reported (Italy) (Cross-sectional, prospective)		35.1 weeks; test for analysis performed within 3 weeks of delivery			before discharge, 5 min Apgar <7, CTG abnormality leading to emergency CS)	of cord, SD>2sd
Szalay (1991)	Unselected INC: Singleton pregnancies EXC: Fetal chromosomal or structural anomalies Age not reported (Germany) (Cohort)	810	Third trimester	BW<10th centile (local/ga)	16.4%	NA	Route not reported, pulsed wave, site not reported, PI>95th centile
Tchirikov (2009)	Mixed risk INC: 181 patients with singleton pregnancies, no fetal malformations Age not reported (Germany) (Cohort, prospective)	181	17-41 weeks (test to delivery interval compromised group mean	NA	NA	Adverse perinatal outcome (arterial cord pH, Apgar at 1 minute, birth weight, duration of gestation, need for respiratory support,	TA, method and site not reported, PI (cut off not reported)

			18.5 (9.5-27.5) days, normal group 79.91 (72.2-87.7 days))			admission to NICU)	
Thiebaugeorges (2006)	High risk INC: Singleton. Population based EPIPAGE cohort study, all births between 22-32 weeks in 9 regions of France; sub-group born 24-32 weeks after a high risk pregnancy defined by antenatally suspected SGA (AC or FL<10th centile) or maternal hypertension. EXC: Multiple pregnancies, congenital malformations Mean age 30.3+/-5.5 (France) (Cohort)	518	24-32 weeks; mean 30.0+/-1.7; test to delivery interval abnormal 2.3 days (3.4), normal 6.8 days (11.1)	NA	NA	Perinatal mortlaity	No details

To (2005)	High risk INC: Singletons, accurate gestation, suspected IUGR (decreased SFH, decreased LV, maternal smoking, PIH or other antenatal disorders) EXC: Fetal chromosomal or structural anomalies Mean age normal Doppler 27.7+/-5.5; abnormal Doppler 27.75+/-3.9 (Hong Kong) (Cohort, prospective, consecutive)	187	>34 weeks; test for analysis performe d within 2 weeks of delivery	NA	NA	Apgar at 1 min≤4, at 5 mins ≤7	Route and method not reported, free loop of cord
Todros (1995)	Low risk INC: Singletons, no pre-pregnancy pathologic condition, no obstetrical risk EXC: Fetal chromosomal or structural anomalies Age: 57%<30 yrs, 38% 30-39 yrs, 5% >40 yrs (Italy) (Cohort)	916	19-24 and 26-31 weeks	BW<10th centile (local/ga)	4.6%	Abnormal perinatal outcome (perinatal death or admission to NICU)	TA, continuous and pulsed wave, site not reported, SD (ROC analysis)
Todros (1996)	High risk INC: Singletons with an ultrasound diagnosis of SGA (AC<10th or weekly increase of AC<5mm) or PIH. EXC: Fetal chromosomal or structural anomalies Age not reported (Italy) (Cohort, prospective)	265	At diagnosis, then every 2-3 weeks	NA	NA	Outcome 1 (IUD or early neonatal death) Outcome 2 (death or Apgar <7 at 5 mins or need for	Route not reported, continuous wave, free floating portion of cord, PI>50th, 60th,

						admission to NICU) Outcome 3 (As outcome 2 or BW<10th centile) Outcome 4 (BW<10th centile and any of outcome 2)	70th, 80th, 90th, 95th centile, AREDF
Torres (1995)	High risk INC: Singletons, hypertension (40 PE, 16 chronic with superimposed PE) (Spain) (Cohort, prospective)	172	Fortnightly from 28 weeks; last before delivery used for analysis	BW<10th centile (local/ga)	27.9%	Fetal death	Route not reported, pulsed wave, site not reported, RI>2sd or AEDF
Trudinger (1991)	High risk INC: Singleton, delivered after 26 weeks, all with obstetric risk factors indicating increased risk of fetal compromise Mean age 27.4+/-5.2 (Australia) (Cohort, prospective)	2178	12.1-43.6 weeks; test to delivery interval mean	BW<10th centile BW<5th centile	27.0% 17.9%	Apgar at 1 and 5 mins≤6, admission to NICU	Route not reported, continuous wave, site not reported, SD>95th centile

			12.5 days+/- 16.5				
Tyrrell (1989)	Mixed risk INC: All women admitted to antenatal ward before elective cs with umbilical artery Doppler performed within 4 hours of delivery Age not reported (UK) (Cohort, prospective)	116	27-42 weeks; test performed within 4 hours of delivery	NA	NA	Hypoxia (umbilical cord artery pO2<2.5th centile) Acidosis (umbilical artery pH<7.25)	Route not reported, continuous wave, site not reported, SD>4.5
Van Asselt (1998)	High risk INC: Singletons, referred for Doppler because of PE (Sweden) (Cohort)	108	Last test before delivery; median interval 3.5 days (0-25); ga at delivery severe PE median 37 (27-	BW mean – 2sd	16.7%	Admission to NICU	Route not reported, pulsed wave, free floating protion of cord, PI>2sd

			42), mild 39 (27-42)				
Vergani (2003)	High risk INC: Antenatally suspected FGR (AC<10th centile), accurate dating, ga>34 weeks at delivery EXC: Fetal chromosomal or structural anomalies, Doppler more than 2 weeks before delivery Mean age adverse outcome 32.5+/-4.5; good outcome 31.9+/04.8 (Italy) (Cohort, prospective, consecutive)	447	Within 2 weeks of delivery	NA	NA	Admission to NICU for reason other than low birth weight alone	TA, pulsed wave, mid section of cord, PI>95th centile
Vintzileos (1991)	Population risk not reported INC: Singletons, consecutive, 25-37 weeks, accurate dates. All delivered by cs prior to onset of labour, umbilical artery Doppler and BPS performed within 3 hrs of delivery EXC: Congenital anomalies, medication during testing Mean age 31.9+/-3.1 weeks (USA) (Cohort, prospective, consecutive)	62	Within 3 hrs of delivery	NA	NA	Umbilical artery pH<7.20	Route not reported, continuous wave, site not reported, SD>3 or AEDF
Weiner	High risk	139	Test	NA	NA	Abnormal	TA, pulsed

(1993)	INC: 142 post term gravid women, all>287 days EXC: Women with complicated pregnancies. Mean age 27.3+/-5.6 (16-39) (Israel) (Cohort, prospective)		performed every 3 days until delivery; mean ga at delivery 41.8+/-0.64 (41-43)			outcome (5 min Apgar score<7, admission to NICU, cs because of felal distress, BW<5th centile)	wave, site not reported, RI>95th centile
Weiner (1996)	High risk INC: Singletons, 98 pregnant women with suspected SGA according to EFW and BW<10th centile EXC: Fetal chromsosomal and structural anomalies Mean age 29.7+/-5.2 (USA) (Cohort)	81	Mean 32.7+/05.1 (26-38.5)	NA	NA	Apgar at 5 mins<7, admission to NICU, perinatal death	TA, continuous wave, site not reported, RI (no threshold)
Wong (2003)	High risk INC: Women with pre-existing diabetes (types 1 and 2) EXC: Gestational diabetes, fetal chromosomal or	104	28, 32, 36 and 38 weeks	BW<10th centile (local/ga)	8.65%	Adverse outcome (one or more of SGA, cs for abnormal CTG,	Route, method and site not reported, PI>95th

	structural anomalies Age not reported (Australia) (Cohort, retrospective)					arterial cord pH<7.2, 1 min Apgar<3, 5 min Apgar<7, HIE, stillbirth or perinatal death)	
Worrell (1991)	High risk INC: Singletons, ultrasound predicted weight<10th centile or a referral for a patient clinically at risk of IUGR, accurate dates Age not reported (USA) (Cohort)	43	Third trimester	BW<10th centile and BW<2500g (local/ga)	20.9%	NA	Route and method not reported, free loop of cord, RI≥0.67
Yildirim (2008)	High risk INC: 310 singleton pregnancies, EFW<10th centile suspected on ultrasound EXC: multiple pregnancies, chromosomal and structural anomalies Abnormal Doppler mean maternal age 28.1 (27.2-29.1) years, normal Doppler 28.4 (27.6-29.3) years (Turkey) (Cohort, retrospective)	310	Normal Doppler test to delivery interval 14.1 (11.2-17.06) days, abnormal	NA	NA	Apgar at 5 mins<7, neonatal mortality and morbidity, perinatal mortality, admission to NICU, intubation, NEC.	TA, method not reported, free loop of cord, AREDF

			Doppler 7.5 (5.8-9.2) days				
Zhou (1991)	Population risk not reported INC: Not reported EXC: Not reported Age not reported (China) (Cohort)	123	37-42 weeks	BW<2500g	4.01%	Apgar at 5 mins ≤7, admission to NICU, adverse perinatal outcome	Route not reported, pulsed wave, site not reported, SD≥3
Zimmerman (1995)	High risk INC: At least 287 days ga, confirmed gestation EXC: Maternal disease, premature rupture of membranes>24 hours, fetal malpresentation, IUGR Age not reported (Finland) (Cross-sectional, prospective)	123	Within 2 days of delivery, median ga at delivery 41.8 (41-42.9)	NA	NA	Asphyxia (Apgar at 1 min or 5 mins≤7 or umbilical artery pH≤7.15 or admission to NICU with signs of asphyxia encephalopathy)	Route not reported, pulsed wave and color, site not reported, RI≥0.62 (ROC analysis)

BP blood pressure; UTI urinary tract infection; Hrs hour; INC inclusion; EXC exclusion; PE preeclampsia; RCT randomised controlled trial; PIH pregnancy induced hypertension; HELLP haemolysis elevated liver enzymes low platelets; FGR fetal growth restriction; SGA small for gestational age; IUGR intrauterine growth restriction; APS antiphospholipid syndrome; SLE systemic lupus erythematous; BW birth weight; PI ponderal index; UK United Kingdom; USA United States of America; NA not applicable; USS ultrasound scan; ga gestational age; sd standard deviation, % percent; NICU neonatal intensive care unit; TA transabdominal; PI pulsatility index; RI resistance index; SD systolic/diastolic ratio; AREDF absent reversed end diastolic flow; AEDF absent end diastolic flow; REDF reversed end diastolic flow; AC abdominal circumference; BPD/AC biparietal diameter/abdominal circumference; EFW estimated fetal weight; CS caesarean section; FHR fetal heart rate; EDD estimated date of delivery; BE base excess; IVH intraventricular haemorrhage; NEC necrotising enterocolitis; RDS respiratory distress syndrome; BPD bronchopulmonary dysplasia; HIE hypoxic ischaemic encephalopathy; PVL (M) peri-ventricular leukomalacia; ICH intracranial haemorrhage; ROP retinopathy of prematurity; VLBW very low birth weight; CTG cardiotocogram; CST contraction stress test; NST non-stress test; mins minutes; cw continuous wave; BP blood pressure; GDM gestational diabetes; IDDM insulin dependent diabetes mellitus; UTI urinary tract infection; APH antepartum haemorrhage; IUD intrauterine death; SFH symphiseal fundal height; MAC/HC mid-arm circumference/head circumference ratio; AFI amniotic fluid index; MPD maximum pool depth; LV liquor volume; BPS biophysical profile; MSAFP maternal serum alpha feto-protein; ROC receiver operating characteristic; IQR interquartile range, TTS twin to twin transfusion syndrome

Appendix 29: Search strategy for systematic review of the accuracy of middle cerebral artery Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing.

Host: Ovid

Date of search: May 2009

Years covered by search: 1950-2009

1. **MEDLINE (inception until May 2009) -1574 citations**
    1. ("Pregnant woman"[MeSH] OR "Pregnancy"[MeSH] OR pregnan*)
    2. ("Prenatal Diagnosis[MeSH] OR "Ultrasonography/Prenatal"[MeSH} OR "Ultrasonography/Doppler"[MeSH])
    3. {(arterial Doppler.mp) OR (Doppler velocimetry.mp) OR (Doppler ultrasound.mp) OR (MCA.mp) OR (Middle cerebral artery[MeSH])}
    4. (1 AND 2)
    5. (4 AND 3)
    6. Limit 5 to animals
    7. (5 NOT 6)

2. **Medline search adapted for EMBASE (inception until May 2009) - 407 citations**

3. **Cochrane library (2009:2) – 85 reviews, 169 clinical trials, 89 DARE**
    1. Pregnant women
    2. Prenatal diagnosis

3. Ultrasonography prenatal
4. Ultrasonography Doppler
5. Arterial Doppler
6. Doppler velocimetry
7. Doppler ultrasound
8. MCA
9. Middle cerebral artery
10. (2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9)
11. (1 AND 10)

4. MEDION – 0 citations

5. Grey literature – 0 citations

Appendix 30: Data extraction form for systematic review of the accuracy of middle cerebral artery Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing.

Section A: Study Information

1)Ref ID:		**4)Publication year:**	
2)Rev name:		**5)First Author:**	
3)Country:		**6)Language:**	

Section B: Data Retrieval for Middle Cerebral Artery Doppler Study

Population

7) Healthcare Centre:

Primary care ☐$_1$ Secondary care ☐$_2$ Mixed ☐$_3$ Other ☐$_4$ Unreported ☐$_5$

8) Setting:

In-patient ☐$_1$ Out-patient ☐$_2$ Mixed ☐$_3$ Unreported ☐$_4$ Other ☐$_5$

9) Number of participating centres: _____

10) Gestation at time of index test:

<20 weeks ☐$_1$ 20-24 weeks ☐$_2$ 24-28 weeks ☐$_3$ 28-34 weeks ☐$_4$ 34-37 weeks ☐$_5$ 37-40 weeks ☐$_6$ > 40 weeks ☐$_7$ Unreported ☐$_8$ Other

10.i) Mean (range)_____ Unreported ☐$_3$

10.ii) Median (range) _____ Unreported ☐$_3$

11) Pregnancy:

Low Risk ☐$_1$ High Risk ☐$_2$ Unselected ☐$_3$ Unreported ☐$_4$

11.i) State high risk conditions: Unreported ☐$_3$

12) Were patients with the following conditions excluded/not included?

12.i) Previous IUGR: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.ii) Insulin dependent diabetes mellitus: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.iii) Chronic renal disease: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.iv) Systemic lupus erythematosus: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.v) Antiphospholipid syndrome: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$ 12.vi) Chronic hypertension: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$ 12.vii) Pre-eclampsia: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.viii) Foetal chromosomal/structural anomalies: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

13) Did all patients have singleton pregnancies?:

Yes ☐$_1$ No ☐$_2$ ----------/---------- Unreported ☐$_3$

14) Were all patients primigravid?:

Yes ☐$_1$ No ☐$_2$ ----------/---------- Unreported ☐$_3$

15) List other eligibility/ in-/exclusion criteria:

Not applicable ☐$_3$

16) Study population: (describe age (mean +/- SD or median/range), ethnicity, smoking, BMI etc.)

17) Start of patient inclusion (year) :

Unreported ☐3

18) End of patient inclusion (year) :

Unreported ☐3

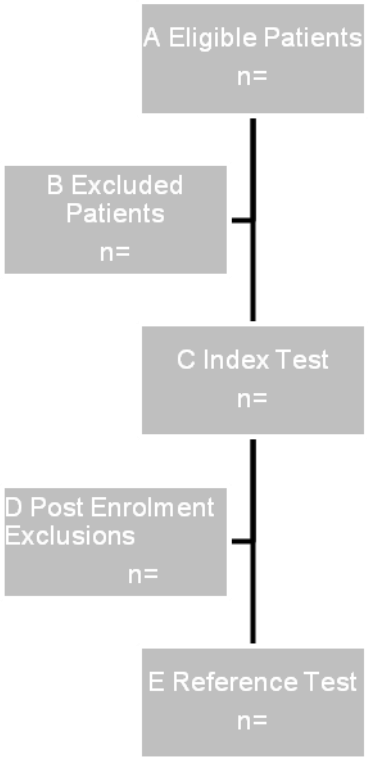
19) Study Design:

cohort ☐1 case control ☐2 RCT/CCT ☐3 cross sectional ☐4 before and after ☐5 case series ☐6 (no _____) other ☐7

19.i) Data collection: prospective ☐1 retrospective ☐2 unreported ☐3 other ☐4

19.ii) Enrolment: consecutive ☐1 arbitary (random) ☐2 unreported ☐3 other ☐4

20) Numbers:



B Reasons _____

D Reasons _____

21) Completeness of Verification:

(= E / C x 100 = %)

> 90% ☐$_1$ 81-90% ☐$_2$ < 81% ☐$_3$

Index Test

22) Description of technique:

Adequate ☐$_1$ Inadequate ☐$_2$

23) Timing of measurement (from delivery):

< 7days ☐$_1$ 7-14 days ☐$_2$ 14 -28 days ☐$_3$ > 28 days ☐$_4$ Mixture ☐$_5$

Unreported ☐$_6$

23.i) Median gestational age at delivery _____unreported ☐$_3$

24) Measurement:

SCANNING:

24.i) Operator:

Single ☐$_1$ Multiple ☐$_2$ Unreported ☐$_3$

24.ii) Operator experience_____

unreported ☐$_3$

24.iii) Scanning Route: Transabdominal ☐$_1$ Transvaginal ☐$_2$ Unreported ☐$_3$

DOPPLER:

24.iv) Method: Continuous wave Doppler ☐$_1$ Pulsed wave Doppler ☐$_2$ Colour mapping ☐$_3$ Unreported ☐$_4$

24.v) Measurement parameter: Resistance index (RI) ☐$_1$ Systolic / diastolic ratio ☐$_2$ Pulsatility index (PI) ☐$_3$ Cerebroplacental ratio ☐$_4$ Unreported ☐$_5$

24.vi) Cut-off level for waveform ratio: > 2 SD ☐$_1$ > 95th centile ☐$_2$ > 90th centile ☐$_3$

> 80th centile ☐$_4$ > 50th centile ☐$_5$ < 10th centile ☐$_6$ < 5th centile ☐$_7$

Unreported/NA ☐$_8$

Other/Threshold data set:

_____

24.vii) Machine: _____

unreported ☐$_3$

24.viii) Probe: _____

unreported ☐$_3$

24.ix) High pass filter: _____ unreported ☐$_3$

24.x) Pulse repetition frequency: _____ unreported ☐3

24.xi) Size of sampling gate: _____unreported ☐3

24.xii) Site : _____ unreported ☐3

24.xiii) Angel of insonation: _____

unreported ☐3

24.xiv) Number of consecutive waveforms: _____unreported ☐3

24.xvi) Other information:

Maximal output power_____

Sample volume _____

Spatial peak temporal intensity _____

Reference Standard / Outcome

25) Measured blind form diagnostic test: Yes ☐1 No ☐2 Unclear ☐3

26) Measurement for FGR: Birthweight ☐1 Neonatal ponderal index ☐2 Skin fold thickness ☐3 MAC / OFC ☐4 Other ☐5 _____

27) Threshold: < 3rd centile ☐1 < 5th centile ☐2 < 10th centile ☐3 < 25th centile ☐4

> 2SD ☐5 Other ☐6 _____ Unclear ☐7

28) What data set was used to define threshold?

_____

unreported ☐3

29)**Timing of measurement:** At delivery ☐1 Within 24 hrs ☐2 > 24 hrs ☐3

Mixture ☐4 Unreported ☐5

30) Marker of wellbeing e.g. Apgar score, perinatal mortality

31) Threshold and data set (if applicable):

32) Measured blind form diagnostic test: Yes ☐1 No ☐2 Unclear ☐3

Results

	Reference Test: Threshold:			
Index test, Measurement:		Positive	Negative	**Total**
	Positive	**TP**	**FP**	
Threshold:	Negative	**FN**	**TN**	
	Total			

Appendix 31: Reference list of included studies in systematic review of accuracy of middle cerebral artery Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing.

Alatas C, Aksoy E, Akarsu C, Yakin K, Bahceci M. Prediction of perinatal outcome by middle cerebral artery Doppler velocimetry. Archives of Gynecology & Obstetrics 1996; 258(3):141-146.

Arduini D, Rizzo G. Prediction of fetal outcome in small for gestational age fetuses: comparison of Doppler measurements obtained from different fetal vessels. Journal of Perinatal Medicine 1992; 20(1):29-38.

Arias F. Accuracy of the middle-cerebral-to-umbilical-artery resistance index ratio in the prediction of neonatal outcome in patients at high risk for fetal and neonatal complications.[see comment]. American Journal of Obstetrics & Gynecology 1994; 171(6):1541-1545.

Cavero A, Merce L, De M, Jr., Mora R, Sancho I, Ceballos C. Predictive capacity of Doppler velocimetry to predict the perinatal outcome in pregnancies with antenatal suspect of intrauterine growth retardation (IUGR). Progresos de Obstetricia y Ginecologia 1996; . 39(6).

Chandran R, Serra-Serra V, Sellers SM, Redman CW. Fetal cerebral Doppler in the recognition of fetal compromise. British Journal of Obstetrics & Gynaecology 1993; 100(2):139-144.

Del Rio M, Martinez JM, Figueras F, Bennasar M, Olivella A, Palacio M et al. Doppler assessment of the aortic isthmus and perinatal outcome in preterm fetuses with severe intrauterine growth restriction.[see comment]. Ultrasound in Obstetrics & Gynecology 2008; 31(1):41-47.

Dubiel M, Gudmundsson S, Gunnarsson G, Marsal K. Middle cerebral artery velocimetry as a predictor of hypoxemia in fetuses with increased resistance to blood flow in the umbilical artery. Early Human Development 47(2):177-84, 1997.

Dubiel M, Breborowicz GH, Marsal K, Gudmundsson S. Fetal adrenal and middle cerebral artery Doppler velocimetry in high-risk pregnancy. Ultrasound in Obstetrics & Gynecology 16(5):414-8, 2000.

Ebrashy A, Azmy O, Ibrahim M, Waly M, Edris A. Middle cerebral/umbilical artery resistance index ratio as sensitive parameter for fetal well-being and neonatal outcome in patients with preeclampsia: case-control study. Croatian Medical Journal 46(5):821-5, 2005.

Fong KW, Ohlsson A, Hannah ME, Grisaru S, Kingdom J, Cohen H et al. Prediction of perinatal outcome in fetuses suspected to have intrauterine growth restriction: Doppler US study of fetal cerebral, renal, and umbilical arteries. Radiology 1999; 213(3):681-689.

Gramellini D, Folli MC, Raboni S, Vadora E, Merialdi A. Cerebral-umbilical Doppler ratio as a predictor of adverse perinatal outcome. Obstetrics & Gynecology 1992; 79(3):416-420.

Gramellini D, Piantelli G, Verrotti C, Fieni S, Chiaie LD, Kaihura C. Doppler velocimetry and non stress test in severe fetal growth restriction. Clinical & Experimental Obstetrics & Gynecology 2001; 28(1):33-39.

Hata T, Aoki S, Manabe A, Kanenishi K, Yamashiro C, Tanaka H et al. Subclassification of small-for-gestational-age fetus using fetal Doppler velocimetry. Gynecologic & Obstetric Investigation 2000; 49(4):236-239.

Hernandez-Andrade E, Figueroa-Diesel H, Jansson T, Rangel-Nava H, Gratacos E. Changes in regional fetal cerebral blood flow perfusion in relation to hemodynamic deterioration in severely growth-restricted fetuses. Ultrasound in Obstetrics & Gynecology 2008; 32(1):71-76.

Hershkovitz R, Kingdom JC, Geary M, Rodeck CH. Fetal cerebral blood flow redistribution in late gestation: identification of compromise in small fetuses with normal umbilical artery Doppler. Ultrasound in Obstetrics & Gynecology 2000; 15(3):209-212.

Joern H, Schroeder W, Sassen R, Rath W. Predictive value of a single CTG, ultrasound and Doppler examination to diagnose acute and chronic placental insufficiency in multiple pregnancies. Journal of Perinatal Medicine 1997; 25(4):325-332.

Jorn H, Funk A, Fendel H. Doppler ultrasound diagnosis in post-term pregnancy. Geburtshilfe und Frauenheilkunde 1993; 53(9):603-608.

Lakhkar BN, Rajagopal KV, Gourisankar PT. Doppler prediction of adverse perinatal outcome in PIH and IUGR. Indian Journal of Radiology & Imaging Vol 16(1)()(pp 109-116), 2006 2006;(1):109-116.

Luzi G, Coata G, Caserta G, Cosmi EV, Di Renzo GC. Doppler velocimetry of different sections of the fetal middle cerebral artery in relation to perinatal outcome. Journal of Perinatal Medicine 1996; 24(4):327-334.

Mari G, Deter R. Middle cerebral artery flow velocity waveforms in normal and small for gestational age fetuses. Am J Obstet Gynecol 1992; 166:1262-1270.

Mari G, Hanif F, Kruger M, Cosmi E, Santolaya-Forgas J, Treadwell MC. Middle cerebral artery peak systolic velocity: A new Doppler parameter in the assessment of growth-restricted fetuses. Ultrasound in Obstetrics & Gynecology Vol 29(3)()(pp 310-316), 2007 2007;(3):310-316.

Maunu J, Ekholm E, Parkkola R, Palo P, Rikalainen H, Lapinleimu H et al. Antenatal Doppler measurements and early brain injury in very low birth weight . Journal of Pediatrics 150(1):51-56 e1, 2007.

Meyberg R, Hendrik HJ, Ertan AK, Friedrich M, Schmidt W. The clinical significance of antenatal pathological Doppler findings in fetal middle cerebral artery compared to umbilical artery and fetal aorta. Clinical & Experimental Obstetrics & Gynecology 2000; 27(2):92-94.

Meyberg R, Tossounidis I, Ertan AK, Friedrich M, Schmidt W. The clinical significance of antenatal pathological Doppler findings in the fetal middle cerebral artery in cases with peripheral reduced diastolic doppler flow but no absence of end-diastolic flow in the umbilical artery or fetal aorta. Clinical & Experimental Obstetrics & Gynecology 28(1):17-9, 2001.

Mimica M, Pejkovi L, Furlan I, Vuli-Mladini D, Praprotnik T. Middle cerebral artery velocity waveforms in fetuses with absent umbilical artery end-diastolic flow. Biology of the Neonate 67(1):21-5, 1995.

Miyashita S, Chiba Y. Doppler studies can predict long-term outcome of growth-restricted fetuses. Journal of Medical Ultrasound 2002; 10(2):86-93.

Ozcan T, Sbracia M, d'Ancona RL, Copel JA, Mari G. Arterial and venous Doppler velocimetry in the severely growth-restricted fetus and associations with adverse perinatal outcome. Ultrasound in Obstetrics & Gynecology 1998; 12(1):39-44.

Ozeren M, Dinc H, Ekmen U, Senekayli C, Aydemir V. Umbilical and middle cerebral artery Doppler indices in patients with preeclampsia. European Journal of Obstetrics, Gynecology, & Reproductive Biology 1999; 82(1):11-16.

Spinillo A, Montanari L, Roccio M, Zanchi S, Tzialla C, Stronati M. Prognostic significance of the interaction between abnormal umbilical and middle cerebral artery Doppler velocimetry in pregnancies complicated by fetal growth restriction. Acta obstetricia et gynecologica Scandinavica 2009; 88(2):159-166.

Strigini FA, De LG, Lencioni G, Scida P, Giusti G, Genazzani AR. Middle cerebral artery velocimetry: different clinical relevance depending on umbilical velocimetry. Obstetrics & Gynecology 1997; 90(6):953-957.

Tchirikov M, Strohner M, Forster D, Huneke B. A combination of umbilical artery PI and normalized blood flow volume in the umbilical vein: Venous-arterial index for the prediction of fetal outcome. European Journal of Obstetrics Gynecology and Reproductive Biology 142(2)()(pp 129-133), 2009 Date of Publication: February 2009 2009;(2):129-133.

Appendix 32: Table of study characteristics of included studies for systematic review of middle cerbral artery Doppler to predict small for gestational age fetuses / compromise of fetal wellbeing.

First Author (year)	**Population Age (country/study design)**	**No of fetuses analysed**	**Gestational age at test (weeks)**	**Reference Standard SGA**	**Incidence of SGA (%)**	**Reference standard Fetal compromise**	Details Index test
Alatas (1996)	High and low risk populations. INC: Previous perinatal death, poor obstetric history, IDDM, previous premature birth, hypertension, recurrent spontaneous abortion, CAH High risk mean age 28.2 (17-41), low risk 25.0 (18-34) (Turkey) (cohort)	237	>24. Mean test to delivery interval high risk 5.5 days (0-12); low risk 7.1 days (0-14)	BW<10th centile	High risk 25% Low risk 14.8%	Apgar score at 1 min<7, 5 min<7, cord pH<7.2, admission to NICU	TA, pulsed+color, site not reported PI<2sd
Arduini (1992)	High risk. INC: singleton, accurate gestation, AC<5th or EFW<10th, no structural or chromosomal anomalies	120	Mean 32.2+/-3.0	NA	NA	Adverse perinatal outcome	TA, pulsed+color, site not

	Mean age 29.4+/-4.3 weeks (Italy) (cohort)					(perinatal death, cs due to abnormal FHR, Apgar score at 5 mins<7, admission to NICU for asphyxia>48hrs)	reported, PI<5th centile
Arias (1994)	High risk. INC: Suspected IUGR, PE, preterm labour, chronic hypertension, APS. Mean age control 31+/-3.8 yrs; study 29.5+/-5.7 yrs. (USA) (Case-control, prospective)	81	24-38 weeks, Doppler within 2 weeks of delivery	BW<10th centile with evidence of FGR (decreased subcutaneous fat, hypoglycaemia, hyperbilirubinaemia, hypocalcaemia, hyperviscosity)	23.4%	NA	TA, method not reported, Circle of Willis, RI-cut-off not reported.
Cavero	High risk.	83	Control	BW<10th	49.4%	Apgar 1	TA,

(1996)	INC: Cases-Antenatally suspected IUGR (AC<2sd), controls matched for maternal age, parity, height, weight, edd. Mean age controls 28.3+/-4.5 yrs; cases 27.6+/-6.8 yrs (Spain) (Case control, prospective)		mean 247 days+/-23.8; cases 247.2 days+/-24.6	centile		min<7, neonatal resuscitation required, admission to NICU.	method and site not reported, parameter and cut-off not reported.
Chandran (1993)	High risk. INC: PE, AC<3rd centile, abnormal umbilical artery Doppler, singleton, delivery by prelabour cs. EXC: fetal chromosomal and structural anomalies Mean age not reported (Malaysia) (Cohort, prospective)	27	24-39 weeks (test for analysis performed within 24 hrs of delivery)	BW<3rd centile	70.4%	Cord pH<7.12 BE>12.0mmol/l; cord pO2<8.9mmHg; IVH, NEC, HMD; adverse outcome.	TA, method not reported, level of BPD, PI<2sd
Del Ri (2008)	High risk INC: 51 singleton fetuses with IUGR (EFW<10th centile) and either an umbilical artery pulsatility index>95th centile or a cerebroplacental ratio<5th centile (MCA/UA<5th centile) EXC: no structural or chromosomal abnormalities Median maternal age in abnormal Doppler 32 (22-40) normal 28 (22-	51	24-36 weeks (test for analysis within 48 hours	NA	NA	APO= any of stillbirth, neonatal mortality, BPD, RDS, grade 3/4	Route not reported, color+pulsed, site not reported, vasodilatat

	37) years. (Spain) (Cohort, prospective)		of delivery)			IVH, NEC, sepsis and stay in NICU>14 days	ion
Dubiel (1997)	High risk. INC: suspected IUGR, PIH, post term, diabetes, decreased fetal movements, increased resistance in umbilical artery. Mean age not reported. (Sweden) (Cohort)	50	31-42 Median 38 (test for analysis performed within 7 days of delivery)	NA	NA	Apgar at 1 or 5 mins<7	Route not reported, color +pulsed, site not reported, PI <mean - 2sd.
Dubiel (2000)	High risk. INC: singletons with PIH. Mean age not reported (Sweden) (Cohort)	102	27-41 Median 36 (test for analysis performed within	BW<5th centile (local values)	55.6%	Apgar at 5 mins<7; cord artery pH<7.15, cord vein pH<7.20; admission to NICU, need	TA, color+pulsed, site not reported, PI<mean - 2sd.

			1-2 days of delivery)			for ventilation, Perinatal mortality	
Ebrashy (2005)	High risk INC: Viable singleton pregnancy, no other obstetric or other morbidity, PE diagnosed according to ISSHP criteria, no medication except for iron and delivered by elective cs not in labour and not for fetal distress. EXC: fetal chromosomal and structural anomalies. Mean age 24.8+/-6.1 years (Egypt) (Case control, prospective)	50	Mean 36.9+/-2.5	NA	NA	Cord pH<7.20	TA, color+pulsed, Cirlce of Willis, RI<0.69
Fong (1999)	High risk. INC: singleton, >24weeks, EFW<10th EXC: congenital or chromosomal abnormality Mean age 32.6+/_3.7 years (Canada) (Cohort, prospective, consecutive)	293	Mean 32.6+/-3.7 (test for analysis within 2 weeks of delivery)	NA	NA	Adverse perinatal outcome (perinatal death, HIE, IVH, PVL, NEC, arterial cord pH<7.1,	TA, pulsed+color, Circle of Willis, PI<5th centile

						Apgar 5 mins<7, cs for fetal distress)	
Gramellini (1992)	Unselected population INC: singletons Age not reported (Italy)(Case control, retrospective)	90	30-41	BW<10th centile, local values	50%	NA	TA, Method not reported, level of thalamus, PI cut-off not reported.
Gramellini (2001)	High risk. INC: Ultrasound dated pregnancy <20 weeks, AC<2.5th centile, Doppler within 2 weeks of birth and a non-stress test within 2 hrs of delivery, singleton. EXC: chromosomal or structural anomalies Age not reported (Italy) (cohort, retrospective)	53	24-35 (test for analysis performed within 2 weeks of delivery)	NA	NA	Neonatal resuscitation required, perinatal mortality.	Route not reported, color, site not reported, PI<5th centile.
Hata (1999)	High risk. INC: singletons, EFW<10th centile EXC: structural and	54	Within 2-3 weeks	NA	NA	Apgar at 5 mins<7,	TA, color+puls

	chromosomal abnormalities Mean age abnormal Doppler 29.4+/-5.1 years, normal Doppler 28.4+/-4.9 years. (Japan)(cohort)		of delivery			umbilical artery pH<7.15, admission to NICU.	ed, level of greater wings of sphenoid, PI cut-off not reported.
Hernandez-Andrade (2008)	High risk INC: 56 fetuses with IUGR (EFW<10th centile and an abnormal PI mean<2sd in umbilical artery). Median maternal age 32 (20-39) years (Spain) (Cohort)	56	Median 29 (26-32) weeks	NA	NA	Apgar at 5 mins<7, umbilical artery pH<7.20, IUD, NND, PND, adverse perinatal outcome	TA, color+pulsed, origin from Circle of Willis, PI mean<2sd.
Hershkovitz (2000)	High risk. INC: singletons, EFW<5th centile EXC: structural and chromosomal abnormalities Mean age abnormal Doppler 28.8+/-9.7 years, normal Doppler 26.27+/-6.7 years. (UK)(case control, retrospective)	47	Median gestation abnormal Doppler 37(35-40),	NA	NA	Apgar 5 mins<7, admission to NICU.	TA, color+pulsed, site not reported, PI<5th centile.

			normal Doppler 38(35-40).				
Joern (1993)	High risk. INC: post term, EDD from LMP and first trimester ultrasound. Age not reported (Germany) (Cohort)	59	>40 weeks	BW<10th centile local values.	10.1%	NA	TA, color+pulsed, site not reported, PI<1.0.
Joern (1997)	High risk. INC: twins and triplets EXC: structural and chromosomal anomalies. Mean age 29 (19-40) (Germany)(cohort)	261	Mean 34 (26-41), (test for analysis performed within mean of 19 days of delivery)	BW<10th centile local values.	31%	Adverse outcome (any of umbilical artery pH<7,20, Apgar at 5 mins<8, admission to NICU)	Route not reported, color, site not reported, PI<1.3.
Lakhkar	High risk.	58	>30	NA	NA	Minor adverse outcome (cs	TA,

(2006)	INC: singletons, PE, EFW<10th centile. EXC: chromosomal and structural anomalies. Mean age 27.3 years. (India) (cohort, prospective)		weeks (test for analysis performed within 10 days of delivery)			for fetal distress or Apgar at 5 mins<7 or admission to NICU) Major (perinatal death, HIE, IVH, PVL, pulmonary haemorrhage, NEC)	pulsed, level of wings of greater sphenoid, PI<5th centile.
Luzi (1996)	High risk. INC: delay in rate of fetal growth>25th centile. Age not reported (Italy) (case control)	37	28-term	BW<10th centile, local values.	8.1%	NA	TA, method and site not reported, PI< 1 sd.
Mari (1992)	High risk. INC: Suspected SGA (HC and AC mean -2sd) Age not reported. (USA) (case-control)	33	Mean 31+/-4.3 weeks (20-37)	NA	NA	Adverse perinatal outcome (admission to NICU>12 hours or	TA, color+pulsed, site not reported, PI mean -2sd.

						perinatal death)	
Mari (2007)	High risk. INC: Singleton, EFW<3rd centile, Umbilical Artery PI>95th, all delivered<33 weeks EXC: chromosomal and structural anomalies. Age not reported (Italy) (cohort, prospective)	30	Median 27+2 weeks (23-32+4) (test for analysis performed within 8 days of delivery)	NA	NA	Perinatal mortality; perinatal morbidity (IVH grade 3 or 4, BPD)	TA, color+pulsed, site not reported, PSV mean +2sd or PI mean -2sd.
Maunu (2007)	High risk. INC: Preterm birth (<37 weeks), VLBW<1500g. Mean age 30,2 (+/-5.2) years (Finland) (Cohort, prospective)	63	Mean 28+2 weeks (24-36) (last Doppler performed within 7 days of	NA	NA	MRI at term abnormal (e.g. IVH, ventriculomegaly, ischaemic lesions)	Route, method and site not reported, PI<5th centile.

			delivery)				
Meyberg (2000)	Unselected. INC: Not reported. Age not reported. (Germany) (Cohort)	144	28-40	BW<10th centile BW<5th centile (local values)	27.1% 13.9%	Adverse outcome (umbilical artery pH<7.2 and/or Apgar at 1 min<7)	Route and site not reported, pulsed+color, SD<10th centile.
Meyberg (2001)	High risk. INC: Abnormal umbilical artery and fetal aorta Doppler. Age not reported (Germany) (case control)	96	28-40	BW<10th centile (local values)	74.0%	Adverse outcome (umbilical artery pH<7.2 and/or Apgar at 1 min<7)	Route and site not reported, pulsed+color, SD<10th centile.
Mimica (1995)	High risk. INC: singletons, suspected SGA., absent end –diastolic flow of umbilical artery Age not reported. (Croatia) (Cohort)	21	Not reported	NA	NA	Apgar 5 mins<7, perinatal death.	TA, pulsed, site not reported, RI<5th

							centile.
Miyashita (2002)	High risk. INC: singletons, EFW<1.5sd EXC: structural and chromosomal anomalies. Age not reported. (Japan) (Cohort, prospective)	119	24-36 (test for analysis within 10 days of delivery)	NA	NA	Adverse outcome (neonatal death, infantile death, cerebral palsy and or developmental retardation)	TA, pulsed, site not reported, RI<4.0sd.
Ozcan (1998)	High risk. INC: EFW<5th centile, Doppler within 2 weeks of delivery EXC: chromosomal and structural anomalies, birth weight>5th centile. Age not reported. (USA) (cohort)	19	Mean 28.2 weeks (27-31.4) (test for analysis performed within a median of 2 days (0-14))	NA	NA	Perinatal death, Apgar 5 mins<7, NEC.	TA, color+pulsed, Circle of Willis, PSV mean +2sd or PI mean -2sd.

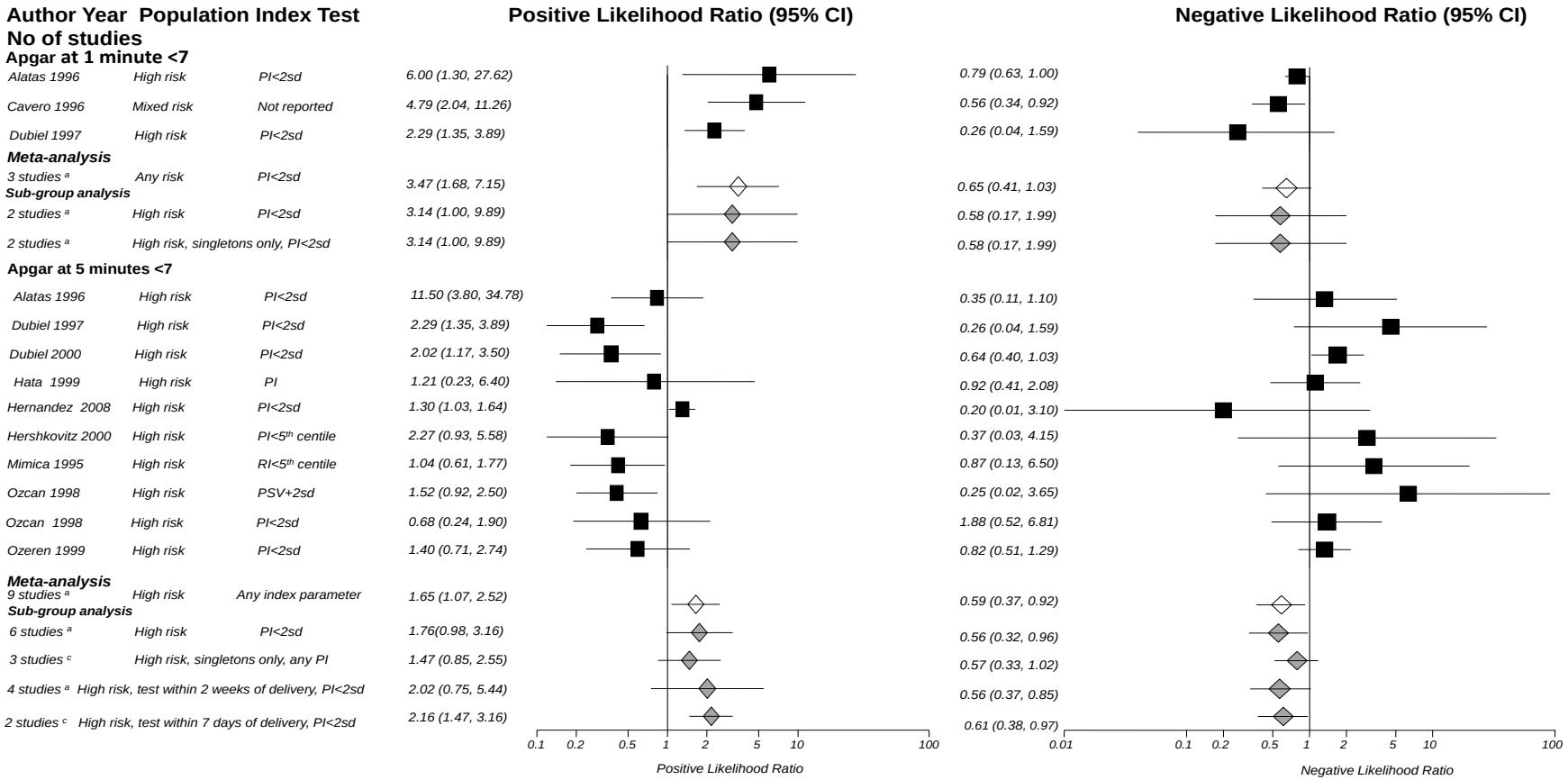
Ozeren (1999)	High risk. INC: singletons, PE EXC: structural and chromosomal anomalies Mean age 27.6+/-5.2 years (Turkey) (case control)	62	Mean 35.8+/-2.6 weeks.	BW<10th centile (local values)	40.3%	Perinatal deaths, NICU≥7 days or neonatal death, Apgar <7 at 5 mins. Combined as adverse outcome.	TA, method not reported, level of thalamus, PI<2sd.
Spinillo (2009)	High risk INC: 184 singleton pregnancies at 24-35 weeks complicated by FGR (AC<10th centile) and abnormal uterine artery Doppler measurements (PI>95th centile or AREDF), ga confirmed by ultrasound scan EXC: fetal chromosomal and structural anomalies Age not reported (Italy) (Cohort, prospective)	184	Median 28.9 (23-34) weeks (test for analysis performed within 48 hrs of delivery)	NA	NA	Adverse perinatal outcome: fetal and neonatal death, severe neonatal brain damage (grade 3 or 4 ICH or cystic leukomalacia). Severe neonatal complicatins:	Route, method and site not reported, PI<10th centile.

						brain damage, ROP, broncho pulmonary dysplasia, NEC)	
Strigini (1997)	High risk. INC: Singleton, suspected FGR, poor obstetric history, preterm labour, hypertension, diabetes, reduced fetal movements, APH EXC: structural and chromosomal anomalies Age not reported. (Italy)	576	Mean 35.1 (25-41) weeks (test for analysis performed within 3 weeks of delivery)	BW<10th centile (local values)	17.9%	Adverse perinatal outcome (fetal death or death before discharge, 5 min Apgar<7, CTG abnormality leading to emergency cs)	TA, color+pulsed, level of BPD, PI mean <1.5sd.
Tchirikov (2009)	Mixed risk INC: 181 patients with singleton pregnancies, confirmed ga. EXC: fetal malformations Age not reported		17-41 weeks	NA	NA	Adverse perinatal outcome	TA, method and site not

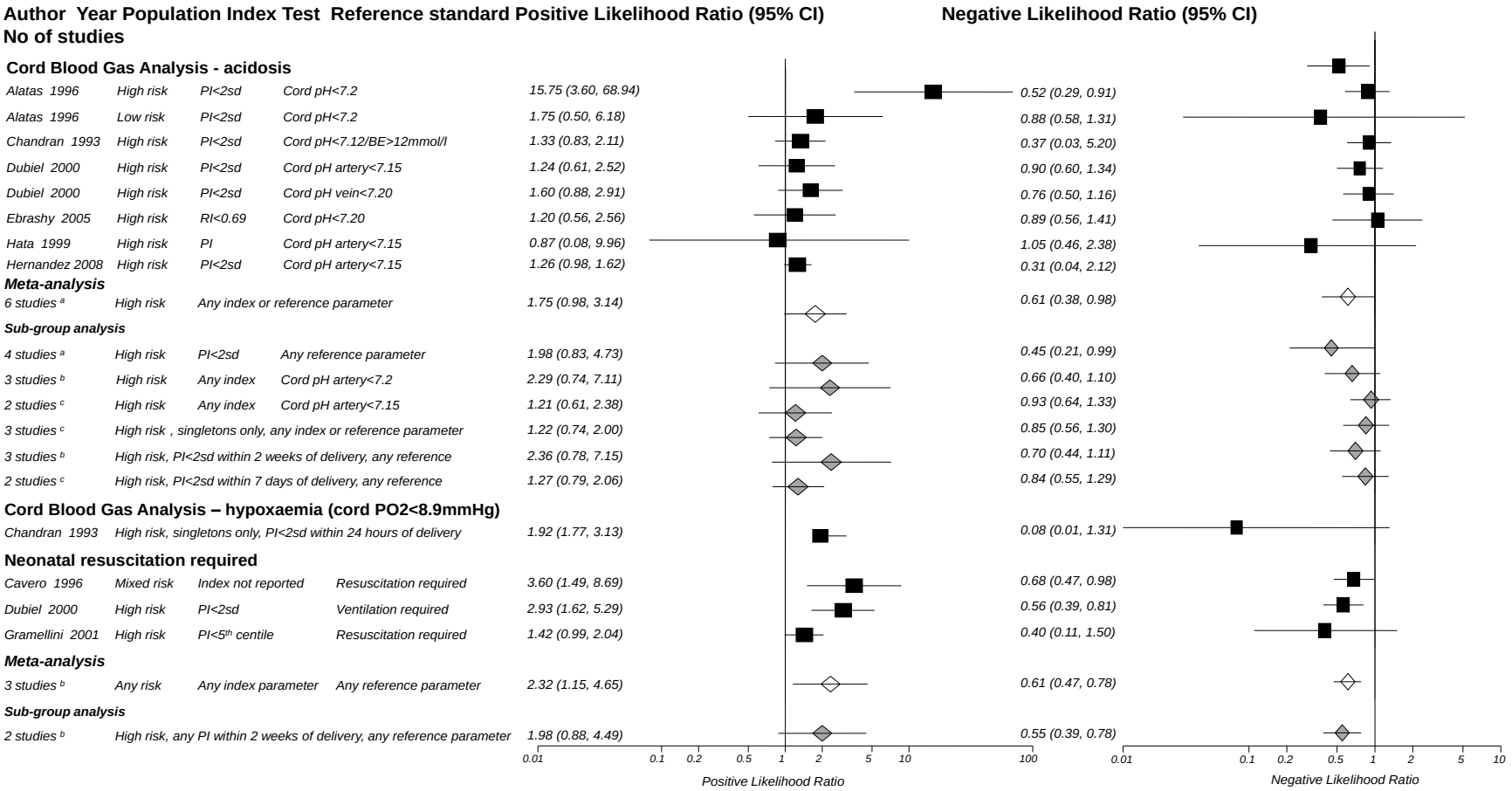
(Germany) (Cohort, prospective)		reported, PI<5th centile

Hrs hour; INC inclusion; EXC exclusion; CAH congenital adrenal hyperplasia; PE preeclampsia; PIH pregnancy induced hypertension; FGR fetal growth restriction; APS antiphospholipid syndrome; BW birth weight; UK United Kingdom; USA United States of America; NA not applicable; SGA small for gestational age; IUGR intrauterine growth restriction; USS ultrasound scan; ISSHP International Society for the Study of Hypertension in Pregnancy; ga gestational age; sd standard deviation, % percent; NICU neonatal intensive care unit; TA transabdominal; PI pulsatility index; RI resistance index; PSV peak systolic velocity; SD systolic/diastolic ratio; AC abdominal circumference; EFW estimated fetal weight; CS caesarean section; FHR fetal heart rate; EDD estimated date of delivery; BE base excess; IVH intraventricular haemorrhage; NEC necrotising enterocolitis; HMD hyaline membrane disease; BPD biparietal diameter; HIE hypoxic ischaemic encephalopathy; PVL peri-ventricular leukomalacia; LMP last menstrual period; CTG cardiotocogram; mins minutes; APO adverse perinatal outcome; MCA middle cerebral artery; UA umbilical artery; RDS respiratory distress syndrome; IUD intrauterine death; NND neonatal death; PND perinatal death; AREDF absent reversed end-diastolic flow; ICH intracranial haemorrhage; ROP retinopathy of prematurity

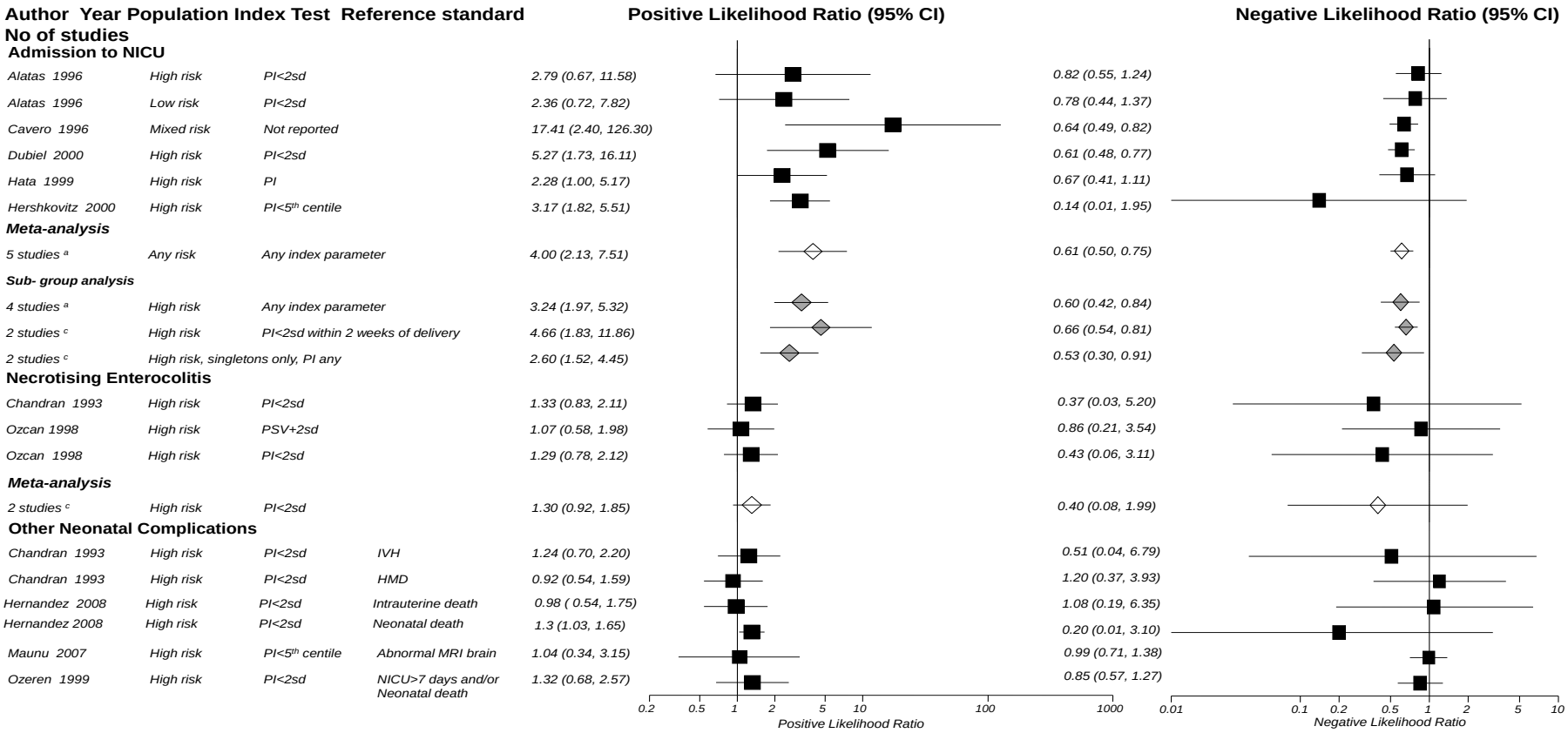
Appendix 33: Forest plot of positive and negative likelihood ratios for middle cerebral artery Doppler to predict compromise of fetal wellbeing (Apgar scores). Single studies are represented by a filled box, pooled results by an open diamond and subgroup analysis by a filled diamond.



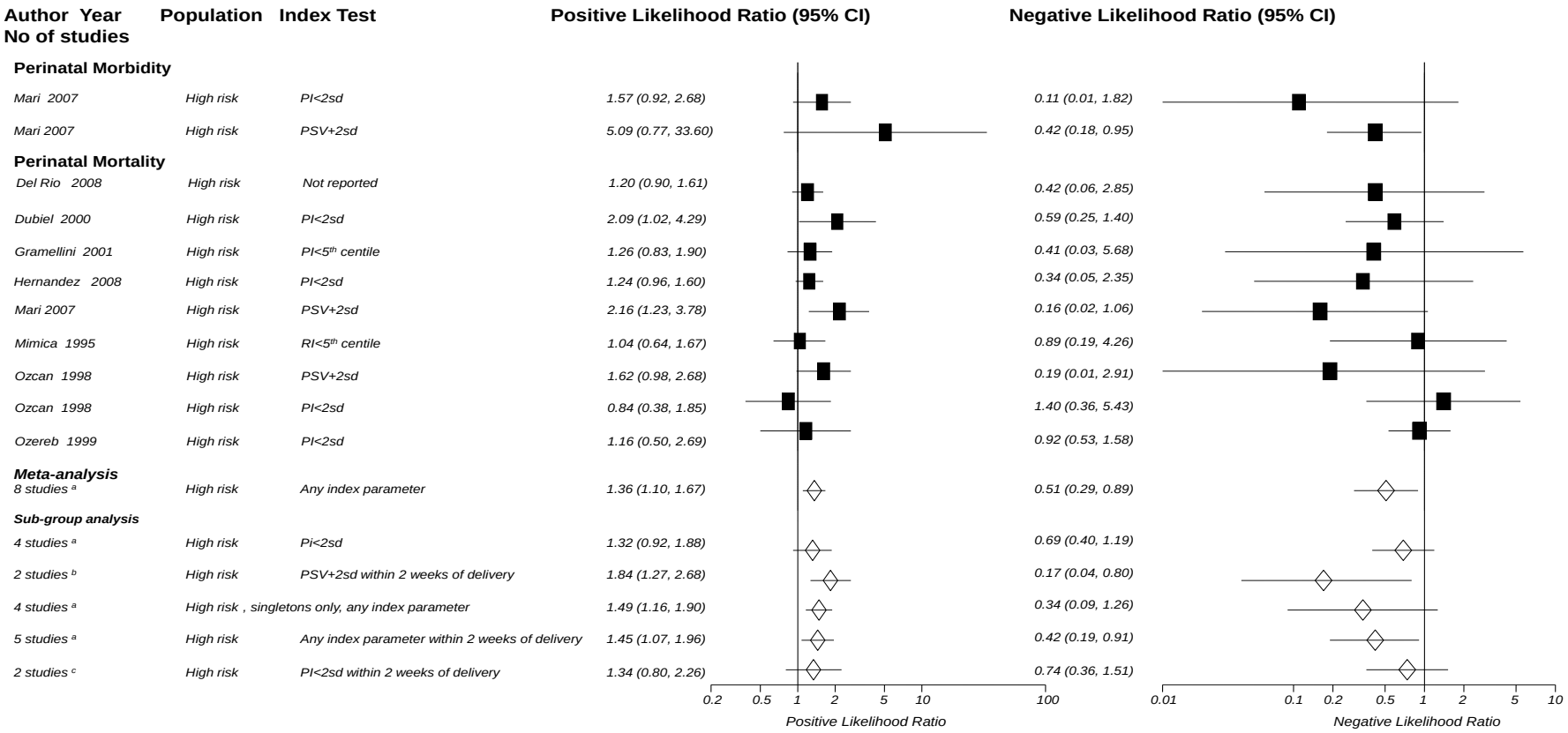
Appendix 34: Forest plot of positive and negative likelihood ratios for middle cerebral artery Doppler to predict compromise of fetal wellbeing (cord blood gas analysis/need for neonatal resuscitation). Single studies are represented by a filled box, pooled results by an open diamond and subgroup analysis by a filled diamond.



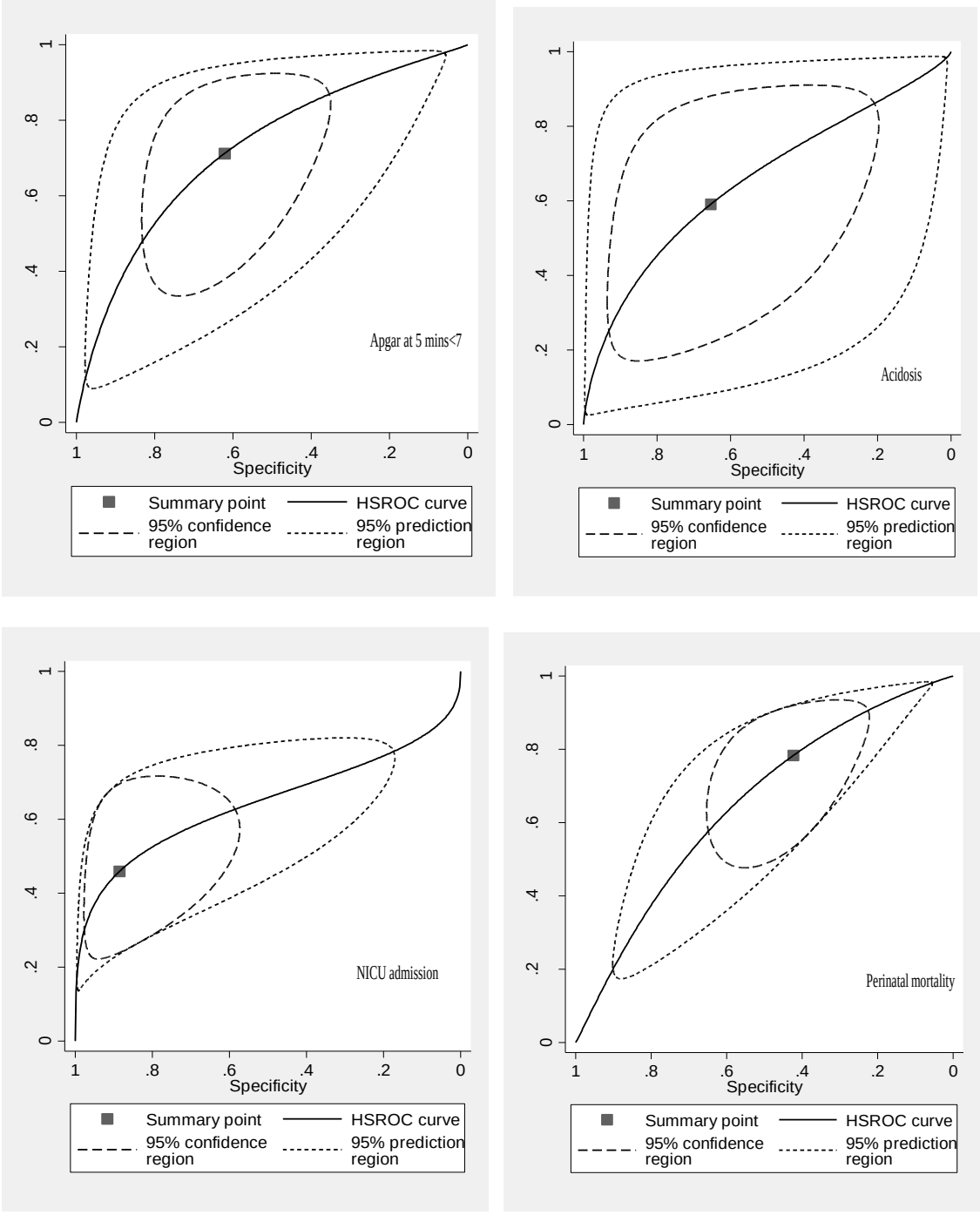
Appendix 35: Forest plot of positive and negative likelihood ratios for middle cerebral artery Doppler to predict compromise of fetal wellbeing (admission to neonatal intensive care/neonatal complications). Single studies are represented by a filled box, pooled results by an open diamond and subgroup analysis by a filled diamond.



Appendix 36: Forest plot of positive and negative likelihood ratios for middle cerebral artery Doppler to predict perinatal morbidity and mortality. Single studies are represented by a filled box, pooled results by an open diamond and subgroup analysis by a filled diamond.



Appendix 37: Summary receiver operating characteristic curves for middle cerebral artery Doppler to predict apgar at 5 mins<7, acidosis at cord pH, admission to neonatal intensive care unit (NICU) and perinatal mortality produced using the bivariate method.



Appendix 38: Search strategy for systematic review of ductus venosus Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing.

Host: Ovid

Date of search: May 2009

Years covered by search: 1950-2009

1. **MEDLINE (inception until May 2009) -614 citations**

   8. ("Pregnant woman"[MeSH] OR "Pregnancy"[MeSH] OR pregnan*)
   9. ("Prenatal Diagnosis[MeSH] OR "Ultrasonography/Prenatal"[MeSH} OR "Ultrasonography/Doppler"[MeSH])
   10. {(venous Doppler.mp) OR (Doppler velocimetry.mp) OR (Doppler ultrasound.mp) OR (DV.mp) OR (Ductus venosus[MeSH])}
   11. (1 AND 2)
   12. (4 AND 3)
   13. Limit 5 to animals
   14. (5 NOT 6)

2. **Medline search adapted for EMBASE (inception until May 2009) - 456 citations**

3. **Cochrane library (2009:2) – 82 reviews, 173 clinical trials, 15 technology assessments, 75 economic evaluations**

   12. Pregnant women
   13. Prenatal diagnosis

14. Ultrasonography prenatal

15. Ultrasonography Doppler

16. Venous Doppler

17. Doppler velocimetry

18. Doppler ultrasound

19. DV

20. Ductus venosus

21. (2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9)

22. (1 AND 10)

4. MEDION – 0 citations

5. Grey literature – 0 citations

Appendix 39: Data extraction form for systematic review of ductus venosus Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing.

Section A: Study Information

1)Ref ID:		**4)Publication year:**	
2)Rev name:		**5)First Author:**	
3)Country:		**6)Language:**	

Section B: Data Retrieval for Ductus venosus Doppler Study

Population

7) Healthcare Centre:

Primary care ☐$_1$ Secondary care ☐$_2$ Mixed ☐$_3$ Other ☐$_4$ Unreported ☐$_5$

8) Setting:

In-patient ☐$_1$ Out-patient ☐$_2$ Mixed ☐$_3$ Unreported ☐$_4$ Other ☐$_5$ _____

9) Number of participating centres: _____

10) Gestation at time of index test:

<20 weeks ☐$_1$ 20-24 weeks ☐$_2$ 24-28 weeks ☐$_3$ 28-34 weeks ☐$_4$ 34-37 weeks ☐$_5$ 37-40 weeks ☐$_6$ > 40 weeks ☐$_7$ Unreported ☐$_8$ Other _____

10.i) Mean (range)_____ Unreported ☐$_3$

10.ii) Median (range) _____ Unreported ☐$_3$

11) Pregnancy:

Low Risk ☐$_1$ High Risk ☐$_2$ Unselected ☐$_3$ Unreported ☐$_4$

11.i) State high risk conditions: Unreported ☐$_3$

12) Were patients with the following conditions excluded/not included?

12.i) Previous IUGR: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.ii) Insulin dependant diabetes mellitus: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.iii) Chronic renal disease: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.iv) Systemic lupus erythematosus: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.v) Antiphospholipid syndrome: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.vi) Chronic hypertension: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.vii) Pre-eclampsia: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

12.viii) Foetal chromosomal/structural anomalies: Yes ☐$_1$ No ☐$_2$ Unreported ☐$_3$

13) Did all patients have singleton pregnancies?:

Yes ☐$_1$ No ☐$_2$ ----------/---------- Unreported ☐$_3$

14) Were all patients primigravid?:

Yes ☐$_1$ No ☐$_2$ ----------/---------- Unreported ☐$_3$

15) List other eligibility/ in-/exclusion criteria:

Not applicable ☐$_3$

16) Study population: (describe age (mean +/- SD or median/range), ethnicity, smoking, BMI etc.)

Unreported ☐$_3$

17) Start of patient inclusion (year) : ☐

Unreported ☐3

18) End of patient inclusion (year) : ☐

Unreported ☐3

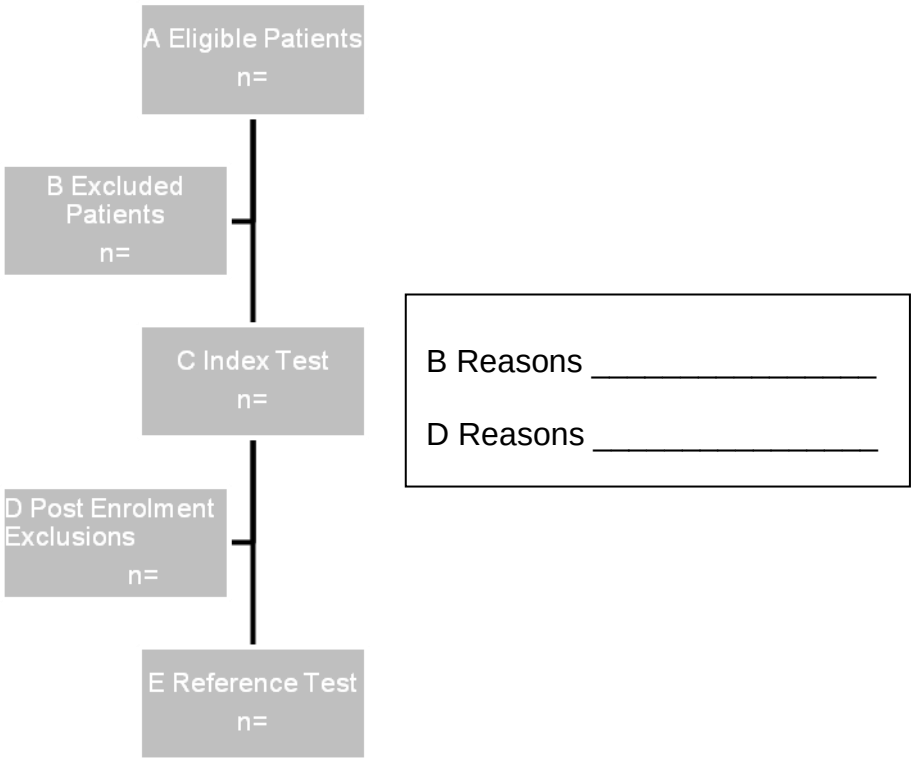
19) Study Design:

cohort ☐1 case control ☐2 RCT/CCT ☐3 cross sectional ☐4 before and after ☐5 case series ☐6 (no _____) other ☐7

19.i) Data collection: prospective ☐1 retrospective ☐2 unreported ☐3 other ☐4

19.ii) Enrolment: consecutive ☐1 arbitary (random) ☐2 unreported ☐3 other ☐4

20) Numbers:



21) Completeness of Verification:

(= E / C x 100 = %)

> 90% ☐$_1$ 81-90% ☐$_2$ < 81% ☐$_3$

Index Test

22) Description of technique:

Adequate ☐$_1$ Inadequate ☐$_2$

23) Timing of measurement (from delivery):

< 7days ☐$_1$ 7-14 days ☐$_2$ 14 -28 days ☐$_3$ > 28 days ☐$_4$ Mixture ☐$_5$

Unreported ☐$_6$

23.i) Median gestational age at delivery _____

unreported ☐$_3$

24) Measurement:

SCANNING:

24.i) Operator:

Single ☐$_1$ Multiple ☐$_2$ Unreported ☐$_3$

24.ii) Operator experience_____

unreported ☐$_3$

24.iii) Scanning Route: Transabdominal ☐$_1$ Transvaginal ☐$_2$ Unreported ☐$_3$

DOPPLER:

24.iv) Method: Continuous wave Doppler ☐$_1$ Pulsed wave Doppler ☐$_2$ Colour mapping ☐$_3$ Unreported ☐$_4$

24.v) Measurement parameter: Resistance index (RI) ☐$_1$ Systolic / diastolic ratio ☐$_2$ Diastolic / systolic ratio ☐$_3$ Unilateral Diastolic notch ☐$_4$ Bilateral diastolic notch ☐$_5$ Pulsatility index (PI) ☐$_6$ Reduced EDF ☐$_7$

Absent EDF ☐$_8$ Reversed EDF ☐$_9$ Peak velocity ☐$_{10}$ Time-averaged maximal velocity ☐$_{11}$ Minimum velocity ☐$_{12}$ Unreported ☐$_{13}$

24.vi) Cut-off level for waveform ratio: > 2 SD ☐$_1$ > 95th centile ☐$_2$ > 90th centile ☐$_3$

> 80th centile ☐$_4$ > 50th centile ☐$_5$ < 10th centile ☐$_6$ < 5th centile ☐$_7$

Unreported/NA ☐$_8$

Other/Threshold data set:

_____

24.vii) Machine: _____

unreported ☐$_3$

24.viii) Probe: _____

unreported ☐$_3$

24.ix) High pass filter: _____

unreported ☐$_3$

24.x) Pulse rePEition frequency: _____

unreported ☐$_3$ _____

24.xi) Size of sampling gate: _____

unreported ☐$_3$

24.xii) Site : _____

unreported ☐$_3$

24.xiii) Angel of insonation: _____unreported ☐$_3$

24.xiv) Number of consecutive waveforms: _____unreported ☐$_3$

24.xv) Other information:

_____

Reference Standard / Outcome

25) Measured blind form diagnostic test: Yes ☐$_1$ No ☐$_2$ Unclear ☐$_3$

26) Measurement for FGR: Birthweight ☐$_1$ Neonatal ponderal index ☐$_2$

Skin fold thickness ☐$_3$ MAC / OFC ☐$_4$ Other ☐$_5$

27) Threshold: < 3rd centile ☐$_1$ < 5th centile ☐$_2$ < 10th centile ☐$_3$ < 25th centile ☐$_4$

> 2SD ☐$_5$ Other ☐$_6$ _____ Unclear ☐$_7$

28) What data set was used to define threshold?

_____ unreported ☐$_3$

29)**Timing of measurement:** At delivery ☐$_1$ Within 24 hrs ☐$_2$ > 24 hrs ☐$_3$

Mixture ☐$_4$ Unreported ☐$_5$

30) Marker of wellbeing e.g. Apgar score, perinatal mortality

31) Threshold and data set (if applicable):

32) Measured blind form diagnostic test: Yes ☐$_1$ No ☐$_2$ Unclear ☐$_3$

Results

	Reference Test: Threshold:			
Index test, Measurement:		Positive	Negative	**Total**
	Positive	**TP**	**FP**	
Threshold:	Negative	**FN**	**TN**	
	Total			

Appendix 40: References of included papers for systematic review of ductus venosus Doppler to predict small for gestational age fetuses and compromise of fetal/neonatal wellbeing.

Alves S, Francisco RP, Miyadahira S, Krebs V, Vaz F, Zugaib M. Ductus venousus Doppler and postnatal outcomes in fetuses with absent or reversed end-diastolic flow in the umbilical arteries. *Eur J Obstet, Gynecol, Repro Biol* 2008; 141:100-103.

Baschat AA, Gembruch U, Weiner CP, Harman CR. Qualitative venous Doppler waveform analysis improves prediction of critical perinatal outcomes in premature growth-restricted fetuses. *Ultrasound Obstet Gynecol* 2003; 22(3):240-245.

Baschat AA, Guclu S, Kush ML, Gembruch U, Weiner CP, Harman CR. Venous Doppler in the prediction of acid-base status of growth-restricted fetuses with elevated placental blood flow resistance. *AJOG* 2004; 191(1):277-284.

Baschat AA, Galan HL, Bhide A, Berg C, Kush ML, Oepkes D et al. Doppler and biophysical assessment in growth restricted fetuses: distribution of test results. *Ultrasound Obstet Gynecol* 2006; 27(1):41-47.

Bilardo CM, Wolf H, Stigter RH, Ville Y, Baez E, Visser GH et al. Relationship between monitoring parameters and perinatal outcome in severe, early intrauterine growth restriction. *Ultrasound Obstet Gynecol* 23(2):119-25, 2004.

Carvalho FH, Moron AF, Mattar R, Santana RM, Murta CG, Barbosa MM et al. Ductus venosus Doppler velocimetry in the prediction of acidemia at birth: which is the best parameter? *Prenat Diag* 25(13):1212-6, 2005.

Cosmi E, Ambrosini G, D'Antona D, Saccardi C, Mari G. Doppler, cardiotocography, and biophysical profile changes in growth-restricted fetuses. *Obstet Gynecol* 106(6):1240-5, 2005.

Del Rio M, Martinez JM, Figueras F, Bennasar A, Olivella M, Palacio M et al. Doppler assessment of the aortic isthmus and perinatal outcome in preterm fetuses with severe intrauterine growth restriction. *Ultrasound Obstet Gynecol* 2008; 31:41-47.

Figueras F, Martinez JM, Puerto B, Coll O, Cararach V, Vanrell JA. Contraction stress test versus ductus venosus Doppler evaluation for the prediction of adverse perinatal outcome in growth-restricted fetuses with non-reassuring non-stress test. *Ultrasound in Obstet Gynecol* 21(3):250-5, 2003.

Figueras RF, Bennasar M, Eixarch E, Martinez JM, Puerto B, Cararach V et al. Arterial, venous and intracardiac parameters in growth-restricted fetuses: Associations with adverse perinatal outcome. *Ultrasound Rev Obstet Gynaecol* 2004; 4(3):179-185.

Gramellini D, Piantelli G, Verrotti C, Fieni S, Chiaie LD, Kaihura C. Doppler velocimetry and non stress test in severe fetal growth restriction. *Clin Exp Obstet Gynecol* 28(1):33-9, 2001.

Hofstaetter C, Gudmundsson S, Dubiel M, Marsal K. Ductus venosus velocimetry in high-risk pregnancies. *Eur J Obstet, Gynecol, Repro Biol* 1996 70(2):135-40.

Hung JH, Fu CY, Hung J. Combination of fetal Doppler velocimetric resistance values predict acidemic growth-restricted neonates. *J Ultrasound Med* 25(8):957-62, 2006.

Maiz N, Valencia C, Emmanuel E, Staboulidou I, Nicolaides K. Screening for adverse pregnancy outcome by ductus venosus Doppler at 11-13+6 weeks gestation. *Obstet Gynecol* 2008; 112(3):598-605.

Maiz N, Staboulidou I, Leal AM, Minekawa R, Nicolaides KH. Ductus venosus Doppler at 11 to 13 weeks of gestation in the prediction of outcome in twin pregnancies. *Obstet Gynecol* 2009; 113(4):860-865.

Manogura A, Turan O, Kush ML, Berg C, Bhide A, Turan S et al. Predictors of necrotising enterocolitis in preterm growth-restricted neonates. *AJOG* 2008; 198(6):638.e1-5.

Muller T, Nanan R, Rehn M, Kristen P, Dietl J. Arterial and ductus venosus Doppler in fetuses with absent or reverse end-diastolic flow in the umbilical artery: correlation with short-term perinatal outcome. *Acta Obstet Gynecol Scand* 81(9):860-6, 2002.

Ozcan T, Sbracia M, d'Ancona RL, Copel JA, Mari G. Arterial and venous Doppler velocimetry in the severely growth-restricted fetus and associations with adverse perinatal outcome. *Ultrasound Obstet Gynecol* 12(1):39-44, 1998.

Schwarze A, Gembruch U, Krapp M, Katalinic A, Germer U, xt-Fliedner R. Qualitative venous Doppler flow waveform analysis in preterm intrauterine growth-restricted fetuses with ARED flow in the umbilical artery--correlation with short-term outcome. *Ultrasound Obstet Gynecol* 25(6):573-9, 2005.

Turan S, Turan OM, Berg C, Moyano D, Bhide A, Bower S et al. Computerized fetal heart rate analysis, Doppler ultrasound and biophysical profile score in the prediction of

acid-base status of growth-restricted fetuses. *Ultrasound Obstet Gynecol* 2007;30(5):750-756.

Appendix 41: Study characteristics of included studies for ductus venosus Doppler to predict compromise of fetal/neonatal wellbeing.

First Author (year)	Population Age (country/study design)	No of women analysed	Gestational age at test (weeks)	Reference Standard SGA	Incidence of SGA (%)	Reference standard Fetal compromise	Details of Index test
Alves (2008)	High risk populations INC: 103 newborns with AREDF of the umbilical artery, singleton, no fetal anomalies, no premature rupture of membranes, fetal wellbeing tests performed on day of delivery EXC: Fetal chromosomal and structural anomalies Mean maternal age 30.08+/-6.7 (16-45) (Brazil) (Cohort, prospective)	103	On day of delivery	BW<10th centile (sex, ga)	74.8%	Hyaline membrane disease, pneumothorax, pulmonary haemorrhage, BPD, arterial canal persistence, septicaemia, NEC, ROP, thrombocytopenia, hypoglycaemia, hyperglycaemia, intracranial	Route not reported, pulsed and colour, origin from umbilical vein, a wave absent or reversed.

						haemorrhage, death	
Baschat (2003)	High risk populations. INC: Singleton, delivery prior to 37 weeks, birth weight<10th centile, umbilical artery PI mean>2sd EXC: Fetal chromosomal and structural anomalies Mean maternal age 28 +/-5.5. (Germany and USA) (cohort)	224	Within 48 hours of delivery	NA	NA	Acidaemia pH<5th percentile for gestational age; birth asphyxia pH<7.0 and/or BE>-13, stillbirths, neonatal deaths, perinatal deaths.	Route not reported, pulsed, at inlet, a wave absent or reversed.
Baschat (2004)	High risk populations. INC: Singleton, EFW<10th centile, umbilical artery PI>2sd, normal anatomy and karyotype, absence of active labour prior to caesarean section. EXC: No structural or chromosomal anomalies Mean maternal age 28+/-5.5. (Germany, USA, Turkey) (cohort, prospective)	122	Within 48 hours of delivery	NA	NA	Umbilical artery pH<7.20; umbilical artery pH<7.00 and/or BE<-13mmol/l	Route and method not reported, inlet, PI, S/A or RAV mean>2SD.
Baschat	High risk populations.	328	23-41	NA	NA	Umbilical artery	Route not

(2006)	INC: Singleton, normal anatomy, AC<5th, elevated umbilical PI, delivery at a viable gestational age. EXC: Fetal infection, chorioamnionitis, fetal anomalies, abnormal fetal karyotype Mean maternal age 29 (14-45). (Germany) (cohort, prospective)		weeks, test within 3 days of delivery			pH<7.20, Apgar at 5 minutes <7, stillbirth, neonatal death, perinatal mortality	reported, pulsed and color, site not reported, PI cut-off not reported, absence or reversal or atrial systolic velocity.
Bilardo (2004)	High risk populations. INC: Singleton, IUGR before 33 weeks (AC<5th)+/-PIH Mean maternal age not reported (Germany and Holland) (cohort)	70	Test within 24 hours of delivery	NA	NA	Adverse perinatal outcome (antenatal death, NND, major neonatal complications before discharge (ICH>grade 2, BPD)	Route, method and site not reported, PIV>/=2sd or absent or reversed diastolic flow.
Carvalho (2005)	High risk populations. INC: Live born, singleton, no chromosomal or structural anomalies, at least 26 weeks of age on	47	>26 weeks (test for analysis	NA	NA	Acidaemia: Umbilical artery pH<7.2 in the	TA, colour and pulsed wave, narrowest

	entry, all had Doppler studies within 24 hours of delivery and cord blood gasometry immediately after birth. EXC: Placental abruption before or at delivery, GA for caesarean section, Mean maternal age 28.9 years (16-44) (Brazil) (Cross-sectional, prospective, consecutive)		performed within 24 hrs of delivery)			absence of uterine contractions and <7.15 with contractions. Had to be mixed or metabolic acidosis (BE<-10 and pCO2>60mmHg in the absence of labour and BE<-11 and pCO2>65 in labour)	isthmic portion, S/A or S-A/S or PIV, thresholds determined by ROC analysis.
Cosmi (2005)	High risk populations. INC: Gestational age established before 20 weeks, normal fetal anatomy, absence of maternal pathology, delivery before 32 weeks, forward umbilical diastole, normal AFI>/=5cm, absence of pulsation in umbilical vein, forward DV flow, last Doppler within 24 hours of delivery. Mean maternal age abnormal Doppler 32 (27-39), normal Doppler 31(24-37)	145	24-30.4 weeks (test for analysis performed within 24 hours of delivery)	NA	NA	Neonatal death	Route not reported, colour and pulsed, origin from umbilical vein, absent or reversed flow at late diastole.

	(Italy and USA) (Cohort, prospective)						
Del Rio (2008)	High risk populations INC: gestational age confirmed by sonography in first trimester, absence of structural malformations or chromosomal abnormalities, estimated birth weight <10th centile, cerebroplacental ratio<5th centile, last Doppler examination performed within 48 hrs before delivery, delivery between 24 and 36 weeks gestation, singleton pregnancy. Median age normal Doppler 32 (22-40); abnormal Doppler 28 (22-37) (Spain) (Cohort, prospective)	51	Within 48hrs of delivery	NA	NA	Adverse perinatal outcome: any of stillbirth, neonatal mortality, BPD, RDS, Grade III/IV IVH, NEC, sepsis and NICU stay longer than 14 days	TA, colour and pulsed, site not reported, absent or reversed a wave
Figueras (2003)	High risk populations. INC: Singleton, no congenital abnormalities EXC: BW>10th centile. Mean age not reported (Spain) (Cohort, consecutive)	68	>26 weeks (test for analysis performed within 3 days of delivery)	NA	NA	Admission to NICU, umbilical artery pH<7.10, neonatal morbidity (IVH, HIE, retinopathy, seizures, NEC, sepsis), intubation.	Route, method and site not reported, PI>95th centile.

Figueras (2004)	High risk populations. INC: Singleton, no congenital abnormalities EXC: BW>10th centile. Median maternal age 30.34 (SD 3.25) years (Spain) (Cohort, consecutive)	108	>26 weeks (test for analysis performed within 3 days of delivery)	NA	NA	Adverse perinatal outcome [admission to NICU, umbilical artery pH<7.10, neonatal morbidity (IVH, HIE, retinopathy, seizures, NEC, sepsis), intubation].	Route, method and isthmic portion, PI>95th centile.
Gramellini (2001)	High risk populations. INC: Pregnancy dated by USS prior to 20 weeks, singleton fetus normal anatomy and karyotype, Doppler within 2 weeks of birth. EXC: chromosomal or structural anomalies Maternal age not reported (Italy) (Cohort, retrospective)	53	24-35 weeks (test for analysis performed within 2 weeks of delivery)	NA	NA	Neonatal resuscitation required, perinatal mortality.	Route not reported, colour, site not reported, S/A>5th centile.
Hofstaetter (1996)	High risk populations. INC: Women with high risk pregnancy referred to ultrasound unit for Doppler.	87	Abnormal Doppler mean	BW mean <2SD	49.4%	Apgars, umbilical artery pH, admission to	Route not reported, colour and

	Maternal age not reported (Sweden and Germany)(Cohort)		gestation 35 (27-39); normal 36 (27-40); test for analysis median interval to delivery 1 day (0-12)			NICU.	pulsed, distal smallest portion, S/A>95th centile.
Hung (2006)	High risk populations. INC: Suspected IUGR and one or more of EH, secondary hypertension, CRD, SLE, PE, eclampsia, DM. Median maternal age 31 (23-36) (Taiwan) (Case-control, retrospective)	97	20-40 weeks (test for analysis performed within 1 week of delivery)	NA	NA	Umbilical artery pH<7.12	Route not reported, colour and pulsed, isthmic portion, PIV>95th centile.
Maiz (2008)	Low risk populations INC: singleton, screening clinic for trisomy 21 Median maternal age 32 (16-49) (UK) (Cohort, prospective)	10490	11-13+6 weeks	NA	NA	Adverse perinatal outcome : miscarriage before 24 weeks, fetal death after	TA, colour and pulsed, above umbilical sinus, reversed a wave

						24 weeks, abnormal fetal karyotype, fetal defects	
Maiz (2009)	High risk population INC: diamniotic twin pregnancies 11-13 weeks, accurate gestation and determination of chorionicity. Median maternal age 33 (29-36) (UK) (Cohort, prospective)	516 dichorionic 179 monochorionic	11-13+6 weeks	NA	NA	Death of one twin	TA, colour and pulsed, above umbilical sinus, reversed a wave
Manogura (2008)	High risk populations INC: accurate assessment of gestational age before 20 weeks, singleton, normal fetal anatomy, fetal AC<5th centile, elevated umbilical artery Doppler, delivery of a live birth at 24-36+6 weeks, last Doppler within 1 week of delivery EXC: Fetal chromosomal and structural anomalies Maternal age not reported (USA, Germany, UK) (Cohort, prospective)	404	Doppler within 1 week of delivery	NA	NA	NEC	Route, method and site not reported. Abnormal ductus venosus or absent or reversed a wave
Muller	High risk populations.	33	Mean	NA	NA	Intubation, NEC,	Route not

(2002)	INC: Singletons, AREDF umbilical artery. EXC: structural and chromosomal abnormalities Mean maternal age not reported. (Germany)(Cohort, prospective)		gestational age 28.5+/-3.4 weeks. Within 24 hours of delivery			IVH.	reported, colour and pulsed, smallest distal portion, absent or reversed flow.
Ozcan (1998)	High risk populations INC: Gestational age established before 20 weeks, normal fetal anatomy, EFW<5th centile on USS between 26-32 weeks, Doppler waveform estimations within 2 weeks of delivery. Mean maternal age not reported. (USA)(Cohort)	18	Median gestation 28.2 (27-31.4) weeks. Median interval 2 days (0-14).	NA	NA	Fetal demise or NND in first 30 days, 5 min Apgar<7mins, stay in NICU>60days, IVH, PVL.	Route not reported, color and pulsed, origin of umbilical vein, absent or reversed a wave.
Schwarze (2005)	High risk populations. INC: Fetuses with suspected IUGR on USS and AREDF in umbilical artery EXC: multiple pregnancies. Maternal age not reported (Germany) (Cohort, retrospective)	74	Within 48 hours of delivery; mean gestational age at	BW<3rd centile local values.	51.5%	Stillbirths, NND, perinatal death, acidaemia, asphyxia	TA, colour, inlet, absent or reversed a wave.

			delivery 28+6 (24+1-33+5)				
Turan (2007)	High risk populations. INC: Singleton, no chromosomal or structural anomalies, elevated umbilical artery PI, delivery at viable gestational age (all delivered by pre-labour CS) EXC: Fetal infection, chorioamnionitis. Median maternal age 30 (16-41) (UK)(Cohort, prospective)	56	Median gestation age at delivery 30+6 weeks (test for analysis performed on day of delivery)	NA	NA	Umbilical artery pH<7.20.	Route not reported, pulsed and colour, site not reported, PI>2sd or absent or reversed a wave.

Hrs hour; INC inclusion; EXC exclusion; PE preeclampsia; PIH pregnancy induced hypertension; IUGR intrauterine growth restriction; BW birth weight; UK United Kingdom; USA United States of America; NA not applicable; SGA small for gestational age; USS ultrasound scan; ga gestational age; sd standard deviation, % percent; NICU neonatal intensive care unit; TA transabdominal; PI pulsatility index; RI resistance index; PIV pulsatility index for vein s; S/A ventricular/atrial systolic ratio; S-A/S ventricular – atrial systole/ventricular systole; AREDF absent reversed end diastolic flow; DV ductus vensosus; AC abdominal circumference; EFW estimated fetal weight; CS caesarean section; BE base excess; IVH intraventricular haemorrhage; NEC necrotising enterocolitis; HIE hypoxic ischaemic encephalopathy; PVL peri-ventricular leukomalacia; NND neonatal death; GA general anaesthetic; mmHg millimetres of mercury; ROC receiver operating characteristic curve; EH essential hypertension ; CRD chronic renal disease; SLE systemic lupus erythematus, AFI amniotic fluid index, BPD bronochopulmonary dysplasia, RDS respiratory distress syndrome, ROP retinopathy of prematurity.

Appendix 42: Search strategy for electronic database identification of systematic reviews of effectiveness for interventions for fetal growth restriction and compromise of fetal wellbeing.

Host: Ovid

Date of search: July 2009

Years covered by search: 1950-2009

Medline: 3228

Embase: 4172

British Nursing Index: 4

Cohrane library: 989

Web of Science: 959

1. exp Infant, Small for Gestational Age/

2. exp Fetal Growth Retardation/

3. exp Infant, Low Birth Weight/

4. exp Placental Insufficiency/

5. exp Asphyxia Neonatorum/

6. exp Fetal Hypoxia/

7. exp Fetal Distress/

8. small for gestational age.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

9. sga.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

10. small for date$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

11. small for gestation$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

12. fetal growth restriction.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

13. fetal growth retardation.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

14. fgr.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

15. intrauterine growth retardation.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

16. intrauterine growth restriction.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

17. iugr.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

18. low birth weight.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

19. low birthweight.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

20. lbw.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

21. fetal wellbeing.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

22. fetal compromise.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

23. fetal distress.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

24. fetal hypoxia.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

25. neonatal wellbeing.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

26. neonatal asphyxia.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

27. neonatal distress.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

28. birth asphyxia.mp. [mp=title, original title, abstract, name of substance word, subject heading word]

29. 6 or 4 or 1 or 3 or 7 or 2 or 5

30. 11 or 21 or 26 or 17 or 22 or 18 or 23 or 16 or 13 or 27 or 25 or 28 or 9 or 12 or 14 or 15 or 20 or 8 or 24 or 10 or 19

31. 30 or 29

32. limit 31 to animals

33. 31 not 32

Then combined with Haynes et al filters[165-167] for (prognosis or therapy) sensitive and reviews (sensitive)

Appendix 43: Quality assessment checklist for methodological quality of included systematic reviews of effectiveness.

Assessed by : _____ Date Assessed:_____ Paper No :_____

	Quality Item	**Code** 1 2 3 4
Did the review ask a clearly structured and focused question? (*utilises PICOS)*	**1**	Yes □ No □ Unclear □ NA □
Were selection criteria clearly described? *(inclusion/exclusion related to question/PICOS)*	**2**	Yes □ No □ Unclear □ NA □
Were all relevant studies identified? *(consider whether search was adequate in the sources and search strategy)*	**3**	Yes □ No □ Unclear □ NA □
Were the included studies synthesised? *(consider whether results of each study are clearly displayed, whether the pooling of results was appropriate/heterogeneity)*	**4**	Yes □ No □ Unclear □ NA □
Was the validity of the included studies assessed? *(was there quality assessment – was this planned? which tools? How many assessors?)*	**5**	Yes □ No □ Unclear □ NA □
Were there sufficient details about the individual included studies presented? *(how are the results summarised and presented? How meaningful/precise are the results?)*	**6**	Yes □ No □ Unclear □ NA □

PICOS – population, intervention, comparator, outcome, study design
Adapted from Oxman AD, Cook DJ, Guyatt GH. User's guide to the medical literature VI. How to use an overview. JAMA 1994; 272(17):1367-1371.

Appendix 44: References of included papers for systematic reviews of reviews of effectiveness for interventions for fetal growth restriction and compromise of fetal wellbeing.

Abalos E, Duley L, Steyn DW, Henderson-Smart DJ. Antihypertensive drug therapy for mild to moderate hypertension during pregnancy. Cochrane Database of Systematic Reviews (1), 2007 Article Number: CD002252 Date of Publication: 2007 2007;(1).

Anotayanonth S, Subhedar N, V, Neilson JP, Harigopal S. Betamimetics for inhibiting preterm labour. Cochrane Database of Systematic Reviews: Reviews 2004 Issue 4 John Wiley & Sons , Ltd Chichester, CD004352 2004.

Askie LM, Duley L, Henderson-Smart DJ, Stewart LA, PARIS Collaborative Group. Antiplatelet agents for prevention of pre-eclampsia: a meta-analysis of individual patient data. Lancet 2007; 369(9575):1791-1798.

Brown HC, Smith HJ. Giving women their own case notes to carry during pregnancy. Cochrane Database of Systematic Reviews: Reviews 2004 Issue 2 John Wiley & Sons, Ltd Chichester, CD002856 pub2 2004.

Charles DH, Ness AR, Campbell D, Smith GD, Whitley E, Hall MH. Folic acid supplements in pregnancy and birth outcome: re-analysis of a large randomised controlled trial and update of Cochrane review. Paediatric and Perinatal

Epidemiology 2005; 19(2):112-124.

Churchill D, Duley L. Interventionist versus expectant care for severe pre-eclampsia before term. Cochrane Database of Systematic Reviews 2002;(3):CD003106.

Crowther CA, Moore V. Magnesium maintenance therapy for preventing preterm birth after threatened preterm labour. Cochrane Database of Systematic Reviews: Reviews 1998 Issue 1 John Wiley & Sons , Ltd Chichester, CD 1998.

Crowther CA, Hiller JE, Doyle LW. Magnesium sulphate for preventing preterm birth in threatened preterm labour. Cochrane Database of Systematic Reviews: Reviews 2002 Issue 4 John Wiley & Sons , Ltd Chichester, CD0 2002.

Dodd JM, Crowther CA, Dare MR, Middleton P. Oral betamimetics for maintenance therapy after threatened preterm labour Cochrane Database of Systematic Reviews: Reviews 2006 Issue 1 John Wiley & Sons , Ltd Chichester..

Dodd JM, Flenady V, Cincotta R, Crowther CA. Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth. Cochrane Database of Systematic Reviews: Reviews 2006 Issue 1.

Doyle LW, Crowther CA, Middleton P, Marret S, Rouse D. Magnesium sulphate for women at risk of preterm birth for neuroprotection of the fetus. Cochrane Database of Systematic Reviews: Reviews 2009 Issue 1 John Wiley & Sons 2009.

Drakeley AJ, Roberts D, Alfirevic Z. Cervical stitch (cerclage) for preventing pregnancy loss in women. Cochrane Database of Systematic Reviews 2003; (1): CD003253.

Duckitt K, Thornton S. Nitric oxide donors for the treatment of preterm labour. Cochrane Database of Systematic Reviews: Reviews 2002 Issue 3 John Wiley & Sons , Ltd Chichester, CD002860 2002.

Duley L, Henderson-Smart DJ, Knight M, King JF. Antiplatelet agents for preventing pre-eclampsia and its complicationsCochrane Database of Systematic Reviews 2004;(1):CD004659.

Duley L, Henderson-Smart DJ, Meher S. Altered dietary salt for preventing pre-eclampsia, and its complications. Cochrane Database of Systematic Reviews: Reviews 2005 Issue 4 John Wiley & Sons , Ltd Chichester.

Duley L, Gülmezoglu AM, Henderson-Smart DJ. Magnesium sulphate and other anticonvulsants for women with pre-eclampsia. Cochrane Database of Systematic Reviews: Reviews 2003 Issue 2 John Wiley & Sons , Ltd Chichester.

Duley L, Williams J, Henderson-Smart DJ. Plasma volume expansion for treatment of pre-eclampsia. Cochrane Database of Systematic Reviews: Reviews 1999 Issue 4

John Wiley & Sons, Ltd Chichester, CD001805.

Duley L, Henderson-Smart DJ. Reduced salt intake compared to normal dietary salt, or high intake, in pregnancy. Cochrane Database of Systematic Reviews: Reviews 1999 Issue 3 John Wiley & Sons , Ltd Chichester, CD001687.

Flenady V, King JF. Antibiotics for prelabour rupture of membranes at or near term. Cochrane Database of Systematic Reviews: Reviews 2002 Issue 3 John Wiley & Sons , Ltd Chichester, CD001807.

Garner P, Gulmezoglu AM. Drugs for preventing malaria-related illness in pregnant women and death in the newborn. Cochrane Database of Systematic Reviews 2003;(1):CD000169.

Grant A, Glazener CM. Elective caesarean section versus expectant management for delivery of the small baby.[update of Cochrane Database Syst Rev. 2001;(1):CD000078.

Gulmezoglu AM, Crowther CA, Middleton P. Induction of labour for improving birth outcomes for women at or beyond term. Cochrane Database of Systematic Reviews 2006;(4).

Haider BA, Bhutta ZA. Multiple-micronutrient supplementation for women during pregnancy. Cochrane Database of Systematic Reviews 2006;(4):CD004905.

Hatem M, Sandall J, Devane D, Soltani H, Gates S. Midwife-led versus other models of care for childbearing women. Cochrane Database of Systematic Reviews (4), 2008 CD004667.

Hodnett ED. Support during pregnancy for women at increased risk.[update in Cochrane Database Syst Rev. 2003;(3):CD000198; PMID: 12917888]. Cochrane Database of Systematic Reviews 2000;(2):CD000198.

Hofmeyr GJ. Abdominal decompression for suspected fetal compromise/pre-eclampsia. Cochrane Database of Systematic Reviews 2000;(2):CD000004.

Hofmeyr GJ, Kulier R. Abdominal decompression in normal pregnancy. Cochrane Database of Systematic Reviews 2000;(2):CD001062.

Hofmeyr GJ, Kulier R. Operative versus conservative management for 'fetal distress' in labour. Cochrane Database of Systematic Reviews 2000;(2):CD001065.

Hofmeyr GJ, Atallah ÁN, Duley L. Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems. Cochrane Database of Systematic Reviews: Reviews 2006 Issue 3 John Wiley & Sons , Ltd Chichester,

UK DOI 2006.

Honest H, Forbes CA, Duree KH, Norman G, Duffy SB, Tsourapas A et al. Screening to prevent preterm birth: systematic reviews of accuracy and effectiveness literature with economic modelling. Health Technol Assess 2009;13(43).

Kenyon S, Boulvain M, Neilson JP. Antibiotics for preterm rupture of membranes. Cochrane Database of Systematic Reviews 2003; (2).

King JF, Flenady V, Murray L. Prophylactic antibiotics for inhibiting preterm labour with intact membranes. Cochrane Database of Systematic Reviews 2002; (4): CD000246. .

King JF, Flenady V, Cole S, Thornton S. Cyclo-oxygenase (COX) inhibitors for treating preterm labour. Cochrane Database of Systematic Reviews: Reviews 2005 Issue 2 John Wiley & Sons , Ltd Chichester, CD001992.

Kramer MS, Kakuma R. Energy and protein intake in pregnancy.[update of Cochrane Database Syst Rev. 2000;(2):CD000032; PMID: 10796092]. Cochrane Database of Systematic Reviews 2003;(4):CD000032.

Lumley J, Oliver S, Waters E. Interventions for promoting smoking cessation during pregnancy.[update in Cochrane Database Syst Rev. 2004;(4):CD001055; PMID:

15495004]. Cochrane Database of Systematic Reviews 2000;(2):CD001055.

Magee LA, Duley L. Oral beta-blockers for mild to moderate hypertension during pregnancy.[update in Cochrane Database Syst Rev. 2003;(3):CD002863; PMID: 12917933]. Cochrane Database of Systematic Reviews 2000;(4):CD002863.

Mahomed K, Bhutta Z, Middleton P. Zinc supplementation for improving pregnancy and infant outcome.[update of Cochrane Database Syst Rev. 2000;(2):CD000230; PMID: 10796187]. Cochrane Database of Systematic Reviews 2007;(2):CD000230.

Mahomed K, Gülmezoglu AM. Vitamin D supplementation in pregnancy. Cochrane Database of Systematic Reviews: Reviews 1999 Issue 1 John Wiley & Sons , Ltd Chichester, CD000228.

Makrides M, Crowther CA. Magnesium supplementation in pregnancy.[update in Cochrane Database Syst Rev. 2001;(4):CD000937; PMID: 11687087]. Cochrane Database of Systematic Reviews 2000;(2):CD000937.

Makrides M, Duley L, Olsen SF. Marine oil, and other prostaglandin precursor, supplementation for pregnancy uncomplicated by pre-eclampsia or intrauterine growth restriction. Cochrane Database of Systematic Reviews 2006; 3:CD003402.

McDonald H, Brocklehurst P, Parsons J. Antibiotics for treating bacterial vaginosis in pregnancy.[update in Cochrane Database Syst Rev. 2007;(1):CD000262; PMID: 17253447][update of Cochrane Database Syst Rev. 2003;(2):CD000262; PMID: 12804393]. Cochrane Database of Systematic Reviews 2005;(1):CD000262.

Meher S, Duley L. Exercise or other physical activity for preventing pre-eclampsia and its complications. Cochrane Database of Systematic Reviews 2006;(2):CD005942.

Meher S, Duley L. Progesterone for preventing pre-eclampsia and its complications. Cochrane Database of Systematic Reviews 2006;(4):CD006175.

Meher S, Abalos E, Carroli G. Bed rest with or without hospitalisation for hypertension during pregnancy. Cochrane Database of Systematic Reviews: Reviews 2005 Issue 4 John Wiley & Sons , Ltd Chichester.

Meher S, Duley L. Garlic for preventing pre-eclampsia and its complications. Cochrane Database of Systematic Reviews: Reviews 2006 Issue 3 John Wiley & Sons , Ltd Chichester, CD006065.

Meher S, Duley L. Nitric oxide for preventing pre-eclampsia and its complications. Cochrane Database of Systematic Reviews: Reviews 2007 Issue 2 John Wiley & Sons , Ltd Chichester, UK CD006490.

Naik GN, Crowther CA. Maintenance therapy with calcium channel blockers for preventing preterm birth after threatened preterm labour. Cochrane Database of Systematic Reviews: Reviews 2004 Issue 3 John Wiley & Sons , Ltd Chichester, 2004.

Papatsonis D, Flenady V, Cole S, Liley H. Oxytocin receptor antagonists for inhibiting preterm labour. Cochrane Database of Systematic Reviews 2005;(3).

Papatsonis D, Flenady V, Liley H. Maintenance therapy with oxytocin antagonists for inhibiting preterm birth after threatened preterm labour Cochrane Database of Systematic Reviews: Reviews 2009 Issue 1 John Wiley & Sons , Ltd Chichester, 2009.

Pena-Rosas JP, Viteri FE. Effects of routine oral iron supplementation with or without folic acid for women during pregnancy. Cochrane Database of Systematic Reviews 2006;(3).

Rahimi R, Nikfar S, Rezaie A, Abdollahi M. A meta-analysis on the efficacy and safety of combined vitamin C and E supplementation in preeclamptic women. Hypertension in Pregnancy 2009;28(4):417-434

Raynes-Greenow CH, Roberts CL, Bell JC, Peat B, Gilbert GL. Antibiotics for ureaplasma in the vagina in pregnancy. Cochrane Database of Systematic Reviews

2004;(1):CD003767.

Roberts D, Dalziel SR. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. Cochrane Database of Systematic Reviews: Reviews 2006 Issue 3 John Wiley & Sons , Ltd Chichester, UK DOI : 10 1002/14 2006.

Rumbold A, Duley L, Crowther C, Haslam R. Antioxidants for preventing pre-eclampsia.[update in Cochrane Database Syst Rev. 2008;(1):CD004227; PMID: 18254042]. Cochrane Database of Systematic Reviews 2005;(4):CD004227.

Rumbold A, Crowther CA. Vitamin C supplementation in pregnancy. Cochrane Database of Systematic Reviews 2005;(2):CD004072.

Rumbold A, Crowther CA. Vitamin E supplementation in pregnancy. Cochrane Database of Systematic Reviews: Reviews 2005 Issue 2 John Wiley & Sons , Ltd Chichester, CD004069.

Say L, Gulmezoglu AM, Hofmeyr GJ. Maternal oxygen administration for suspected impaired fetal growth.[update of Cochrane Database Syst Rev. 2000;(2):CD000137; PMID: 10796154]. Cochrane Database of Systematic Reviews 2003;(1):CD000137.

Say L, Gülmezoglu AM, Hofmeyr GJ. Bed rest in hospital for suspected impaired fetal growth. Cochrane Database of Systematic Reviews: Reviews 1996 Issue 1 John Wiley & Sons , Ltd Chichester, CD000034.

Say L, Gülmezoglu AM, Hofmeyr GJ. Betamimetics for suspected impaired fetal growth. Cochrane Database of Systematic Reviews: Reviews 2001 Issue 4 John Wiley & Sons , Ltd Chichester, CD000036.

Say L, Gülmezoglu AM, Hofmeyr GJ. Calcium channel blockers for potential impaired fetal growth. Cochrane Database of Systematic Reviews: Reviews 1996 Issue 1 John Wiley & Sons , Ltd Chichester, CD000049.

Say L, Gülmezoglu AM, Hofmeyr GJ. Hormones for suspected impaired fetal growth. Cochrane Database of Systematic Reviews: Reviews 2003 Issue 1 John Wiley & Sons , Ltd Chichester, CD000109.

Say L, Gülmezoglu AM, Hofmeyr GJ. Maternal nutrient supplementation for suspected impaired fetal growth. Cochrane Database of Systematic Reviews: Reviews 2003 Issue 1 John Wiley & Sons , Ltd Chichester, CD000148.

Say L, Gülmezoglu AM, Hofmeyr GJ. Plasma volume expansion for suspected impaired fetal growth Cochrane Database of Systematic Reviews: Reviews 1996 Issue 4 John Wiley & Sons , Ltd Chichester, CD000167.

Say L, Gülmezoglu AM, Hofmeyr GJ. Transcutaneous electrostimulation for suspected placental insufficiency (diagnosed by Doppler studies). Cochrane Database of Systematic Reviews: Reviews 1996 Issue 1 John Wiley & Sons , Ltd Chichester, CD00007.

Shah PS, Ohlsson A, Knowledge Synthesis Group on Determinants of Low Birth Weight and Preterm Births. Effects of prenatal multimicronutrient supplementation on pregnancy outcomes: a meta-analysis. CMAJ 2009; 180(12):E99-108.

Smaill F. Antibiotics for asymptomatic bacteriuria in pregnancy.[update in Cochrane Database Syst Rev. 2007;(2):CD000490; PMID: 17443502][update of Cochrane Database Syst Rev. 2000;(2):CD000490; PMID: 10796206]. Cochrane Database of Systematic Reviews 2001;(2):CD000490.

Su LL, Samuel M, Chong YS. Progestational agents for treating threatened or established preterm labour. Cochrane Database of Systematic Reviews: Protocols 2007 Issue 4 John Wiley & Sons , Ltd Chichester, CD006770.

Thaver D, Saeed MA, Bhutta ZA. Pyridoxine (vitamin B6) supplementation in pregnancy. Cochrane Database of Systematic Reviews (2), 2006 Article Number: CD000179 Date of Publication: 2006 2006;(2).

Thinkhamrop J, Hofmeyr GJ, Adetoro O, Lumbiganon P. Prophylactic antibiotic administration in pregnancy to prevent infectious morbidity and mortality. Cochrane Database of Systematic Reviews 2002;(4):CD002250.

Whitworth M, Quenby S. Prophylactic oral betamimetics for preventing preterm labour in singleton pregnancies. Whitworth Melissa , Quenby Siobhan Prophylactic oral betamimetics for preventing preterm labour in singleton pregnancies Cochrane Database of Systematic Reviews: Reviews 2008 Issue 1 John Wiley & Sons , Ltd Chichester,CD006395.

Appendix 45: Study characteristics of included studies in review of systematic reviews of effectiveness of interventions for prevention o

fetal growth restriction and compromise of wellbeing.

Author and year	Population	Intervention	Comparator	Outcomes	No of inc RCTs
Abalos 2007	Pregnant women with mild to moderate hypertension	Antihypertensives	Placebo or none	Death, SGA, PTL, Apgars, NICU admission, RDS, impaired long term growth and development in infancy and childhood and maternal outcomes	46
Anotayanonth 2004	Pregnant women in spontaneous preterm labour	Betamimetics	Placebo or none	Maternal and perinatal	17
Askie 2007	Women at risk of PE, GH, IUGR based on previous pregnancy history, pre-existing medical condition or obstetric risk factors in current pregnancy	One or more antiplatelet agents	Placebo or none	PE, IUD, death before discharge, PTL, SGA, maternal death, APH, abruption, maternal morbidity, PPH, NICU, ventilation, neonatal bleeding	63
Brown 2004	Pregnant women	Carrying own case notes	Usual care	Maternal and perinatal	3
Charles 2005	Pregnant women	Folic acid	Placebo	Birth weight, PTL, APH, PE, stillbirths and neonatal deaths	6
Churchill 2002	Women with early onset	Early elective delivery	Expectant management	Maternal and perinatal	2

	pre-eclmapsia				
Crowther 1998	Pregnant women with at least one episode of threatened preterm labour that settled without delivery	Magnesium maintenance therapy any route	Placebo or no treatment	Preterm birth, perinatal mortality, neurological disability	3
Crowther 2002	Women in threatened PTL	Magnesium sulphate IV or oral	Placebo or no treatment	Preterm birth, IVH or PVL, death, apgars	2 3
Dodd 2006	Pregnant women at risk of preterm birth	Progesterone any route	Placebo	Perinatal mortality, preterm birth, neurodevelopmental, birth weight, apgar	1 1
Dodd 2006	Women with at least one episode of threatened preterm labour	Oral betamimetics	Placebo or none	Maternal and fetal	1 1
Doyle 2009	Women at risk of preterm birth	Magnesium sulphate IV, IM or oral	Placebo or none	Perinatal mortality, neurological, IVH, apgar	5
Drakeley 2003	Women with confirmed or suspected cervical incomPEence or women who present as an emergency with potential cervical incomPEence	Cervical cerclage	No intervention	Maternal and perinatal	6
Duckitt 2002	Pregnant women assessed as being in preterm labour and suitable for tocolysis	Nitric oxide donors	Placebo or none	Maternal and perinatal	5

Duley 2004	Women at risk of PE	Antiplatelets	Placebo or none	Death, PE, bleeding, maternal morbidity, perinatal mortality and morbidity	59
Duley 2005	Women who had normal or high blood pressure without proteinuria in pregnancy	Altered dietary salt intake	Normal salt intake	PE, death, morbidity, APH, abruption, side effects, PTL, SGA, Apgars.	2
Duley 2003	Women with PE	Anticonvulsants	Placebo or none	Maternal and neonatal mortality and morbidity	6
Duley 1999	Women with hypertension in pregnancy, with or without proteinuria	Plasma volume expansion	No expansion	Maternal and neonatal mortality and morbidity	3
Duley 1999	Pregnant women	Dietary advice to alter salt intake	No alteration	Maternal and neonatal mortality and morbidity	2
Flenady 2002	Women with PROM > 36 weeks	Antibiotics	Placebo or none	Fetal and maternal mortality and morbidity	2
Garner 2003	Pregnant women living in endemic malaria areas	Antimalarial drugs	None	Fetal and maternal mortality and morbidity	17
Grant 2001	Women at high risk of delivering a small or immature baby	Elective caesarean section	Expectant management	Fetal and maternal mortality and morbidity	6
Gulmezoglu 2006	Pregnant women at or beyond term	Induction of labour	Expectant management	Maternal and perinatal mortality and morbidity	19
Haider 2006	Pregnant women	Multiple micronutrients (three	Placebo or none	PTL, SGA, LBW, PROM, PE, miscarriage, perinatal mortality	9

		or more)		and morbidity.	
Hatem 2008	Pregnant women low and mixed risk	Midwifery led models of care	Other models	Fetal and maternal mortality and morbidity	1 1
Hodnett 2003	Pregnant women at risk of preterm labour or IUGR	Standardised or individualised programs of additional social support	Routine care	FGR, neonatal morbidity and mortality	1 8
Hofmeyr 1996	Healthy pregnant women	Abdominal decompression antenatally or during labour	None or dummy decompression	PE, FGR, perinatal morbidity and mortality	3
Hofmeyr 1996	Women with PE, fetal compromise	Antenatal abdominal decompression	None or dummy decompression	Perinatal morbidity and mortality	3
Hofmeyr 2000	Pregnant women	Calcium	Placebo	Perinatal morbidity and mortality plus long term outcomes	1 2
Hofmeyr 2006	Pregnant women with fetal distress	Operative management	Conservative management	Maternal and perinatal	1
Honest 2009	Asymptomatic low risk women with singleton gestation and low-risk women symptomatic for threatened preterm labour with singleton pregnancy	Home uterine activity monitoring	None	Maternal and perinatal	3
		Periodontal treatment	None	Maternal and perinatal	1

Kenyon 2003	Women with preterm (<37 weeks) rupture of membranes	Any antibiotic	placebo	Maternal and fetal outcomes	2 2
King 2002	Women assessed as being in PTL	Cyclo-oxygenase inhibitors	Placebo or none	PTL, gestational age at delivery, birth weight	1 3
King 2005	Women in preterm labour with intact membranes	Antibiotics	Placebo or none	Maternal, perinatal or paediatric benefit	1 1
Kramer 2003	Pregnant women	Advice to increase dietary energy and protein intakes, energy and or protein supplementation or low energy diet	Usual diet	Pregnancy outcome	2 3
Lumley 2004	Pregnant smokers	Smoking cessation	No intervention	PTL, LBW, perinatal morbidity and mortality	7 2
Magee 2003	Women with mild to moderate hypertension	Oral beta-blockers	Placebo or none	Maternal, perinatal mortality or morbidity	2 9
Mahomed 2007	Pregnant women	Zinc	No treatment	Maternal, perinatal mortality or morbidity	1 7
Mahomed 1999	Pregnant women at risk of vitamin D deficiency	Vitamin D	No treatment	LBW, perinatal mortality	2
Makrides 2001	Normal or high risk pregnancies	Oral magnesium prior to 25th week	No treatment	Neonatal mortality, maternal morbidity	7

Makrides 2006	Pregnant women	Marine oil or other prostaglandin precursors	Placebo or none	PE, PTL, LBW	6
McDonald 2007	Pregnant women with bacterial vaginosis	Antibiotic treatment	Placebo or none	PTL, LBW	15
Meher 2006	Women at risk of PE	Exercise or increased physical activity	Maintenance or normal activity	Maternal and perinatal	2
Meher 2006	Pregnant women with normal or high blood pressure	Progestogen	None or placebo	Maternal and perinatal	2
Meher 2005	Women with hypertension	Bed rest	Normal activity	Maternal and perinatal	4
Meher 2006	Pregnant women	Garlic	Placebo or none	Maternal, perinatal mortality or morbidity	1
Meher 2007	Pregnant women	Nitric oxide	Placebo or none	Maternal, perinatal mortality or morbidity	6
Naik 2004	Pregnant women after threatened PTL	Maintenance with calcium channel blockers	No treatment	Maternal, perinatal mortality or morbidity	1
Papatsonis 2005	Women in PTL	Oxytocin receptor antagonists	Placebo or no treatment	Perinatal mortality, neonatal morbidity	2
Papatsonis 2009	Pregnant women with at least one episode of threatened preterm labour that settled without delivery	Oxytocin antagonists administered as maintenance therapy	Placebo or none	PTL, perinatal or neonatal outcome	1

Pena-Rosas 2006	Pregnant women	Iron and iron plus folic acid	Placebo or none	LBW	49
Rahimi 2009	Women at risk of PE	Vitamin C and vitamin E	Placebo	Gestational hypertension, PE, PTL, SGA and LBW	7
Raynes-Greenow 2004	Pregnant women with ureaplasma in the vagina	Antibiotics	Placebo or none	Perinatal mortality, neonatal morbidity	1
Roberts 2006	Pregnant women expected to deliver preterm	Steroids	Placebo or none	Maternal and perinatal	21
Rumbold 2008	Pregnant women any risk	Antioxidants	Placebo or none	Maternal and perinatal	10
Rumbold 2005	Pregnant women	Vitamin c	Placebo or none	Maternal and perinatal	7
Rumbold 2005	Pregnant women	Vitamin e	Placebo or none	Maternal and perinatal	4
Say 2003	Suspected impaired fetal growth	Maternal oxygen therapy	Placebo or none	Fetal growth, perinatal mortality, neonatal morbidity, adverse effects	3
Say 1996	Suspected impaired fetal growth	Bed rest in hospital	Ambulatory management	Fetal and neonatal outcome	1
Say 2001	Suspected impaired fetal growth	Betamimetic	Placebo or none	Perinatal mortality, neonatal morbidity	2
Say 1996	Suspected impaired fetal growth	Calcium channel blockers	Placebo or none	Neonatal morbidity and mortality	1
Say 2003	Suspected impaired fetal growth	Hormones	Placebo or none	Perinatal death, neonatal morbidity, fetal growth, adverse effects	0
Say 2003	Suspected impaired fetal	Nutrient	Placebo or none	Fetal growth, perinatal mortality,	4

	growth	administration		neonatal morbidity, adverse effects	
Say 1996	Suspected impaired fetal growth	Plasma volume expansion	None	Fetal growth, pregnancy duration, neonatal condition, maternal complications	0
Say 1996	Suspected impaired fetal growth or placental insufficiency	Transcutaneous electrostiumlation	Dummy or no treatment	Fetal growth, perinatal mortality, neonatal morbidity, adverse effects	0
Shah 2009	Pregnant women	Multimicronutrients	Placebo	LBW<2500g, SGA,	13
Smaill 2007	Pregnant women asymptomatic bacteriuria	Any antibiotics	None	LBW	14
Su 2007	Preterm labour	Progestational agents	Placebo or none	LBW	4
Thaver 2006	Pregnant women	Vitamin B6	None	LBW	5
Thinkamrop 2002	Women in second or third trimester before pregnancy or delivery	Prophylactic antibiotics	Placebo or none	Maternal and perinatal	6
Whitworth 2008	Pregnant women at high risk of PTL	Oral betamimetics	Placebo	Perinatal mortality, neonatal morbidity	1

RCT randomised controlled trial; PE pre-eclampsia; GH gestational hypertension; IUGR intra-uterine growth restriction; PTL preterm labour; SGA small for gestational age; NICU neonatal intensive care unit; RDS respiratory distress syndrome; IUD intra-uterine death; APH ante-partum haemorrhage; PPH post-partum haemorrhage; IV intravenous; PVL periventricular leukomalacia; LBW low birth weight; IVH intraventricular haemorrhage; IM intramuscular; PROM preterm rupture of membranes; FGR fetal growth restriction

Appendix 46: Table of effectiveness data for outcomes relating to fetal growth restriction

Intervention	Number of included trials	Number of participants	Population	Outcome	Relative risk	95% confidence intervals	Z p value
Abdominal decompression							
Abdominal decompression	1	253	Normal pregnancy	LBW	0.69	0.27-1.77	0.44
	2	304	Suspected FGR and/or PE	LBW	0.5	0.40-0.63	<0.00 001
Antibiotics							
Inhibiting preterm labour with intact membranes	5	6628	Suspected preterm labour with intact membranes 20-36 weeks	BW<2500g	1.04	0.95-1.13	0.38
Preterm rupture of membranes	2	4876	Preterm rupture of membranes	BW<2500g	1	0.96-1.04	0.96
Bacterial vaginosis	4	3151	General population	LBW	1	0.80-1.24	0.99
	1	80	High risk	LBW	0.41	0.17-0.95	0.037
Ureaplasma	1	825	Women with ureaplasma	BW<2500g	0.7	0.46-1.07	0.1
Prophylactic antibiotics to prevent infectious morbidity and mortality	2	555	Unselected	LBW	0.83	0.30-2.32	0.87
	1	229	Unselected	SGA	1.29	0.45-3.77	0.65
	1	253	High risk	LBW	0.48	0.27-0.84	0.01
Asymptomatic bacteriuria	7	1502	Asymptomatic bacteriuria	BW<2500g	0.66	0.49-0.89	0.0059
Antihypertensives							
Antihypertensive drug therapy for mild to moderate hypertension during pregnancy	19	2437	Mild to moderate hypertension	SGA	1.04	0.84-1.27	0.74
	9	1116		BW<10th	1.1	0.86-1.42	0.45

				centile			
	3	287		BW<5th centile	3.04	1.25-7.40	0.014
Oral beta-blockers for mild to moderate hypertension during pregnancy	7	485	Mild to moderate hypertension	SGA	1.36	1.02-1.82	0.035
Antioxidants							
Antioxidants (vitamin C and E)	3	3582	At risk of PE	LBW	1.13	1.00-1.27	
	5	5621		SGA	1.04	0.94-1.15	
Antioxidants for preventing pre-eclampsia	5	5271	At risk of PE	BW<10th centile	0.83	0.62-1.11	0.21
	1	2784		BW<5th centile	1.13	0.98-1.32	0.1
	1	1853		BW<3rd centile	0.64	0.38-1.08	0.092
Antiplatelets							
Antiplatelets agents for preventing pre-eclampsia and its complications	36	23638	All pregnant women	SGA	0.90	0.83-0.98	0.02
	23	19399	At moderate risk of PE	SGA	0.91	0.83-0.99	0.04
	13	4239	At high risk of PE	SGA	0.89	0.74-1.08	0.02
	16	8945		BW<10th centile	0.92	0.82-1.04	0.2
	5	1962		BW<5th centile	0.97	0.78-1.21	0.8
	8	13002		BW<3rd centile	0.92	0.81-1.06	0.2
	6	7512		BW<2500g	0.93	0.83-1.05	0.2
Antiplatelet agents for prevention of pre-eclampsia: a meta-analysis of individual patient data	20	21426	At risk of PE	SGA	0.9	0.81-1.01	
Bed rest							
Bed rest with or without hospitalisation for hypertension during pregnancy	1	218	At risk of PE	SGA	0.98	0.51-1.91	0.96

Betamimetics							
Prophylactic oral betamimetics for preventing preterm labour in singleton pregnancies	1	64	At risk of PTL	BW<2500g	1.60	0.53-4.89	0.43
Betamimetics for suspected impaired fetal growth	1	98	Suspected impaired fetal growth	LBW	1.17	0.75-1.83	0.49
Calcium							
Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems	8	14359	Pregnant women	BW<2500g	0.84	0.68-1.03	0.097
	3	13091		SGA	1.1	0.88-1.37	0.39
Calcium channel blockers							
Maintenance therapy with calcium channel blockers for preventing preterm birth after threatened preterm labour	1	74	Pregnant women after threatened PTL	SGA	1.5	0.27-8.46	0.65
Delivery							
Interventionist versus expectant care for severe pre-eclampsia before term	1	95	Women with severe PE before term	SGA	0.36	0.14-0.90	0.029
Energy and protein intake							
Nutritional advice during pregnancy	1	404	Pregnant women	SGA	0.97	0.45-2.11	0.94
Balanced protein/energy supplementation	6	3396	Pregnant women	SGA	0.68	0.56-0.84	0.00026
High protein supplementation	1	505	Pregnant women	SGA	1.58	1.03-2.41	0.036
Isocaloric balanced protein supplementation	1	782	Pregnant women	SGA	1.35	1.12-1.61	0.0013
Exercise							
Exercise or other physical activity for preventing pre-eclampsia and its complications	1	16	At risk of PE	SGA	3	0.14-64.26	0.48
Fish oils							

Marine oil, and other prostaglandin precursor, supplementation for pregnancy uncomplicated by pre-eclampsia or intrauterine growth restriction	1	1374	All pregnant women	SGA	1.13	0.96-1.34	0.15
	1	1111	Low/moderate risk	SGA	1.12	0.93-1.35	0.23
	1	263	High risk	SGA	1.17	0.81-1.69	0.4
	5	2302	All pregnant women	BW<2500g	1	0.88-1.12	0.94
	2	1413	Low/moderate risk	BW<2500g	0.99	0.87-1.13	0.91
	3	789	High risk	BW<2500g	1.03	0.80-1.33	0.8
Folic acid and iron							
Folic acid supplementation	6	NA	NA	LBW	0.81	0.63-1.04	0.11
Effects and safety of preventive oral iron or iron and folic acid supplementation for women during pregnancy	9	6275	Pregnant women with iron deficiency anaemia	BW<2500g	0.79	0.61-1.03	0.08
	5	2687		VLBW <1500g	0.73	0.31-1.74	0.48
	4	2511		BW<10th centile	0.87	0.58-1.30	0.48
Home uterine monitoring							
Home uterine monitoring	1	133	Asymptomatic women at risk of preterm labour	BW<2500g	1.11	0.56-2.18	
	1	279		BW<2500g	0.47	0.28-0.78	
	1	133		BW<1500g	0.69	0.20-2.33	
Magnesium supplementation							
Orally administered magnesium prior to 25 weeks	4	1954	High and low risk women	BW<2500g	0.67	0.46-0.96	0.031
	1	568		BW<1500g	0.52	0.13-2.07	0.35
	3	1741		BW<10th centile	0.7	0.53-0.93	0.014
Malaria							

Drugs for preventing malaria in pregnant women	2	1438	Women of all parity	LBW	1.06	0.83-1.34	0.66
Midwifery-led care							
Midwife-led versus other models of care for childbearing women	5	8009	Pregnant women	BW<2500g	0.99	0.83-1.17	0.87
Multiple micronutrient supplementation							
Multiple micronutrient supplementation for women during pregnancy	2	2826	Pregnant women	SGA	0.92	0.86-0.99	0.036
	5	5110	Pregnant women	LBW	0.83	0.76-0.91	0.000091
Prenatal multimicronutrient supplementation	4	6097	Pregnant women	BW<2500g	0.81	0.73-0.91	
	3	5140	Pregnant women	SGA<10th centile or 2sd	0.85	0.71-1.02	
Nitric oxide donors							
Nitric oxide for preventing pre-eclampsia and its complications	2	108	Pregnant women	SGA	0.78	0.36-1.70	0.62
Nutrient supplementation							
Maternal nutrient supplementation for suspected impaired fetal growth (calf blood extract)	1	31	Women with suspected impaired fetal growth	BW<5th centile	0.54	0.20-1.47	0.22
Periodontal care							
Periodontal therapy to prevent preterm birth	1	351	Pregnant women	BW<2500g	0.16	0.02-1.33	
Plasma volume expansion							
Plasma volume expansion for treatment of pre-eclampsia	1	10	Women with hypertension during pregnancy	BW<2500g	1.57	0.77-3.22	0.22
Prenatal care							

Support during pregnancy for women at increased risk of low birth weight babies	13	10235	Pregnant women at risk of preterm or growth restricted babies	BW<2500g	0.98	0.89-1.08	0.69
	3	2428		BW<1500g	0.72	0.47-1.09	0.12
	2	3523		SGA	1.05	0.88-1.26	0.58
Progesterone							
Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth	2	501	Pregnant women at risk of preterm birth	BW<2500g	0.64	0.49-0.83	0.00068
Progesterone for preventing pre-eclampsia and its complications	1	168	Pregnant women with normal blood pressure or high blood pressure without proteinuria	SGA	0.83	0.19-3.57	0.8
Salt							
Altered dietary salt for preventing pre-eclampsia and its complications (low versus normal salt intake)	1	242	Pregnant women with normal blood pressure or high blood pressure without proteinuria	SGA	1.5	0.73-3.07	0.27
Reduced salt intake compared to normal dietary salt, or high intake, in pregnancy (low versus normal salt intake)	1	361	Normal pregnant women	BW<2500g	0.84	0.42-1.67	0.62
Smoking cessation							
Interventions for promoting smoking cessation during pregnancy	16	9916	Pregnant women	BW<2500g	0.83	0.73-0.95	0.0079
	4	5496		BW<1500g	1.16	0.69-1.96	0.57
Steroids							
Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth	3	378	Women expected to deliver preterm	SGA	0.96	0.63-1.44	

Vitamins							
Vitamin C supplementation in pregnancy	2	383	All pregnant women	BW<10th centile	0.72	0.49-1.04	0.079
Vitamin D supplementation in pregnancy	1	128	Pregnant women at risk of vitamin D deficiency	BW<2500g	0.5	0.20-1.26	0.14
Vitamin E supplementation in pregnancy	2	383	All pregnant women	BW<10th centile	0.72	0.49-1.04	0.079
Zinc							
Zinc supplementation for improving pregnancy outcome and infant outcome	5	3469	Normal pregnant women	SGA	1.04	0.96-1.13	0.3
	11	4860		LBW	1.03	0.94-1.13	0.51

NA not available

BW birth weight; LBW low birth weight; VLBW very low birth weight; SGA small for gestational age; FGR fetal growth restriction; PE pre-eclampsia; PTL preterm labour

Appendix 47: Table of effectiveness data for outcome of perinatal mortality

Intervention	Number of included trials	Number of participants	Population	Relative risk	95% confidence intervals	Z p value
Abdominal decompression						
Abdominal decompression	2	709	Normal pregnancy	2.47	0.77-7.92	0.13
	3	367	Suspected FGR and/or PE	0.39	0.22-0.71	0.0021
Antibiotics						
Inhibiting preterm labour with intact membranes	9	7208	Suspected preterm labour with intact membranes 20-36 weeks	1.22	0.88-1.70	0.24
Prelabour rupture of membranes at or near term	2	838	Spontaneous rupture of membranes >36 weeks	0.98	0.14-6.89	0.98
Preterm rupture of membranes	13	6411	Preterm rupture of membranes	0.9	0.74-1.10	0.32
Bacterial vaginosis	3	2666	General population	0.96	0.53-1.73	0.89
Prophylactic antibiotics to prevent infectious morbidity and mortality	1	229	Unselected	0.19	0.00-1.76	0.14
	1	253	High risk	0.71	0.20-2.58	0.37
Anticonvulsants						
Magnesium sulphate and other anticonvulsants for women with pre-eclampsia	2	9259	Women with PE	0.98	0.88-1.10	0.78
Antihypertensives						

Antihypertensive drug therapy for mild to moderate hypertension during pregnancy	20	2382	Mild to moderate hypertension	0.96	0.60-1.54	0.87
Oral beta-blockers for mild to moderate hypertension during pregnancy	13	1429	Mild to moderate hypertension	1.01	0.46-2.22	0.97
Antiplatelets						
Antiplatelets agents for preventing pre-eclampsia and its complications	23	28655	At moderate risk of PE	0.92	0.80-1.07	0.3
	17	4443	At high risk of PE	0.69	0.53-0.90	0.006
Antiplatelet agents for prevention of pre-eclampsia: a meta-analysis of individual patient data	23	30672	At risk of PE	0.91	0.81-1.03	
Bed rest						
Bed rest with or without hospitalisation for hypertension during pregnancy	2	145	At risk of PE	1.07	0.52-2.19	0.86
Betamimetics						
Betamimetics for inhibiting preterm labour	11	1332	Pregnant women in spontaneous PTL	0.84	0.46-1.55	0.58
Oral betamimetics for maintenance therapy after threatened preterm labour	6	681	Women with at least one episode of threatened PTL	2.41	0.86-6.74	0.093
Prophylactic oral betamimetics for preventing preterm labour in singleton pregnancies	1	64	At risk of PTL	4.74	0.50-45.00	0.18
Betamimetics for suspected impaired fetal growth	1	98	Suspected impaired fetal growth	0.24	0.01-4.96	0.36
Calcium						
Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems	10	15141	Pregnant women	0.89	0.73-1.09	0.25
Calcium channel blockers						

Calcium channel blockers for potential impaired fetal growth	1	100	Women at high risk or with suspected impaired fetal growth	0.33	0.00-3.78	0.32
Case notes						
Giving women their own case notes to carry during pregnancy	1	212	Pregnant women	1.04	0.15-7.24	0.97
Cervical cerclage						
Cervical stitich for preventing pregnancy loss in women	4	2059	Women with confirmed or suspected incomPEence	0.8	0.48-1.36	0.41
Cyclo-oxygenase inhibitors						
Cyclo-oxygenase inhibitors for treating preterm labour	3	106	Women in PTL	0.8	0.25-2.58	0.71
Delivery						
Elective caesarean section versus expectant management for delivery of the small baby	5	122	Women in labour with suspected small baby-breech	0.30	0.08-1.18	0.041
Induction of labour for improving birth outcomes for women at or beyond term	2	584	Pregnant women at or beyond term- 37-40 weeks	0.32	0.03-3.09	0.33
	10	5643	41 weeks	0.25	0.05-1.18	0.081
	2	296	42 weeks	0.41	0.06-2.73	0.36
Interventionist versus expectant care for severe pre-eclampsia before term	1	38	Women with severe PE before term	0.45	0.04-4.55	0.5
Operative versus conservative management for fetal distress in labour	1	350	Pregnant women with evidence of fetal distress	1.18	0.56-2.48	0.66
Folic acid and iron						
Folic acid supplementation	2	NA	NA	1.18	0.74-1.91	0.47

Effects and safety of preventive oral iron or iron and folic acid supplementation for women during pregnancy	3	5036	Pregnant women with iron deficiency anaemia	0.93	0.67-1.29	0.66
Magnesium sulphate						
Magnesium sulphate for preventing preterm birth in threatened preterm labour	7	727	Women though to be in preterm labour	2.82	1.20-6.62	0.017
Magnesium maintenance therapy for preventing preterm birth after threatened preterm birth	1	50	Pregnant women with at least one episode of threatened PTL	5	0.25-99.16	0.29
Magnesium sulphate for women at risk of preterm birth for neuroprotection of the fetus	5	6145	Women at risk of PTL	1.04	0.92-1.17	0.57
Malaria						
Drugs for preventing malaria in pregnant women	4	2890	Women of all parity	1.02	0.73-1.43	0.9
Midwifery-led care						
Midwife-led versus other models of care for childbearing women	9	11604	Pregnant women	1.01	0.67-1.53	0.95
Multiple micronutrient supplementation						
Multiple micronutrient supplementation for women during pregnancy	7	11956	Pregnant women	1.05	0.90-1.23	0.65
Nitric oxide donors						
Nitric oxide for preventing pre-eclampsia and its complications	2	114	Pregnant women	0.25	0.03-2.34	0.23
Nutrient supplementation						
Maternal nutrient supplementation for suspected impaired fetal growth (calf blood extract)	1	31	Women with suspected impaired fetal growth	0.19	0.01-3.63	0.27
Oxygen						

Maternal oxygen administration for suspected impaired fetal growth	3	94	Women with suspected impaired fetal growth	0.5	0.32-0.81	0.0041
Oxytocin receptor antagonists						
Oxytocin receptor antagonists for inhibiting preterm labour	1	583	Women in PTL	2.25	0.79-6.40	0.13
Maintenance therapy with oxytocin antagonists for inhibiting preterm birth after threatened preterm labour	1	512	Pregnant women with at least one episode of threatened PTL	0.77	0.21-2.83	0.69
Plasma volume expansion						
Plasma volume expansion for treatment of pre-eclampsia	1	32	Women with hypertension during pregnancy	3.5	0.18-67.45	0.41
Prenatal care						
Support during pregnancy for women at increased risk of low birth weight babies	11	9507	Pregnant women at risk of preterm or growth restricted babies	1.15	0.89-1.51	0.29
Progesterone						
Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth	3	1114	Pregnant women at risk of preterm birth	0.65	0.38-1.11	0.11
Progesterone for preventing pre-eclampsia and its complications	2	296	Pregnant women with normal blood pressure or high blood pressure without proteinuria	0.72	0.21-2.51	0.61
Salt						
Altered dietary salt for preventing pre-eclampsia and its complications (low versus normal salt intake)	2	409	Pregnant women with normal blood pressure or high blood pressure without proteinuria	1.92	0.18-21.03	0.59

Reduced salt intake compared to normal dietary salt, or high intake, in pregnancy (low versus normal salt intake)	2	409	Normal pregnant women	1.92	0.18-21.03	0.59
Smoking cessation						
Interventions for promoting smoking cessation during pregnancy	3	4335	Pregnant women	1.13	0.72-1.77	0.59
Steroids						
Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth	13	3627	Women expected to deliver preterm	0.77	0.67-0.89	0.00035
Vitamins						
Vitamin C supplementation in pregnancy	2	238	All pregnant women	1.16	0.61-2.18	0.65
Vitamin E supplementation in pregnancy	1	56	All pregnant women	1.29	0.67-2.48	0.45
Zinc						
Zinc supplementation for improving pregnancy outcome and infant outcome	1	1555	Normal pregnant women	1.03	0.71-1.51	0.87

NA not available

FGR fetal growth restriction; PE pre-eclampsia; PTL preterm labour; PPROM preterm prelabour rupture of membranes

Appendix 48: Search strategy for cost data related to outcome for economic evaluation

((("small-for-gestational-age") OR (small-for-gestational-age) OR (lbw) OR (small for gestational age) OR (sga) OR (small for date*) OR (small for gestation*) OR (fgr) OR (iugr) OR (intrauterine growth retard*) OR (intrauterine growth restrict*) OR (fetal growth retard*) OR (fetal growth restrict*) OR (growth restrict*) OR (growth retard*) OR ("Placental Insufficiency"[MeSH]) OR ("Fetal Growth Retardation"[MeSH]) OR ("Infant, Low Birth Weight"[MeSH]) OR (low birth weight) OR (low birthweight))) AND (economic* OR cost OR costs OR cost-effectiveness OR cost analysis OR resource*)

Appendix 49: Reference list

(1) http://.ncchta.org/about/whatishta.html. 2010

(2) Stevens et al. Evaluating health care technologies. Oxford Handbook of Public Health Practice. Oxford: Oxford University Press, 2006.

(3) Maulik D. Management of fetal growth restriction: an evidence based approach. Clinical Obstetrics and Gynecology 2006; 49(2):320-334.

(4) Deeks J, Morris J. Evaluating diagnostic tests. Baillieres Clinical Obsetrics and Gynaecology 1996;(10):613-630.

(5) Honest H, Khan KS. Reporting of measures of accuracy in systematic reviews of diagnostic literature. BMC Health Services Research 2002; 2:4.

(6) Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. Methods for the economic evaluation of healthcare programmes. 3rd edition ed. Oxford: Oxford University Pres, 2005.

(7) Hay WW, Jr., Catz CS, Grave GD, Yaffe SJ. Workshop summary: fetal growth: its regulation and disorders. Pediatrics 1997; 99(4):585-591.

(8) McIntire DD, Bloom SL, Casey BM, Leveno KJ. Birth weight in relation to morbidity and mortality among newborn . N Engl J Med 1999; 340(16):1234-1238.

(9) Taylor DJ, Howie PW. Fetal growth achievement and neurodevelopmental disability. British Journal of Obstetrics & Gynaecology 1989; 96(7):789-794.

(10) Confidential Enquiry into Maternal and Child Health (CEMACH) Perinatal Mortality 2007: United Kingdom. London: CEMACH, 2009.

(11) Rich-Edwards JW, Colditz GA, Stampfer MJ, Willett WC, Gillman MW, Hennekens CH et al. Birthweight and the risk for type 2 diabetes mellitus in adult women. Ann Intern Med 1999; 130(4 Pt 1):278-284.

(12) Barker DJ. The developmental origins of chromic adult disease. Acta Paediatr Suppl 2004; 93(446):26-33.

(13) Chang TC, Robson SC, Boys RJ, Spencer JA. Prediction of the small for gestational age infant: which ultrasonic measurement is best? Obstet Gynecol 1992; 80(6):1030-1038.

(14) Ott WJ. The diagnosis of altered fetal growth. Obstetrics & Gynecology Clinics of North America 1988; . 15(2).

(15) Wilcox AJ. Intrauterine growth retardation: beyond birthweight criteria. Early Hum Dev 1983; 8(3-4):189-193.

(16) Scott K, Usher R. Fetal malnutrition: its incidence, causes and effects. American Journal Obstetrics and Gynecology 1966; 94(7):951-963.

(17) Clausson B, Gardosi J, Francis A, Cnattingius S. Perinatal outcome in SGA births defined by customised versus population-based birthweight standards. BJOG 2001; 108(8):830-834.

(18) Hill R, Verniuad W, Deter R, Tennyson L, Rettig G, Zion T et al. The effect of intrauterine malnutrition on the term infant. A 14 year progressive study. Acta Paedatri Scand 1984; 73:482-487.

(19) Khong T, De Wolf F, Robertson W, Brosens I. Inadequate maternal vascular response to placentation in pregnancies complicated by preeclampsia and small for date gestational age . BJOG 1986; 93:1049-1059.

(20) Brooke O, Anderson H, Bland J, et al. Effects on birth weight of smoking, alcohol, caffeine, socioeconomic factors, and psychosocial stress. BMJ 1989; 298(6676):795-801.

(21) Coomarasamy A, Gee H, Robson S. The investigation and management of the small for gestational age fetus. Royal College of Obstetricians and Gynaecologists 2002; No 31.

(22) Fahey J, King T. Intrauterine asphyxia: clinical implications for providers of intrapartum care. Journal of Midwifery and Women's Health 2005; 50(6):498-506.

(23) MacLennan A. A template for defining a causal relation between acute intrapartum events and cerebral palsy: international consensus statement. BMJ 1999; 319:1054-1059.

(24) Malin GL, Morris RK, Khan KS. What is the relationship between umbilical cord pH and perinatal and long term outcome? A systematic review and meta-analysis to evaluate strength of association. BMJ 2010; 340:c1471.

(25) Papageorghiou AT, Yu CK, Cicero S, Bower S, Nicolaides KH. Second-trimester uterine artery Doppler screening in unselected populations: a review. J Matern Fetal Neonatal Med 2002; 12(2):78-88.

(26) Chien PF, Arnott N, Gordon A, Owen P, Khan KS. How useful is uterine artery Doppler flow velocimetry in the prediction of pre-eclampsia, intrauterine growth retardation and perinatal death? An overview. BJOG 2000; 107(2):196-208.

(27) Farrell T, Chien PF, Gordon A. Intrapartum umbilical artery Doppler velocimetry as a predictor of adverse perinatal outcome: a systematic review. Br J Obstet Gynaecol 1999; 106(8):783-792.

(28) Sherer DM. Prenatal ultrasonographic assessment of the middle cerebral artery: a review. Obstet Gynecol Surv 1997; 52(7):444-455.

(29) Sherer DM, Fromberg RA, Divon MY. Prenatal ultrasonographic assessment of the ductus venosus: a review. Obstet Gynecol 1996; 88(4 Pt 1):626-632.

(30) Sherer DM, Divon MY. Prenatal ultrasonographic assessment of the ductus arteriosus: A review. Obstetrics & Gynecology 1996; 87(4):630-637.

(31) Cox S. Intrauterine Growth Restriction. ACOG Practice Bulletin 2000; Number 12.

(32) Khan KS, Khan K, Nwosu C, Arnott N, Chien PF. Misleading authors' inferences in obstetric diagnostic test literature. American Journal Obstetrics and Gynecology 1999; 181(1):112-115.

(33) Cochrane Methods Working Group on Systematic Reviews of Screening and Diagnostic Tests: Recommended methods. 1996.

(34) Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. [Review] [19 refs]. Annals of Internal Medicine 2003; 138(1):40-44.

(35) Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Medical Research Methodology 2003; 3:25.

(36) Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Medical Research Methodology 2006; 6:9.

(37) 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. [Review] [46 refs]. Journal of Clinical Epidemiology 2005; 58(10):982-990.

(38) NICE. Antenatal Care: routine care for the healthy pregnant woman. NICE Clinical Guidance No 62 . 2008. London, NICE.

(39) Gulmezoglu M, Villar J. Effectiveness of interventions to prevent or treat impaired fetal growth. Obstet and Gynecol Surv 1997; 52(2):139-149.

(40) -The-Editorial-Team. Cochrane Pregnancy and Childbirth Group. 2010.

(41) Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 ed. The Cochrane Collaboration, 2008.

(42) Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000; 283(15):2008-2012.

(43) Khan KS, Dinnes J, Kleijnen J. Systematic reviews to evaluate diagnostic tests. European Journal of Obstetrics, Gynecology, & Reproductive Biology 2001; 95(1):6-11.

(44) Deeks J, Khan KS, Song F, Popay J, Nixon J, Kleijnen J. Data synthesis. In: Khan KS, ter Riet G, Glandville J, Sowden A, Kleijnen J, editors. Undertaking Systematic Reviews of Research on Effectiveness: CDR's Guidance for

carrying out or commissioning reviews. York, England: University of York, 2001.

(45) Deville W, Buntinx F. Guidelines for conducting systematic reviews of studies evaluating the accuracy of diagnostic studies. In: Knottnerus JA, editor. The evidence base of clinical diagnosis. London: BMJ Books, 2002: 145-165.

(46) Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC et al. Guidelines for meta-analyses evaluating diagnostic tests.[see comment]. Annals of Internal Medicine 1994; 120(8):667-676.

(47) Morris R, Khan KS, Coomarasamy A, Robson S, Kleijnen J. The value of predicting restriction of fetal growth and compromise of its wellbeing: Systematic quantitative overviews (meta-analysis) of test accuracy literature. BMC Pregnancy and Childbirth 7, 3. 2007.

(48) Khan KS. Systematic reviews of diagnostic tests: a guide to methods and application. [Review] [20 refs]. Best Practice & Research in Clinical Obstetrics & Gynaecology 2005; 19(1):37-46.

(49) Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Compoiste outcome measures in randomized trials - Greater precision but with greater uncertainty? JAMA 2003; 289(19):2554-2559.

(50) Freemantle N, Calvert M. Composite and surrogate outcomes in randomised trials. BMJ 2007; 334:756-757.

(51) Rutjes AW, Reitsma J, Di N, Smidt N, van Rijn J, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. Canadian Medical Association Journal 2006; 174(4):469-476.

(52) Bachmann LM, Coray R, Estermann P, ter RG. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. J Am Med Inform Assoc 2002; 9(6):653-658.

(53) Haynes R, Wilczynski N, McKibbon KA, Walker C, Sinclair J. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. J Am Med Inform Assoc 1994; 1(6):447-458.

(54) Meads C, Cnossen J, Meher S, Juarez-Garcia A, ter Riet G, Duley L et al. Methods of prediction and prevention of pre-eclampsia: systematic reviews of accuracy and effectiveness literature with economic modelling. Health Technology Assessment 2008; 12(6).

(55) Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med Res.Methodol. 5, 19. 2005.

(56) Sankey s, Weistfiels L, Fine M, Kapoor W. An assessment of the use of the continuity correction for sparse data in meta analysis. Commun Stat Simulation Computation 1996; 25:1031-1056.

(57) Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-Disc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 2006; 6:31.

(58) Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA 1994; 271(9):703-707.

(59) Deeks J, Altman D. Diagnostic Tests 4: Likelihood ratios. BMJ 2004; 2004(329):7458-168.

(60) Van Houwelingen H, Arends L, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med 2002; 21(4):589-624.

(61) Van Houwelingen H, Zwinderman K, Stijnen T. A bivariate approach to meta-analysis. Stat Med 1993; 12(24):2273-2284.

(62) Song F, Khan KS, Dinnes J, Sutton AJ. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. International Journal of Epidemiology 2002; 31(1):88-95.

(63) Deeks JJ, Macaskill P IL. The performance of tests of publication bias and other sample size effects in sytematic reviews of diagnostic test accuracy was assessed. Journal of Clinical Epidemiology 58[9], 882-893. 2005.

(64) METABIAS: Stata module to test for small-study effects in meta-analysis. Boston College Department of Economics, 2009.

(65) Khan K, Chien P. Seizure prophylaxis in hypertensive pregnancies: a framework for making clinical decisions. British Journal of Obstetrics and Gynaecology 1997; 104(10):1173-1179.

(66) Askie L, Duley L, Henderson-Smart D, Stewart L. Antiplatelet agents for preventing pre-eclampsia: a meta-analysis of individual patient data. Lancet 2007; 369(9575):1791-1798.

(67) Wapner R, Thom E, Simpson J, Pergament E, Silver E, Filkins R. First trimester screening for trisomies 21 and 18. N Engl J Med 2003; 349(15):1405-1413.

(68) Yaron Y, Cherry M, Kramer RL, O'Brien JE, Hallak M, Johnson MP et al. Second-trimester maternal serum marker screening: maternal serum alpha-

fetoprotein, beta-human chorionic gonadotropin, estriol, and their various combinations as predictors of pregnancy outcome. Am J Obstet Gynecol 1999; 181(4):968-974.

(69) Aquilina J, Maplethorpe R, Ellis P, Harrington K. Correlation between second trimester maternal serum inhibin-A and human chorionic gonadotrophin for the prediction of pre-eclampsia. Placenta 2000; 21(5-6):487-492.

(70) Ong CY, Liao AW, Spencer K, Munim S, Nicolaides KH. First trimester maternal serum free beta human chorionic gonadotrophin and pregnancy associated plasma protein A as predictors of pregnancy complications. BJOG 2000; 107(10):1265-1270.

(71) Morssink L, Komman L, Hallahan T, Kloosterman M, Beckhuis J, De Wolf B et al. Maternal serum levels of free beta-hCG and PAPP-A in the first trimester of pregnancy are not associated with subsequent fetal growth retardation or preterm delivery. Prenatal Diagnosis 1998; 18(2):147-152.

(72) Conde-Agudelo A, Villar J, Lindheimer M. World health organization systematic review of screening tests for preeclampsia. Obstet Gynecol 2004; 104(6):1367-1391.

(73) Cnossen JC, Mol BW, van der Post JAM, Bossuyt PM, ter RG. World health organization systematic review of screening tests for pre-eclampsia. Obstet.Gynecol. 105, 1151-1152. 2005.

(74) Zwinderman AH, Bossuyt PM. We should not pool diagnostic likelihood ratios in systematic reviews. Stat Med.(in press).

(75) Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282(11):1061-1066.

(76) Waller DK, Lustig LS, Cunningham GC, Feuchtbaum LB, Hook EB. The association between maternal serum alpha-fetoprotein and preterm birth, small for gestational age , preeclampsia, and placental complications. Obstet Gynecol 1996; 88(5):816-822.

(77) Wenstrom KD, Owen J, Davis RO, Brumfield CG. Prognostic significance of unexplained elevated amniotic fluid alpha-fetoprotein. Obstet Gynecol 1996; 87(2):213-216.

(78) Milunsky A, Jick SS, Bruell CL, MacLaughlin DS, Tsung YK, Jick H et al. Predictive values, relative risks, and overall benefits of high and low maternal serum alpha-fetoprotein screening in singleton pregnancies: new epidemiologic data. Am J Obstet Gynecol 1989; 161(2):291-297.

(79) Gonen R, Perez R, David M, Dar H, Merksamer R, Sharf M. The association between unexplained second-trimester maternal serum hCG elevation and pregnancy complications. Obstet Gynecol 1992; 80(1):83-86.

(80) Konchak P, Bernstein I, Capeless E. Uterine artery Doppler velocimetry in the detection of adverse obstetric outcomes in women with unexplained elevated maternal serum alpha-fetoprotein levels. American Journal Obstetrics and Gynecology 1995; 173(4):1115-1119.

(81) Steel SA, Pearce JM, Chamberlain G. Doppler ultrasound of the uteroplacental circulation as a screening test for severe pre-eclampsia with intra-uterine growth retardation. Eur J Obstet Gynecol Reprod Biol 1988; 28(4):279-287.

(82) Bolte A, Dekker G. Uterine artery Doppler as screening tool for preeclampsia. In: Wildschut H, Weiner C, editors. When to screen in Obstetrics and Gynecology. Philadelphia: Saunders Elsevier, 2006: 408-419.

(83) Khan KS, Bachmann LM, ter RG. Systematic reviews with individual patient data meta-analysis to evaluate diagnostic tests. [Review] [12 refs]. European Journal of Obstetrics, Gynecology, & Reproductive Biology 2003; 108(2):121-125.

(84) Gagnon R, Van Den HM, Diagnostic Imaging Committee EaCotSoOaGoC. The use of fetal Doppler in obstetrics. Journal of Obstetrics & Gynaecology Canada: JOGC 2003; 25(7):601-614.

(85) Alfirevic Z, Neilson JP. Doppler ultrasonography in high-risk pregnancies: Systematic review with meta-analysis. American Journal of Obstetrics & Gynecology 1995; 172(5):1379-1387.

(86) Goffinet F, Paris-Llado J, Nisand I, Breart G. Umbilical artery doppler velocimetry in unselected and low risk pregnancies: a review of randomised controlled trials. British Journal of Obstetrics & Gynaecology 1997; 104(4):425-430.

(87) Bricker L, Neilson JP. Routine Doppler ultrasound in pregnancy. The Cochrane Library 2005;(Oxford):2005.

(88) Westergaard HB, Langhoff-Roos J, Lingman G, Marsal K, Kreiner S. A critical appraisal of the use of umbilical artery Doppler ultrasound in high-risk pregnancies: use of meta-analyses in evidence-based obstetrics.[see comment]. Ultrasound in Obstetrics & Gynecology 2001; 17(6):466-476.

(89) Bossuyt PM, Lijmer JG, Mol BW. Randomised comparisons of medical tests: sometimes invalid, not always efficient. Lancet 2000; 356(25):1844-1847.

(90) Neilson JP. Doppler ultrasound in high risk pregnancies. Cochrane Database of Systematic Reviews 1994; Review no: 03889.

(91) Neilson JP, Alfirevic Z. Doppler ultrasound for fetal assessment in high risk pregnancies. The Cochrane Library 2005;(Oxford):2005.

(92) Cohn H, Sacks E, Heymann M, Rudolph A. Cardiovascular responses to hypoxemia and acidemia in fetal lambs. Am J Obstet Gynecol 1974; 120:817-824.

(93) Sheldon R, Peeters L, Jones M, Makowski E, Meschia G. Redistribution of cardiac output and oxygen delivery in the hypoxemic fetal lamb. Am J Obstet Gynecol 1979; 135:1071-1077.

(94) Wladimiroff JW, van den Wijngaard J, Degani S, Noordam M, van Eyck J, Tonge H. Cerebral and umbilical blood flow velocity waveforms in normal and growth retarded pregnancies. Obstet Gynecol 1987;(69):613-616.

(95) Mari G, Deter R. Middle cerebral artery flow velocity waveforms in normal and small for gestational age fetuses. Am J Obstet Gynecol 1992; 166:1262-1270.

(96) Gembruch U, Baschat A. Demonstration of fetal coronary blood flow by color-coded and pulsed wave Doppler sonography: a possible indicator of severe compromise and impending demise in intrauterine growth retardation. Ultrasound Obstet Gynecol 1996; 7:10-5.

(97) Mari G, Uerpairojkit B, Abuhamad A, Copel J. Adrenal artery velocity waveforms in the appropriate and small for gestational age fetus. Ultrasound Obstet Gynecol 1996; 8:82-6.

(98) Dubiel M, Gudmundsson S, Gunnarsson G, Marsal K. Middle cerebral artery velocimetry as a predictor of hypoxemia in fetuses with increased resistance to blood flow in the umbilical artery. Early Human Development 47(2):177-84, 1997.

(99) Ozeren M, Dinc H, Ekmen U, Senekayli C, Aydemir V. Umbilical and middle cerebral artery Doppler indices in patients with preeclampsia. European Journal of Obstetrics, Gynecology, & Reproductive Biology 82(1):11-6, 1999.

(100) Hershkovitz R, Kingdom JC, Geary M, Rodeck CH. Fetal cerebral blood flow redistribution in late gestation: identification of compromise in small fetuses with normal umbilical artery Doppler. Ultrasound in Obstetrics & Gynecology 15(3):209-12, 2000.

(101) Mari G, Hanif E, Kruger M, Cosmi E, Santolaya-Forgas J, Treadwell M. Middle cerebral artery peak systolic velocity: a new Doppler parameter in the assessment of growth restricted fetuses. Ultrasound in Obstetrics & Gynecology 2007; 29(3):310-316.

(102) Baschat AA, Gembruch U, Weiner CP, Harman CR. Qualitative venous Doppler waveform analysis improves prediction of critical perinatal outcomes in premature growth-restricted fetuses. Ultrasound in Obstetrics & Gynecology 2003; 22(3):240-245.

(103) Baschat AA, Cosmi E, Bilardo CM, Wolf H, Berg C, Rigano S et al. Predictors of neonatal outcome in early-onset placental dysfunction.[see comment]. Obstetrics & Gynecology 2007; 109(2 Pt 1):253-261.

(104) Baschat AA, Guclu S, Kush ML, Gembruch U, Weiner CP, Harman CR. Venous Doppler in the prediction of acid-base status of growth-restricted fetuses with elevated placental blood flow resistance. Am J Obstet Gynecol 2004; 191(1):277-284.

(105) Baschat AA, Galan HL, Bhide A. Doppler and biophysical assessment in growth restricted fetuses: distribution of test results. Ultrasound in Obstetrics & Gynecology 2006; 27(1):41-47.

(106) Manogura A, Turan O, Kush ML, et al. Predictors of necrotising enterocolitis in preterm growth-restricted neonates. American Journal Obstetrics and Gynecology 2008; 198(6):638-e1-5.

(107) Turan S, Turan OM, Berg C, et al. Comupterized fetal heart rate analysis, Doppler ultrasound and biophysical profile score in the prediction of acid-base status of growth-restricted fetuses. Ultrasound in Obstetrics & Gynecology 2007; 30(5):750-756.

(108) Diemart A, Hecher K. Severely abnormal flow patterns in the ductus venosus in the presence of otherwise normal fetal Doppler parameters in two cases of severe intrauterine growth restriction. Ultrasound in Obstetrics & Gynecology 2009; 34:605-608.

(109) McCormick MC. The contribution of low birth weight to infant mortality and childhood morbidity. New England Journal of Medicine 312(2)()(pp 82-90), 1985 Date of Publication: 1985 1985;(2):82-90.

(110) Higgins JP, Thompson SG. Controlling the risk of spurious finding from meta-regression. Statistics in Medicine 2004; 23(11):1663-1682.

(111) Bachmann LM, ter Riet G, Wim EJ, Weber A, Kessels GH. Multivariable adjustments counteract spectrum and test review bias in accuracy studies. Journal of Clinical Epidemiology 2009; 62(4):357-361.

(112) McKenna H. The Delphi technique: a worthwhile approach to nursing? Journal of Advanced Nursing 1994; 19:1221-1225.

(113) Sackman H. Delphi Critique. Boston: Lexington Books, 1975.

(114) Fink A, Kosecoff J, Chassin M, Brook R. Consensus Methods: Characterisitcs and Guidelines for Use. Santa Monica, California: RAND, 1991.

(115) Nabhan AF, Abdelmoula YA. Amniotic fluid index versus single deepest vertical pocket as a screening test for preventing adverse pregnancy outcome. Cochrane Database of Systematic Reviews 2008;(3).

(116) Lalor JG, Fawole B, Alfirevic Z, Devane D. Biophysical profile for fetal assessment in high risk pregnancies. Cochrane Database of Systematic Reviews (1):CD000038, 2008 2008;(1):CD000038.

(117) Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evalaution of random effects meta-analysis. J R Statist Soc 2009; 172:137-159.

(118) Bachmann LM, Rama K, ter Riet G, Bossuyt PM. Sample sizes of studies on diagnostic accuracy: literature survey. BMJ 2006; 332:1127-1129.

(119) Irwig L, Bossuyt PM, Glasziou PP, Gatsonis C, Lijmer JG. Designing studies to ensure that estimates of test accuracu are transferrable. BMJ 2002; 324:669-671.

(120) Khan KS, Bakour SH, Bossuyt PM. Evidence-based obstetric and gynaecologic diagnosis: the STARD checklist for authors, peer-reviewers and readers of test accuracy studies. BJOG: An International Journal of Obstetrics & Gynaecology 2004; 111(7):638-640.

(121) Selman T, Khan K, Mann C. An evidence based approach to test accuracy studies in gynecologic onlcology: the STARD checklist. Gynecologic Oncology 2005; 96:575-578.

(122) Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. Consolidated Standards of Reporting Trials. JAMA. 1998;279:1489-91. JAMA 279, 1489-1491. 1998.

(123) Coppus S, van der Veen F, Bossuyt P, Mol B. Quality of reporting of test accuracy studies in reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative. Fertil Steril 2006; 86(5):1321-1329.

(124) Smidt N, Rutjes A, van der Windt D, Ostelo R, Bossuyt P, Reitsma J et al. The quality of diagnostic accuracy studies since the STARD statement: has it improved? Neurology 2006; 67(5):792-797.

(125) Zafar A, Khan G, Siddiqui M. The quality of reporting of diagnostic accuracy studies in diabetic retinopathy screening: a systematic review. Clin Experiment Opthalmol 2008; 36(6):537-542.

(126) Rama K, Poovali S, Apsingi S. Quality of reporting of orthopaedic diagnostic accuracy studies is suboptimal. Clin Orthop Relat Res 2006; 447:237-246.

(127) Gomez Saez N, Hernandez-Aguado I, Lumbreras B. Evaluation of the diagnostic research methodology in Spain after STARD publication. Med Clin (Barc) 2009; 133(8):302-310.

(128) Morris RK, Cnossen J, Langejans M, Robson S, Kleijnen J, ter RG et al. Serum screening with Down's Syndrome markers to predict pre-eclampsia and small for gestational age: Systematic review and meta-analysis. BMC Pregnancy and Childbirth 2008; 8(1):33.

(129) Cnossen J, Morris RK, Mol BW, ter RG, van der Post JAM, Bindels P et al. Uterine artery Doppler to predict pre-eclampsia and intrauterine growth restriction: a systematic review and bivariable meta-analysis. Canadian Medical Association Journal 178[6], 701-711. 2008.

(130) Falagas M, Michalopoulos A, Bliziotis I, Sotereiades E. A bibliometric analysis by geographic area of published research in several biomedical fields, 1995-2003. Canadian Medical Association Journal 175[11], 1389-1390. 2006.

(131) Schulz KF, Chalmers I, Hayes R, Altman DG. Emperical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 273[408], 412. 1995.

(132) Moher D, Cook D, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta- analyses. Lancet 354[9193], 1896-1900. 1999.

(133) de Vet HC, van der WT, Muris JW, Heyrman J, Buntinx F, Knottnerus JA. Systematic reviews of diagnostic research. Considerations about assessment

and incorporation of methodological quality. European Journal of Epidemiology 2001; 17(4):301-306.

(134) Westwood ME, Whiting P, Kleijnen J. How does study quality affect the results of a diagnostic meta-analysis. BMC Med Res.Methodol. 5[1], 20. 2005.

(135) Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Statistics in Medicine 2002; 21(11):1513-1524.

(136) McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet 352, 1228-1231. 2000.

(137) Moher D, Pham B, Jones A, Cook D, Jadad A, Moher M et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 352, 609-613. 1998.

(138) Gardosi J, Mul T, Mongelli M, Fagan D. Analysis of birthweight and gestational age in antepartum stillbirths. Br J Obstet Gynaecol 1998; 105(5):524-530.

(139) Kamoji VM, Dorling JS, Manktelow BN et al. Extremely growth retarded : is there a viability centile? Pediatrics 2006; 118(2):758-763.

(140) Spinillo A, Capuzzo E, Piazzi G et al. Maternal high-risk factors and severity of growth deficit in small for gestationl age . Early Hum Dev 1994; 38(1):35-43.

(141) NHS Centre for Reviews and Dissemination. Undertaking Systematic Reviews of Research on Effectiveness: CRD Report number 4. 2nd Edition ed. York: York Publishing Services Ltd, 2010.

(142) Oxman AD, Cook DJ, Guyatt GH. User's guide to the medical literature VI. How to use an overview. JAMA 1994; 272(17):1367-1371.

(143) Duley L, Henderson-Smart DJ, Knight M, King JF. Antiplatelet agents for preventing pre-eclampsia and its complications.[update in Cochrane Database Syst Rev. 2007;(2):CD004659; PMID: 17443552]. [Review] [139 refs]. Cochrane Database of Systematic Reviews 2004;(1):CD004659.

(144) Askie LM, Duley L, Henderson-Smart DJ, Stewart LA, PARIS Collaborative Group. Antiplatelet agents for prevention of pre-eclampsia: a meta-analysis of individual patient data.[see comment]. [Review] [59 refs]. Lancet 2007; 369(9575):1791-1798.

(145) Kramer MS, Kakuma R. Energy and protein intake in pregnancy.[update of Cochrane Database Syst Rev. 2000;(2):CD000032; PMID: 10796092]. [Review] [91 refs]. Cochrane Database of Systematic Reviews 2003;(4):CD000032.

(146) Makrides M, Crowther CA. Magnesium supplementation in pregnancy.[update in Cochrane Database Syst Rev. 2001;(4):CD000937; PMID: 11687087]. [Review] [9 refs]. Cochrane Database of Systematic Reviews 2000;(2):CD000937.

(147) Haider BA, Bhutta ZA. Multiple-micronutrient supplementation for women during pregnancy. [Review] [98 refs]. Cochrane Database of Systematic Reviews 2006;(4):CD004905.

(148) Magee LA, Duley L. Oral beta-blockers for mild to moderate hypertension during pregnancy.[update in Cochrane Database Syst Rev. 2003;(3):CD002863; PMID: 12917933]. [Review] [112 refs]. Cochrane Database of Systematic Reviews 2000;(4):CD002863.

(149) Hofmeyr GJ. Abdominal decompression for suspected fetal compromise/pre-eclampsia. [Review] [3 refs]. Cochrane Database of Systematic Reviews 2000;(2):CD000004.

(150) Thinkhamrop J, Hofmeyr GJ, Adetoro O, Lumbiganon P. Prophylactic antibiotic administration in pregnancy to prevent infectious morbidity and mortality. [Review] [22 refs]. Cochrane Database of Systematic Reviews 2002;(4):CD002250.

(151) Honest H, Forbes CA, Duree K, Normal G, Duffy S, Tsourapas A et al. Screening to prevent spontaneous preterm birth: systematic reviews of accuracy and effectiveness literature with economic modelling. Health Technology Assessment 2009; 13(43).

(152) Grant A, Glazener CM. Elective caesarean section versus expectant management for delivery of the small baby.[update of Cochrane Database Syst Rev. 2001;(1):CD000078; PMID: 11279676]. [Review] [21 refs]. Cochrane Database of Systematic Reviews 2001;(2):CD000078.

(153) Say L, Gulmezoglu AM, Hofmeyr GJ. Maternal oxygen administration for suspected impaired fetal growth.[update of Cochrane Database Syst Rev. 2000;(2):CD000137; PMID: 10796154]. [Review] [15 refs]. Cochrane Database of Systematic Reviews 2003;(1):CD000137.

(154) Roberts D, Dalziel SR. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. Cochrane Database of Systematic Reviews: Reviews 2006 Issue 3 John Wiley & Sons , Ltd Chichester, UK DOI : 10 1002/14 2006.

(155) Crowther CA, Hiller JE, Doyle LW. Magnesium sulphate for preventing preterm birth in threatened preterm labour. Cochrane Database of Systematic Reviews: Reviews 2002 Issue 4 John Wiley & Sons , Ltd Chichester, UK DOI : 10 1002/14651858 CD0 2002.

(156) Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002; 21:2313-2324.

(157) Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004; 23:3105-3124.

(158) DATATreeage. Williamstown, MA; USA: 2009.

(159) Curtis L, Netten A. Unit costs of health and social care. Canterbury: PSSRU, 2006.

(160) Petrou S, Henderson J, Roberts T, Martin MA. Recent economic evaluations of antenatal screening: a systematic review and critique. Journal of Medical Screening 2000; 7:59-73.

(161) Roberts T, Henderson J, Mugford M, Bricker L, Neilson J, Garcia J. Antenatal ultrasound screening for fetal abnormalatoes: a systematic review of cost and cost effectiveness studies. BJOG 2002; 109(1):44-56.

(162) Health Resource Group 5 2009. The Information Centre for NHA and Social Care . 2009.

(163) Briggs A. Economic notes. Handling uncertainty in economic evaluation. BMJ 1999; 319:120.

(164) Coomarasamy A, Braunholtz D, Song F, Taylor R, Khan KS. Individualising use of aspirin to prevetn pre-eclampsia: a framework for clinical decision making. BJOG 2003; 110:882-888.

(165) Haynes R, McKibbon KA, Wilczynski NL, Walker SD, Werre SR. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. BMJ 2005; 330(7501):1179.

(166) Montori VM, Wilczynski NL, Morgan D, Haynes R. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. BMJ 2005; 330(7482):68.

(167) Haynes R, Wilczynski NL. Optimal search strategies for detecting clinically sound prognostic studies from MEDLINE: an analytic survey. BMC Med 2004; 1:23.