

Universal late pregnancy ultrasound screening to predict adverse outcomes in nulliparous women

Smith, Gordon Cs; Moraitis, Alexandros A; Wastlund, David; Thornton, Jim G; Papageorghiou, Aris; Sanders, Julia; Heazell, Alexander Ep; Robson, Stephen C; Sovio, Ulla; Brocklehurst, Peter; Wilson, Edward Cf

DOI:
[10.3310/hta25150](https://doi.org/10.3310/hta25150)

License:
Creative Commons: Attribution (CC BY)

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

Citation for published version (Harvard):
Smith, GC, Moraitis, AA, Wastlund, D, Thornton, JG, Papageorghiou, A, Sanders, J, Heazell, AE, Robson, SC, Sovio, U, Brocklehurst, P & Wilson, EC 2021, 'Universal late pregnancy ultrasound screening to predict adverse outcomes in nulliparous women: a systematic review and cost-effectiveness analysis', *Health Technology Assessment*, vol. 25, no. 15, pp. 1-190. <https://doi.org/10.3310/hta25150>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

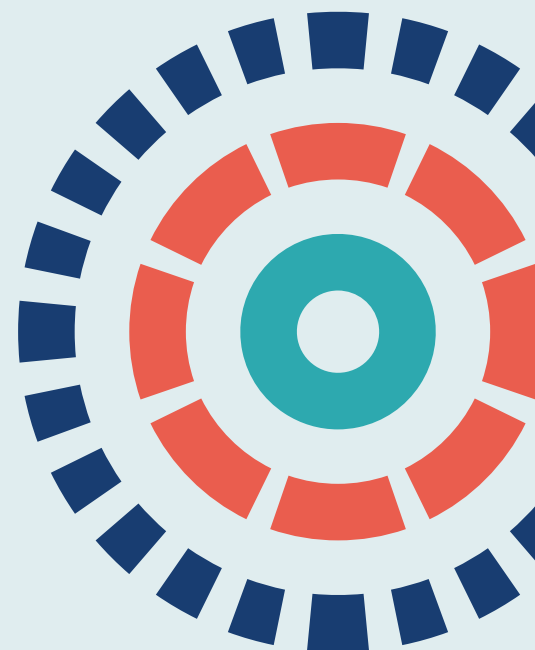
Health Technology Assessment

Volume 25 • Issue 15 • February 2021

ISSN 1366-5278

Universal late pregnancy ultrasound screening to predict adverse outcomes in nulliparous women: a systematic review and cost-effectiveness analysis

Gordon CS Smith, Alexandros A Moraitis, David Wastlund, Jim G Thornton, Aris Papageorghiou, Julia Sanders, Alexander EP Heazell, Stephen C Robson, Ulla Sovio, Peter Brocklehurst and Edward CF Wilson



Universal late pregnancy ultrasound screening to predict adverse outcomes in nulliparous women: a systematic review and cost-effectiveness analysis

Gordon CS Smith^{1*}, Alexandros A Moraitis¹,
David Wastlund², Jim G Thornton³,
Aris Papageorghiou⁴, Julia Sanders⁵,
Alexander EP Heazell⁶, Stephen C Robson⁷,
Ulla Sovio¹, Peter Brocklehurst⁸
and Edward CF Wilson^{2,9}

¹Department of Obstetrics and Gynaecology, NIHR Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK

²The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

³Division of Child Health, Obstetrics and Gynaecology, School of Medicine, University of Nottingham, Nottingham, UK

⁴Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford, UK

⁵School of Healthcare Sciences, Cardiff University, Cardiff, UK

⁶Faculty of Biology, Medicine and Health, School of Medical Sciences, University of Manchester, Manchester, UK

⁷Reproductive and Vascular Biology Group, The Medical School, Newcastle University, Newcastle upon Tyne, UK

⁸Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, UK

⁹Health Economics Group, Norwich Medical School, University of East Anglia, Norwich, UK

*Corresponding author

Declared competing interests of authors: Gordon CS Smith reports grants from the Medical Research Council, National Institute for Health Research (NIHR) Health Technology Assessment (HTA), the Wellcome Trust and Wellbeing of Women (London, UK); grants and personal fees from GlaxoSmithKline Research and Development Ltd (GlaxoSmithKline plc, Brentford, UK), grants from Sera Prognostics Inc. (Salt Lake City, UT), non-financial support from Illumina Inc. (San Diego, CA), and personal fees from Roche Diagnostics Ltd (Basel, Switzerland) outside the submitted work. Jim G Thornton is a member of the NIHR HTA and Efficacy and Mechanism Evaluation Editorial Board. Aris Papageorghiou reports personal fees from educational events/lectures, clinical services in the private sector and consultancy via Oxford University Innovation, royalties from published works, and editorial work for *Ultrasound in Obstetrics and Gynecology* and *British Journal of Obstetrics and Gynaecology* outside the submitted work. Ulla Sovio reports grants from the NIHR Cambridge Biomedical Research Centre during the conduct of the study. Peter Brocklehurst reports grants and personal fees from the Medical Research Council, grants from the NIHR Health Services and Delivery Research programme, the NIHR HTA programme and the Wellcome Trust, and personal fees from AG Biotest (Dreieich, Germany) outside the submitted work. In addition, Gordon CS Smith and Ulla Sovio have a patent in preparation for a novel predictive test for fetal growth restriction pending.

Published February 2021

DOI: 10.3310/hta25150

This report should be referenced as follows:

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

Health Technology Assessment is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 3.370

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, the Cochrane Library and Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 15/105/01. The contractual start date was in March 2017. The draft report began editorial review in August 2019 and was accepted for publication in March 2020. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

Copyright © 2021 Smith *et al.* This work was produced by Smith *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaptation in any medium and for any purpose provided that it is properly attributed. See: <https://creativecommons.org/licenses/by/4.0/>. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

NIHR Journals Library Editor-in-Chief

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

NIHR Journals Library Editors

Professor John Powell Chair of HTA and EME Editorial Board and Editor-in-Chief of HTA and EME journals. Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK, and Professor of Digital Health Care, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals) and Editor-in-Chief of HS&DR, PGfAR, PHR journals

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Consultant Advisor, Wessex Institute, University of Southampton, UK

Ms Tara Lamont Senior Scientific Adviser (Evidence Use), Wessex Institute, University of Southampton, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Emeritus Professor of Wellbeing Research, University of Winchester, UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Great Ormond Street Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: journals.library@nihr.ac.uk

Abstract

Universal late pregnancy ultrasound screening to predict adverse outcomes in nulliparous women: a systematic review and cost-effectiveness analysis

Gordon CS Smith^{1*}, Alexandros A Moraitis¹, David Wastlund²,
Jim G Thornton³, Aris Papageorghiou⁴, Julia Sanders⁵,
Alexander EP Heazell⁶, Stephen C Robson⁷, Ulla Sovio¹,
Peter Brocklehurst⁸ and Edward CF Wilson^{2,9}

¹Department of Obstetrics and Gynaecology, NIHR Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK

²The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

³Division of Child Health, Obstetrics and Gynaecology, School of Medicine, University of Nottingham, Nottingham, UK

⁴Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford, UK

⁵School of Healthcare Sciences, Cardiff University, Cardiff, UK

⁶Faculty of Biology, Medicine and Health, School of Medical Sciences, University of Manchester, Manchester, UK

⁷Reproductive and Vascular Biology Group, The Medical School, Newcastle University, Newcastle upon Tyne, UK

⁸Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, UK

⁹Health Economics Group, Norwich Medical School, University of East Anglia, Norwich, UK

*Corresponding author gcss2@cam.ac.uk

Background: Currently, pregnant women are screened using ultrasound to perform gestational aging, typically at around 12 weeks' gestation, and around the middle of pregnancy. Ultrasound scans thereafter are performed for clinical indications only.

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

Design: We performed diagnostic test accuracy reviews of five ultrasonic measurements in late pregnancy. We conducted cost-effectiveness and value-of-information analyses of screening for fetal presentation, screening for small for gestational age fetuses and screening for large for gestational age fetuses. Finally, we conducted a survey and a focus group to determine the willingness of women to participate in a future randomised controlled trial.

Data sources: We searched MEDLINE, EMBASE and the Cochrane Library from inception to June 2019.

Review methods: The protocol for the review was designed a priori and registered. Eligible studies were identified using keywords, with no restrictions for language or location. The risk of bias in studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool.

Health economic modelling employed a decision tree analysed via Monte Carlo simulation. Health outcomes were from the fetal perspective and presented as quality-adjusted life-years. Costs were from the perspective of the public sector, defined as NHS England, and the costs of special educational needs. All costs and quality-adjusted life-years were discounted by 3.5% per annum and the reference case time horizon was 20 years.

Results: Umbilical artery Doppler flow velocimetry, cerebroplacental ratio, severe oligohydramnios and borderline oligohydramnios were all either non-predictive or weakly predictive of the risk of neonatal morbidity (summary positive likelihood ratios between 1 and 2) and were all weakly predictive of the risk of delivering a small for gestational age infant (summary positive likelihood ratios between 2 and 4). Suspicion of fetal macrosomia is strongly predictive of the risk of delivering a large infant, but it is only weakly, albeit statistically significantly, predictive of the risk of shoulder dystocia. Very few studies blinded the result of the ultrasound scan and most studies were rated as being at a high risk of bias as a result of treatment paradox, ascertainment bias or iatrogenic harm. Health economic analysis indicated that universal ultrasound for fetal presentation only may be both clinically and economically justified on the basis of existing evidence. Universal ultrasound including fetal biometry was of borderline cost-effectiveness and was sensitive to assumptions. Value-of-information analysis indicated that the parameter that had the largest impact on decision uncertainty was the net difference in cost between an induced delivery and expectant management.

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

Conclusions: Universal screening for presentation at term may be justified on the basis of current knowledge. The current literature does not support universal ultrasonic screening for fetal growth disorders.

Future work: We describe proof-of-principle randomised controlled trials that could better inform the case for screening using ultrasound in late pregnancy.

Study registration: This study is registered as PROSPERO CRD42017064093.

Funding: This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 25, No. 15. See the NIHR Journals Library website for further project information.

Contents

List of tables	xiii
List of figures	xv
List of abbreviations	xix
Plain English summary	xxi
Scientific summary	xxiii
Chapter 1 Background	1
Screening for pregnancy complications	1
Use of ultrasound in pregnancy screening	1
Use of ultrasound in late pregnancy	1
Coupling interventions to scan results	2
Evidence for screening using universal late pregnancy ultrasound	3
Critical analysis of the Cochrane review	4
Parity and the risk of adverse outcome	5
Summary of the rationale for the focus on nulliparous women in late pregnancy	5
The health economics of screening and intervention	5
Value-of-information analysis	6
Designing a randomised controlled trial	6
Chapter 2 Objectives	7
Chapter 3 Identifying the research questions	9
Chapter 4 Systematic review of the diagnostic effectiveness of universal ultrasonic screening using late pregnancy umbilical artery Doppler flow velocimetry in the prediction of adverse perinatal outcome	11
Methods	11
<i>Analysis of data from the Pregnancy Outcome Prediction study</i>	11
<i>Sources for meta-analysis</i>	11
<i>Study selection</i>	12
<i>Study quality assessment and data extraction</i>	12
<i>Statistical and meta-analysis methods</i>	12
Results	12
<i>The Pregnancy Outcome Prediction study</i>	12
<i>Meta-analysis</i>	13
Discussion	17
Chapter 5 Systematic review of the diagnostic effectiveness of universal ultrasonic screening using late pregnancy cerebroplacental ratio in the prediction of adverse perinatal outcome	19
Methods	19
<i>Sources for meta-analysis</i>	19
<i>Study selection</i>	19

<i>Study quality assessment and data extraction</i>	20
<i>Statistical and meta-analysis methods</i>	20
Results	20
Discussion	21
Chapter 6 Systematic review of the diagnostic effectiveness of universal ultrasonic screening using severe oligohydramnios in the prediction of adverse perinatal outcome	27
Methods	27
<i>Sources for meta-analysis</i>	27
<i>Study selection</i>	27
<i>Study quality assessment and data extraction</i>	28
<i>Statistical and meta-analysis methods</i>	28
Results	28
Discussion	30
Chapter 7 Systematic review of the diagnostic effectiveness of universal ultrasonic screening using borderline oligohydramnios in the prediction of adverse perinatal outcome	35
Methods	35
<i>Analysis of data from the Pregnancy Outcome Prediction study</i>	35
<i>Sources for meta-analysis</i>	35
<i>Study selection</i>	35
<i>Study quality assessment and data extraction</i>	36
<i>Statistical and meta-analysis methods</i>	36
Results	36
<i>The Pregnancy Outcome Prediction study</i>	36
<i>Meta-analysis</i>	36
Discussion	38
Chapter 8 Systematic review of the diagnostic effectiveness of universal ultrasonic screening using fetal macrosomia in the prediction of adverse perinatal outcome	41
Methods	41
<i>Sources for meta-analysis</i>	41
<i>Study selection</i>	42
<i>Study quality assessment and data extraction</i>	42
<i>Statistical and meta-analysis methods</i>	42
Results	42
Discussion	45
Chapter 9 Conclusions regarding the evidence around universal ultrasound screening of nulliparous women in late pregnancy	47
Chapter 10 Evidence-based protocol for the care of screen-positive women	49
Management plan for breech presentation	49
Management plan for diagnosis of a small for gestational age fetus	49
Management plan following diagnosis of a large for gestational age fetus	50
Chapter 11 Economic analysis of universal versus selective ultrasound screening in late-stage pregnancy: cost-effectiveness and value-of-information analyses	53
Introduction	53
Methods	54
<i>Scope and population</i>	54
<i>Comparators and interventions</i>	54

Outcomes	55
Model structure	55
Data	62
Results	73
Stability testing	73
Cost-effectiveness results	73
One-way and scenario analyses	75
Value-of-information analysis	79
Discussion	82
Main findings	82
Strengths and limitations	83
Comparison with other studies	85
Implementation considerations	85
Conclusions	86
Chapter 12 The views of recently delivered and currently pregnant women on universal ultrasound screening in late pregnancy	87
Aims	87
Methods	87
Results	88
Survey	88
Focus group	89
Discussion and conclusions	90
Reflections/clinical perspective	90
Chapter 13 Designing a randomised controlled trial of screening and intervention	91
Implications of the health economic analysis	91
Case for considering a randomised controlled trial of screening and intervention	91
Candidate primary outcomes	92
Proxies	92
Subgroups	92
Early delivery and iatrogenic harm	93
Current status of screening tests	93
Possible trial designs	93
Acceptability of the 'screen-all' approach	94
Power calculations	95
Implications of sample size calculations	96
Chapter 14 Overall conclusions and assessment of evidence required for a national screening programme	97
Overall conclusions	97
Consultation with the National Screening Committee	97
Acknowledgements	99
References	101
Appendix 1 Supporting data for the systematic review of the diagnostic effectiveness of universal ultrasonic screening using late pregnancy umbilical artery Doppler flow velocimetry in the prediction of adverse perinatal outcome	115

CONTENTS

Appendix 2 Supporting data for the systematic review of the diagnostic effectiveness of universal ultrasonic screening using late pregnancy cerebroplacental ratio in the prediction of adverse perinatal outcome	123
Appendix 3 Supporting data for the systematic review of the diagnostic effectiveness of universal ultrasonic screening using severe oligohydramnios in the prediction of adverse perinatal outcome	133
Appendix 4 Supporting data for the systematic review of the diagnostic effectiveness of universal ultrasonic screening using borderline oligohydramnios in the prediction of adverse perinatal outcome	141
Appendix 5 Supporting data for the systematic review of the diagnostic effectiveness of universal ultrasonic screening using macrosomia in the prediction of adverse perinatal outcome	151
Appendix 6 Derivation of input parameters for economic simulation model	167
Appendix 7 Brief summary of economic analyses of universal screening for breech presentation, large for gestational age fetuses and small for gestational age fetuses	185
Appendix 8 Questionnaire for attitudes towards universal ultrasound screening in late pregnancy	189

List of tables

TABLE 1 Diagnostic performance of umbilical artery PI > 90th centile in predicting adverse pregnancy outcome in the POP study ($n = 3615$)	13
TABLE 2 Summary diagnostic results of meta-analysis of umbilical artery Doppler for predicting adverse pregnancy outcome	14
TABLE 3 Diagnostic accuracy of CPRs in predicting adverse pregnancy outcome	21
TABLE 4 Summary diagnostic performance of low AFI (< 5 cm) in predicting adverse pregnancy outcome	29
TABLE 5 Diagnostic performance of borderline AFI (5–8 cm) in predicting adverse pregnancy outcome at term in the POP study ($n = 3387$)	37
TABLE 6 Summary diagnostic performance of borderline AFI in predicting adverse pregnancy outcome	37
TABLE 7 Summary diagnostic performance of suspected LGA in predicting LGA at birth and shoulder dystocia	43
TABLE 8 Diagnostic effectiveness of ultrasonic screening at 36 weeks' gestation for subsequent delivery of a SGA infant associated with either maternal pre-eclampsia or perinatal morbidity or mortality	48
TABLE 9 Comparator strategies for economic simulation model	55
TABLE 10 Model inputs for diagnostic performance	64
TABLE 11 Model inputs for probabilities	65
TABLE 12 Model inputs for costs and related probabilities	69
TABLE 13 Results from stability testing	73
TABLE 14 Cost-effectiveness results (per woman scanned)	74
TABLE 15 The expected value of partial perfect information for individual parameters and groups of parameters	80
TABLE 16 Results of the survey of low-risk pregnant women ($n = 100$)	88
TABLE 17 Sample size calculations for different outcomes, screening tests and trial designs	95
TABLE 18 Maternal characteristics and birth outcomes of POP study	116
TABLE 19 Characteristics of studies included in the meta-analysis	119

TABLE 20 Characteristics of studies included in the meta-analysis of CPRs to predict adverse pregnancy outcome	126
TABLE 21 Characteristics of studies included in the meta-analysis of severe oligohydramnios	136
TABLE 22 Patient characteristics and birth outcomes of POP study	143
TABLE 23 Characteristics of studies included in the meta-analysis of borderline oligohydramnios	146
TABLE 24 Characteristics of studies included in the meta-analysis of macrosomia	154
TABLE 25 Prevalence of no, moderate and severe neonatal morbidity in the POP study by fetal size diagnosis	170
TABLE 26 Risk of respiratory morbidity from emergency caesarean section	173
TABLE 27 Risk of acidosis from emergency caesarean section	173
TABLE 28 Risk of perinatal mortality from emergency caesarean section	173
TABLE 29 Baseline risk of CP by 5-minute Apgar score	177
TABLE 30 Relative risk of CP by 5-minute Apgar score	177

List of figures

FIGURE 1 Summary ROC curves for umbilical artery Doppler for predicting (a) neonatal intensive care unit admission; (b) neonatal metabolic acidosis; (c) SGA (< 10th centile); and (d) severe SGA (< 3rd centile)	14
FIGURE 2 Meta-analysis of DORs of umbilical artery Doppler at predicting (a) neonatal intensive care unit admission; (b) neonatal metabolic acidosis; (c) 5-minute Apgar score of < 7; (d) severe adverse perinatal outcome; (e) caesarean section for fetal distress; (f) pre-eclampsia; (g) SGA (< 10th centile); and (h) severe SGA (< 3rd centile)	15
FIGURE 3 Summary ROC curves for the diagnostic performance of abnormal CPRs at predicting adverse pregnancy outcomes	22
FIGURE 4 The diagnostic odds ratios for the diagnostic performance of abnormal CPRs at predicting adverse pregnancy outcomes	24
FIGURE 5 Summary ROC curves for AFI < 5 cm at predicting adverse pregnancy outcome	29
FIGURE 6 Meta-analysis of DORs for AFI < 5 cm at predicting adverse pregnancy outcome: (a) NICU admission; (b) 5-minute Apgar score of < 7; (c) neonatal metabolic acidosis; (d) caesarean section for fetal distress; (e) SGA (< 10th centile); and (f) neonatal death	31
FIGURE 7 Summary ROC curves of borderline AFI at predicting (a) SGA < 10th centile; (b) NICU admission; (c) 5-minute Apgar score of < 7; and (d) caesarean section for fetal distress	38
FIGURE 8 The diagnostic odds ratios of borderline AFI at predicting: (a) SGA < 10th centile; (b) NICU admission; (c) 5-minute Apgar score of < 7; and (d) caesarean section for fetal distress	39
FIGURE 9 Summary ROC curves for the diagnostic performance of EFW > 4000 g (or 90th centile) at predicting (a) LGA at birth (birthweight > 4000 g or > 90th centile); and (b) shoulder dystocia	44
FIGURE 10 The diagnostic odds ratios for the diagnostic performance of EFW > 4000 g (or 90th centile) at predicting (a) LGA at birth (birthweight > 4000 g or > 90th centile); and (b) shoulder dystocia	44
FIGURE 11 Summary of the management plan following the 36 weeks' gestation scan	51
FIGURE 12 Model overview	56
FIGURE 13 Outcomes associated with breech	58
FIGURE 14 Outcomes associated with LGA	60
FIGURE 15 Outcomes associated with SGA	61

FIGURE 16 Outcomes associated with AGA	63
FIGURE 17 Cost-effectiveness acceptability curve for the chance that each strategy will be the most cost-effective as a function of WTP for an additional QALY	75
FIGURE 18 One-way sensitivity analysis of model time horizon	76
FIGURE 19 One-way sensitivity analysis of the cost of a scan for fetal presentation only	76
FIGURE 20 One-way sensitivity analysis of baseline risk of (a) perinatal mortality; (b) severe morbidity; and (c) moderate morbidity	77
FIGURE 21 One-way sensitivity analysis on relative risk of SEN from IOL	79
FIGURE 22 Per-patient EVPI as a function of the WTP for an additional QALY	80
FIGURE 23 Population EVSI for a study on the cost of IOL	81
FIGURE 24 Flow charts of possible trial designs: (a) screen vs. no screen; and (b) screen all	94
FIGURE 25 The POP study inclusion flow chart	116
FIGURE 26 Literature search PRISMA flow diagram for the systematic review on umbilical artery Doppler	118
FIGURE 27 Risk of bias and applicability concerns using the QUADAS-2 tool for the studies included in the meta-analysis of umbilical artery Doppler	118
FIGURE 28 Deeks' funnel plot for publication bias for umbilical artery Doppler for the prediction of neonatal unit admission	122
FIGURE 29 Literature search PRISMA flow diagram for the systematic review on CPRs	124
FIGURE 30 Risk of bias and applicability concerns using the QUADAS-2 tool for the studies included in the meta-analysis of CPRs	125
FIGURE 31 Deeks' funnel plot for publication bias for CPRs for the prediction of neonatal unit admission	131
FIGURE 32 The PRISMA flow diagram for the systematic review of severe oligohydramnios	134
FIGURE 33 Risk-of-bias graph of included studies for systematic review of severe oligohydramnios using the QUADAS-2 tool	135
FIGURE 34 Deeks' funnel plot for publication bias for severe oligohydramnios for the prediction of neonatal unit admission	140
FIGURE 35 The POP study inclusion flow chart	142
FIGURE 36 The PRISMA flow diagram for the systematic review of borderline oligohydramnios	144

FIGURE 37 Risk-of-bias and applicability concerns for included studies in systematic review of borderline oligohydramnios using the QUADAS-2 tool	145
FIGURE 38 Deeks' funnel plot for publication bias for borderline oligohydramnios for the prediction of SGA < 10th centile	149
FIGURE 39 The PRISMA flow diagram for the systematic review of macrosomia	152
FIGURE 40 Risk-of-bias applicability concerns for included studies for systematic review of macrosomia	153
FIGURE 41 Deeks' funnel plot for publication bias for the prediction of LGA (birthweight > 4000 g or > 90th centile)	166

List of abbreviations

AC	abdominal circumference	LR	likelihood ratio
AFI	amniotic fluid index	MCA	middle cerebral arteries
AGA	appropriate for gestational age	NICE	National Institute for Health and Care Excellence
BMI	body mass index	NICU	neonatal intensive care unit
BPI	brachial plexus injury	NIH	National Institutes of Health
CDSR	Cochrane Database of Systematic Reviews	NMB	net monetary benefit
CENTRAL	Cochrane Central Register of Controlled Trials	NSC	National Screening Committee
CI	confidence interval	PAC	peripheral arterial chemoreceptor
CP	cerebral palsy	PAPP-A	pregnancy associated plasma protein-A
CPR	cerebroplacental ratio	PI	pulsatility index
CUHFT	Cambridge University Hospitals NHS Foundation Trust	POP	Pregnancy Outcome Prediction
DOR	diagnostic odds ratio	PPI	patient and public involvement
DTA	diagnostic test accuracy	PPV	positive predictive value
ECV	external cephalic version	PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
EFW	estimated fetal weight	QALY	quality-adjusted life-year
ENGS	expected net gain of sampling	QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies 2
EVPI	expected value of perfect information	RCOG	Royal College of Obstetricians and Gynaecologists
EVPPi	expected value of partial perfect information	RCT	randomised controlled trial
EVSI	expected value of sample information	RI	resistance index
FGR	fetal growth restriction	ROC	receiver operating characteristic
GMFCS	Gross Motor Function Classification System	SAVI	Sheffield accelerated value of information
HCHS	Hospital & Community Health Services	SDP	single deepest pocket
HTA	Health Technology Assessment	SEN	special educational needs
ICER	incremental cost-effectiveness ratio	SGA	small for gestational age
INMB	incremental net monetary benefit	SNM	severe neurological morbidity
IOL	induction of labour	VOI	value of information
LGA	large for gestational age	WTP	willingness to pay

Plain English summary

Ultrasound scans allow doctors to check on the health of an unborn infant. Usually, all pregnant women receive a scan at about 3 months and about 5 months of pregnancy. After that, women are offered a scan during birth only if they have risk factors or if a problem develops. Problems can arise in the later stages of pregnancy, including issues with the infant's growth or whether or not the infant is breech. Some of these problems may be prevented if a scan is carried out, but scans can also be inaccurate. When they are, a woman may receive unnecessary treatment, which could even harm her or her infant.

In this study we set out to review previous research about how good ultrasound scanning is at detecting infants who may be born with a condition. This study focused on detecting if the infant was too big or too small. Unfortunately, much of the previous research was not carried out to a high standard. Scanning can detect the size of a infant relatively well, but it is much less clear if scanning can predict complications that may harm the infant during birth. We also studied the costs and outcomes of scanning. We calculated the extra cost required to scan every woman and compared this with the extra benefits from preventing complications. One thing that ultrasound scans detect is whether the infant is presenting head first or bottom first (a 'breech presentation'), as infants presenting breech have high risks of complications. Scanning all women to check whether or not their infant is presenting breech seems to be cost-effective and the cost savings may even be higher than the cost of implementation, although this depends on how much the scan would cost.

Whether or not it is worthwhile scanning all infants to see if they are above or below the thresholds for normal size is less clear. A larger research study could provide more reliable numbers from which to draw a conclusion. We show how such a study could be designed, so that a single study could tell us both how well scans can predict adverse outcomes and how helpful this information is.

Scientific summary

Background

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

Objectives

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

- to assess the diagnostic effectiveness of late pregnancy ultrasound in nulliparous women based on the existing research literature
- having identified the key ultrasonic findings that define women as high risk, to review the existing literature and current guidelines to identify a management plan for women with high-risk characteristics
- to conduct a health economic analysis of the likely cost-effectiveness of screening and intervention based on the best available evidence of the costs, diagnostic effectiveness of ultrasound and clinical effectiveness of intervention
- to perform a value-of-information analysis to determine whether or not there is a strong economic case for funding future research in this area
- depending on the above, to outline the design of a randomised controlled trial that could strengthen the evidence base relating to the issues above.

Methods

We identified the following as key ultrasound measurements that might be used in late pregnancy screening: (1) the infant is suspected to be small for gestational age, (2) the baby is suspected to be large for gestational age, (3) high-resistance pattern of umbilical artery Doppler flow velocimetry, (4) low cerebroplacental ratio, (5) severe oligohydramnios and (6) borderline oligohydramnios. We found that there was an ongoing Cochrane Diagnostic Test Accuracy review for infants suspected to be small for gestational age, so we focused on the other five measures. The protocol for the reviews was designed a priori and registered with the International Prospective Register of Systematic Reviews PROSPERO (CRD42017064093). We searched MEDLINE, EMBASE and the Cochrane Library from inception. The studies were identified using a combination of keywords. Selection criteria included cohort or cross-sectional studies including women with singleton pregnancies who had an ultrasound performed at ≥ 24 weeks' gestation. Case-control studies were excluded. We included all studies in which the ultrasound was performed as part of universal ultrasound screening (i.e. the ultrasound was offered to all women regardless of indication), studies that were carried out in low-risk populations (i.e. those that excluded pregnancies with any maternal or fetal complications) and studies with a mixed-risk population

(i.e. the ultrasound was offered selectively based on current clinical indications). We excluded studies that focused on high-risk populations only. The literature search, study selection and analysis were performed independently by two researchers using RevMan 5.3 (The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark). Any differences were resolved by discussion with the senior author. The risk of bias in each included study was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool as described in the Cochrane Handbook of Diagnostic Test Accuracy Studies (Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529–36.). We used a pre-designed data extraction form to extract information on study characteristics (e.g. year of publication, country, setting, study design and blinding), patient characteristics (e.g. inclusion and exclusion criteria, and sample size), the index test (e.g. gestation at scan, Doppler indices and cut-off values used), and reference standard (e.g. pregnancy outcome, gestation at delivery and interval from scan to delivery).

From each study we extracted the 2×2 tables for all combinations of index tests and outcomes and we calculated the sensitivity, specificity, and positive and negative likelihood ratios. For the data synthesis we used a hierarchical summary receiver operating characteristic curve model. Whenever four or more studies were available, estimates of mean sensitivity and specificity and their respective variances at a specific threshold were additionally generated using the bivariate logit-normal model. We also pooled the diagnostic odds ratios using the method described by Deeks *et al.* (Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;**58**:882–93.) and used the Deeks' funnel plot asymmetry test for publication bias in which a p -value of < 0.05 was defined as significant asymmetry. For the statistical analyses we used the *metandi*, *metan* and *midas* packages in Stata® version 14 (StataCorp LP, College Station, TX, USA).

We included studies regardless of blinding of the ultrasound to the clinicians but this was reported in the study characteristics. However, revealing the scan result has the potential for multiple biases. We had access to the original data from the Pregnancy Outcome Prediction study [Sovio U, White IR, Dacey A, Pasupathy D, Smith GCS. Screening for fetal growth restriction with universal third trimester ultrasonography in nulliparous women in the Pregnancy Outcome Prediction (POP) study: a prospective cohort study. *Lancet* 2015;**386**:2089–97]. This is the larger of only two studies that performed blinded ultrasonic assessment near term in nulliparous women. The other study (Galvin DM, Burke N, Burke G, Breathnach F, McAuliffe F, Morrison J, *et al.* 94: Accuracy of prenatal detection of macrosomia $> 4,000$ g and outcomes in the absence of intervention: results of the prospective multicenter genesis study. *Am J Obstet Gynecol* 2017;**216**:S68.) has not yet been widely reported. Given the importance of blinding, we carried out a number of new analyses of the Pregnancy Outcome Prediction study data set.

Health economic modelling employed a decision tree analysed via Monte Carlo simulation (repeated sampling from input parameter distributions) and coded in R (The R Foundation for Statistical Computing, Vienna, Austria) (an open-source statistical software package). Health outcomes were from the fetal perspective and presented as quality-adjusted life-years. The perspective used was the public sector, defined as NHS England, and special educational needs. All costs and quality-adjusted life-years were discounted by 3.5% per annum and the reference case time horizon was 20 years. The health economic analysis evaluated three different strategies for ultrasound screening in late pregnancy, defined as a scan between 36⁺⁰ and 36⁺⁶ weeks' gestation: (1) 'selective ultrasound' (i.e. where ultrasound is performed only if clinically indicated), the current standard of care in England; (2) 'universal ultrasound for presentation only' (i.e. scanning with the sole purpose of detecting breech presentation); and (3) 'universal ultrasound for fetal size' (i.e. a scan to assess fetal weight plus assessment of presentation).

We assumed that in all identified cases of breech presentation the woman would be offered an external cephalic version unless contraindicated, in line with guidelines from the Royal College of Obstetricians and Gynaecologists. We also assumed that mothers of infants identified as small for gestational age (whether or not these infants were correctly diagnosed) would be given early induction

of labour at 37 weeks' gestation. However, for infants diagnosed as large for gestational age, there is uncertainty about whether or not intervention (i.e. induction of labour) is beneficial. For this reason, expectant management of suspected large for gestational age fetuses was also an option. We assumed that selective scanning (i.e. only where clinically indicated) with a policy of offering external cephalic version for suspicion of breech presentation and induction of labour for suspicion of small for gestational age or large for gestational age fetuses represents an approximation of the status quo from which estimates of incremental net benefit are calculated.

Results

We identified 13 studies of umbilical artery Doppler flow velocimetry that met our inclusion criteria, which comprised 67,764 patients. Umbilical artery Doppler flow velocimetry had weak/moderate predictive accuracy for detecting fetuses who are small for gestational age or severely small for gestational age (< 3rd percentile) (positive likelihood ratio of about 2.5 and 3.0, respectively). However, it did not predict neonatal morbidity at term. The results were very similar in both the Pregnancy Outcome Prediction study and the meta-analysis (which included the Pregnancy Outcome Prediction study), the only notable difference being that the association with a fetus being severely small for gestational age was slightly stronger in the Pregnancy Outcome Prediction study.

We identified 16 studies of cerebroplacental ratio that met our inclusion criteria, which resulted in a total of 121,607 patients. Meta-analysis demonstrated that the cerebroplacental ratio may be slightly more predictive than umbilical artery Doppler flow velocimetry scanning in identifying pregnancies at an increased risk of adverse outcome. In the case of a fetus being small for gestational age, the positive likelihood ratios were in the region of 3.5–4.0. Moreover, unlike umbilical artery Doppler flow velocimetry, a low cerebroplacental ratio was associated with an increased risk of neonatal morbidity. However, the association with morbidity was weaker with positive likelihood ratios of < 2.0. Furthermore, in both analyses, there was very significant heterogeneity in relation to both small for gestational age fetuses and neonatal morbidity. Consequently, the 95% confidence intervals for the positive likelihood ratio are wide and include the point estimates observed for umbilical artery Doppler flow velocimetry for both small for gestational age fetuses and severely small for gestational age fetuses.

We identified 14 studies of severe oligohydramnios that met our inclusion criteria, which involved a total of 109,679 patients. Diagnosis of severe oligohydramnios was associated with a positive likelihood ratio for small for gestational age fetuses of between 2.5 and 3.0. It was also associated with positive likelihood ratios for admission to a neonatal intensive care unit and emergency caesarean section for fetal distress of between 1.5 and 2.5. However, these associations are more difficult to interpret. First, for both of these outcomes, the association was weaker than it was for fetuses who were small for gestational age. Second, in both cases the associations could be a consequence of the scan rather than an outcome predicted by the scan, as the authors of only two studies comprised < 5% of the patients in the meta-analysis blinded the results of the scan.

We identified 11 studies of borderline oligohydramnios (including the Pregnancy Outcome Prediction study) that met our inclusion criteria and involved a total of 37,848 patients. Borderline oligohydramnios was weakly/moderately predictive of a fetus being small for gestational age (positive likelihood ratio 2.5–3.0). This was observed in the meta-analysis of multiple studies of variable quality. A comparable association was also seen between borderline oligohydramnios and fetuses being severely small for gestational age in the only study in which the scan result was blinded, namely the Pregnancy Outcome Prediction study.

We identified 40 studies of large for gestational age fetuses that met our inclusion criteria, which comprised 66,187 patients. Suspicion of fetal macrosomia on ultrasound was strongly predictive of the risk of delivering a large infant, but it was only weakly, albeit statistically significantly, predictive of the

risk of shoulder dystocia. In the case of delivering a large for gestational age infant, using the Hadlock formula (Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology* 1991;**181**:129–33.), the positive likelihood ratios were quite strong, in the region of 7 to 12; whereas in relation to the diagnosis of shoulder dystocia, the positive likelihood ratio was ≈ 2 . The forest plot of diagnostic odds ratios indicates significant heterogeneity between the studies in the ability to predict a large for gestational age infant.

Based on current information, and assuming a willingness to pay threshold of £20,000 per quality-adjusted life-year, offering a universal ultrasound presentation-only scan is, on average, the most cost-effective strategy. This is associated with an incremental net monetary benefit of £87.36 (95% confidence interval £4.88 to £205.68) per pregnancy compared with current practice. Scaled up to the English population, this equates to a net benefit of £17.1M or 857 quality-adjusted life-years per annual birth cohort. This is the present value of the future flows of expected costs and benefits over a time horizon of 20 years. Owing to uncertainties in the evidence base (parameter uncertainty), there is only a 44.19% probability that this conclusion is correct (i.e. there is a 55.81% probability that this conclusion is incorrect, in which case a loss will be incurred). The expected loss associated with this decision uncertainty is £31.56 per pregnancy. Equivalently, this is the expected gain if uncertainty were to be eliminated (expected value of perfect information). Scaled up to the population of England who could benefit from the information provided by any future studies, this equates to an expected value of perfect information of £53.3M. If it is assumed the results of any future study are generalisable to all pregnancies in England, the expected value of perfect information is £172.9M.

The parameter that has the biggest impact on decision uncertainty is the cost of induction of labour (specifically, the difference in cost between an induced delivery and expectant management). It should be noted that this does not relate simply to the cost of a procedure to induce delivery; included in this definition is uncertainty about the timing of induction of labour and the impact on, for example, antenatal appointments, as well as the cost of the delivery itself. A study of 'reasonable size' to reduce the uncertainty regarding this parameter is likely to yield a positive return on investment. For example, the expected value of sample information of a study of 1000 mothers in each arm is worth in excess of £11M. If this were to be delivered for a cost of £1M, it would yield a > 10-fold return on investment. Of note is that studies on the outcomes of small for gestational age fetuses or macrosomic deliveries are unlikely to yield a positive return on investment. The results described above relate to a willingness-to-pay threshold of £20,000 per quality-adjusted life-year. At a threshold of £30,600 per quality-adjusted life-year (just above the upper end of the National Institute for Health and Care Excellence's stated acceptable range of £20,000–£30,000), universal scanning becomes the most cost-effective option. Furthermore, our one-way sensitivity analyses suggest that there is scope for universal scanning to be cost-effective under other assumptions; for example, the most cost-effective option remains a breech-only scan only as long as the time horizon of the analysis is < 45 years.

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

Conclusions

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

Study registration

This study is registered as PROSPERO CRD42017064093.

Funding

This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 25, No. 15. See the NIHR Journals Library website for further project information.

Chapter 1 Background

Screening for pregnancy complications

Complications of pregnancy are a major contributor to the global burden of disease as a result of the effects on both the mother and the infant.¹ Identifying and managing the risk of complications is a key element of antenatal care that aims to reduce the number and severity of adverse outcomes. Current clinical guidelines² describe multiple methods of identifying high-risk women, including (1) identification of maternal risk factors associated with disease (e.g. obesity, being aged > 40 years), (2) assessment of complications in previous pregnancies, (3) identification of pre-existing medical conditions (e.g. diabetes mellitus) and (4) clinical presentation with symptoms that are associated with an increased risk of adverse outcome (e.g. antepartum haemorrhage, reduced fetal movements). In addition, multiple tests are given to pregnant women to assess their risk. Taking the example of screening for Down syndrome, a woman's risk is first assessed by maternal age; this background risk is then adjusted for the results of ultrasonic imaging (nuchal translucency) and biomarkers (pregnancy-associated plasma protein A and free beta subunit of human chorionic gonadotrophin), and the summative risk is used to inform the use of invasive testing.³

Use of ultrasound in pregnancy screening

The first trimester ultrasound scan used to screen for Down syndrome is an example of a scan that is offered to all pregnant women as part of their risk assessment. Routine pregnancy care in the UK also involves a second screening ultrasound scan, performed at or after 18 weeks' gestation but before 21 weeks' gestation, the primary purpose of which is to identify fetuses with structural abnormalities.³ A positive result from this scan might inform decisions around termination of pregnancy (e.g. some women may choose to terminate a pregnancy if the fetus has a severe neural tube defect) or it might inform the need for targeted follow-up and changes to the perinatal care of the infant. For example, identifying a congenital diaphragmatic hernia could lead to invasive testing for aneuploidy, prenatal discussions with the paediatric surgery team and modification to neonatal resuscitation (e.g. early intubation to avoid expansion of the stomach with air).

In the UK and the USA, universal ultrasound is not recommended after the mid-pregnancy anomaly scan.^{2,4} Instead, it is recommended that ultrasound be offered in a targeted manner and only to women in whom there is a clinical indication. Such indications could include presentation with symptoms (e.g. antepartum haemorrhage), relevant medical history (e.g. antiphospholipid antibody syndrome) and relevant medical history [e.g. previous fetal growth restriction (FGR)], or result from physical examination [e.g. the fetus is small for gestational age (SGA)] on clinical examination.

Use of ultrasound in late pregnancy

When ultrasound scans are performed in late pregnancy, a number of features are commonly reported. Ultrasound allows the estimation of the size (length and circumference) of fetal parts, termed fetal biometry. A variety of methods exist for converting these measurements to an estimated fetal weight (EFW)⁵ and a number of reference ranges exist for EFW in relation to the exact gestational age.^{6,7} The interpretation of EFW and the individual biometric measurements generally focuses on two properties: the position of the value on the distribution for the given gestational age and the change in the value over serial measurements. Taking the first of these, infants in the smallest 10% of measurements for gestational age are referred to as SGA and infants in the largest 10% are referred to as large for

gestational age (LGA). The second property examines the growth velocity across the pregnancy. For example, if a fetus is on the 9th percentile at 36 weeks' gestation and it had also been on the 9th percentile at 20 weeks' gestation, it would be regarded as SGA but with normal fetal growth velocity. SGA infants with normal growth velocity are often constitutionally small. SGA combined with evidence of reduced fetal growth velocity is regarded as indicating FGR.⁸

Another major category of measurement in ultrasound in late pregnancy is Doppler flow velocimetry (referred to as 'Doppler').⁹ In brief, a blood vessel is imaged and electronic callipers on the screen are placed over the vessel. The machine then plots out the velocity of flow on the y-axis, with time on the x-axis. The resultant plot is termed a flow-velocity waveform. Different blood vessels have different patterns of flow-velocity waveform and the pattern is analysed both qualitatively and quantitatively. One of the key blood vessels for study is the umbilical artery. Flow is characterised qualitatively by the direction of flow in end-diastole (i.e. immediately prior to the rise in flow that occurs with a heartbeat, i.e. systole). The normal state is forward flow, but there can be absent flow or even reversed flow. The waveform can also be analysed mathematically, and a number of indices have been described, such as the pulsatility index (PI) and resistance index (RI). The derivation, calculation and detailed interpretation of these indices are described in detail elsewhere.⁹ However, both values correlate positively with the presumed resistance to flow in the vascular bed supplied by the artery. Hence, high values of PI and RI in the umbilical arteries are interpreted as indicating a high resistance to flow in the fetal vascular tree of the placenta. Correlative studies of umbilical artery Doppler and placental microscopy support this interpretation in cases of FGR occurring before 36 weeks' gestation.¹⁰

The four most common sites for Doppler are the umbilical arteries, the maternal uterine arteries, the fetal middle cerebral arteries (MCAs) and the ductus venosus.⁹ In contrast to the other three, it is low resistance in the fetal MCAs that is thought to indicate compromise. The interpretation is that a reduced level of oxygen in the fetal blood leads to cerebral vasodilation and, hence, reduced measures of resistance in the arteries supplying the brain.

Other features that are examined in late pregnancy include the placenta, the amniotic fluid and the fetal presentation. Reporting on the placenta generally focuses on its site in relation to the cervix. Implantation of the placenta over the cervix is called placenta praevia and it can cause massive haemorrhage during labour. Reduced amniotic fluid is called 'oligohydramnios' and increased amniotic fluid is called 'polyhydramnios'. Amniotic fluid volume is quantitatively assessed by measuring the biggest single pool (deepest vertical pool) or by summing the four deepest pools in each quadrant of the uterus (amniotic fluid index) (AFI). One of the simplest findings on scan is the presentation of the fetus. Near term, > 95% of fetuses present head first. Women are examined close to term to assess presentation, but this approach frequently misses infants presenting breech.¹¹ Ultrasound unambiguously establishes the presentation at the time of a scan.

Coupling interventions to scan results

A limited number of disease-modifying interventions can be coupled with ultrasound performed in late pregnancy to alter the outcome of pregnancy. Most of the interventions involve modifications to either the timing of delivery [e.g. induction of labour (IOL)] or the mode of delivery (e.g. delivery by pre-labour caesarean section). One exception to this is breech presentation. It has been known for many years that vaginal breech delivery, although safe for the majority of women, can be associated with complications that could have severe consequences for the infant. Breech delivery is associated with a number of specific complications, such as increased risk of umbilical cord compression and entrapment of the fetal head after delivery of the fetal body. Vaginal breech birth in the UK has been shown to be associated with an absolute risk of death during labour or in the first 4 weeks of life of 8.3 per 1000. Although the absolute risk is low, it is much higher than the risk associated with

a planned caesarean delivery of 0.3 per 1000.¹² The risks associated with vaginal breech birth (an awareness of which has long predated the epidemiological study confirming the higher risk of death) were the basis for the procedure to turn the infant from breech to a cephalic presentation using manual manipulation by a clinician, called external cephalic version (ECV). If this procedure is unsuccessful, generally, delivery by planned caesarean section is recommended.¹³ This is based both on the observational data of increased risks associated with vaginal breech birth and on the results of randomised controlled trials (RCTs) of planned caesarean section, which have confirmed the reduced risk of perinatal death with this procedure, compared with planned vaginal breech birth.¹⁴

In the case of most of the other diagnoses that may be made by ultrasound, the primary disease-modifying intervention in the second half of pregnancy is to deliver the infant either by IOL or by planned caesarean section. However, screening may also be used to inform the assessment of fetal well-being to help inform the timing of this intervention. For example, if an infant is found to be SGA and FGR is suspected, there are multiple ways to assess the well-being of the infant. However, these simply constitute another layer of diagnostic and prognostic tests, and ultimately, are used to target the timing of disease-modifying interventions in delivery. The primary reason for expediting delivery is that IOL removes the subsequent risk of stillbirth (i.e. intrauterine fetal death followed by the delivery of an infant showing no signs of life). Most cases of stillbirth are due to complications that can occur to the fetus only in utero (e.g. placental abruption or placental failure); hence, delivering the fetus removes the risk of stillbirth.¹⁵ This is confirmed by RCTs that demonstrate that IOL at term is associated with a 67% reduction in stillbirth risk.¹⁶

Although early delivery can be performed safely at term, this is not the case preterm. A Cochrane review¹⁶ described exactly the same reduction in the risk of perinatal death with IOL at term as was observed for stillbirth. Perinatal deaths include both stillbirths and neonatal deaths, and hence the favourable effect of IOL on stillbirth was not cancelled out by an unfavourable effect on the risk of neonatal death. However, preterm birth is one of the major determinants of neonatal death, and, therefore, if women are routinely induced preterm, reducing the risk of stillbirth will be outweighed by the increased risks of intrapartum stillbirth and neonatal death associated with prematurity. The inflection point (i.e. where the risks balance out) has previously been estimated as between 38 and 39 weeks' gestation.¹⁷ Hence, although 37 weeks' gestation is, strictly, term, routinely delivering all women at 37 weeks' gestation could increase overall perinatal mortality as a result of higher rates of intrapartum stillbirth and neonatal death.¹⁸ It follows, therefore, that screening using a test with a high false-positive rate has the potential to cause net harm by increasing iatrogenic prematurity (or early term delivery) in false positives.¹⁹

Evidence for screening using universal late pregnancy ultrasound

There is strong evidence to support the use of ultrasound scanning in high-risk pregnancies. A systematic review of umbilical artery Doppler has shown that this procedure reduces perinatal mortality by about 30% in high-risk pregnancies.²⁰ The mechanism of the effect is likely to be explained by the fact that its use is also associated with lower rates of IOL and caesarean delivery. Hence, it is likely that the use of Doppler reduces the risk of perinatal death overall by reducing unnecessary intervention. However, there is also a strong trend towards a reduced risk of stillbirth, indicating that Doppler may also be useful for targeting intervention to the highest-risk pregnancies.

The fundamental role of ultrasound scanning in the care of high-risk women led researchers to explore whether or not routinely using the same approaches might improve outcomes in low-risk women. Disappointingly, a meta-analysis of 13 RCTs comprising \approx 35,000 women did not demonstrate any evidence that routine ultrasound scanning improved outcome.²¹ It is this finding that has led to the recommendation that ultrasound should not routinely be performed in the second half of pregnancy in

the UK and the USA. The cautious approach is supported by some evidence from countries where universal late pregnancy ultrasound has been introduced, despite the lack of strong evidence supporting its clinical effectiveness. A seminal study²² from France reported rates of adverse perinatal outcome in relation to women's screening status for SGA. Each woman's screening status was identified [screened positive for SGA or screened normal, i.e. appropriate for gestational age (AGA)] and the actual status of the infant at birth was also assessed (SGA or AGA by actual birthweight). The authors subsequently described rates of perinatal morbidity and mortality by true-positive and false-positive status. As one might have predicted, false positives had higher rates of multiple adverse outcomes than AGA infants that were true negatives, and this was explained primarily by higher rates of iatrogenic prematurity among the false positives. Interestingly, the true-positive SGA infants also had higher rates of adverse outcomes that were missed by scanning than SGA infants (false negatives). The former observation confirms that screening has the potential to result in iatrogenic harm to false positives. The latter observation questions the rationale for screening for SGA infants in late pregnancy at all.

Critical analysis of the Cochrane review¹⁶

Although it is generally accepted that a systematic review of RCTs represents the highest level of evidence, a number of features of the systematic review of RCTs of universal ultrasound²¹ undermine its main conclusions.

- All of the 13 studies in the meta-analysis used different definitions of 'screen positive'. Moreover, some of the ultrasound findings were completely divergent. For example, whereas multiple studies analysed some variant of an estimation of fetal size, one large study assessed placental calcification without assessing any other features of the scan. An implicit assumption around combining these studies is that these different ultrasonic tests all had comparable effectiveness, which a subsequent systematic review of diagnostic test accuracy (DTA) studies has demonstrated is not the case.²³
- None of the studies was preceded by a high-quality assessment of the diagnostic effectiveness of the test in a low-risk population. This is problematic for a number of reasons. A key element of study design is a power calculation. It is impossible to perform a power calculation without quantitative information on the diagnostic effectiveness of a test. Moreover, the tests had generally been developed for and evaluated in high-risk populations. It is well recognised in screening that test performance differs according to the risk status of the population. One of the key outcomes of a screening test is the positive predictive value (PPV) (i.e. the proportion of women screening positive who experience the outcome). The PPV of a test is determined by the prior risk of disease multiplied by the positive likelihood ratio ($LR+ =$ the proportional increase in the odds among screen-positive women compared with the whole population). Hence, the higher the prior risk of disease, the higher the PPV for a given positive likelihood ratio. Consequently, it is typical that a positive screening test is associated with a much lower PPV in a low-risk population. As the PPV determines the ratio of true positives to false positives, this will have a major impact on trials of screening.
- None of the 13 RCTs coupled the screening test with an intervention. In all 13 studies the result was revealed to the attending clinicians but no specific intervention was planned. It is self-evident that a screening test could have an impact on an outcome only if it is coupled with an intervention. Moreover, the tests were performed at a wide range of gestational ages. Given that the primary intervention available to the attending clinicians would have been delivery of the infant, the potential for this to result in benefit or harm would vary according to the gestational age at which the scan was performed. Hence, a positive effect of late pregnancy ultrasound and delivery could have been masked by a negative effect of preterm pregnancy ultrasound scan with higher rates of iatrogenic harm.

- Although the meta-analysis included 35,000 women, it was still underpowered for the key outcome of interest: perinatal death. The risk ratio for perinatal death from the meta-analysis was 1.01 with a 95% confidence interval (CI) of 0.67 to 1.54. Although this CI may seem quite narrow, the capacity for reducing the rate of an outcome with a screening trial is different from interventional trials in women with established disease. If we identified a screening test for perinatal death with a positive likelihood ratio of 10 and a 5% screen-positive rate, and if we applied an intervention that reduced the risk by 50%, the estimated relative risk would be 0.76, which is within the 95% CI of the systematic review. Hence, the Cochrane review¹⁶ is underpowered to detect the effect of a highly effective screening test coupled with a highly effective intervention. If we use the 5.8 per 1000 perinatal mortality rate in the control group of the Cochrane review, a power calculation indicates that a sample size of 110,000 women would be required to detect this effect with 90% power.

Parity and the risk of adverse outcome

One of the most important determinants of adverse pregnancy outcome is obstetric history (i.e. the outcome of previous pregnancies). Many conditions of pregnancy have quite high risks of recurring in subsequent pregnancies, such as pre-eclampsia,²⁴ preterm birth,²⁵ stillbirth²⁶ and FGR.²⁷ Hence, women who have experienced complications in previous pregnancies generally receive enhanced antenatal care. Conversely, uncomplicated previous pregnancies are strongly predictive of a normal outcome in future pregnancies. Hence, women who have had a previous vaginal delivery of a normally grown liveborn infant at term following an uncomplicated pregnancy have a low absolute risk of complications in future pregnancies.²⁸ Obstetric history is, necessarily, not available for women who have not had previous births. Although maternal characteristics, as described above, are associated with the risk of pregnancy complications, the associations are generally rather weak and perform poorly as a screening test in isolation.²⁹ Moreover, first pregnancies, collectively, have higher rates of complications than second pregnancies. This increased rate of complications has identified first pregnancies as a priority area for research. Quoting a National Institutes of Health (NIH) study description of nulliparous women:

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

Haas et al.³⁰

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

The characteristics above provide the rationale for the focus of this review. Screening and intervention near term has less potential to cause harm than screening and intervention in the preterm period, as the primary intervention, delivery of the infant, is less likely to lead to iatrogenic injury. The need for screening is greatest in the nulliparous population because their background suggests that they are at higher risk of an adverse outcome and they lack one of the key discriminating characteristics of risk assessment: knowledge of the outcome of prior births.

The health economics of screening and intervention

A critical consideration in relation to screening and intervention using universal ultrasound is whether or not this screening is cost-effective. It is possible that, for the individual woman and infant, having a screening ultrasound scan and associated intervention leads to a better outcome but that the cost of providing the screening test and intervention results in net societal harm as it removes resources from other more cost-effective elements of the health-care system. The capacity of all health-care systems is finite; however, systems differ in their willingness to pay (WTP). These questions are addressed

quantitatively in health economic analyses by calculating the sum of money required to gain one additional quality-adjusted life-year (QALY), a subject that is discussed in detail elsewhere.³¹ In NHS England, interventions are considered cost-effective if the cost of each QALY is below a given threshold, and this is typically between £20,000 and £30,000.

Providing a late pregnancy ultrasound scan will clearly incur direct costs. Managing women who are assessed as high risk after screening will clearly incur further costs. However, these additional costs then have to be set against the reduction in harm (i.e. the QALYs gained by the mother or child because of being screened). Many of the individual elements required for these calculations are associated with uncertainty. Hence, these health economic analyses frequently employ a probabilistic approach, running large numbers of simulations where the different parameters for the models are sampled from the presumed plausible range of values from the literature. These methods and their interpretation are discussed in more detail in *Chapter 11*.

Value-of-information analysis

The health economic analyses described above relate to the economic case for implementing a given programme of screening and information. Value-of-information (VOI) analysis addresses the economic case for funding research to try to reduce the uncertainty in the evidence base. Generally speaking, a research question that will be identified as being cost-effective from this perspective will have uncertain input values (i.e. the CIs for the given parameter in the literature are wide). Moreover, questions identified as being cost-effective in a VOI analysis will often generate highly variable results in sensitivity analyses in which the input value of the parameter is varied within the range of uncertainty. This subject is again dealt with in detail in *Chapter 11*.

Designing a randomised controlled trial

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

We previously reviewed the approach to screening in pregnancy³² and highlighted an alternative, namely that all women in a population be screened and that randomisation is to either revealing or masking the result. Those that have the result revealed will have an intervention as required, and those that have the result masked will receive routine care. Using this design, randomisation is performed in a group that has a higher rate of complications (by virtue of the positive screening test) and a greater proportion of the adverse events will be related to disease being screened for. This approach has the advantages that the overall number needed to screen for statistical power is substantially reduced and that the screening test can be validated in the same study design by comparing screen-negatives with screen-positives randomised to have the result masked. These issues are discussed further in *Chapter 13*.

Chapter 2 Objectives

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

- to assess the diagnostic effectiveness of late pregnancy ultrasound in nulliparous women based on the existing research literature
- having identified the key ultrasonic findings that identified women as high risk, to review the existing literature and current guidelines to identify a management plan for women with high-risk characteristics
- to conduct a health economic analysis of the likely cost-effectiveness of screening and intervention based on the best available evidence of the costs, diagnostic effectiveness of ultrasound and clinical effectiveness of intervention
- to perform a VOI analysis to determine whether or not there is a strong economic case for funding future research in this area
- conditional on the above, to outline the design of a RCT that could strengthen the evidence base relating to the issues above.

Chapter 3 Identifying the research questions

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

The survey was completed by 54 respondents: 20 consultant obstetricians, eight obstetricians in training, 18 midwives, five sonographers and three consultant neonatologists. All replies were anonymous.

The first question was about identifying the most important ultrasonography findings for universal screening in late pregnancy. The most important findings (ranked by frequency of response) were abnormal fetal biometry or growth velocity (83%), malpresentation (63%), abnormal amniotic fluid volume (63%), high-resistance pattern of umbilical artery Doppler flow velocimetry (32%) and abnormal cerebroplacental ratio (CPR) or MCA Doppler (22%).

The second question was about identifying the most important adverse pregnancy outcomes (apart from perinatal death). The most important outcomes (ranked by frequency of response) were hypoxic–ischaemic encephalopathy (69%), fetal asphyxia (low umbilical cord blood pH plus a base deficit consistent with metabolic acidosis) (64%), SGA or severe SGA (51%), severe shoulder dystocia (46%), breech presentation diagnosed in labour (41%), admission to neonatal intensive care unit (28%) and a low 5-minute Apgar score (21%).

Having completed the survey, we then searched relevant databases (MEDLINE, EMBASE and the Cochrane Library) to identify any other systematic reviews of DTA that might overlap with our aims. This yielded a protocol for a Cochrane DTA review of ultrasonic diagnosis of SGA (which was subsequently published in 2019).²³ Hence, we did not include this in our own plans. We also identified a previously published systematic review³³ of DTA on severe oligohydramnios that was published in 2014 and included publications up to 2011. We selected the studies in this review that were performed in low- and mixed-risk pregnancies and then we performed a literature search for eligible studies published after the search date of the 2014 paper. We then performed a meta-analysis of all relevant studies.

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

1. high-resistance pattern of umbilical artery Doppler flow velocimetry
2. low CPR
3. severe oligohydramnios
4. borderline oligohydramnios
5. suspected fetal macrosomia.

All five of these priority subjects were written up in a single study protocol and the analyses were registered on the International Prospective Register of Systematic Reviews PROSPERO as CRD42017064093.

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

High-resistance patterns of umbilical artery Doppler flow velocimetry are thought to reflect placental vascular resistance. This method is currently in widespread clinical use to monitor high-risk pregnancies, including those with suspected FGR. A Cochrane review of RCTs has demonstrated that use of umbilical artery Doppler ultrasound in high-risk pregnancies appears to reduce the number of perinatal deaths and the number of obstetric interventions (risk ratio 0.71, 95% CI 0.52 to 0.98).²⁰ However, a Cochrane review of RCTs in low-risk pregnancies failed to demonstrate any difference in outcome between pregnancies screened using umbilical artery Doppler and control pregnancies (risk ratio 0.80, 95% CI 0.35 to 1.83).³⁴ This review included five studies that compared routine Doppler with no Doppler, but there was no consistent management plan for the women who had abnormal results. Moreover, although the review comprised 14,185 women, it was underpowered to detect an effect on perinatal death using clinically plausible estimates of screening performance and the clinical effectiveness of intervention.³² The authors concluded that there is no adequate evidence that the routine use of umbilical artery Doppler ultrasound benefits either the mother or the infant and they recommended that future studies should be designed to detect smaller changes in adverse perinatal outcome. The aim of this chapter was to provide level 1 evidence on the diagnostic accuracy of third-trimester umbilical artery Doppler to predict adverse pregnancy outcome at term. We conducted a systematic review and meta-analysis of all studies focusing on low- and mixed-risk populations. In this analysis, we also included data from a prospective cohort study of nulliparous women, the Pregnancy Outcome Prediction (POP) study.^{8,35}

Methods

Analysis of data from the Pregnancy Outcome Prediction study

In the systematic review we included data from a prospective cohort study, the POP study,³⁵ which was conducted at the Rosie Hospital, Cambridge, UK, between 2008 and 2012 and previously has been described in detail.³⁶ In brief, the study included nulliparous women only, and all women who agreed to participate underwent two research ultrasound scans, one at 28 weeks' gestation and one at 36 weeks' gestation, the results of which were not disclosed to the women and the clinicians. About 40% of the women had clinically indicated ultrasound scans in the third trimester, based on local and national guidelines. In the present analysis we included women who attended their 36 weeks' gestation research scan and had a live birth at the Rosie Hospital. Women who delivered prior to their 36 weeks' gestation scan appointment were excluded. Screen positive was defined as an umbilical artery PI > 90th percentile. A full description of the study, including definition of outcome data and the results on the diagnostic effectiveness of ultrasound as a screening test for SGA, has been published in *The Lancet*.⁸

Sources for meta-analysis

The protocol for the review was designed a priori and registered with the International Prospective Register of Systematic Reviews PROSPERO (registration number CRD42017064093). We searched MEDLINE, EMBASE and the Cochrane Library from inception to March 2019. The studies were identified using a combination of words related to 'ultrasound', 'Doppler', 'umbilical artery', 'pregnancy' and 'prenatal diagnosis' (see *Appendix 1*). No restrictions on language or geographical location were applied.

Study selection

Selection criteria included cohort or cross-sectional studies including women with singleton pregnancies who had an ultrasound performed at ≥ 24 weeks' gestation. Case-control studies were excluded as these overestimate the effect size. We included all studies in which the ultrasound was performed as part of universal ultrasound screening (ultrasound was offered to all women regardless of indication), studies that were carried out in low-risk populations (those that excluded pregnancies with any maternal or fetal complication) and studies in a mixed-risk population (ultrasound was offered selectively based on current clinical indications). We excluded studies that were focused only on high-risk populations, such as pregnancies with FGR. We included all reported indices of umbilical artery Doppler, such as the PI, the RI or the systolic-diastolic ratio, as well as all reported cut-off values. In addition, we included studies regardless of whether or not the clinicians were blinded to the ultrasound results but this was reported in the study characteristics.

Study quality assessment and data extraction

The literature search, study selection and analysis were performed independently by two authors (AM and TB) using Review Manager 5.3 (The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark). Any differences were resolved in discussion with the senior author (GS). The risk of bias in each included study was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool,³⁷ which is the recommended tool by the Cochrane Handbook of Diagnostic Test Accuracy Studies. We used a predesigned data extraction form to extract information on study characteristics (i.e. year of publication, country, setting, study design, blinding), patient characteristics (i.e. inclusion and exclusion criteria, sample size), the index test (i.e. gestational age at scan, Doppler indices and cut-off values used) and reference standard (i.e. pregnancy outcome, gestational age at delivery and interval from scan to delivery).

Statistical and meta-analysis methods

From each study we extracted the 2×2 tables for all combinations of index tests and outcomes and we calculated the sensitivity, specificity and positive and negative likelihood ratios (LRs). For the data synthesis we used the hierarchical summary receiver operating characteristic curve model of Rutter and Gatsonis.³⁸ Whenever four or more studies were available, estimates of mean sensitivity and specificity and their variances at a specific threshold were additionally generated using the bivariate logit-normal model.³⁹ We also pooled the diagnostic odds ratios (DORs) using the method described by Deeks.⁴⁰ For the assessment of publication bias we used the Deeks' funnel plot asymmetry test, in which a p -value of < 0.05 was defined as significant asymmetry.⁴¹ As this method requires a large number of studies, we used the most commonly reported outcome for the analysis. For the statistical analyses we used the *metandi*, *metan* and *midas* packages in Stata[®] version 14 (StataCorp LP, College Station, TX, USA).

Results

The Pregnancy Outcome Prediction study

Initially, we analysed the data from the POP study.³⁵ The analysis included 3615 women who met the inclusion criteria (see *Appendix 1, Figure 25*). All women had a blinded umbilical artery ultrasound scan at 36 weeks' gestation and 346 (9.6%) had an umbilical artery PI > 90 th percentile (see *Appendix 1, Figure 25*). Maternal age, socioeconomic status, ethnicity, body mass index (BMI), and rates of alcohol consumption and smoking were similar in the two groups (see *Appendix 1, Table 18*). Moreover, the groups had similar rates of pre-existing hypertension, pre-eclampsia, type 1 and 2 diabetes and gestational diabetes. Gestational age at delivery and rate of IOL were similar in both groups, which can be attributed to the blinding of the ultrasound. The screening performance of umbilical artery PI > 90 th centile is presented in *Table 1*. A high-resistance pattern of umbilical artery Doppler was associated with an increased risk of delivering a SGA infant or a severely SGA infant and the

TABLE 1 Diagnostic performance of umbilical artery PI > 90th centile in predicting adverse pregnancy outcome in the POP study (*n* = 3615)

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

NICU, neonatal intensive care unit.

a See Sovio *et al.*⁸ for definitions.

association was stronger for the latter outcome. However, the finding was not strongly predictive, with positive LRs between 2.5 and 3.5. A high-resistance pattern of umbilical artery Doppler was not associated with an increased risk of a range of indicators of neonatal morbidity in the POP study.

Meta-analysis

The literature search Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram is presented in *Appendix 1, Figure 26*. We identified 13 studies^{35,42-53} that met our inclusion criteria and these involved a total of 67,764 patients. The study characteristics are presented in *Appendix 1, Table 19*. Five studies^{35,42,48,51,52} (*n* = 63,436) included unselected pregnancies as part of universal screening, four studies^{43,46,47,53} (*n* = 2634) included low-risk pregnancies only and four studies^{44,45,49,50} (*n* = 1694) included mixed-risk pregnancies. Three of the studies^{42,51,52} that were done in the same hospitals may have had short periods of overlap. Nine studies^{35,43,44,46-50,53} (*n* = 8097) were prospective and four^{42,45,51,52} (*n* = 59,687) were retrospective. Studies varied in relation to the gestational age at scan (ranging from 28 to 41 weeks' gestation), as well as in the indices and the cut-off points used. The majority of patients in the included studies delivered at term. The assessment of study quality is presented in *Appendix 1, Figure 27*. Overall, the quality was variable. The main risk of bias was that only six studies^{35,43,44,46,48,50} (*n* = 5777) blinded clinicians to the umbilical artery Doppler result. However, five of these six studies revealed other features of the scan result, such as fetal biometry. Only the POP study³⁵ blinded participants to the results of both the uteroplacental Doppler and fetal biometry.

The summary results of the meta-analysis are presented in *Table 2*. The pattern of results was very similar to that in the POP study. A high-resistance pattern detected by Doppler was associated with an increased risk of delivering a SGA infant or a severely SGA infant. However, the finding was not strongly predictive, with positive LRs between 2.5 and 3.0. A high-resistance pattern of umbilical artery Doppler was not associated with an increased risk of a range of indicators of neonatal morbidity. The summary receiver operating characteristic (ROC) curves are presented in *Figure 1*. For some outcomes, such as 5-minute Apgar score of < 7, caesarean section for fetal distress and pre-eclampsia, the Rutter-Gatsonis model could not produce summary results despite an adequate number of studies. We also pooled DORs for all the reported outcomes (*Figure 2*) and illustrated the variation between studies using forest plots.

TABLE 2 Summary diagnostic results of meta-analysis of umbilical artery Doppler for predicting adverse pregnancy outcome

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

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

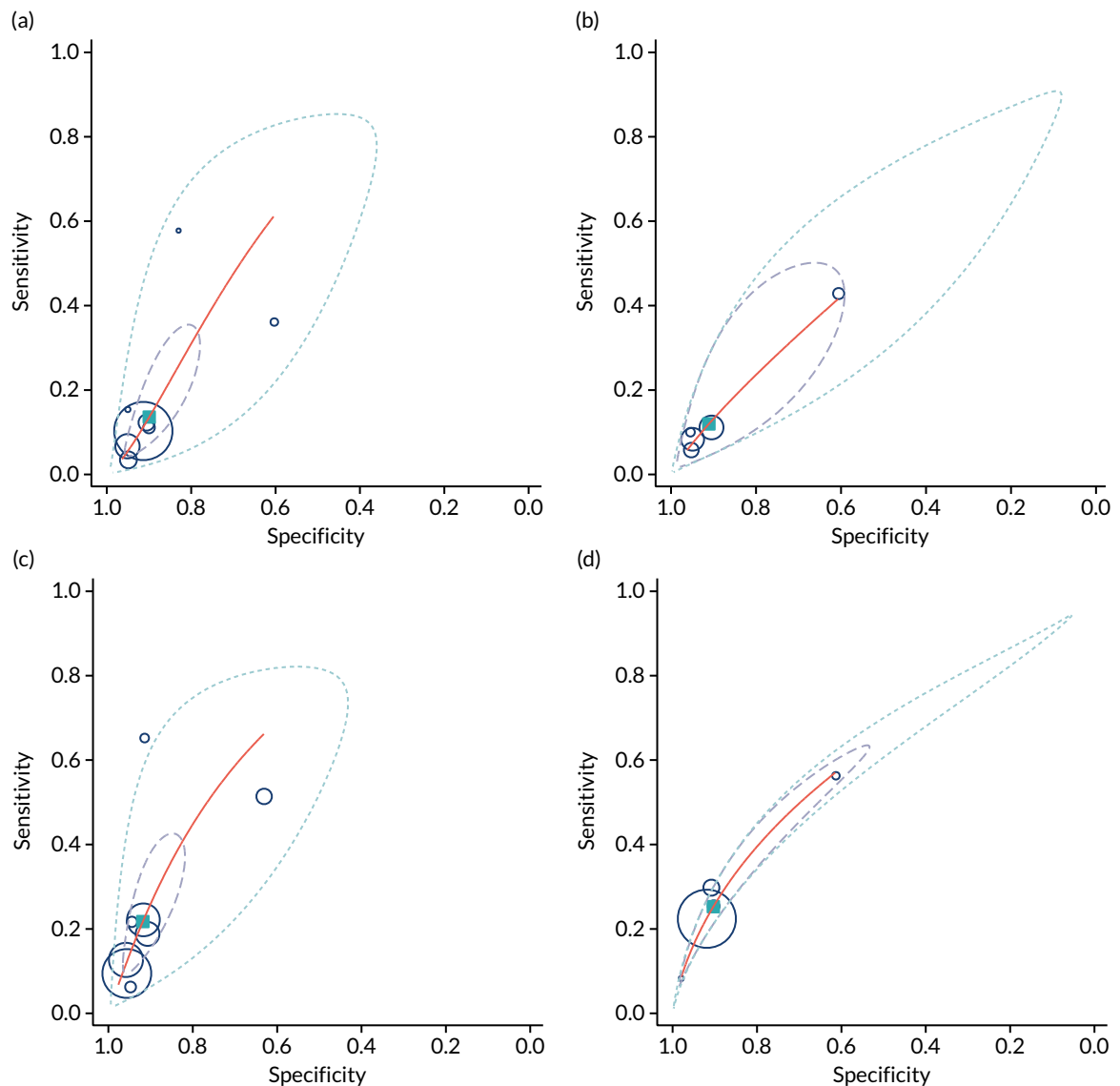


FIGURE 1 Summary ROC curves for umbilical artery Doppler for predicting (a) neonatal intensive care unit admission; (b) neonatal metabolic acidosis; (c) SGA (< 10th centile); and (d) severe SGA (< 3rd centile).

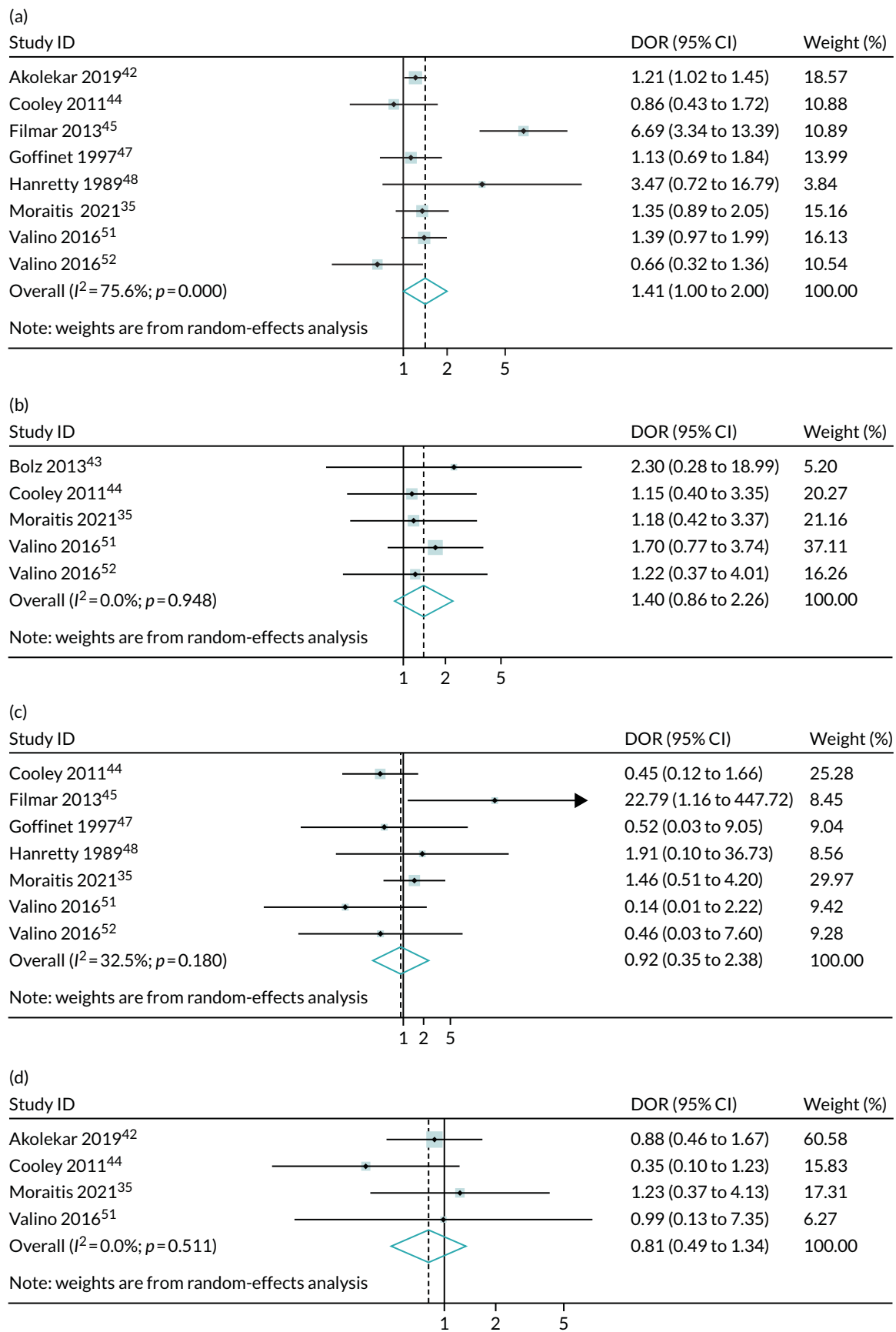


FIGURE 2 Meta-analysis of DORs of umbilical artery Doppler at predicting (a) neonatal intensive care unit admission; (b) neonatal metabolic acidosis; (c) 5-minute Apgar score of < 7; (d) severe adverse perinatal outcome; (e) caesarean section for fetal distress; (f) pre-eclampsia; (g) SGA (< 10th centile); and (h) severe SGA (< 3rd centile). (continued)

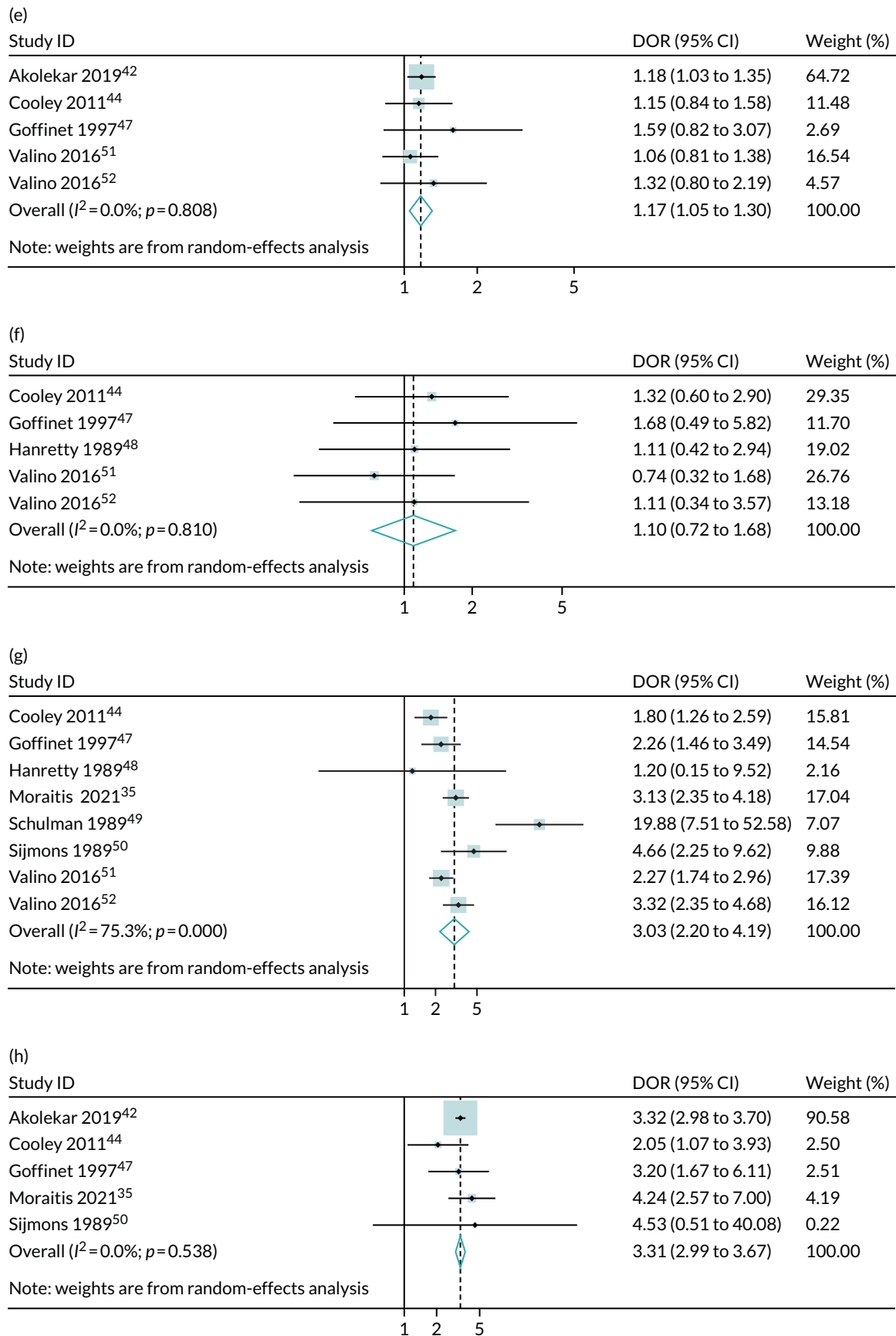


FIGURE 2 Meta-analysis of DORs of umbilical artery Doppler at predicting (a) neonatal intensive care unit admission; (b) neonatal metabolic acidosis; (c) 5-minute Apgar score of < 7; (d) severe adverse perinatal outcome; (e) caesarean section for fetal distress; (f) pre-eclampsia; (g) SGA (< 10th centile); and (h) severe SGA (< 3rd centile).

Finally, we used Deeks' funnel plot asymmetry test to assess the risk of publication bias using the outcome of neonatal unit admission for the analysis (see *Appendix 1, Figure 28*). The test showed no evidence of publication bias ($p = 0.52$).

Discussion

The main finding of this study was that the umbilical artery Doppler has moderate predictive accuracy in detecting SGA and severely SGA infants. However, it did not predict neonatal morbidity at term. The results were very similar in both the POP study and the meta-analysis that included the POP study and other published studies. The only notable difference between the analysis of the POP study and the meta-analysis including the POP study is that the association in the former was slightly stronger for severe SGA. The outcome of SGA is used as a proxy for FGR. As discussed in *Chapter 1*, FGR is a theoretical concept with no gold standard. SGA is used as a proxy for FGR but it is recognised that only a proportion of SGA infants are small because of FGR. As the threshold for defining SGA is lowered, the proportion of cases that are truly FGR increases. Hence, the stronger association with severe SGA is most likely explained by a true association between high-resistance patterns of umbilical artery Doppler and FGR.

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

The lack of association between umbilical artery Doppler and adverse neonatal outcome is likely to be explained by two reasons. First, a minority of term SGA infants have abnormal umbilical artery Doppler. This study showed that about one in five of the SGA infants born below the tenth birthweight centile and one in four of those born below the third birthweight centile had an abnormal umbilical artery Doppler. Second, only a small percentage of overall morbidity at term is associated with abnormal fetal growth. For example, previous studies of perinatal death at term have demonstrated that only one in three stillbirths at term is associated with abnormal fetal growth.⁵⁴ This association would probably be even weaker for other outcomes, such as neonatal intensive care unit (NICU) admission, which includes morbidity for various reasons not related to fetal size, such as neonatal infection. It is plausible that umbilical artery Doppler would be more strongly predictive of adverse neonatal outcome in fetuses who were actually SGA, and this has been confirmed in a previous analysis of the POP study.⁸

Given that umbilical artery Doppler appears to be predictive of FGR in low-risk women, it might be regarded as surprising that the RCTs of its use as a screening test failed to demonstrate any benefit. However, a previous analysis of required sample sizes of screening and intervention to prevent stillbirth demonstrated that, even if a test had a positive LR of 5 for perinatal death, and was observed in 5% of women, and even if the test was coupled with an intervention that reduced the risk of perinatal death by 50%, a RCT of screen versus no screen would need to recruit $\approx 300,000$ women to achieve

90% power (see supplementary figure 10 in Flenady *et al.*⁵⁵). Thus, the Cochrane meta-analysis of low-risk pregnancies is significantly underpowered to identify a reduction in perinatal death.

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

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

Chapter 4 has detailed the fact that a high-resistance pattern of flow in the umbilical artery is most strongly associated with severe SGA, which is thought to be most indicative of FGR. The abnormal flow in the umbilical artery is thought to be related to the pathophysiology of FGR, reflecting impaired perfusion of the placenta due to placental dysfunction. The placenta is the site of gaseous exchange for the fetus and, hence, a consequence of placental dysfunction is that the fetus may have low levels of oxygen in the arterial blood. Physiologically, low levels of oxygen are detected by the central and peripheral arterial chemoreceptors (PACs).⁵⁶ Activation of these receptors initiates compensatory responses, but these differ in fetuses and in adults as, in a fetus, there is no capacity for reversing the low levels of oxygen by increasing ventilation of the lungs (the chemoreceptors stimulate increased depth and frequency of ventilation in extrauterine life). In fetal life, one of the key effects of PAC activation is to reduce the resistance of blood flow to the brain. Clinically, this process is manifested by reduced indices of vascular resistance using Doppler flow velocimetry of the fetal middle cerebral artery due to the cerebral vasodilation caused by the hypoxia.

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

Methods

Sources for meta-analysis

A systematic search was performed using MEDLINE, EMBASE, the Cochrane Database of Systematic Reviews (CDSR) and Cochrane Central Register of Controlled Trials (CENTRAL). The initial search was carried out in June 2017 and was updated on 30 May 2019. No restrictions on language or geographical location were applied. The protocol for the review was designed a priori and registered with the International Prospective Register of Systematic Reviews PROSPERO (registration number CRD42017064093). The studies were identified using a combination of words related to 'ultrasound', 'pregnancy', 'cerebroplacental', 'cerebro-umbilical', 'middle cerebral artery' and 'fetal brain Doppler'. We defined the CPR as the ratio of MCA PI to umbilical artery PI.

Study selection

Selection criteria allowed the inclusion of cohort or cross-sectional studies involving singleton pregnancies in which an ultrasound scan was performed at ≥ 24 weeks' gestation. We included all studies in which the ultrasound was performed as part of universal screening, studies that included low-risk populations only and studies with mixed-risk populations. We excluded studies that were focused on high-risk patients, such as those with FGR, and studies in which ultrasound scanning was performed during labour. We included studies regardless of the threshold used to define abnormality of the CPR and regardless of whether or not clinicians were blinded to the result.

We included studies that reported the following outcomes: severe adverse perinatal outcome (which included stillbirth, neonatal death and hypoxic–ischaemic encephalopathy); fetal growth abnormalities such as SGA (defined as birthweight < 10th centile) and severe SGA (birthweight < 3rd or < 5th centile); adverse neonatal outcomes such as neonatal unit admission, 5-minute Apgar score of < 7, and neonatal metabolic acidosis (as defined in each study); and caesarean section or operative delivery (including both caesarean section and instrumental delivery) for fetal compromise in labour. In cases of significant population overlap between studies that reported the same outcomes, we included the larger study in the meta-analysis. However, if the studies reported different outcomes or performed the ultrasound at different gestational ages, we included both in the meta-analysis.

Study quality assessment and data extraction

The literature search, study selection and analysis were performed independently by two authors (AM and TB) using Review Manager 5.3. Any differences were resolved in discussion with the senior author (GS). The risk of bias in each included study was assessed using the QUADAS-2 tool as outlined in the Cochrane Handbook of Diagnostic Test Accuracy Studies. This tool assesses the included studies for potential bias in four domains: patient selection, index test, reference standard, and flow and timing. We assessed the risk for flow and timing from the perspective of universal ultrasound screening at 36 weeks' gestation. We used a predesigned data extraction form to extract information on study characteristics (i.e. year of publication, country, setting, study design, blinding), patient characteristics (i.e. inclusion and exclusion criteria, sample size), index test (i.e. gestational age at scan, cut-off values used) and reference standard (i.e. pregnancy outcome, gestational age at delivery, and interval from scan to delivery). We also collected information such as parity and rates of IOL, when reported.

Statistical and meta-analysis methods

The statistical and meta-analysis methods employed are described in *Chapter 4*.

Results

The literature search flow chart is presented in *Appendix 2, Figure 29*. We identified 16 studies^{42,57–71} that met the inclusion criteria, which involved a total of 121,607 patients. The study characteristics are presented in *Appendix 2, Table 20*. Four studies^{42,57,58,68} ($n = 85,059$) included unselected pregnancies, seven studies^{59,60,62,63,66,67,70} ($n = 12,929$) included low-risk pregnancies only and five studies^{61,64,65,69,71} ($n = 23,619$) included mixed-risk pregnancies. Nine studies ($n = 87,208$) were prospective and seven ($n = 34,399$) were retrospective. There was population overlap between the Akolekar *et al.*,⁵⁷ Akolekar *et al.*⁴² and Bakalis *et al.*⁵⁸ studies. For the first two we reported different outcomes and for those outcomes that were the same we employed the data from the larger Akolekar *et al.*⁴² study in the meta-analysis. In the study by Bakalis *et al.*,⁵⁸ ultrasound was performed at 32 weeks' gestation, compared with the two Akolekar *et al.*^{42,57} studies, in which ultrasound was performed at around 36 weeks' gestation. There was also population overlap between the Khalil *et al.*,⁶² Monaghan *et al.*,⁶⁴ and Morales-Roselló *et al.*⁶⁵ studies, which reported different outcomes at the same tertiary maternity unit. Moreover, there was population overlap between the Flatley and Kumar,⁶¹ Sabdia *et al.*⁶⁹ and Twomey *et al.*⁷¹ studies. In the study by Twomey *et al.*,⁷¹ ultrasound was performed at 32 weeks' gestation, and the other two studies, in which ultrasound was performed between 35 and 38 weeks' gestation, reported different rates of nulliparity and different gestational age at delivery (Sabdia *et al.*⁶⁹ included preterm deliveries), which indicates that the potential population overlap was not significant. Furthermore, there was a complete population overlap between the studies by Bligh *et al.*,^{59,60} but the two studies reported different outcomes.

The assessment of study quality was performed using the QUADAS-2 tool and is summarised in *Appendix 2, Figure 30*. The main risk of bias was for reference standard because of the lack of blinding in the majority of studies. Only five studies^{59,60,66–68} ($n = 3079$) blinded the clinicians to results. The second most common risk of bias was for flow and timing because of the different gestational ages at which ultrasound was performed. In the studies by Bakalis *et al.*,⁵⁸ Rial-Crestelo *et al.*⁶⁸ and Twomey *et al.*,⁷¹ ultrasound was

performed at around 32–33 weeks' gestation, and in Prior *et al.*^{66,67} and Stumpfe *et al.*,⁷⁰ it was performed prior to IOL (interval between ultrasound and delivery of < 72 hours). Hence, the results of the above studies might not be applicable to universal screening at 36 weeks' gestation. One study⁶³ had unclear risk of selection bias as it did not specify whether the selection of patients was consecutive or random.

The summary results for the diagnostic accuracy of CPR at predicting adverse pregnancy outcomes are presented in *Table 3*. Overall, the strongest associations were with the risk of delivering a SGA or severely SGA infant and the positive LRs were in the region of 3.5–4.0, which was stronger than for umbilical artery on its own. Moreover, unlike umbilical artery Doppler in *Chapter 4*, a low CPR was associated with a statistically significantly increased risk of neonatal morbidity. However, the strength of prediction was weak, with positive LRs of between 1.5 and 3.0.

The summary ROC curves are presented in *Figure 3*. Generally, the larger studies reported lower sensitivities and higher specificities for all the outcomes. We also present the pooling of the DORs in *Figure 4*. These demonstrate that, for many of the outcomes, there was a very high level of heterogeneity between the studies.

Furthermore, we used Deeks' funnel plot asymmetry test to assess the risk of publication bias using the outcome of neonatal unit admission for the analysis. The test showed no significant risk of publication bias ($p = 0.28$; see *Appendix 2, Figure 31*).

Discussion

The meta-analysis demonstrated that the CPR may be slightly more predictive than umbilical artery Doppler in identifying pregnancies at an increased risk of adverse outcome. In the case of SGA, the positive LRs were in the region of 3.5–4.0, compared with 2.5–3.0 for umbilical artery Doppler. Moreover, unlike umbilical artery Doppler, a low CPR was associated with an increased risk of neonatal morbidity.

TABLE 3 Diagnostic accuracy of CPRs in predicting adverse pregnancy outcome

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

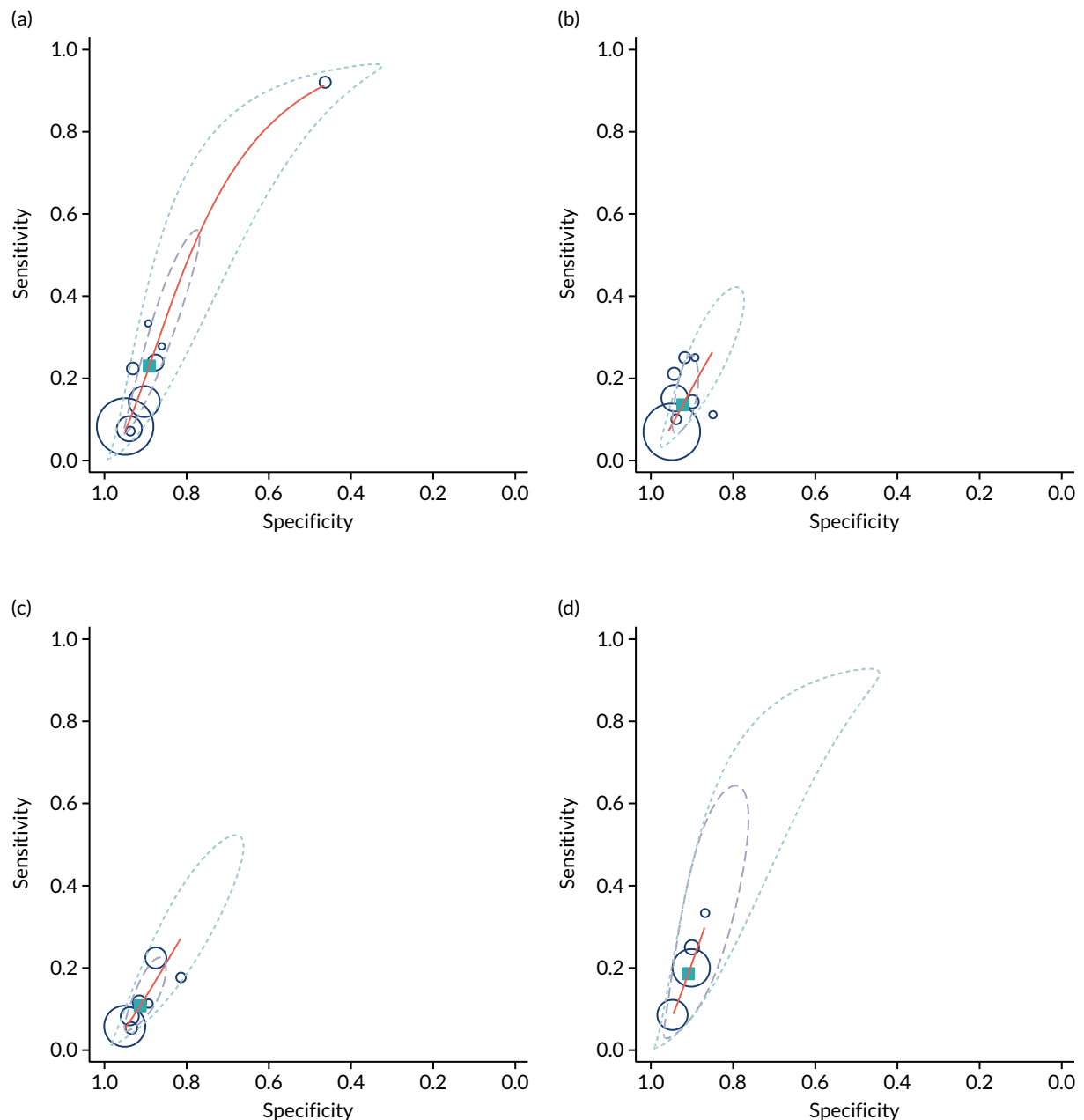


FIGURE 3 Summary ROC curves for the diagnostic performance of abnormal CPRs at predicting adverse pregnancy outcomes. (a) Neonatal unit admission; (b) 5-minute Apgar score of < 7; (c) neonatal metabolic acidosis; (d) severe adverse perinatal outcome (including stillbirth, neonatal death and hypoxic-ischaemic encephalopathy); (e) SGA (birthweight < 10th centile); (f) severe SGA (< 3rd or < 5th centile); (g) caesarean section for fetal distress; and (h) operative delivery for fetal distress (including both caesarean section and instrumental delivery). (*continued*)

However, in this case the strength of prediction was weaker, with positive LR of < 2.0. Moreover, in both analyses, there was very significant heterogeneity in relation to both birthweight-based outcomes and neonatal morbidity. Consequently, the 95% CIs for the positive LR are wide and include the point estimates observed for umbilical artery Doppler for both SGA and severe SGA. Furthermore, given that many of the studies were not blinded, it is possible that the associations with neonatal morbidity were a result of bias. However, the association between the CPR and SGA fetuses indicates that the ratio is likely to predict FGR. Overall, this analysis indicates that the CPR is indeed predictive of adverse pregnancy outcome. However, it is not clear from the present analysis whether or not the

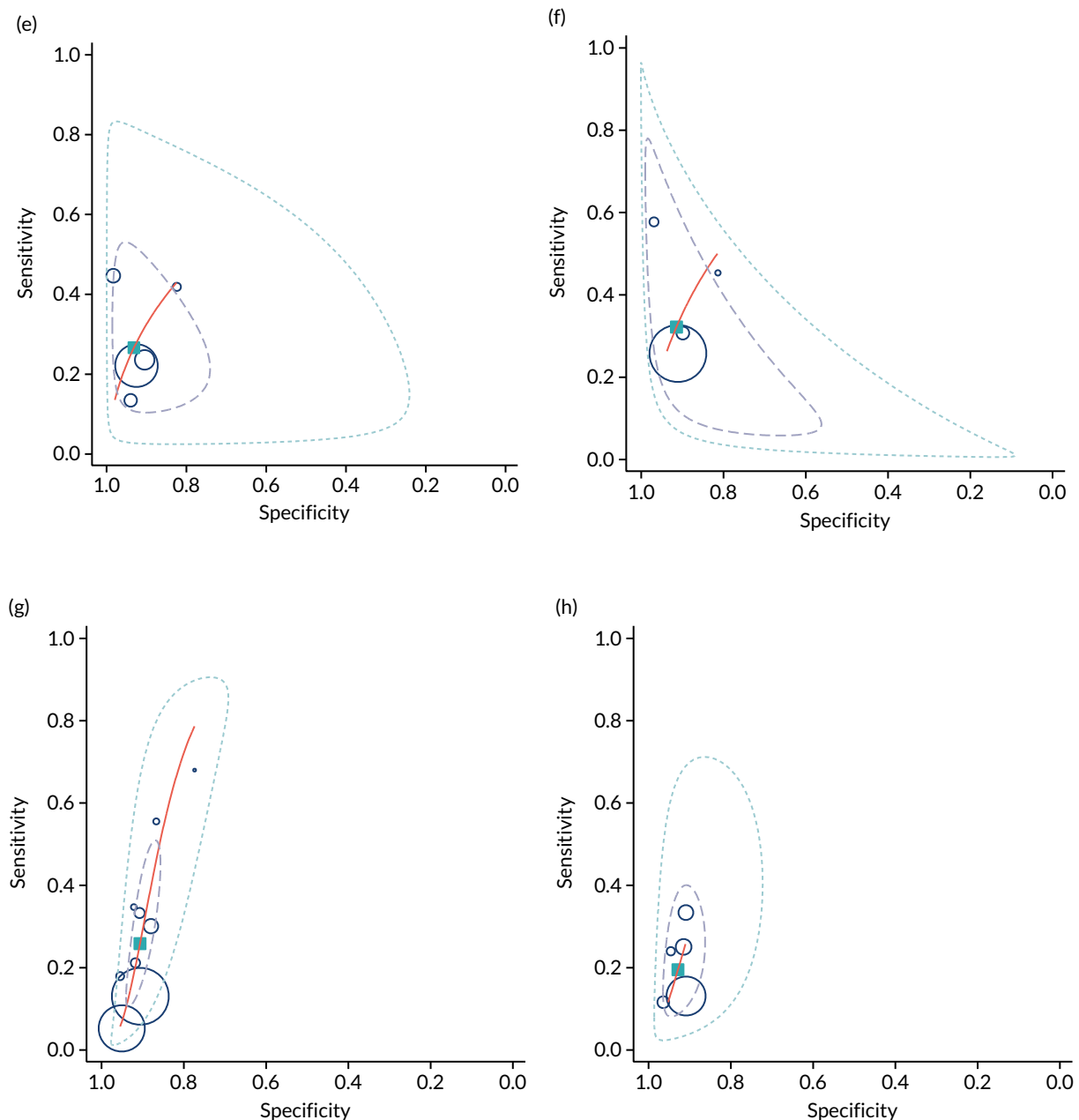


FIGURE 3 Summary ROC curves for the diagnostic performance of abnormal CPRs at predicting adverse pregnancy outcomes. (a) Neonatal unit admission; (b) 5-minute Apgar score of < 7; (c) neonatal metabolic acidosis; (d) severe adverse perinatal outcome (including stillbirth, neonatal death and hypoxic–ischaemic encephalopathy); (e) SGA (birthweight < 10th centile); (f) severe SGA (< 3rd or < 5th centile); (g) caesarean section for fetal distress; and (h) operative delivery for fetal distress (including both caesarean section and instrumental delivery).

ratio performs better than simply assessing the result of umbilical artery Doppler, which is used in its calculation anyway. Of the indices assessed in these sections of the report, only MCA Doppler was not measured in the POP study; hence, unlike in the other chapters, we are unable to compare the strength of association in the POP study with the meta-analysis. Our findings contradict the previously published systematic review,⁷² which concluded that the CPR at term has a strong association with adverse obstetric and perinatal outcomes. We believe that this is because the systematic review by Dunn *et al.*⁷² included studies carried out in mostly high-risk populations, did not include some large, recently published studies that offered ultrasound as part of universal screening^{42,57,58} and did not produce any pooled analyses.

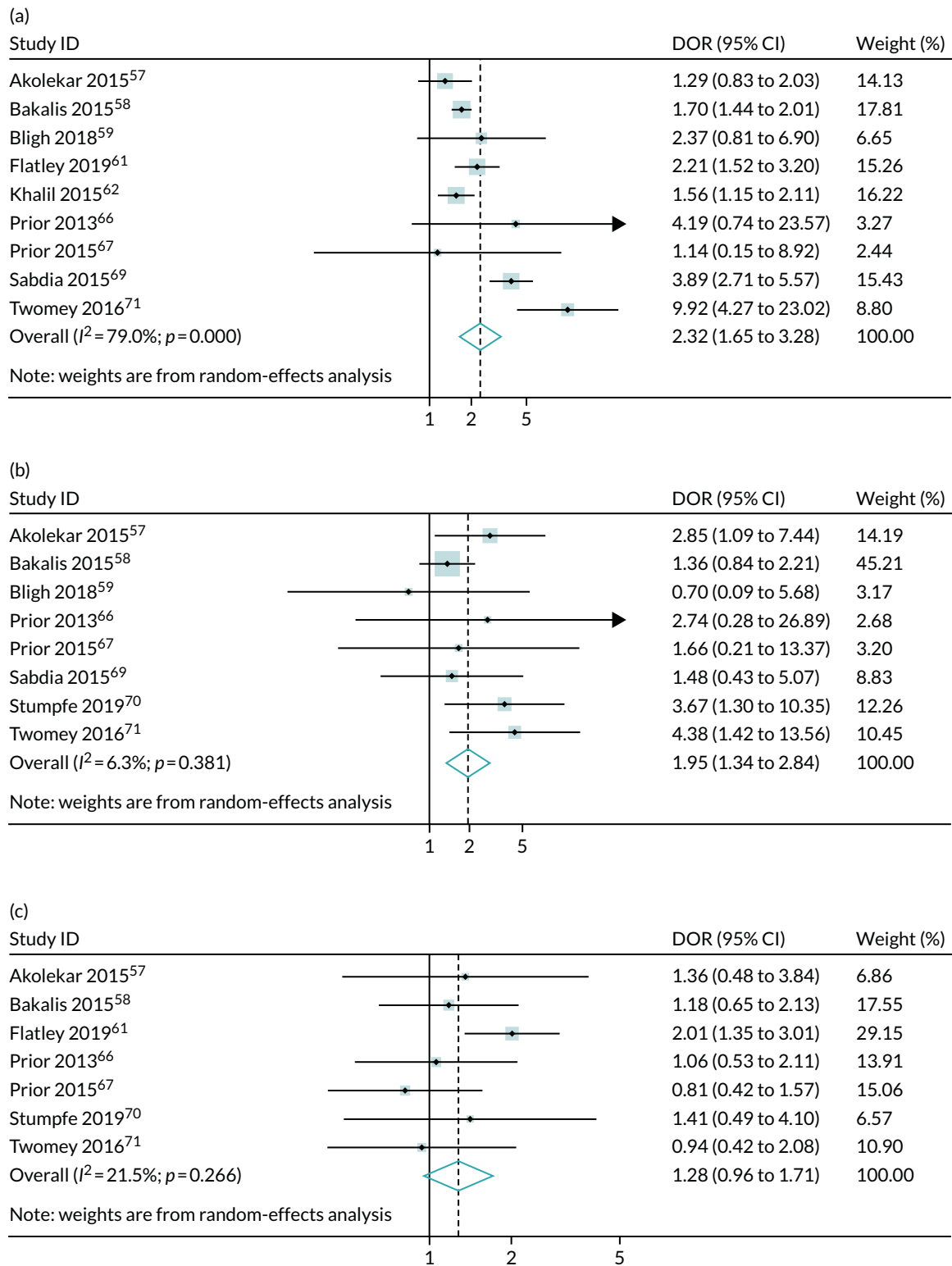


FIGURE 4 The diagnostic odd ratios for the diagnostic performance of abnormal CPRs at predicting adverse pregnancy outcomes. (a) neonatal unit admission; (b) 5-minute Apgar score of < 7; (c) neonatal metabolic acidosis; (d) severe adverse perinatal outcome (including stillbirth, neonatal death and hypoxic-ischaemic encephalopathy); (e) SGA (birthweight < 10th centile); (f) severe SGA (< 3rd or < 5th centile); (g) caesarean section for fetal distress; and (h) operative delivery for fetal distress (including both caesarean section and instrumental delivery). (continued)

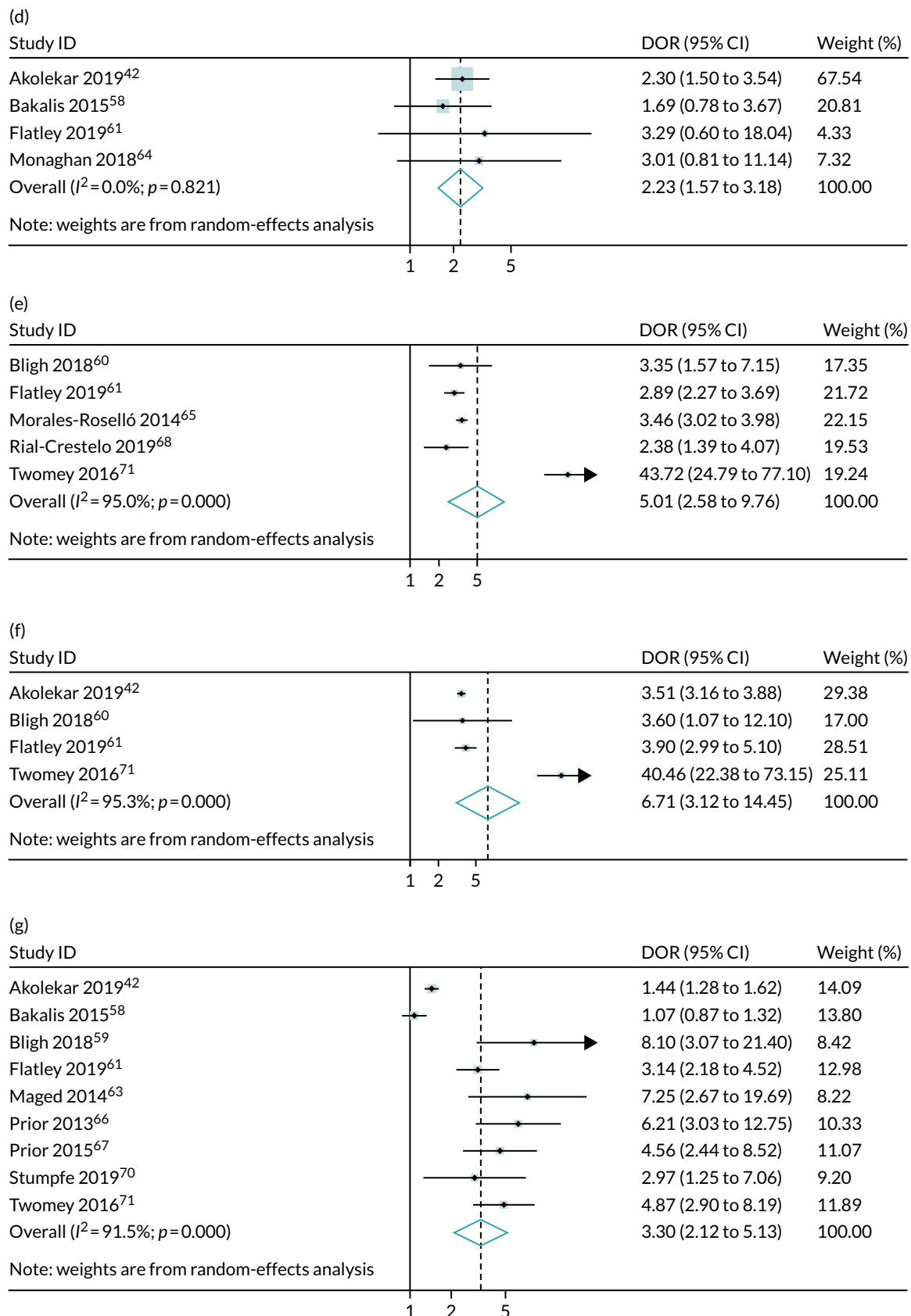


FIGURE 4 The diagnostic odd ratios for the diagnostic performance of abnormal CPRs at predicting adverse pregnancy outcomes. (a) neonatal unit admission; (b) 5-minute Apgar score of < 7; (c) neonatal metabolic acidosis; (d) severe adverse perinatal outcome (including stillbirth, neonatal death and hypoxic-ischaemic encephalopathy); (e) SGA (birthweight < 10th centile); (f) severe SGA (< 3rd or < 5th centile); (g) caesarean section for fetal distress; and (h) operative delivery for fetal distress (including both caesarean section and instrumental delivery). (*continued*)

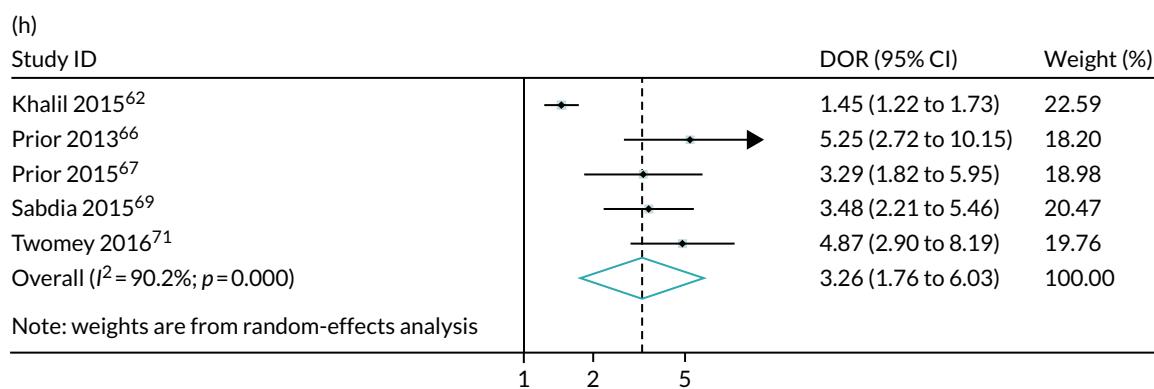


FIGURE 4 The diagnostic odd ratios for the diagnostic performance of abnormal CPRs at predicting adverse pregnancy outcomes. (a) neonatal unit admission; (b) 5-minute Apgar score of < 7; (c) neonatal metabolic acidosis; (d) severe adverse perinatal outcome (including stillbirth, neonatal death and hypoxic-ischaemic encephalopathy); (e) SGA (birthweight < 10th centile); (f) severe SGA (< 3rd or < 5th centile); (g) caesarean section for fetal distress; and (h) operative delivery for fetal distress (including both caesarean section and instrumental delivery).

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

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

Amniotic fluid evaluation is routinely performed as part of the assessment of fetal well-being in the third trimester using ultrasound. Reduced amniotic fluid is called oligohydramnios and increased amniotic fluid is called polyhydramnios. In the second half of pregnancy, the amniotic fluid comes from the fetal urine. Fetuses with no kidneys (renal agenesis) typically have no amniotic fluid at the time of the routine 20 weeks' gestation scan and it remains absent thereafter. However, congenital anomaly is a rare cause of oligohydramnios. One of the common causes of oligohydramnios is rupture of the fetal membranes; in this event, the overall level of fluid is reduced through vaginal loss. In such cases, the normal fetal production of urine in such cases can be confirmed by filling and emptying the fetal bladder. However, fetal distress is thought to be a potential cause of oligohydramnios as a result of reduced fetal urine production. Stress, for example because of arterial hypoxaemia, results in the activation of a number of compensatory responses.⁵⁶ These include increased release of arginine vasopressin (also known as antidiuretic hormone), which has a direct effect on the kidneys. Fetal hypoxia leads to a chemoreceptor-mediated cardiovascular response that increases blood supply to the vital organs (e.g. the heart and brain) but reduces blood flow to the fetal trunk, including the kidneys. The combination of increased arginine vasopressin and reduced renal blood flow will reduce fetal urine output and lead to oligohydramnios. Hence, checking for oligohydramnios has been a feature of ultrasonic assessment of fetal well-being for many years.

The most common methods of quantitative assessment of amniotic fluid volume are the AFI (the sum of the four deepest pockets of amniotic fluid in the four quadrants of the uterus)⁷³ and the single deepest pocket (SDP). Severe oligohydramnios is commonly defined as AFI < 5 cm or SDP < 2 cm. Given the known association between oligohydramnios and fetal stress, the aim of the present study was to produce level 1 evidence of diagnostic effectiveness of severe oligohydramnios in predicting adverse pregnancy outcomes at, or near, term, and so we performed a systematic review and meta-analysis of the literature.

Methods

Sources for meta-analysis

We identified a previous systematic review³³ that was published in 2014 and included source material from publications up to 2011. However, the review did not limit searches to low- or mixed-risk pregnancies. We updated the systematic review to include studies published from 1 January 2011 up to the latest search date of 5 June 2019. The systematic search was performed using MEDLINE, EMBASE, CDSR and CENTRAL. No restrictions on language or geographical location were applied. The studies were identified using a combination of words related to 'ultrasound', 'pregnancy', 'amniotic fluid volume', 'AFI', 'oligohydramnios' and 'single deepest pocket'.

Study selection

Selection criteria allowed the inclusion of cohort or cross-sectional studies involving singleton pregnancies in which an ultrasound scan was performed at ≥ 24 weeks' gestation. We included all studies in which the ultrasound was performed as part of universal screening, studies that included low-risk populations only and studies in mixed-risk populations. These criteria were applied to the studies included in the previously published review and to the studies published subsequent to that review.

We excluded studies that were focused on high-risk patients, such as those with suspected FGR, studies that included pregnancies with preterm premature rupture of membranes, and studies in which ultrasound was performed intrapartum. We included studies that reported the following outcomes: stillbirth; neonatal death; fetal growth abnormalities, such as SGA (defined as birthweight < 10th centile) and severe SGA (i.e. birthweight < 3rd of < 5th centile); adverse neonatal outcomes, such as neonatal unit admission, 5-minute Apgar score of < 7, and neonatal metabolic acidosis (as defined in each study); and caesarean section or operative delivery (including both caesarean section and instrumental delivery) for fetal compromise in labour.

Study quality assessment and data extraction

The literature search, study selection and analysis were performed independently by two authors (AM and DW) using Review Manager 5.3. Any differences were resolved in discussion with the senior author (GS). The risk of bias in each included study was assessed using the QUADAS-2 tool as outlined in the Cochrane Handbook of Diagnostic Test Accuracy Studies.³⁷ This tool assesses studies for potential bias in four domains: patient selection, index test, reference standard, and flow and timing. We assessed the risk of bias for flow and timing from the perspective of universal ultrasound screening at 36 weeks' gestation. We used a predesigned data extraction form to extract information on study characteristics (i.e. year of publication, country, setting, study design, blinding), patient characteristics (i.e. inclusion and exclusion criteria, sample size), the index test (i.e. gestational age at scan, cut-off values used) and reference standard (i.e. pregnancy outcome, gestational age at delivery and interval from scan to delivery). We also collected information such as parity and rates of IOL when reported.

Statistical and meta-analysis methods

The statistical and meta-analysis methods employed are described in *Chapter 4*.

Results

The literature search flow chart is presented in *Appendix 3, Figure 32*. We identified 14 studies⁷⁴⁻⁸⁷ that met our inclusion criteria, which involved a total of 109,679 patients. The study characteristics are presented in *Appendix 3, Table 21*. Two studies^{77,78} ($n = 30,555$) included unselected pregnancies, 10 studies^{74-76,80-85,87} ($n = 61,047$) included low-risk pregnancies only and two studies^{79,86} ($n = 18,077$) included mixed-risk pregnancies. Six studies^{75,78,79,81,82,84} ($n = 5740$) were prospective, six studies^{74,77,80,83,85,86} ($n = 97,022$) were retrospective, one study⁷⁶ ($n = 260$) was cross-sectional and one study⁸⁷ ($n = 6657$) was carried out as part of a clinical trial.

The assessment of study quality was performed using the QUADAS-2 tool and is summarised in *Appendix 3, Figure 33*. The main risk of bias was for reference standard because of the lack of blinding in the majority of studies. Only two studies^{81,84} ($n = 1892$) blinded the results to clinicians, one of which blinded only the AFI result and not the other aspects of the ultrasound. The second, more common, risk of bias was for flow and timing. Two studies^{75,85} performed ultrasound prior to IOL or within 4 days of delivery. Two other studies^{77,82} did not report gestational age at either ultrasound or delivery. Hence, these results may not be applicable for universal third-trimester screening at 36 weeks' gestation. Two studies were rated as having unclear risk of selection bias^{79,86} as they did not report how the patients had been selected and one study⁷⁶ was rated as having high applicability concerns for patient selection as it included prolonged (> 41 weeks' gestation) pregnancies only.

The summary results for the diagnostic accuracy of oligohydramnios at predicting adverse pregnancy outcomes are presented in *Table 4*. The most commonly reported outcomes were neonatal unit admission and caesarean section for fetal distress (11 and 10 studies respectively). The stronger statistically significant association was with SGA < 10th centile, with a positive LR of 2.8 (see *Table 4*).

TABLE 4 Summary diagnostic performance of low AFI (< 5 cm) in predicting adverse pregnancy outcome

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

There were also statistically significant associations with NICU admission and caesarean section for fetal distress, with positive LR of 1.7 and 2.2 respectively. The positive LR for neonatal death was 3.7 but, because of the small number of events, the CIs were very large and include unity. The summary ROC curves are presented in Figure 5. Generally, the larger studies reported lower sensitivities and higher specificities for all outcomes. Figure 6 illustrates forest plots of DORs. Finally, we used Deeks' funnel plot asymmetry test to assess the risk of publication bias using the outcome of neonatal unit admission for the analysis (see Appendix 3, Figure 34). The test showed no evidence of publication bias ($p = 0.54$).

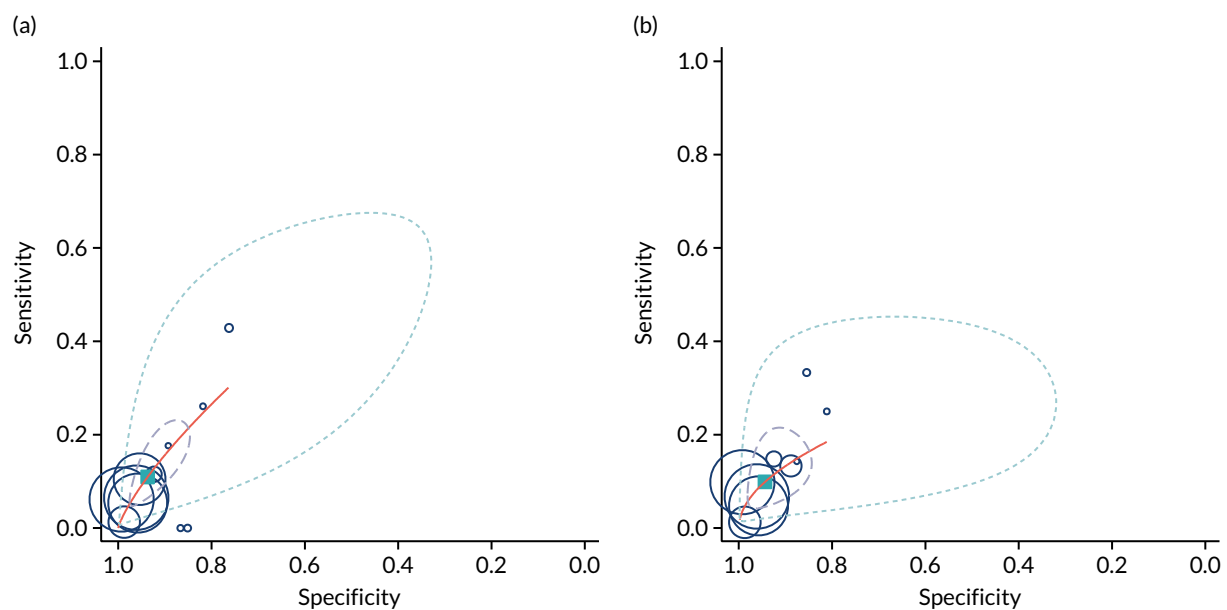


FIGURE 5 Summary ROC curves for AFI < 5 cm at predicting adverse pregnancy outcome. (a) NICU admission; (b) 5-minute Apgar score of < 7; (c) neonatal metabolic acidosis; (d) caesarean section for fetal distress; (e) SGA (< 10th centile); and (f) neonatal death. (continued)

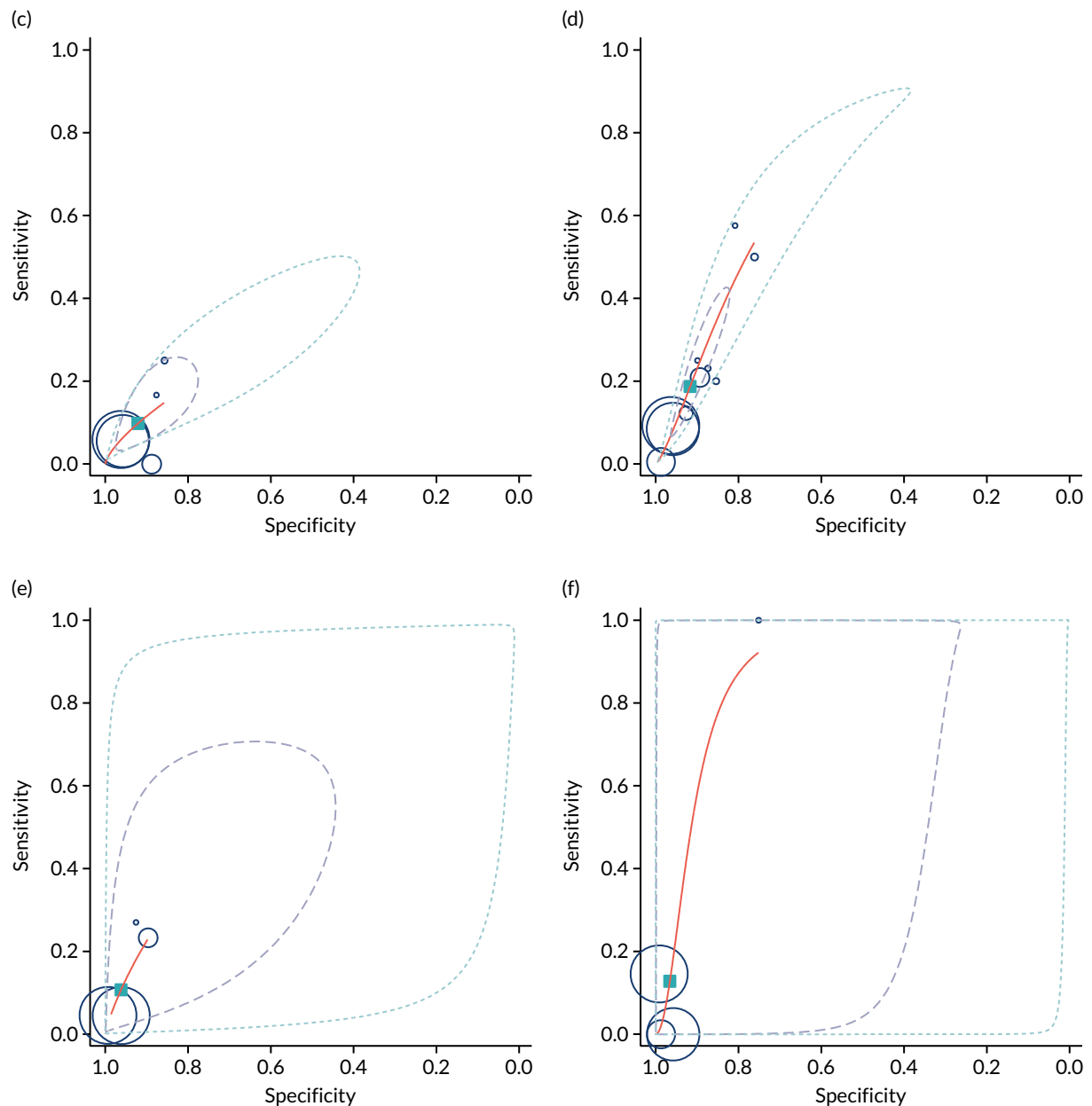


FIGURE 5 Summary ROC curves for AFI < 5 cm at predicting adverse pregnancy outcome. (a) NICU admission; (b) 5-minute Apgar score of < 7; (c) neonatal metabolic acidosis; (d) caesarean section for fetal distress; (e) SGA (< 10th centile); and (f) neonatal death.

Discussion

This meta-analysis confirms that a diagnosis of severe oligohydramnios is associated with adverse pregnancy outcome. The key finding was that severe oligohydramnios had a positive LR for SGA of between 2.5 and 3.0. The associations with admission to NICU and emergency caesarean section for fetal distress are more difficult to interpret. First, for both of these outcomes, the association was weaker than it was for SGA. Second, in both cases the association could have been a consequence of the scan rather than an outcome predicted by the scan. Only two studies, containing < 5% of the patients included in the meta-analysis, blinded the results of the scan. Revealing the results of the scan could explain both associations. In the case of NICU admission, revealing the scan result could lead to a decision to deliver the infant as a result of suspected fetal distress. If this occurs preterm or at early term gestation it could lead to NICU admission as a result of iatrogenic prematurity. In the case of caesarean delivery for fetal distress, revealing the result that there is severe oligohydramnios could be

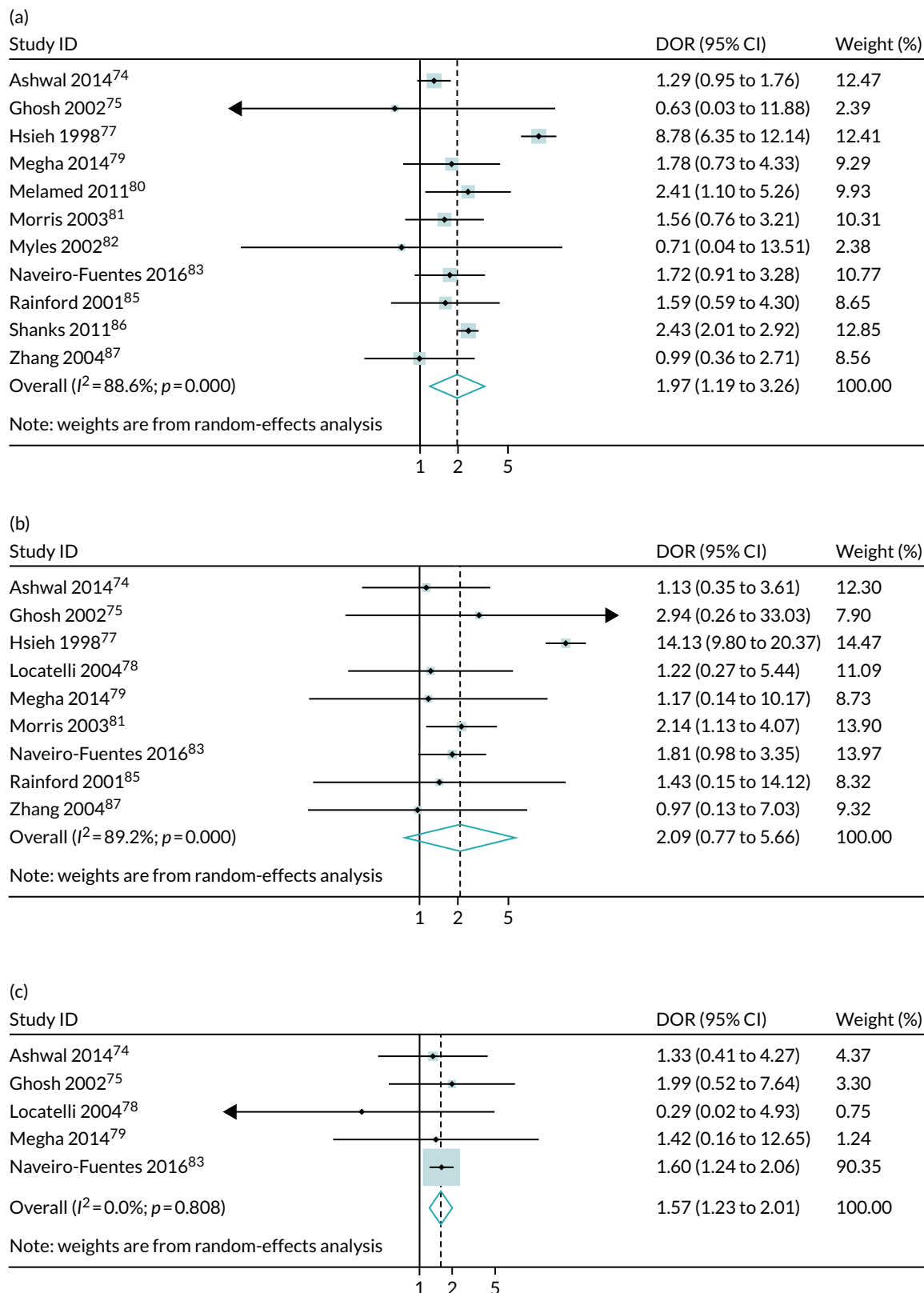


FIGURE 6 Meta-analysis of DORs for AFI < 5 cm at predicting adverse pregnancy outcome: (a) NICU admission; (b) 5-minute Apgar score of < 7; (c) neonatal metabolic acidosis; (d) caesarean section for fetal distress; (e) SGA (< 10th centile); and (f) neonatal death. (continued)

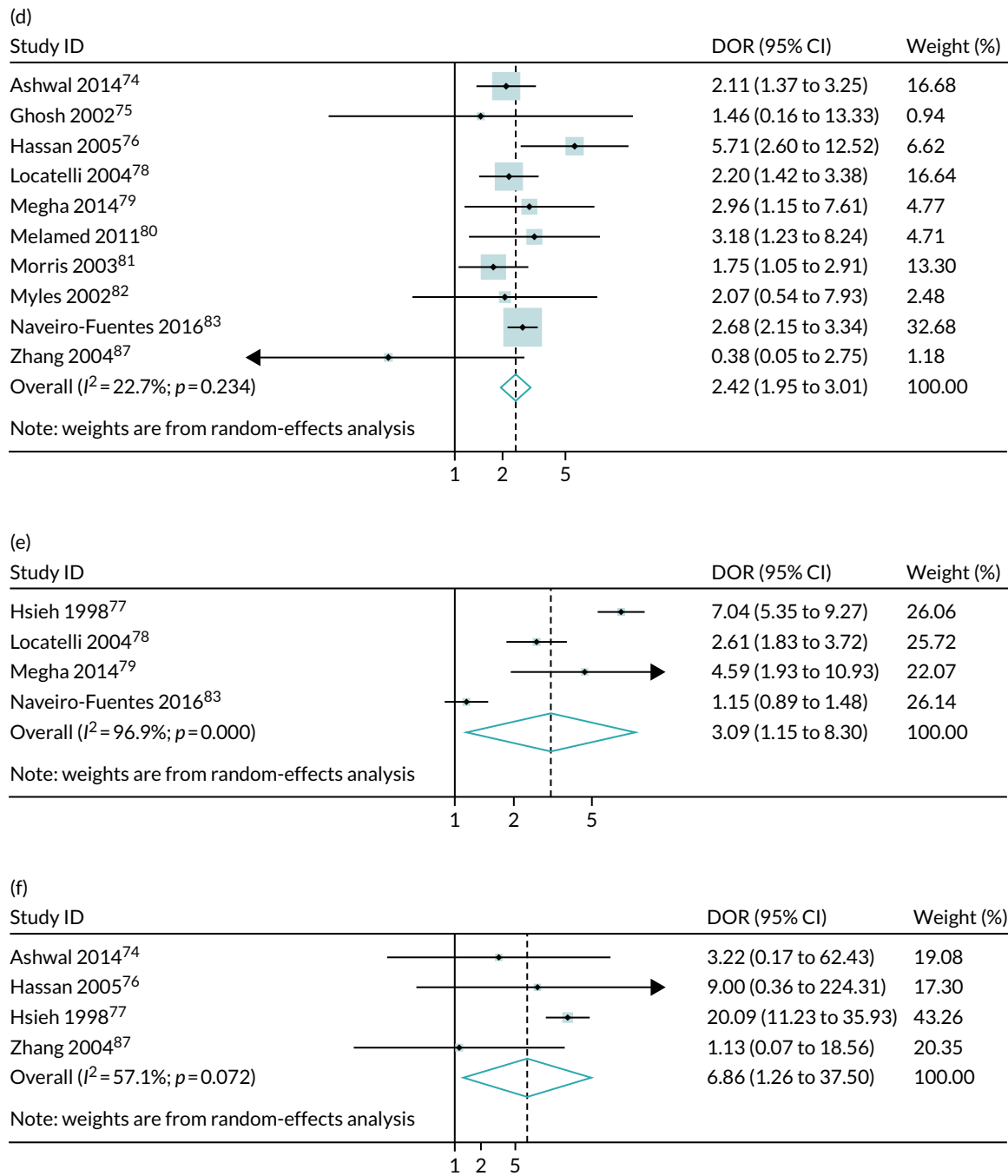


FIGURE 6 Meta-analysis of DORs for AFI < 5 cm at predicting adverse pregnancy outcome: (a) NICU admission; (b) 5-minute Apgar score of < 7; (c) neonatal metabolic acidosis; (d) caesarean section for fetal distress; (e) SGA (< 10th centile); and (f) neonatal death.

used as an indication (in whole or in part) to perform a caesarean section for suspected fetal distress. Alternatively, if a caesarean section was performed for failure to progress, it is possible that the operator may include suspected fetal distress in the indication given the scan finding.

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

In conclusion, this analysis confirms that severe oligohydramnios is associated with adverse pregnancy outcome. This can confidently be stated, as there was an association with SGA, which is much less likely to arise from biases. However, the association between oligohydramnios and neonatal morbidity is less clear. Despite the association with SGA, the positive LR was not very high, and its capacity to act as a screening test in unselected nulliparous women at 36 weeks' gestation is limited.

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

In *Chapter 6*, we assessed the association between severe oligohydramnios and the risk of adverse pregnancy outcome. Although the finding was associated with the risk of SGA, it was not strongly predictive of SGA, and associations with neonatal morbidity were difficult to assess as > 95% of the patients included in the meta-analysis participated in studies in which the results of the ultrasound scan were revealed. The aim of this element of the work was to determine the association between borderline oligohydramnios and adverse pregnancy outcome. First, we aimed to determine whether there was indeed a gradient in the strength of association comparing severe with borderline oligohydramnios. Second, we were able to analyse previously unpublished data obtained from the POP study of unselected nulliparous women using a blinded assessment of the presence or absence of borderline oligohydramnios. This allowed us to address the true association between the finding and the risk of adverse outcome while avoiding associated biases, for example treatment paradox and ascertainment bias.

As severe oligohydramnios is defined as AFI of < 5 cm, borderline oligohydramnios can be defined as AFI of 5–8 cm or 5–10 cm. To establish the predictive associations, we analysed unpublished data from the POP study (as described in *Chapter 4*) and a systematic review of other studies of diagnostic effectiveness.

Methods

Analysis of data from the Pregnancy Outcome Prediction study

In the systematic review we included unpublished data from a prospective cohort study, the POP study, as described in *Chapter 4*. The present analysis excluded women who delivered prior to their 36 weeks' gestation scan appointment. Screen positive was defined as an AFI between 5 and 8 cm and screen negative was defined as an AFI between 8 and 24 cm. Outcome data have been defined previously.⁸

Sources for meta-analysis

The protocol for the review was designed a priori and registered with the international Prospective Register of Systematic Reviews PROSPERO (registration number CRD42017064093). We searched MEDLINE, EMBASE, CDSR and CENTRAL from inception to June 2019. The studies were identified using a combination of words related to 'ultrasound', 'pregnancy', 'amniotic fluid index', 'AFI', 'liquor volume' and 'prenatal diagnosis'. No restrictions on language or geographical location were applied.

Study selection

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

We excluded studies that were focused on high-risk populations only (e.g. pregnancies known to be complicated by FGR) and those in which the scan was performed during labour.

Study quality assessment and data extraction

The literature search, study selection and analysis were performed independently by two authors (AM and IA) using Review Manager 5.3. Any differences were resolved in discussion with the senior author (GS). The risk of bias in each included study was assessed using the QUADAS-2 tool³⁷ as outlined in the Cochrane Handbook of Diagnostic Test Accuracy Studies. We used a predesigned data extraction form to extract information on study characteristics (i.e. year of publication, country, setting, study design, blinding), patient characteristics (i.e. inclusion and exclusion criteria, sample size), the index test (i.e. gestational age at scan, cut-off values used) and reference standard (i.e. pregnancy outcome, gestation at delivery, and interval from scan to delivery).

Statistical and meta-analysis methods

The statistical and meta-analysis methods employed are described in *Chapter 4*.

Results

The Pregnancy Outcome Prediction study

Initially, we analysed the previously unpublished data from the POP study.⁸⁸ Applying the inclusion criteria described above yielded a total of 3387 women with a blinded scan at 36 weeks' gestation out of the 4512 women recruited (see *Appendix 4, Figure 35*), and 108 (3.2%) of these women had borderline oligohydramnios (AFI of 5–8 cm, *Appendix 4*). Maternal age, socioeconomic deprivation, ethnicity, BMI, and rates of alcohol consumption and smoking were similar in the two groups (see *Appendix 4, Table 22*). Moreover, the groups had similar rates of pre-existing hypertension and pre-eclampsia. The median birthweight was 200 g lower in the cases of borderline oligohydramnios, with a small difference in the gestational age at delivery. The rates of IOL were similar in both groups but women with borderline oligohydramnios had higher rates of spontaneous vaginal delivery. The screening performance of borderline AFI in the POP study⁸⁸ is presented in *Table 5*. Borderline AFI was associated with an increased risk of delivering a severely SGA infant but was not associated with SGA or an increased risk of a range of indicators of neonatal morbidity in the POP study.⁸⁸

Meta-analysis

The literature search flow chart is presented in *Appendix 4, Figure 36*. We identified 11 studies^{88–98} (including the POP study) that met our inclusion criteria, which involved a total of 37,848 patients. The study characteristics are presented in *Appendix 4, Table 23*. Only the POP study⁸⁸ ($n = 3387$) included unselected pregnancies, three studies^{91,97,98} ($n = 1890$) included low-risk pregnancies only and seven studies^{89,90,92–96} ($n = 32,571$) included mixed-risk pregnancies. Two studies⁹⁷ ($n = 3817$) were prospective and nine studies^{89–96,98} ($n = 34,031$) were retrospective. Seven studies^{91,93–97} ($n = 36,293$) defined borderline oligohydramnios as AFI of between 5 and 8 cm and four studies^{89,90,92,98} ($n = 1555$) defined it as between 5 and 10 cm. The majority of patients in all the studies delivered at term. However, four studies^{89,92,95,97} reported a significantly higher rate of preterm delivery among those with borderline oligohydramnios.

The assessment of study quality was performed using the QUADAS-2 tool and is summarised in *Appendix 4, Figure 37*. The main risk of bias was from the lack of blinding of the ultrasound result (which we defined as high risk for reference standard), which affected all studies except the POP study.⁸⁸ We classified one study⁹³ as being at high risk for selection bias as it used only low-risk patients for the comparison group and we classified two studies^{89,90} as being at unclear risk of selection bias as they did not specify whether they enrolled a consecutive or random sample of patients. Moreover, we classified five studies^{89,92,94,96,98} as having an unclear risk of bias for flow and timing because they did not report gestational age at ultrasound or delivery.

TABLE 5 Diagnostic performance of borderline AFI (5–8 cm) in predicting adverse pregnancy outcome at term in the POP study ($n = 3387$)

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

a One or more of the following: 5-minute Apgar score of < 7, delivery with metabolic acidosis (defined as a cord blood pH of < 7.1 and a base deficit of > 10mmol/l) and/or NICU admission.

b Term live birth associated with neonatal death, hypoxic–ischaemic encephalopathy, use of inotropes, mechanical ventilation or severe metabolic acidosis (defined as a cord blood pH of < 7.0 and a base deficit of > 12mmol/l).

The summary diagnostic performance of borderline AFI at predicting adverse pregnancy outcome is presented in Table 6. The most commonly reported outcomes were SGA < 10th centile (nine studies), NICU admission (eight studies), 5-minute Apgar score of < 7 (eight studies), meconium-stained amniotic fluid (seven studies) and caesarean section for fetal distress (six studies). The meta-analysis demonstrated a statistically significant association between borderline oligohydramnios and all of the outcomes, and the strongest association was with delivery of a SGA infant (positive LR = 2.6). The summary ROC curves are presented in Figure 7. Forest plots of the DORs (Figure 8) demonstrated statistically significant heterogeneity for SGA and NICU admission. Two studies (POP and Petrozella *et al.*⁹⁵) reported SGA below the third centile and three studies reported perinatal death. However, we could not generate summary results for outcomes that were reported in fewer than four studies. Finally we used Deeks' funnel plot asymmetry test to assess the risk of publication bias using the outcome of SGA < 10th centile for the analysis (see Appendix 4, Figure 38). The test showed no evidence of publication bias ($p = 0.33$).

TABLE 6 Summary diagnostic performance of borderline AFI in predicting adverse pregnancy outcome

Outcome	Number of studies	Number of patients	Summary sensitivity, % (95% CI)	Summary specificity, % (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)
SGA < 10th centile	9	37,132	31.6 (13.0 to 58.7)	87.9 (71.9 to 95.3)	2.60 (1.83 to 3.69)	0.78 (0.61 to 0.99)
NICU admission	8	9747	34.8 (15.9 to 60.1)	82.6 (69.1 to 91.0)	2.00 (1.41 to 2.85)	0.79 (0.61 to 1.02)
5-minute Apgar score of < 7	8	9666	34.0 (17.4 to 55.8)	82.0 (68.8 to 90.4)	1.89 (1.47 to 2.42)	0.80 (0.66 to 0.98)
Caesarean section for fetal distress	6	33,517	21.2 (7.5 to 47.2)	90.0 (74.5 to 96.5)	2.13 (1.56 to 2.90)	0.87 (0.75 to 1.02)
Meconium-stained in amniotic fluid	7	2885	42.1 (28.7 to 56.9)	74.9 (67.7 to 81.0)	1.68 (1.24 to 2.28)	0.77 (0.62 to 0.96)

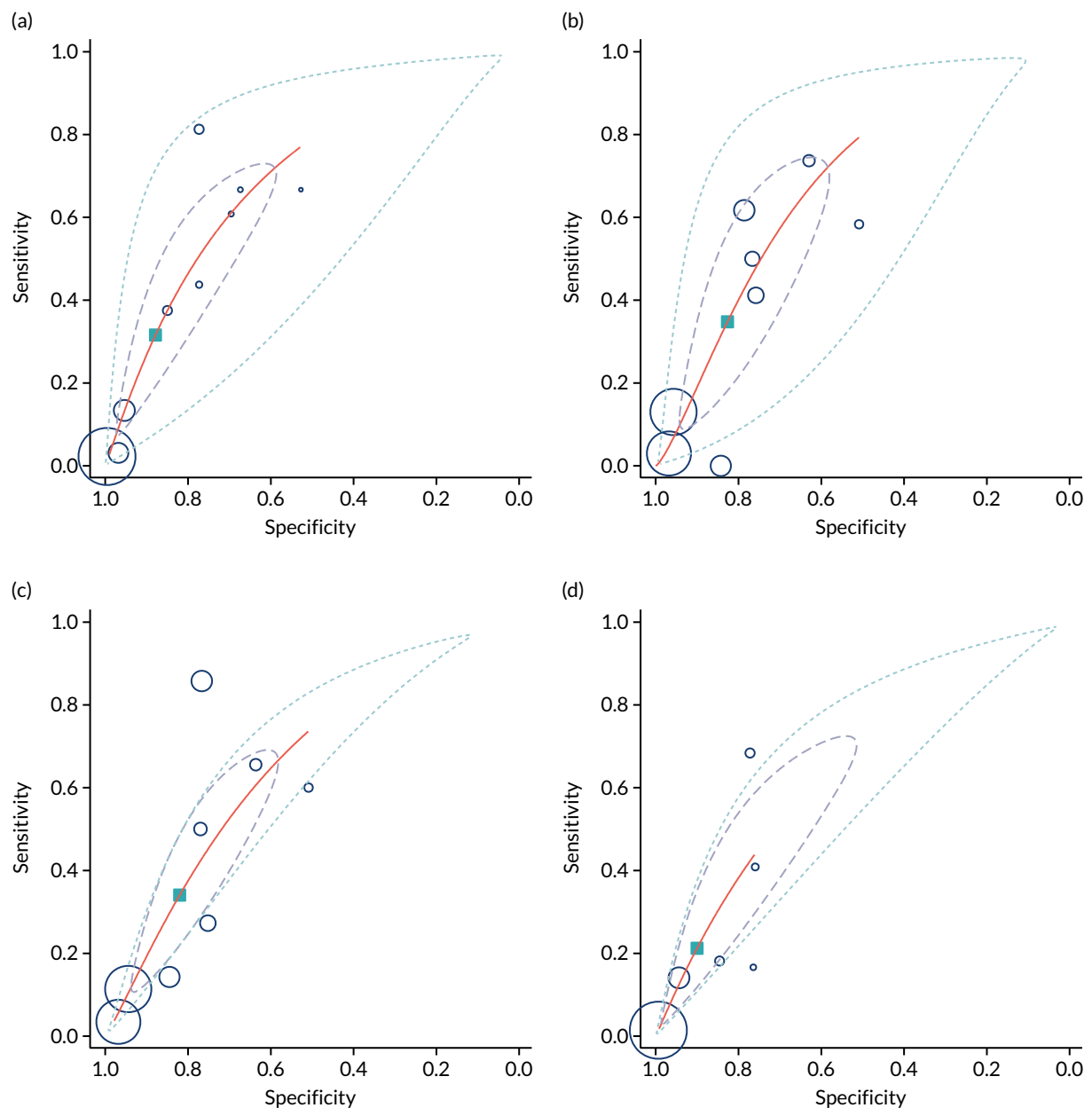


FIGURE 7 Summary ROC curves of borderline AFI at predicting (a) SGA < 10th centile; (b) NICU admission; (c) 5-minute Apgar score of < 7; and (d) caesarean section for fetal distress.

Discussion

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

The observation that borderline oligohydramnios was associated with severe SGA only in the POP study is of interest. One possible explanation for this is that the scan result was not revealed; hence, the finding did not lead to changes in clinical management. The success from blinding the result is evidenced by the fact that borderline oligohydramnios was not associated with increased rates of IOL in the POP study. A previous RCT of routine early term induction compared with expectant management of pregnancies in which ultrasonic fetal biometry indicated a SGA infant demonstrated

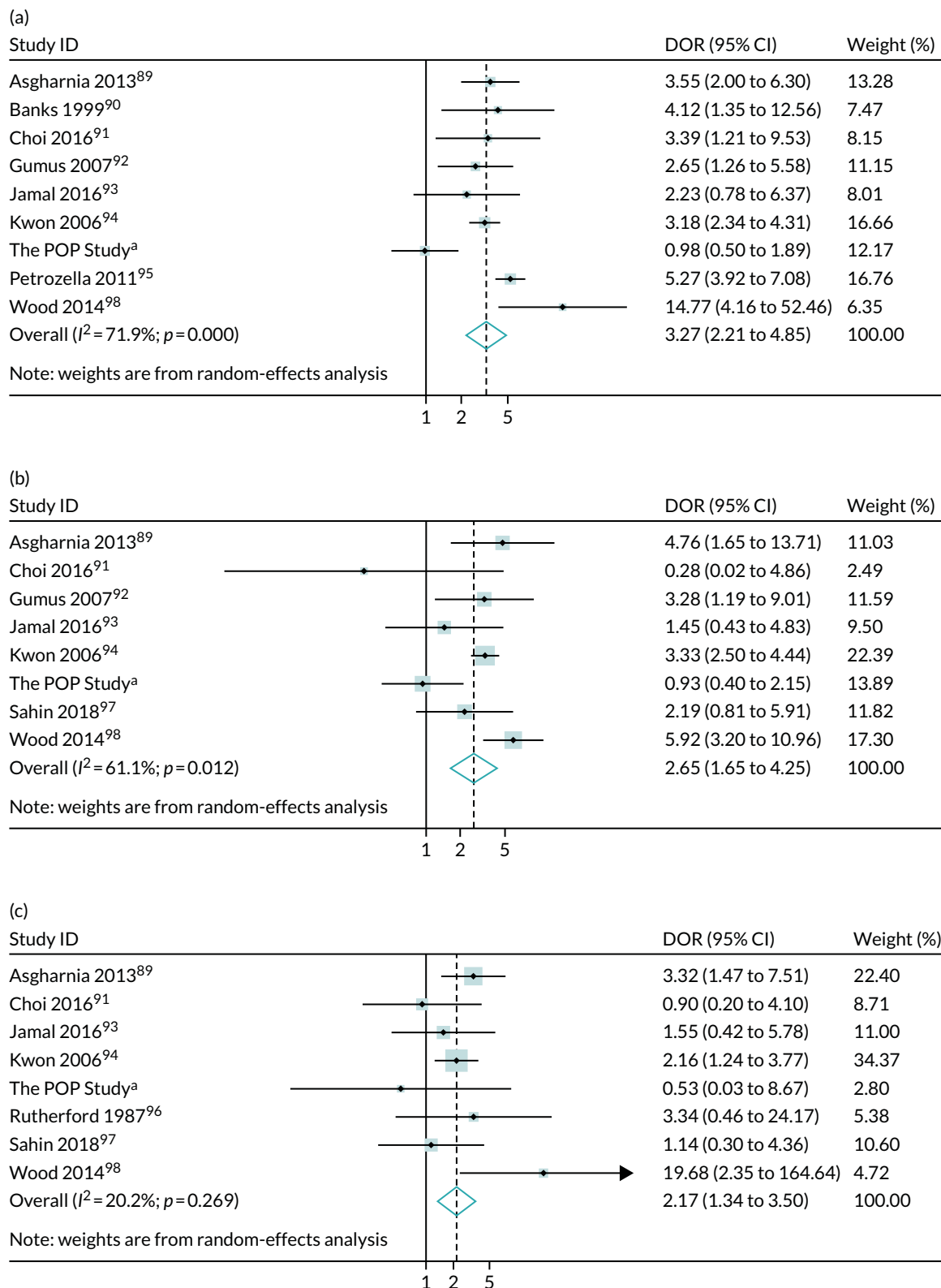


FIGURE 8 The diagnostic odd ratios of borderline AFI at predicting: (a) SGA < 10th centile; (b) NICU admission; (c) 5-minute Apgar score of < 7; and (d) caesarean section for fetal distress. a, Alexandros A Moraitis, Ilianna Armata, Ulla Sovio, Peter Brocklehurst, Alexander EP Heazell, Jim G Thornton, Stephen C Robson, Aris Papageorghiou and Gordon CS Smith, University of Cambridge, 2021. (continued)

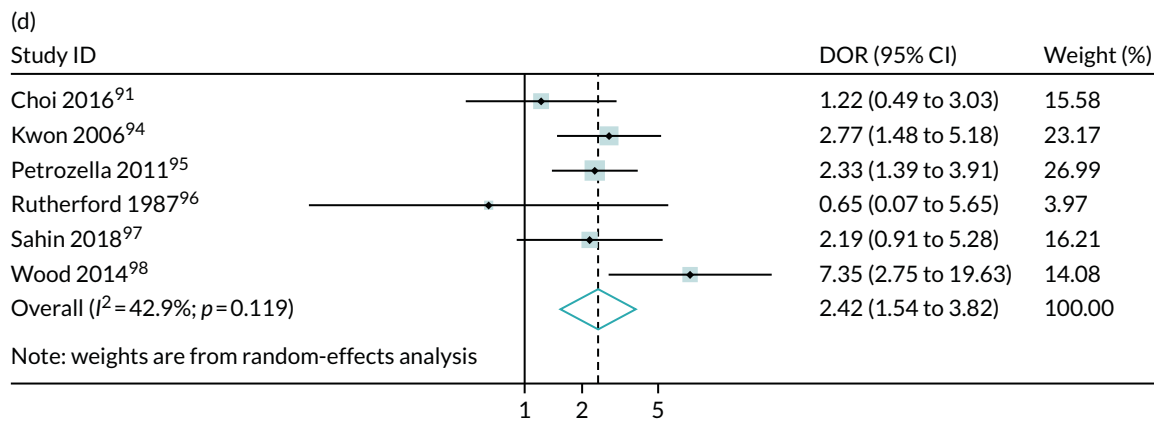


FIGURE 8 The diagnostic odd ratios of borderline AFI at predicting: (a) SGA < 10th centile; (b) NICU admission; (c) 5-minute Apgar score of < 7; and (d) caesarean section for fetal distress. a, Alexandros A Moraitis, Ilianna Armata, Ulla Sovio, Peter Brocklehurst, Alexander EP Heazell, Jim G Thornton, Stephen C Robson, Aris Papageorgiou and Gordon CS Smith, University of Cambridge, 2021.

that early delivery was associated with a significantly decreased risk of the infant being delivered with a birthweight < 3rd percentile.⁹⁹ A possible explanation for the POP study's association with severe SGA and the meta-analysis association with all SGA is that a finding of borderline oligohydramnios may have led to increased rates of early delivery in studies in which the result was revealed, whereas the lack of intervention in the POP study led to growth-restricted fetuses becoming progressively smaller for gestational age as the pregnancy advanced.

The other major difference between the meta-analysis and the POP study may also relate to the lack of blinding in the other studies. Borderline oligohydramnios was associated with increased rates of neonatal morbidity in the meta-analysis but none of the outcomes of neonatal morbidity was associated with this finding in the POP study. However, the CIs were wide and one explanation could be the lower statistical power of the POP study. However, plotting the DORs demonstrates that, in relation to NICU admission, the 95% CI observed in the POP study excluded the point estimate of the meta-analysis. This result could also be explained by the absence of blinding in the other studies. If the scan result is revealed, the only disease-modifying intervention available in late pregnancy is early delivery, and this could be late preterm or early term. It is well recognised that both are associated with increased rates of neonatal morbidity and NICU admission. Hence, the association between borderline oligohydramnios and neonatal morbidity in the meta-analysis could be because the finding led to iatrogenic prematurity and the absence of the finding in the POP study could be due to the lack of this effect. Assessment of individual studies in the meta-analysis is consistent with this interpretation. Gumus *et al.*⁹² reported higher rates of IOL in women with borderline oligohydramnios, which was associated with higher rates of preterm and early term delivery, and higher rates of NICU admission. Similarly Asgharnia *et al.*,⁸⁹ who offered screening after 28 weeks' gestation, found that those with borderline AFI had a rate of preterm delivery of 40.4% (compared with 14.9% for those with normal AFI) and this is the likely explanation for the strong association between borderline oligohydramnios and NICU admission. This association was not found in studies that offered ultrasound later in pregnancy such as that by Sahin *et al.*⁹⁷

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

Chapter 8 Systematic review of the diagnostic effectiveness of universal ultrasonic screening using fetal macrosomia in the prediction of adverse perinatal outcome

Birthweight is a basic characteristic that defines an individual; the weight and sex of an infant are key themes in discussion following a birth. Similarly, when considering pregnancy outcome and its associations with the subsequent health of the infant, birthweight is critical. Much of the focus on birthweight is on infants who are SGA because of the association of being SGA with perinatal mortality. The diagnostic effectiveness of ultrasound in this context was the subject of a Cochrane review of diagnostic effectiveness,²³ and this is discussed extensively in *Chapter 9*. However, being born LGA is also a predictor of adverse outcomes, including perinatal mortality and morbidity arising from traumatic delivery, which is the focus of this chapter.

Ultrasonic EFW was first described > 40 years ago.¹⁰⁰ The most widely employed equation for EFW was published by Hadlock *et al.*⁵ in 1985, and a reference range for EFW was published in 1991.⁶ A subsequent multicountry study by the World Health Organization⁷ derived very similar EFW percentiles, as described by Hadlock in Houston, Texas, USA, in the early 1990s. Hence, the diagnostic tools have been available for many years to identify SGA and LGA fetuses. Moreover, a RCT¹⁰¹ has indicated that routine IOL in the presence of suspected macrosomia may prevent shoulder dystocia, which is one of the key adverse outcomes associated with an infant being LGA.

Despite the widely available diagnostic tools, it is still not clear whether or not screening and intervention for suspected fetal macrosomia is clinically effective. The Health Technology Assessment (HTA) programme is currently funding a RCT of intervention in women diagnosed with a LGA infant ('Induction of labour for predicted macrosomia: the Big Baby trial'; ISRCTN18229892). However, as universal ultrasound in late pregnancy is not recommended in the UK, these women will have received a clinically indicated scan. Hence the results of the study may not be applicable to low-risk women, because the diagnostic effectiveness of the test will vary between women who are scanned routinely and those scanned for a clinical indication. Hence, the aim of the present study was to quantify the diagnostic effectiveness of universal ultrasound in late pregnancy in predicting delivery of a large infant and one of the major associated complications, namely shoulder dystocia.

Methods

Sources of meta-analysis

A systematic search was performed in MEDLINE, EMBASE, CDSR and CENTRAL. The search was carried out on 22 October 2018. No restrictions on language or geographical location were applied. The protocol for the review was designed a priori and registered with the International Prospective Register of Systematic Reviews PROSPERO (registration number CRD42017064093). The studies were identified using a combination of words related to 'ultrasound', 'pregnancy', 'estimated fetal weight', 'EFW', 'birthweight', 'macrosomia', 'large for gestational age', 'shoulder dystocia' and 'brachial plexus injury'.

Study selection

Selection criteria allowed the inclusion of cohort or cross-sectional studies involving singleton pregnancies in which an ultrasound scan was performed at ≥ 24 weeks' gestation. We included all studies in which the ultrasound was performed as part of universal screening, studies that used low-risk populations only and studies with mixed-risk populations. We excluded studies that were focused on high-risk patients, such as patients with pre-existing or gestational diabetes, and studies in which the ultrasound was performed intrapartum. We included studies regardless of the formula and threshold they used to define macrosomia. We also included studies regardless of whether the result was blinded to clinicians. We included studies that reported the following outcomes: LGA (defined as birthweight > 4000 g or > 90 th centile) and severe LGA (birthweight > 4500 g or > 97 th centile); shoulder dystocia; and adverse neonatal outcomes, such as neonatal unit admission, 5-minute Apgar score of < 7 and neonatal metabolic acidosis.

Study quality assessment and data extraction

The literature search, study selection and analysis were performed independently by two authors (AM and NS) using Review Manager 5.3. Any differences were resolved in discussion with the senior author (GS). The risk of bias in each included study was assessed using the QUADAS-2 tool as outlined in the Cochrane Handbook of Diagnostic Test Accuracy Studies.³⁷ This tool assesses the included studies for potential bias in four domains: patient selection, index test, reference standard, and flow and timing. We assessed the risk for flow and timing from the perspective of universal ultrasound screening at about 36 weeks' gestation. We used a predesigned data extraction form to extract information on study characteristics (i.e. year of publication, country, setting, study design and blinding), patient characteristics (i.e. inclusion and exclusion criteria, and sample size), the index test (i.e. gestational age at scan, formula and cut-off values used) and reference standard (i.e. pregnancy outcome, gestational age at delivery and interval from scan to delivery). We also collected information, such as inclusion or exclusion of patients with pre-existing or gestational diabetes.

Statistical and meta-analysis methods

The statistical and meta-analysis methods employed are described in *Chapter 4*.

Results

The literature search flow chart is presented in *Appendix 5, Figure 39*. We identified 40 studies¹⁰²⁻¹⁴¹ that met our inclusion criteria, which involved a total of 66,187 patients. The study characteristics are presented in *Appendix 5, Table 24*. Five studies^{105,114,120,123,138} ($n = 8088$) included unselected pregnancies, nine studies^{110,116,118,119,122,129,131,139,140} ($n = 6436$) included only low-risk pregnancies and 26 studies^{102-104,106-109,111-113,115,117,121,124-128,130,132-137,141} ($n = 51,663$) included mixed-risk pregnancies.

The assessment of study quality was performed using the QUADAS-2 tool and is summarised in *Appendix 5, Figure 40*. The main risk of bias was for reference standard because only two studies^{116,138} blinded the results to the clinicians. The second most common risk of bias was for flow and timing. This is because six studies^{106,111,123,125,133,142} had a very short interval between ultrasound and delivery (the ultrasound was carried out either prior to IOL or < 72 hours from delivery), two studies^{105,114} had a long interval (the ultrasound was carried out prior to 33 weeks' gestation) and two studies^{104,107} did not specify the gestational age at delivery. Finally, three studies^{110,134,140} included prolonged (> 41 weeks' gestation) pregnancies only, which were classified as having 'high applicability concerns because of patient selection'.³⁷

The most commonly reported outcomes were birthweight of > 4000 g (29 studies^{103-106,110-113,118-123,125-135,138-141}) followed by birthweight > 90 th centile (seven studies^{102,107,109,114,115,124,138}), both of which we classified as LGA. We defined severe LGA as a birthweight of > 4500 g (six studies^{113,117,131,137,138,141}) or > 95 th or 97th centiles (two studies^{114,138}). Shoulder dystocia was reported in six studies.^{108,112,116,136,138,141}

Finally, neonatal morbidity (any related outcomes) was reported in only two studies,^{112,138} and consequently we could not produce summary results for this outcome. The most commonly used formulas for EFW were those described by Hadlock *et al.*,⁵ followed by Shepard *et al.*¹⁴³ The most common thresholds for suspected LGA on scan were 4000 g (21 studies^{103,104,106,108,110,118,119,121,125,126,128–135,139–141}) and 90th centile for the gestational age (nine studies). The abdominal circumference was used in nine studies,^{102,105,107,109,111,112,114,115,138} with the most common threshold applied being 36 cm (five studies^{122,123,125,127,137}).

We present the summary diagnostic performance in *Table 7*. An estimated EFW of > 4000 g or > 90th centile had > 50% sensitivity for predicting LGA at birth and this was similar regardless of the formula used. The positive LR ranged between 7.5 and 12 for the Hadlock formulas^{5,6} and was about 5 for the Shepard formula.¹⁴³ The abdominal circumference (AC) had similar performance with the EFW. Suspected LGA also had about 70% sensitivity at predicting severe LGA at birth. Finally, an EFW of > 4000 g or 90th centile had 22% sensitivity at predicting shoulder dystocia with a statistically significant positive LR of 2.1.

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

TABLE 7 Summary diagnostic performance of suspected LGA in predicting LGA at birth and shoulder dystocia

Diagnostic test	Number of studies	Number of patients	Summary sensitivity, % (95% CI)	Summary specificity, % (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)
Outcome: birthweight of > 4000 g (or > 90th centile)						
EFW (any) > 4000 g (or > 90th centile)	29	34,198	53.5 (47.3 to 59.6)	93.9 (91.8 to 95.5)	8.82 (6.83 to 11.4)	0.49 (0.44 to 0.56)
EFW (Hadlock – AC/FL/HC/BPD)	9	22,073	63.1 (49.1 to 75.2)	94.3 (90.9 to 96.5)	11.13 (8.24 to 15.04)	0.39 (0.28 to 0.55)
EFW (Hadlock – AC/FL/BPD)	10	17,110	55.1 (44.1 to 65.7)	92.9 (89.7 to 95.2)	7.77 (5.55 to 10.89)	0.48 (0.38 to 0.61)
EFW (Hadlock – AC/FL/HC)	6	14,801	57.3 (47.0 to 67.0)	95.2 (92.3 to 97.0)	11.89 (7.81 to 18.10)	0.45 (0.36 to 0.56)
EFW (Hadlock – AC/FL)	9	16,736	60.5 (50.7 to 69.5)	92.0 (89.4 to 93.7)	7.54 (6.13 to 9.29)	0.43 (0.34 to 0.54)
EFW (Hadlock – AC/BPD)	6	13,617	62.9 (36.1 to 83.5)	93.7 (85.9 to 97.3)	9.99 (6.40 to 15.58)	0.40 (0.21 to 0.75)
EFW (Shepard)	7	14,060	73.7 (54.4 to 86.9)	85.1 (76.5 to 90.9)	4.96 (3.29 to 7.48)	0.31 (0.17 to 0.56)
AC > 36 cm (or > 90th centile)	5	10,543	57.8 (39.6 to 74.2)	92.3 (88.7 to 94.9)	7.56 (5.85 to 9.77)	0.46 (0.30 to 0.68)
Outcome: birthweight of > 4500 g (or > 95th centile)						
EFW (any) > 4000 g (or > 90th centile)	4	5839	70.2 (42.6 to 88.2)	89.2 (74.4 to 95.9)	6.49 (2.2 to 19.1)	0.33 (0.14 to 0.78)
Outcome: shoulder dystocia						
EFW (any) > 4000 g (or > 90th centile)	6	26,264	22.0 (9.9 to 42.0)	89.6 (80.8 to 94.6)	2.12 (1.34 to 3.35)	0.87 (0.74 to 1.02)

BPD, biparietal diameter; FL, femur length; HC, head circumference.

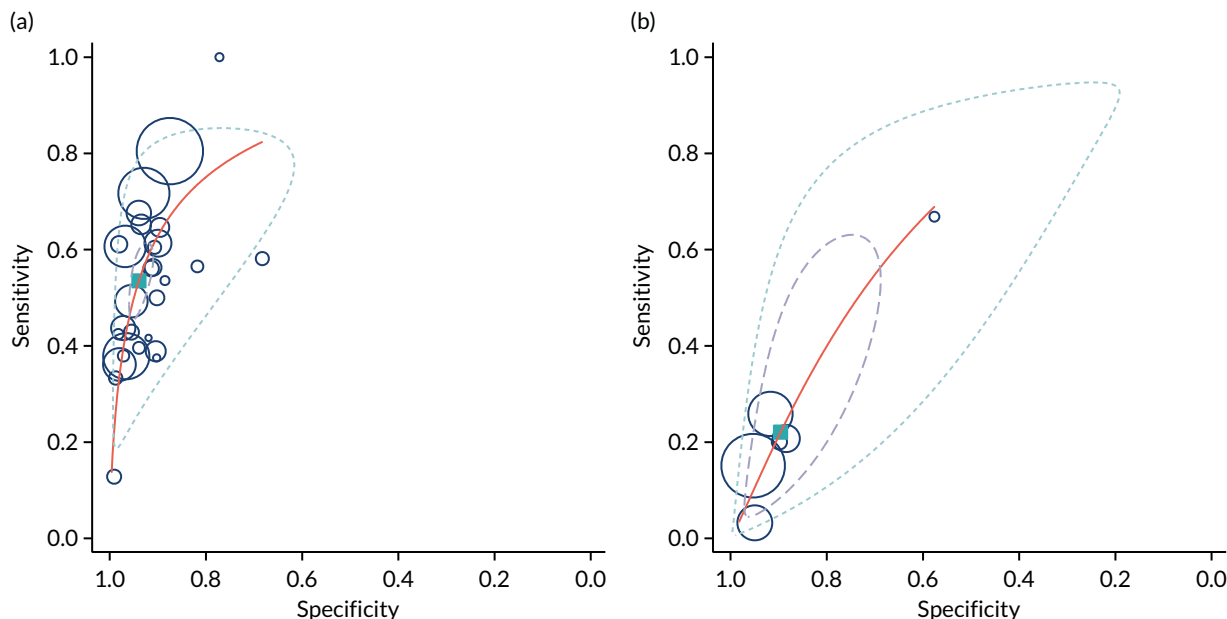


FIGURE 9 Summary ROC curves for the diagnostic performance of EFW > 4000 g (or 90th centile) at predicting (a) LGA at birth (birthweight > 4000 g or > 90th centile); and (b) shoulder dystocia.

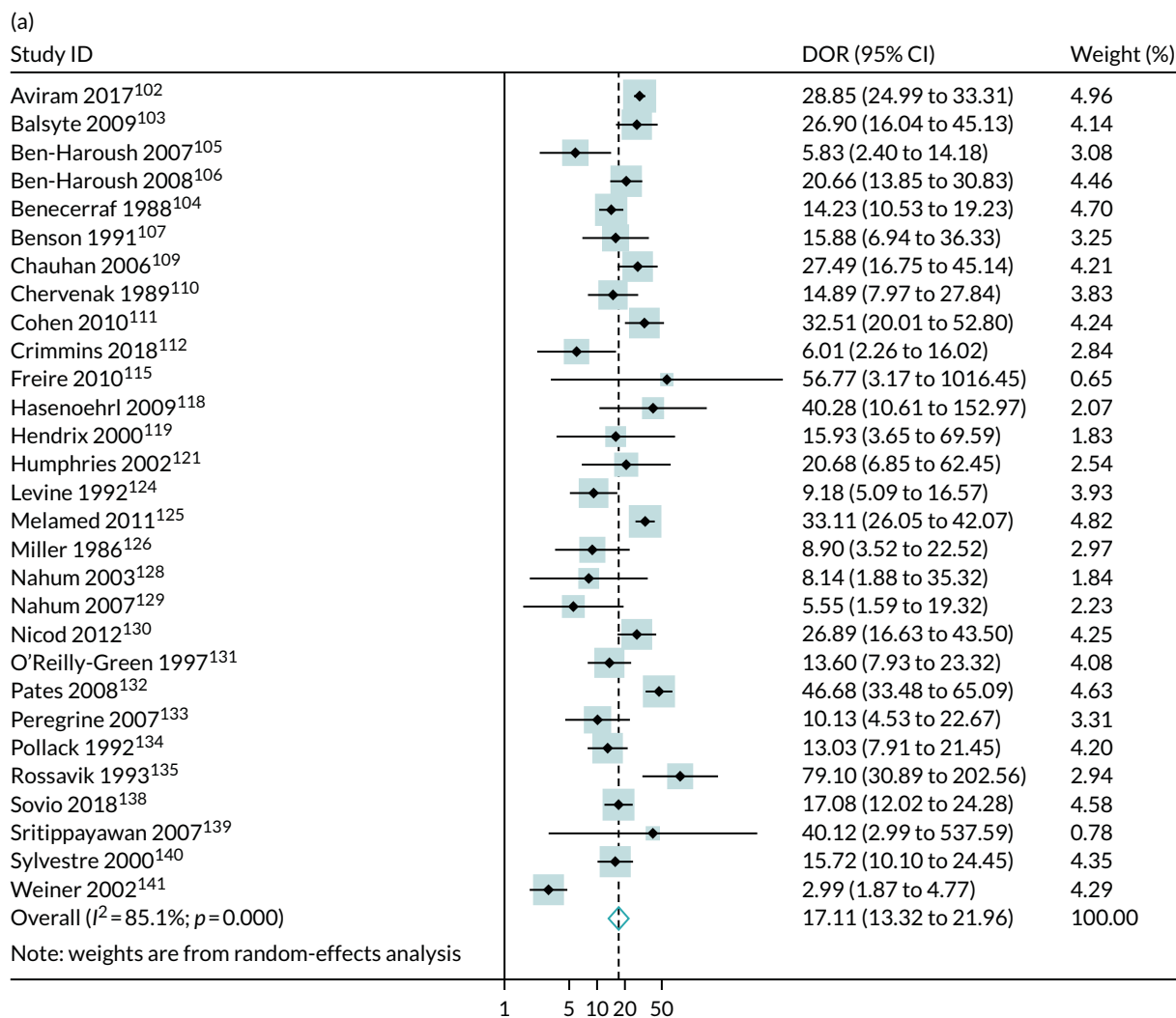


FIGURE 10 The diagnostic odds ratios for the diagnostic performance of EFW > 4000 g (or > 90th centile) at predicting (a) LGA at birth (birthweight > 4000 g or > 90th centile); and (b) shoulder dystocia. (continued)

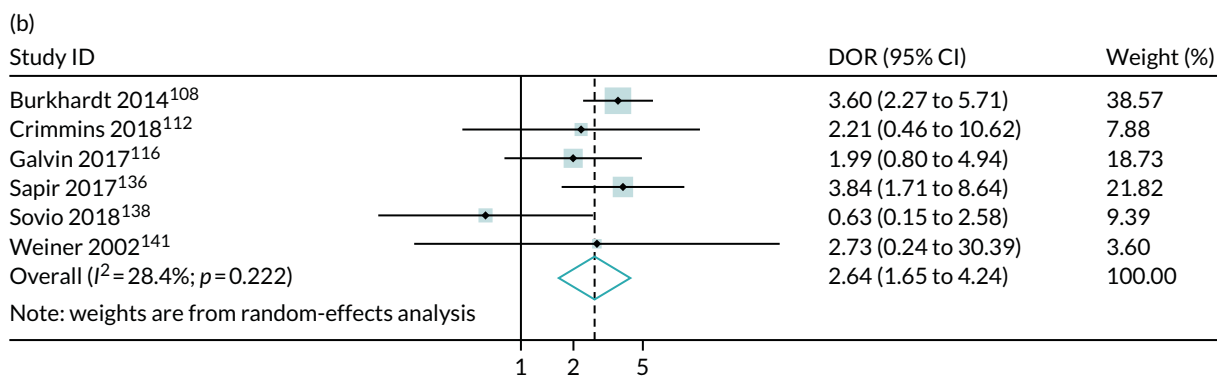


FIGURE 10 The diagnostic odds ratios for the diagnostic performance of EFW > 4000 g (or > 90th centile) at predicting (a) LGA at birth (birthweight > 4000 g or > 90th centile); and (b) shoulder dystocia.

Discussion

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

In this chapter and in *Chapters 4* and *7* we have focused analysis on data from the POP study, as these data are particularly applicable to the research question addressed in this report, given that late-pregnancy ultrasound was performed in a large number of nulliparous women using contemporary equipment and staff trained using the standards of NHS England. The POP study analysis of a 36 weeks' gestation scan in the diagnosis of macrosomia had previously been published¹³⁸ and this was incorporated into the meta-analysis. Interestingly, the DOR from the POP study was 17.1 (95% CI 12.0 to 24.3) and this was virtually identical to the summary estimate from all of the other studies, which was also 17.1 but with a slightly narrower 95% CI (13.3 to 22.0). These data suggest that the results from the POP study are likely to be generalisable.

A recurrent theme in all chapters has been the lack of blinding in studies of the diagnostic effectiveness of ultrasound of pregnancy screening research. Hence, generally, the POP study has been unique as a contemporary study of late pregnancy in nulliparous women. However, in this analysis there is a second comparable study: the Genesis study. This was a prospective cohort study of 2772 nulliparous pregnant women recruited across seven centres in Ireland between 2012 and 2015. Women had the ultrasound scan between ≥ 39 weeks' gestation and < 41 weeks' gestation (i.e. ≈ 3 –4 weeks later than in the POP study). Although the scan was carried out slightly later than stated in the research question of the current report, the design makes the study particularly useful.

The analysis of fetal macrosomia from the Genesis study has been published in abstract form only. It did not report the diagnostic effectiveness of EFW as a predictor of LGA birthweight, but it did report shoulder dystocia. Interestingly, the POP study and the Genesis study were the only two large studies (i.e. comprising > 1000 women) not to demonstrate a statistically significant association between macrosomic EFW and the risk of shoulder dystocia. Overall, the meta-analysis indicated that ultrasound may be weakly predictive of shoulder dystocia. However, as with other analyses in *Chapters 4*–*7*, these findings could be explained by ascertainment bias. Specifically, if a scan is performed and the fetus is suspected to be macrosomic, the clinical staff attending the birth may be more likely to institute

manoeuvres for shoulder dystocia in the event of any delay, or to document a given delay as being due to shoulder dystocia. The potential for such biases may explain why the studies with blinded ultrasound were not significantly associated with shoulder dystocia and why the meta-analysis as a whole was only weakly predictive of shoulder dystocia, whereas it was strongly predictive for macrosomia. A weak association between ultrasonic EFW and the risk of shoulder dystocia is not surprising given that the actual birthweight of the infant is not strongly predictive of shoulder dystocia and that the majority of cases of shoulder dystocia do not involve a macrosomic infant.¹⁴⁴

Finally, the relationship between fetal macrosomia and pregnancy outcome is an area where there is good evidence that revealing the scan result changes the experience of complications of women who are false positives. Multiple studies have demonstrated that a false-positive diagnosis of fetal macrosomia is an independent risk factor for emergency caesarean delivery.¹⁴⁵⁻¹⁴⁷ These observations underline the potential of screening low-risk women to cause harm and that designing a study where the results are revealed to the attending physician could lead to an association that is iatrogenic (because the knowledge of the result may change clinical decision-making) rather than because of a true prediction.

Chapter 9 Conclusions regarding the evidence around universal ultrasound screening of nulliparous women in late pregnancy

Chapters 4–7 have outlined the association between umbilical artery Doppler, the CPR, severe oligohydramnios, borderline oligohydramnios, and fetal macrosomia and the risk of adverse pregnancy outcome. The main overall conclusions are as follows:

- Umbilical artery Doppler, the CPR, severe oligohydramnios, borderline oligohydramnios and fetal macrosomia were all either non-predictive or weakly predictive of the risk of neonatal morbidity.
- Umbilical artery Doppler, the CPR, severe oligohydramnios and borderline oligohydramnios were all weakly predictive of the risk of delivering a SGA infant.
- The vast majority of the studies did not blind the result of the index test. Hence, interpreting the results in relation to prediction of adverse neonatal outcome could be biased against not seeing associations where true associations exist (e.g. through treatment paradox) or biased towards seeing associations where no true associations exist (e.g. through ascertainment bias or iatrogenic harm).
- Only the POP study¹³⁸ has reported the range of ultrasonic findings in late pregnancy in unselected nulliparous women, which is the optimal study design, and was conducted in the target population. In a second study conducted in Ireland (Genesis)¹¹⁶ blinded ultrasound scanning were also carried out in late pregnancy in nulliparous women but the results have not been published widely.
- The results of the POP study in relation to both SGA and LGA (outcomes that are objectively defined and less prone to biases) were comparable with the summary estimates across all studies.

During the current project, a systematic review of DTA in relation to ultrasonic diagnosis of SGA using EFW was published.²³ In this review, the authors reported that in the majority of studies clinicians were not blinded to test results or this was not reported.²³

The Weiner *et al.*¹⁴² study was carried out on 405 women during active labour and compared clinical assessment of fetal size with ultrasonic EFW. Hence, the conclusion of the Heazell *et al.*²³ systematic review is that the POP study is only study in which blinded ultrasonic assessment of SGA was performed that was relevant for population screening in the antenatal period.

We were aware of the Heazell *et al.*²³ review and did not therefore address ultrasonic diagnosis of SGA in the present review. The authors reported detection of SGA (birthweight < 10th percentile) as follows:

- For a specificity of 88%, ultrasonic suspicion of SGA had a sensitivity of 74% (95% CI 64% to 83%). In the POP study, the sensitivity was 57% (95% CI 51% to 62%) for a specificity of 90% (95% CI 89% to 91%).

The meta-analysis reported detection of severe SGA (birthweight < 3rd percentile) as follows:

- For a specificity of 87%, ultrasonic suspicion of SGA had a sensitivity of 66% (95% CI 56% to 76%). In the POP study, the sensitivity was 77% (95% CI 68% to 86%) for a specificity of 87% (95% CI 86% to 88%).

We had expected a similar prediction of the more severe outcome in the Heazell *et al.*²³ review. The inconsistency between these two analyses^{8,23} may reflect inclusion of different studies that may have included different populations. However, the review does suggest that the data observed in the POP study were generally comparable to those obtained in the studies included in the Heazell *et al.*²³ review.

A further level of complexity in considering these issues is that, generally, an ultrasonic assessment of a fetus typically includes the measurement of multiple parameters simultaneously. Hence, a further issue when trying to apply the findings of the Heazell *et al.*²³ review, and our own reviews, to health economic analysis and trial design is that none of the reviews completely captures what may be expected to happen clinically. This issue is affected by another layer of complexity, namely defining the features on a scan that the majority of clinicians would accept as indicating FGR. This last question has been addressed by researchers employing the Delphi consensus method to generating an agreed ultrasonic diagnosis of FGR. The paper arising from this process was published in 2016.¹⁴⁸ These authors described the following criteria for diagnosis of late FGR (32 weeks' gestation or later): EFW or abdominal circumference (AC) < 3rd percentile or two or more of the following – (1) EFW/AC < 10th percentile, (2) EFW/AC falling two quartiles, or (3) a CPR < 5th percentile or umbilical artery Doppler > 95th percentile.

In a paper in *The Lancet Child & Adolescent Health* in 2018,¹⁴⁹ the POP study data were used to compare the Delphi definition of late FGR using the blinded 36 weeks' gestation scan with simply an EFW of < 10th percentile as a predictor of the risk of delivering a SGA infant with complications. The results are presented in *Table 8*.

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

Taking the totality of the data, the approach we took for the health economic analysis was that we defined screen positive as either ultrasonic EFW of < 10th percentile (suspected SGA) or ultrasonic EFW of > 90th percentile (suspected LGA). The Heazell *et al.*²³ review demonstrated good diagnostic effectiveness for SGA and the analysis in *Chapter 8* demonstrated that ultrasonic suspicion of macrosomia was strongly associated with the risk of delivering a LGA infant. The attractiveness of this approach was underlined by the fact that there were Cochrane reviews^{150,151} that reported meta-analyses of RCTs of IOL in both situations and there are extensive epidemiological data on the outcome of SGA and LGA pregnancies. There was one additional further exposure that is detectable by scan and where management is informed by RCT evidence, namely breech presentation near term. Ultrasound establishes fetal presentation with 100% accuracy at the time of the scan (although presentation will sometimes change spontaneously after the scan). Hence, we included this in subsequent analyses.

TABLE 8 Diagnostic effectiveness of ultrasonic screening at 36 weeks' gestation for subsequent delivery of a SGA infant associated with either maternal pre-eclampsia or perinatal morbidity or mortality

Screening test	Positive LR (95% CI)	Negative LR (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Ultrasonic EFW < 10th	5.1 (4.2 to 6.3)	0.38 (0.26 to 0.54)	67.2 (53.8 to 78.3)	86.9 (85.8 to 88.0)
Delphi definition of late FGR	5.9 (4.7 to 7.4)	0.43 (0.31 to 0.60)	61.4 (47.9 to 73.4)	89.6 (88.6 to 90.6)

Reproduced with permission from Gaccioli *et al.*¹⁴⁹ © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>. The table includes minor additions and formatting changes to the original table.

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

Chapter 9 identified three elements of a late pregnancy ultrasound scan that constituted evidence of a high-risk fetus (i.e. in breech presentation), a SGA fetus or a LGA fetus. We next sought to determine the evidence base that existed to inform interventions for women whose scan revealed these features, and used the search engine of the National Institute for Health and Care Excellence (NICE), at www.evidence.nhs.uk/.

Management plan for breech presentation

This search identified an existing UK-based guideline from the Royal College of Obstetricians and Gynaecologists (RCOG), *Management of Breech Presentation (Green-top Guideline No. 20a)*.¹³ In brief, women who do not have a contraindication to ECV are offered this procedure (turning of the fetus by manual manipulation without anaesthetic). Where the procedure is contraindicated, declined or unsuccessful women would then have a discussion regarding attempting vaginal breech birth. Where vaginal breech birth was contraindicated or declined, a planned caesarean section would be scheduled at 39 weeks' gestation (in the absence of a clinical indication for earlier delivery) with the proviso that the infant would be delivered by emergency caesarean section if the woman presented in labour before the scheduled date. Women who had a successful ECV would have routine care thereafter, but with midwife checks to ensure that the infant had not reverted to breech. In practice, given that the target population is nulliparous, it would be a small minority who would opt for vaginal breech birth and no women took up this option in the POP study.¹¹ For the purposes of the Markov chain modelling and health economic analysis we used the effect estimates of a Cochrane review that quantified 'the effects of planned Caesarean section for singleton breech presentation at term on measures of pregnancy outcome'.¹⁴ Other parameters were obtained from the literature and are detailed in *Chapter 11*.

Management plan for diagnosis of a small for gestational age fetus

We next used the NICE evidence search engine to identify existing guidelines for the management of a SGA fetus. This search identified an existing UK-based guideline from the RCOG, *The Investigation and Management of the Small-for-Gestational-Age Fetus, Investigation and Management (Green-top Guideline No. 31)*.¹⁵² Much of this guideline focuses on the identification of risk factors in early pregnancy and the management of the preterm SGA fetus. The RCOG recommendations are: (1) to take into consideration and abnormal umbilical artery or MCA Doppler to time delivery, (2) to offer delivery of the SGA fetus at 37 weeks' gestation even if the umbilical artery Doppler is normal, (3) to recommend caesarean section in the SGA fetus with umbilical artery AREDV and (4) to offer IOL and continuous fetal heart monitoring in the SGA fetus with normal umbilical artery Doppler or with abnormal umbilical artery PI but end-diastolic velocities present.

The same search also identified an NHS England care bundle that aimed to reduce rates of perinatal death, *Saving Babies' Lives Version Two: A Care Bundle for Reducing Perinatal Mortality*.¹⁵³ This guideline has a section on the management of SGA fetuses at term, and the following are key recommendations: (1) in the cases of severe SGA < 3rd centile and with no other concerning features, delivery should be offered at 37⁺⁰ weeks' gestation and no later than 37⁺⁶ weeks' gestation. (2) Fetuses between the 3rd and 10th centile should be assessed individually and the risk assessment should include Doppler investigations, the presence of any other high-risk features, for example, recurrent reduced fetal movements. In the absence of any high-risk features IOL should be offered at 39⁺⁰ weeks.

However, the context for both the RCOG and the NHS England guidelines was the management of women who were identified through the current approach of targeting ultrasound to high-risk women. As outlined in *Chapter 9*, we have not found evidence that these additional ultrasound tests are diagnostically effective when used as screening tests. Hence, the management protocol for SGA infants employed in the health economic analysis is to offer IOL. For the purposes of the health economic analysis we used the effect estimates of a Cochrane review that quantified ‘the effects of immediate delivery versus expectant management of the term suspected compromised infant on neonatal, maternal and long-term outcomes’.¹⁵⁰ In practice, 90% of the women included in the review came from a trial of IOL for suspected FGR.⁹⁹ IOL took place in the intervention group of this trial at an average of 38 weeks’ gestation and we have incorporated this into our management protocol (see section below). This does not represent an extreme intervention as a large-scale NIH-funded RCT demonstrated no adverse effect of routine IOL at 39 weeks’ gestation in nulliparous women who did not have risk factors.¹⁵⁴ Other parameters were obtained from the observational literature and are detailed in *Chapter 11*.

Management plan following diagnosis of a large for gestational age fetus

We next used the NICE evidence search engine to identify existing guidelines for the management of a LGA fetus. The only guidelines that we identified using this search related to women with diabetes. These women are routinely scanned during pregnancy and have specific issues, and the recommendations for this group are not generalisable to the population of interest in the current report. However, the search did identify a number of systematic reviews that addressed IOL, and one of these was a Cochrane review.¹⁵¹ The Cochrane review concluded that IOL for suspected fetal macrosomia results in a lower mean birthweight, fewer birth fractures and shoulder dystocia. They concluded that to prevent one fracture it would be necessary to induce labour in 60 women and that induction of labour does not appear to alter the rate of caesarean delivery or instrumental delivery. However they suggested that further trials of induction shortly before term for suspected fetal macrosomia are needed.¹⁵¹

Consistent with this recommendation, the HTA programme has funded a RCT [‘Induction of labour for predicted macrosomia: the Big Baby trial’ (ISRCTN18229892)]. Given the uncertainty in the evidence base, it is not possible to develop a robust plan for management following a diagnosis of macrosomia. For the purposes of the Markov chain modelling and health economic analysis, we addressed this uncertainty by comparing multiple strategies, including expectant management, early-term IOL and planned caesarean section. The effects in relation to IOL were taken from the Cochrane review,¹⁵¹ as this was assessed as the highest-quality evidence available at the time of writing. About 70% of the women came from a single trial¹⁰¹ in which the most common week for IOL was 38 weeks’ gestation. Other parameters for the modelling and health economic analysis were obtained from the observational literature and are detailed in *Chapter 11*. A summary of the management plan is outlined in *Figure 11*.

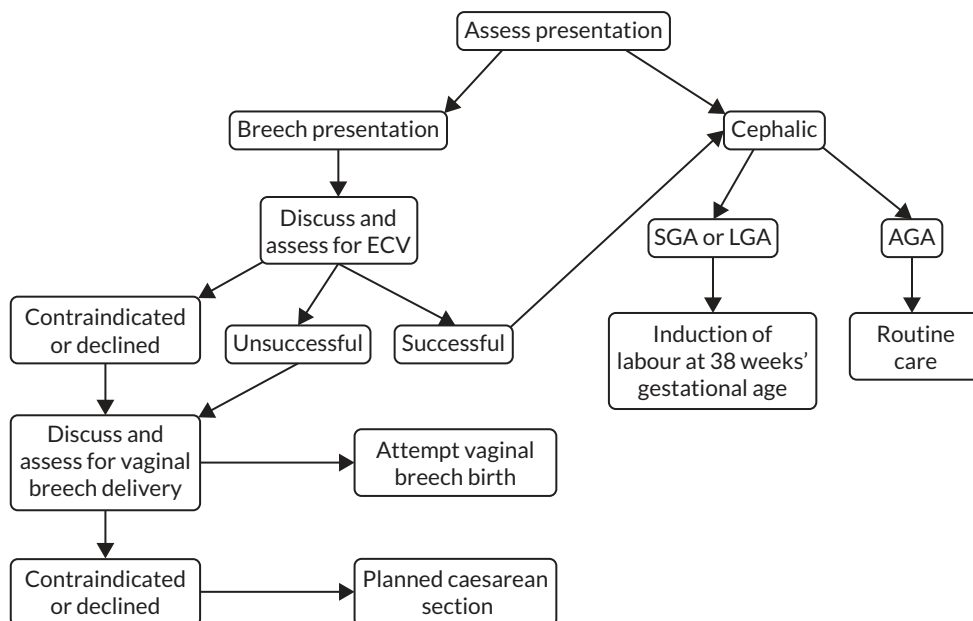


FIGURE 11 Summary of the management plan following the 36 weeks' gestation scan.

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

Introduction

This study was commissioned to evaluate the current evidence base on the costs and clinical effectiveness of performing a routine ultrasound scan in late pregnancy in all nulliparous women combined with appropriate management plans, to identify evidence gaps, and to predict whether or not future research to fill those gaps is likely to be a cost-effective use of health-care resources. In this analysis, we use decision modelling to assess the likely outcomes from universal ultrasound screening and determine whether or not its potential benefits can be clinically and economically justified.

We present a cost-utility analysis focusing on three of the main conditions detectable by ultrasound screening that may warrant intervention: breech presentation, the fetus being SGA and the fetus being LGA. The cost-effectiveness of universal ultrasound screening for each of these conditions individually has been explored previously.^{11,155} However, here we evaluate the cost-effectiveness of screening for all of these conditions at the same session. Furthermore, we use decision uncertainty to predict the expected return on further research. We have applied the simplified management plan outlined in *Figure 11*. In essence, women are first assessed for presentation. If the infant is in breech presentation, ECV is offered. If this is successful, the woman reverts to receiving expectant management, and, if it is unsuccessful, the baby is delivered by planned caesarean section. If the infant is in a cephalic presentation and the EFW is in the normal range, the woman receives expectant management. If the infant is either SGA or LGA, IOL is offered. However, we also compare combined assessment for presentation and fetal biometry with a scan simply for presentation. The rationale for this is that a presentation scan may be readily implemented and relatively inexpensive, and there is much less uncertainty about the usefulness of knowing the infant's presentation than there is about the usefulness of estimating the infant's size.

The structure of this chapter is as follows. In *Methods*, we first introduce the general methodology for our economic evaluation. We then summarise the clinical definitions used, as well as the competing strategies evaluated, in this study before introducing the structure of the economic simulation model underlying the analysis. Once the model structure and mechanics have been explained, we discuss how we populated the model with the best available data; complete technical details regarding how individual parameters were derived are presented in *Appendix 6*. Finally, we describe the base-case analyses, sensitivity analyses and VOI analysis to guide how future research in this area could be prioritised.

In *Results*, we present the results of the baseline economic evaluation and sensitivity analyses. The results of the VOI analysis are then presented, which include the results for the expected value of perfect information (EVPI), the expected value of partial perfect information (EVPPPI) and, finally, the expected value of sample information (EVSII).

In *Discussion*, we summarise the key findings, explain the interpretation of our results and discuss what impact our methodological limitations may have had on the results.

Methods

To compare long-term health and cost outcomes associated with different strategies of screening in third-trimester pregnancy, we constructed an economic simulation model. We focused the model on two features for which late-pregnancy ultrasound is amenable to detect: fetal presentation and fetal size. We used a decision tree model consisting of four subtrees, one each for breech presentation, LGA, SGA and AGA. The model structure is based largely on previous economic analyses of screening for these conditions individually, and the development and key characteristics of these submodels' models have previously been described^{11,155} (a brief summary is provided in *Appendix 7*). *Chapter 10* dealt with the diagnostic effectiveness of ultrasound in this setting and outlined how a positive result on scan could influence subsequent care. This chapter focuses on how these submodels were incorporated into a joint framework, enabling a cost-effectiveness analysis of simultaneous screening for all of these conditions.

Scope and population

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

Comparators and interventions

This analysis evaluated three different strategies for ultrasound screening in late pregnancy, defined as a scan between 36⁺⁰ weeks' gestation and 36⁺⁶ weeks' gestation. 'Selective ultrasound' (i.e. when ultrasound is performed only if clinically indicated) is the current standard in England.¹⁵² 'Universal ultrasound for fetal size' would mean routinely offering a third-trimester ultrasound assessment of fetal weight in every pregnancy. Given the simplicity of detecting fetal presentation during an ultrasound scan, this screening strategy would also identify breech presentation. A third option would be to offer 'universal ultrasound for presentation only' (i.e. a simpler ultrasound scan with the sole purpose of detecting pregnancies with breech presentation). Compared with a standard antenatal ultrasound for which, typically, multiple measurements are made, an ultrasound scan for fetal presentation alone is technically simple. We theorised that such a scan could be carried out by an attending midwife during a standard antenatal visit in primary care, using basic ultrasound equipment.

We assumed that all women identified with breech presentation would be offered an ECV unless contraindicated, in line with RCOG guidelines.¹⁵⁶ We further assumed that pregnancies in which the fetus is identified as SGA (whether or not correctly diagnosed) would be given early IOL. However, for pregnancies in which the fetus is diagnosed as LGA, there is uncertainty about the benefits of the intervention (IOL). For this reason, expectant management of suspected LGA pregnancies was also an option. We had previously considered also including elective caesarean section for the management of macrosomia, but we ruled this out because it was inferior to IOL in our cost-effectiveness analysis of ultrasound assessment for macrosomia alone.¹⁵⁵ This conclusion was consistent with a previous decision model analysis.¹⁵⁷ We therefore compare six discrete strategies in the analysis (*Table 9*).

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

As discussed in *Chapter 10*, there is more uncertainty in relation to the management of LGA than of SGA. However, performing fetal biometry will yield a percentile of EFW and, hence, a scan involving fetal biometry can yield three possible outcomes: AGA, SGA or LGA. Consequently, we considered two possible approaches to screening involving fetal biometry. Both approaches included IOL for SGA; however, one also included IOL for LGA, whereas the other dictated expectant management, given the uncertainty.

TABLE 9 Comparator strategies for economic simulation model

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

+, positive diagnosis of the condition; Exp, expectant management.

Outcomes

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

Model structure

As stated, the model structure is a decision tree. It was coded in R (The R Foundation for Statistical Computing, Vienna, Austria) version 3.4.1, using the packages *BCEA*, *FinCal*, *ggplot2*, *gtools*, *readxl*, *tidyr* and *SAVI*.^{158,159} The code for the model is available from the corresponding author on request.

Figure 12 shows the structure of the first stages of the decision model. The [+] indicates sub-branches that have been collapsed for clarity. Nodes are named to show their relationship to one another; nodes with the same letter have identical structures to the branches of the tree beyond, whereas a different number and/or a lower-case letter indicates a different set of probabilities. The prefixes B, L and S denote nodes with probability sets specific to breech presentation or large or small for gestational age infants, respectively.

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

A fetus will be in either breech or cephalic presentation (node A1), or be LGA, SGA or AGA (node A2). For ease of modelling, we assume that all four possibilities are mutually exclusive and structured hierarchically, beginning with presentation (breech or cephalic) and followed by size (LGA, SGA or AGA). The implications of this are considered in *Discussion*. The probability of breech is the prevalence of breech at the time of screening (approximately 4.6%).¹¹ If the scan policy is universal ultrasound (whether for fetal biometry or for presentation only), then, given the ease of interpretation of such a scan, we assume all breeches are detected (i.e. 100% sensitivity and specificity, node B_B). However, under the selective scan policy, approximately 45% of breeches will be undetected¹¹ owing to the mother not having undergone a scan at all (for consistency with the rest of the model, we label these 'false negatives'). Further outcomes relating to breech presentation are described in *Outcomes relating to breech*.

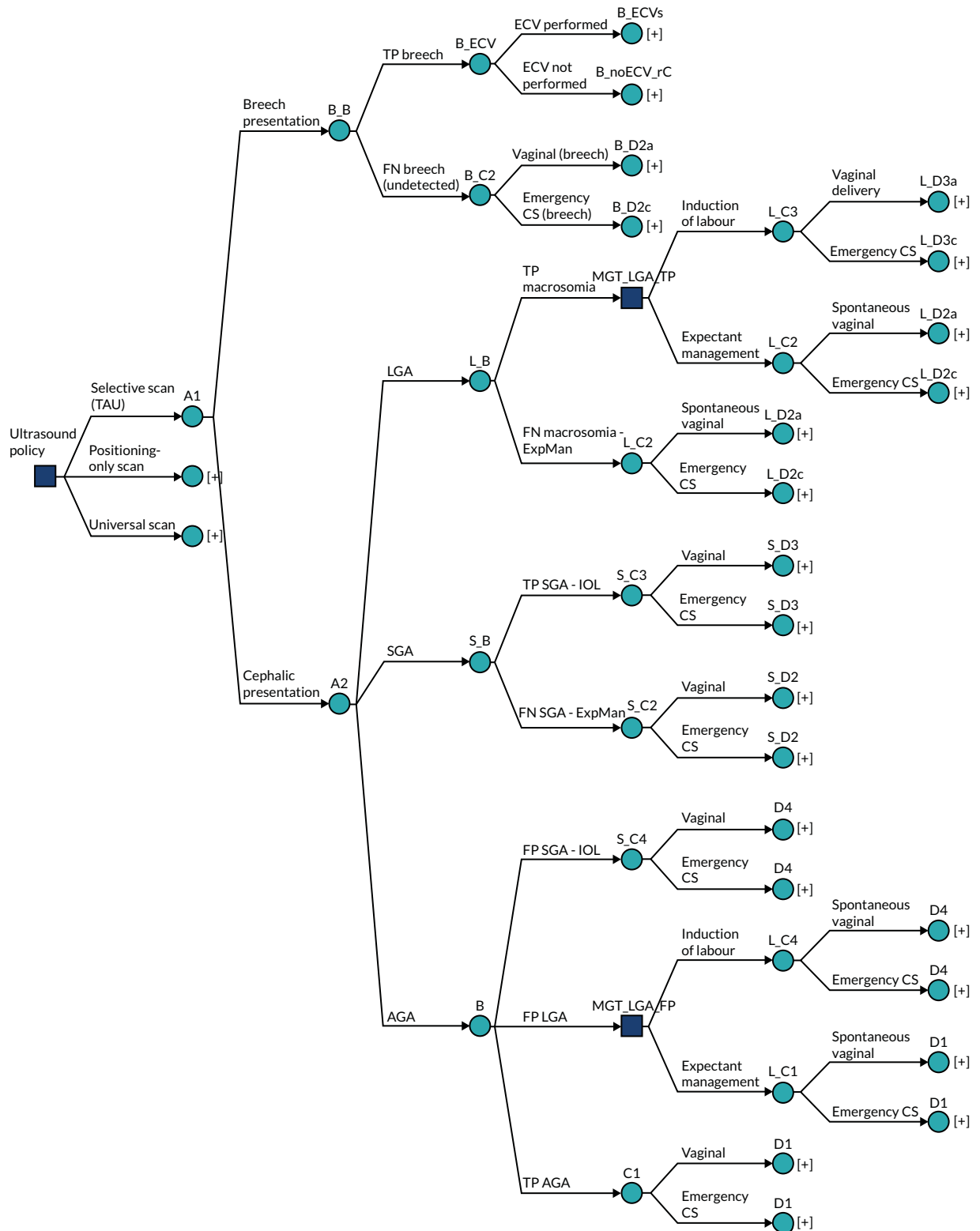


FIGURE 12 Model overview. [+], sub-branches of model collapsed for clarity.

If the infant is in cephalic presentation, it may be LGA, SGA or AGA. The probabilities of each is the prevalence of the condition (node A2, by definition 10% for each). If an infant is LGA or SGA, the probability of detection is a function of the sensitivity of the scanning policy (nodes L_B and S_B; LGA: 26.55% under selective and presentation-only scan, 37.85% under universal scan for fetal size¹³⁸ SGA: 19.6% under selective and presentation-only scan, 56.53% under universal scan for fetal size⁸).

The sensitivity and specificity of ultrasound for detecting SGA and LGA were derived from the POP study.^{8,138} The rationale for using the POP study values is that this study was conducted in NHS England, it involved nulliparous women being scanned at 36 weeks' gestation, it is the only level 1 study of the diagnostic effectiveness of ultrasound to predict SGA and LGA (i.e. where the test result was blinded) and the values of sensitivity and specificity for SGA were similar to those in a 2019 Cochrane review of DTA.²³ In addition, the DOR from the POP study for macrosomia was identical to the DOR in the meta-analysis presented in *Chapter 8*.

If a LGA infant is correctly diagnosed as LGA, the pregnancy is managed in accordance with the defined LGA policy of either IOL or expectant management (node 'MGT_LGA_TP'), in either case leading to either vaginal delivery or emergency caesarean section (nodes L_C3 and L_C2a; odds ratio of emergency caesarean section compared with otherwise healthy infant, 1.79¹⁴⁶). If a LGA infant is misdiagnosed as AGA (i.e. false-negative scan), delivery can be either vaginal or by emergency caesarean section. Further outcomes relating to LGA babies are described in *Outcomes relating to large for gestational age infants*.

If the infant is SGA and is correctly diagnosed as such, labour is induced, leading to either vaginal delivery or emergency caesarean section (node S_C3). False negatives may lead to vaginal delivery or emergency caesarean section (node S_C2). Further outcomes relating to SGA pregnancies are described in *Outcomes relating to small for gestational age infants*.

An AGA infant may be misdiagnosed as SGA or LGA (false-positive SGA and LGA, respectively), or correctly diagnosed as AGA (node B). A false-positive SGA infant will be induced unnecessarily, leading to either vaginal delivery or emergency caesarean section (node S_C4). A false-positive LGA infant will be managed in accordance with the defined LGA policy namely either IOL or expectant management (node 'MGT_LGA_FP'). IOL and expectant management can lead to either spontaneous vaginal or emergency caesarean section delivery (nodes L_C4 and L_C1 respectively). Finally, a correctly diagnosed AGA infant (true negative) can be delivered vaginally or by emergency caesarean section (node C1).

Short- and long-term outcomes

For all parts of the model, different levels of neonatal morbidity and mortality are possible, although these outcomes are structured slightly differently between the model's subtrees. For the breech, SGA and AGA models, delivery outcomes include no, moderate and severe neonatal morbidity, as well as perinatal death. The risks of each level of adverse outcome differ between specific branches (i.e. are affected by the true status of the infant, the mode of delivery and whether or not labour was induced early). Long-term outcomes are then modelled as a function of the level of neonatal morbidity at delivery. For the LGA model, delivery and long-term outcomes are modelled differently. This is explained in detail in *Outcomes relating to large for gestational age infants*.

Long-term outcomes include 'no long-term complications', 'SEN', 'severe neurological morbidity' (SNM) and 'neonatal/infant mortality'. The risk of long-term complications increases with the level of neonatal morbidity (nodes E1, E2 and E3). Unlike delivery outcomes, long-term outcomes are not affected by the actual status of the infant prior to delivery, only by the level of neonatal morbidity at delivery. Importantly, this means that all screening and management options affect long-term outcomes indirectly only as a result of the impact that they have on the outcomes at delivery.

Outcomes relating to breech

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

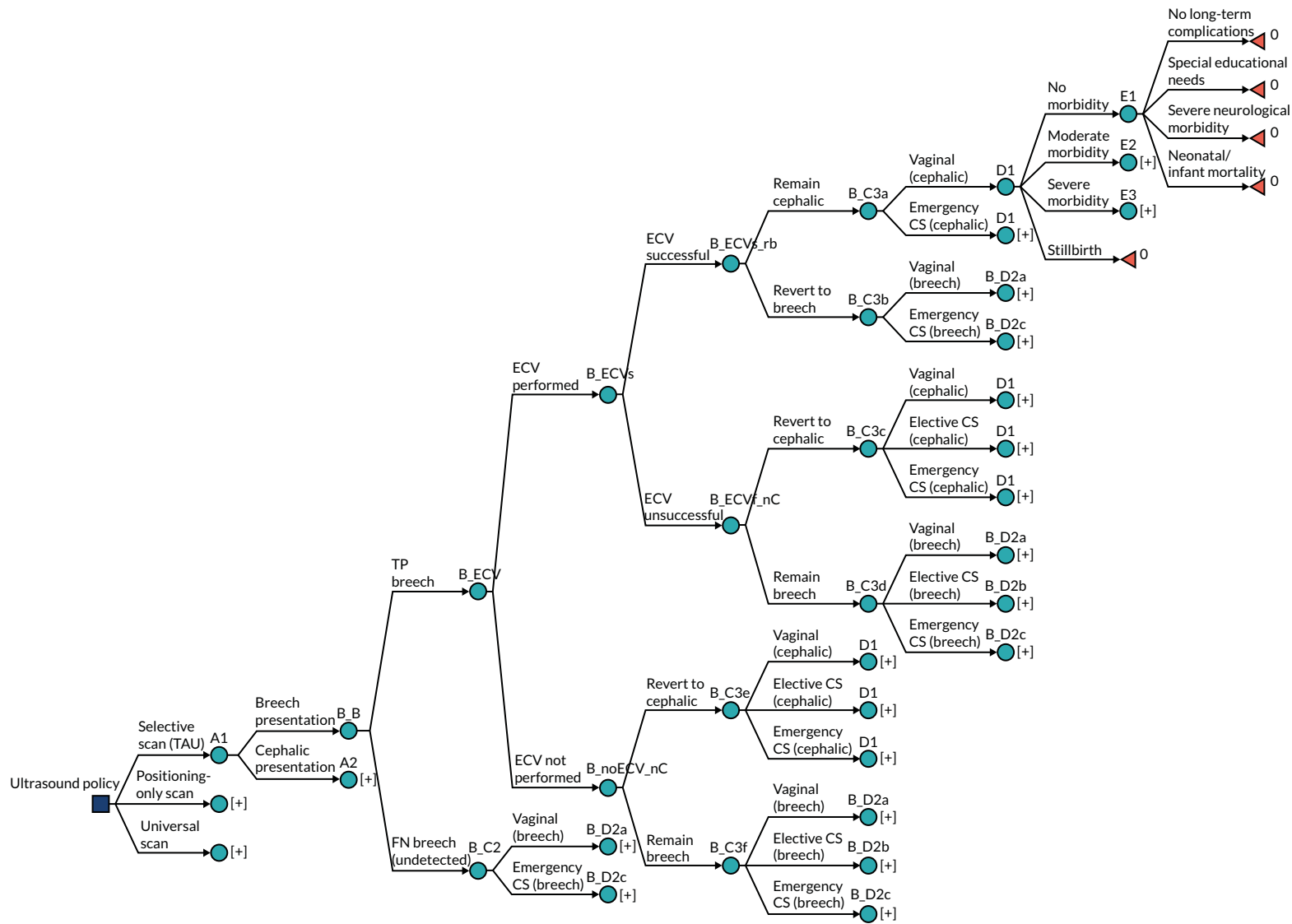


FIGURE 13 Outcomes associated with breech. [+], collapsed sections of the decision tree.

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

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

Outcomes relating to large for gestational age infants

Figure 14 shows the decision tree with outcomes relevant to LGA expanded and remaining branches collapsed. When LGA is suspected, the intervention given will be in accordance with the predetermined management strategy (IOL or expectant management) for both true-positive and false-positive LGA diagnoses. The management option will affect the likelihood of the delivery outcome, as well as the mode of delivery, which can be either vaginal or by emergency caesarean section. When LGA is not suspected, delivery can be either vaginal or by emergency caesarean section.

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

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

Outcomes relating to small for gestational age infants

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

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

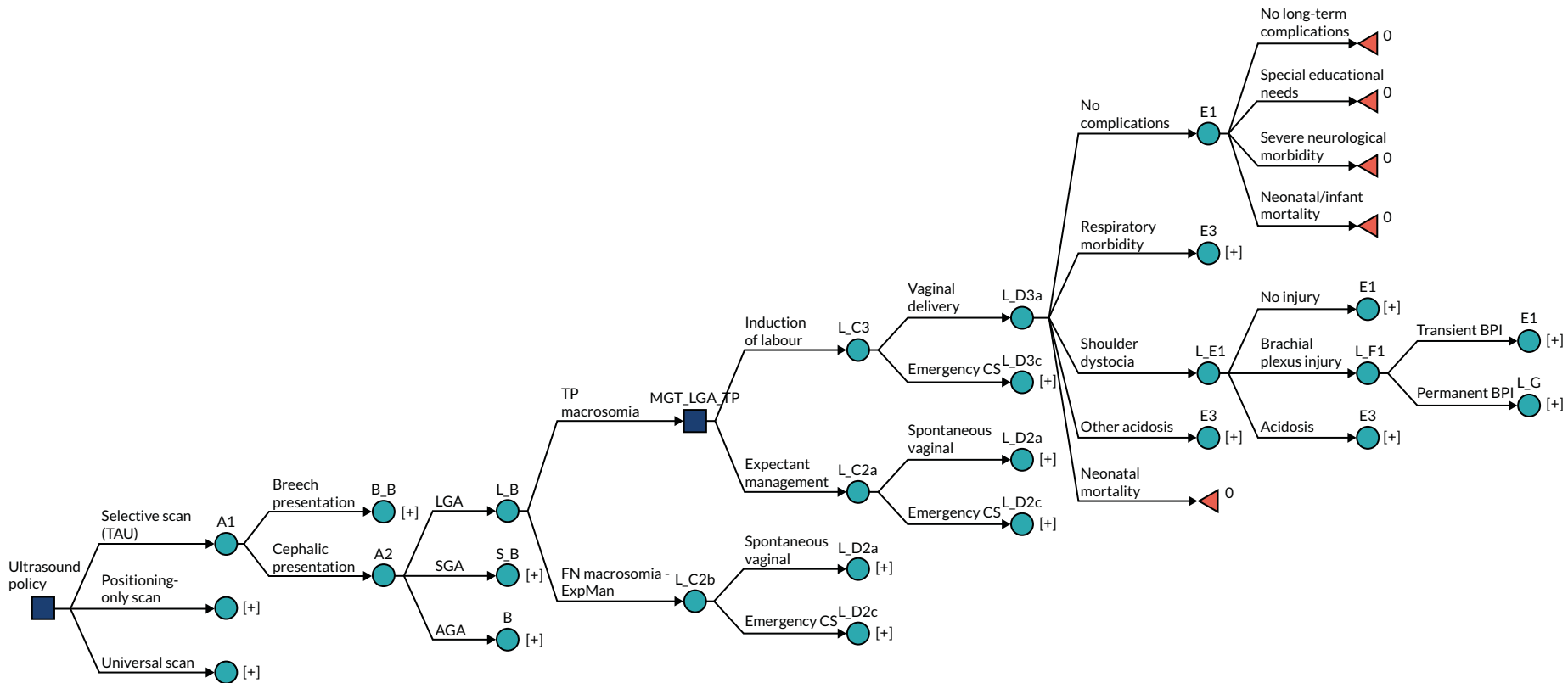


FIGURE 14 Outcomes associated with LGA. [+], collapsed sections of the decision tree.

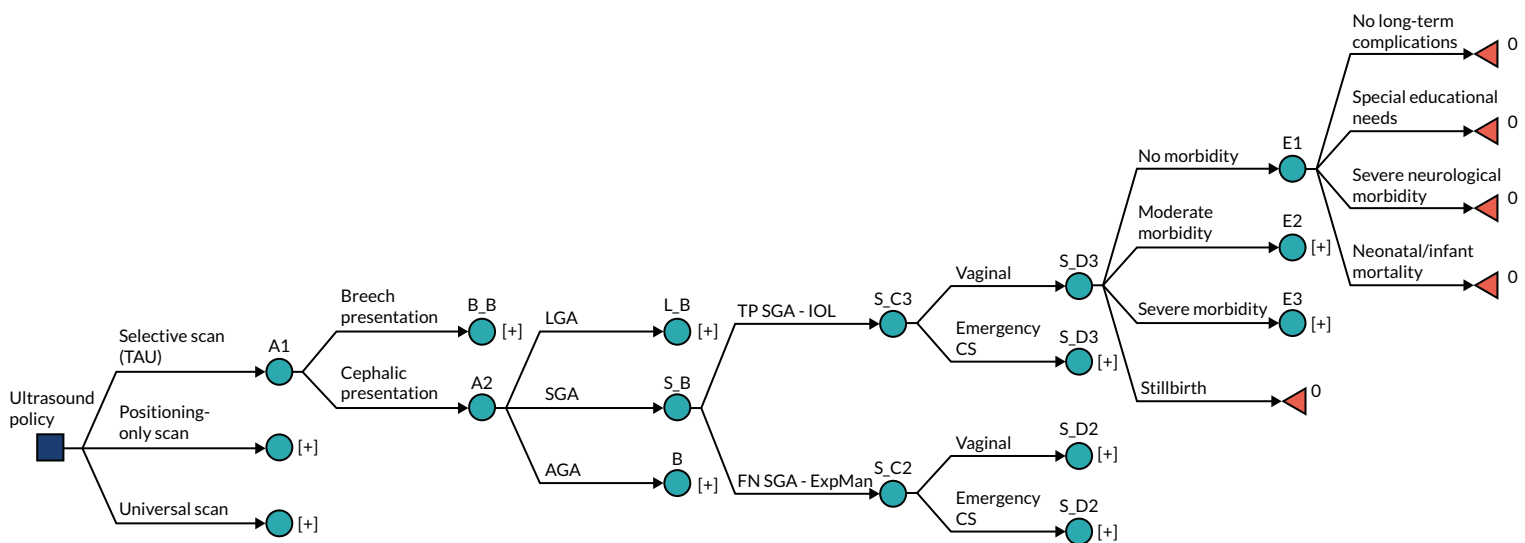


FIGURE 15 Outcomes associated with SGA. [+], collapsed sections of the decision tree.

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

Outcomes relating to appropriate for gestational age infants

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

If an AGA fetus is falsely diagnosed as SGA, early IOL is offered. Unlike in the case of a true SGA, early labour induction of AGA pregnancies increases the risk of morbidity; however, the risk of perinatal death is still reduced.¹⁶⁰ Short- and long-term outcomes will then follow as described in *Short- and long-term outcomes*. If, instead, an AGA fetus is misdiagnosed as LGA, the short- and long-term outcomes depend on the management strategy. Compared with expectant management, early IOL decreases the risk of emergency caesarean section and perinatal death but increases the risk of neonatal morbidity.

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

Data

We populated the model with data from multiple sources from the literature. Where possible, we prioritised the inclusion of good-quality systematic reviews and meta-analyses, followed by large, good-quality clinical trials or cohort studies, as appropriate. When there was no objective evidence for a parameter, we relied on expert opinion either to judge whether or not a study in a related area provided a sufficient proxy or to provide a central estimate and credible interval representing beliefs about plausible values for the parameter. Data sources were subjectively graded as high, moderate or low, where high represented directly relevant data (i.e. providing the required parameter) from a good-quality source (e.g. RCT for relative effects and high-quality epidemiological study for baseline risks). A low grade represents instances in which evidence on the required parameter was absent from the literature and so is sourced from a related parameter, used as indirect evidence and revised reflecting expert opinion as to the plausible values. Full details of the derivation of model inputs are provided in *Appendix 6, Tables 25–30*, and all parameters are listed in *Tables 10–12*.

Probabilities

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

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

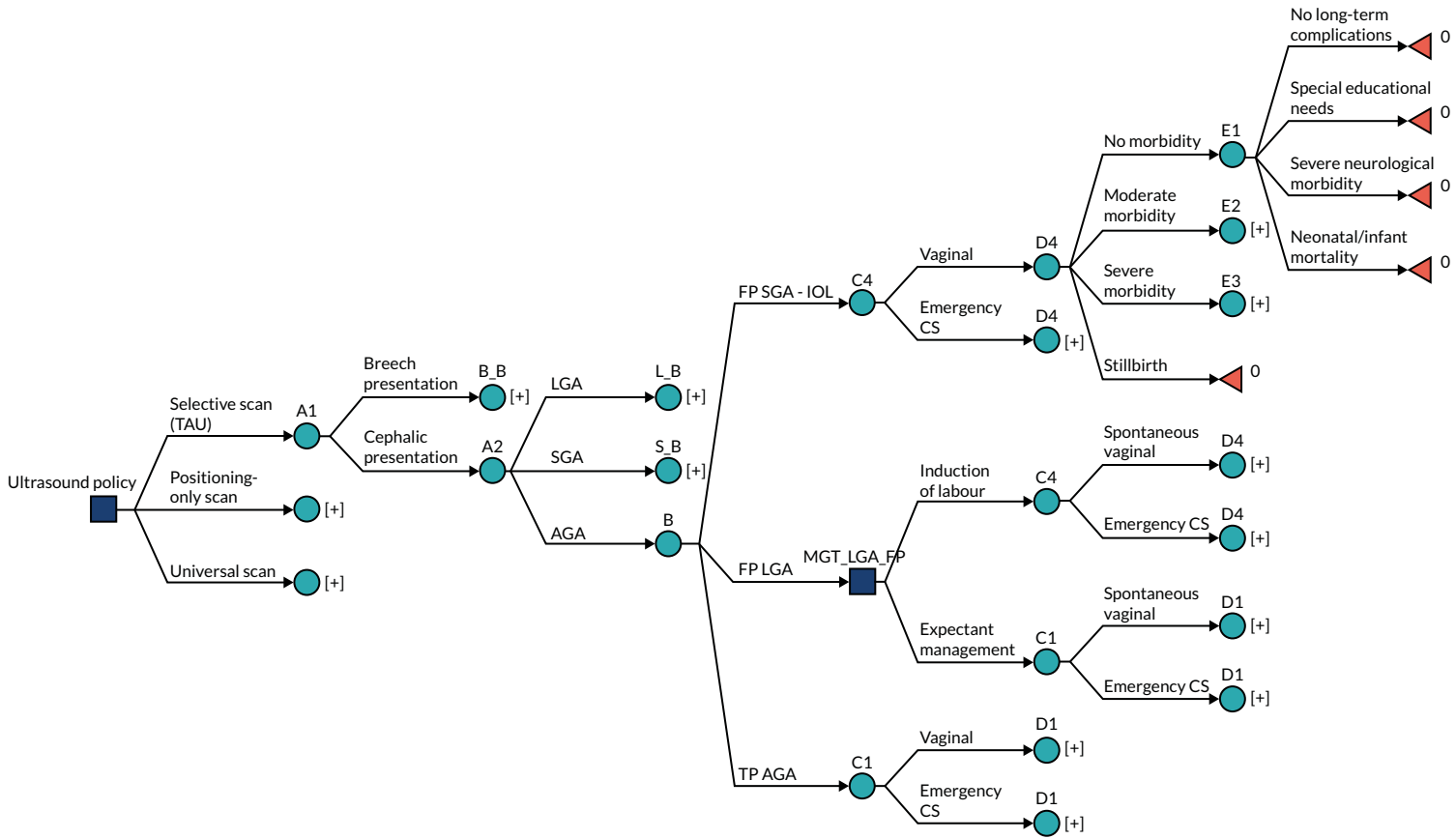


FIGURE 16 Outcomes associated with AGA. [+], collapsed sections of the decision tree.

TABLE 10 Model inputs for diagnostic performance

Parameter	Mean (%) (95% CI)	Distribution summary ^a	Node	Source	Quality of evidence ^b
Prevalence of breech	4.60 (3.98 to 5.30)	~B(179, 3700)	A1	Wastlund <i>et al.</i> ¹¹	High
Prevalence of LGA	10.00 (10 to 10)	N/A	A2	By definition	High
Prevalence of SGA	10.00 (10 to 10)	N/A	A2	By definition	High
Selective ultrasound					
Specificity SGA – selective ultrasound	98.10 (97.63 to 98.52)	~B(3556, 69)	B	Sovio <i>et al.</i> ⁸	High
Specificity LGA – selective ultrasound	98.67 (98.28 to 99.02)	~B(3640, 49)	B	Sovio <i>et al.</i> ¹³⁸	High
Sensitivity SGA – selective ultrasound	19.60 (15.63 to 23.90)	~B(69, 283)	S_B	Sovio <i>et al.</i> ⁸	High
Sensitivity LGA – selective ultrasound	26.55 (20.33 to 33.28)	~B(47, 130)	L_B	Sovio <i>et al.</i> ¹³⁸	High
Sensitivity breech – selective ultrasound	45.10 (37.85 to 52.54)	~B(79, 96)	B_B	Wastlund <i>et al.</i> ¹¹	High
Universal ultrasound for fetal size and presentation					
Specificity SGA – universal ultrasound	89.99 (88.99 to 90.94)	~B(3262, 363)	B	Sovio <i>et al.</i> ⁸	High
Specificity LGA – universal ultrasound	96.56 (95.95 to 97.12)	~B(3562, 127)	B	Sovio <i>et al.</i> ¹³⁸	High
Sensitivity SGA – universal ultrasound	56.53 (52.33 to 61.67)	~B(199, 153)	S_B	Sovio <i>et al.</i> ⁸	High
Sensitivity LGA – universal ultrasound	37.85 (30.87 to 45.10)	~B(67, 110)	L_B	Sovio <i>et al.</i> ¹³⁸	High
Sensitivity breech – universal ultrasound	100 (100 to 100)	N/A	B_B	Assumption	N/A
Universal ultrasound for fetal presentation only					
Specificity SGA – positioning scan	98.10 (97.63 to 98.52)	~B(3556, 69)	B	Sovio <i>et al.</i> ⁸	High
Specificity LGA – positioning scan	98.67 (98.28 to 99.02)	~B(3640, 49)	B	Sovio <i>et al.</i> ¹³⁸	High
Sensitivity SGA – positioning scan	19.60 (15.63 to 23.90)	~B(69, 283)	S_B	Sovio <i>et al.</i> ⁸	High
Sensitivity LGA – positioning scan	26.55 (20.33 to 33.28)	~B(47, 130)	L_B	Sovio <i>et al.</i> ¹³⁸	High
Sensitivity breech – positioning scan	100 (100 to 100)	N/A	B_B	Assumption	N/A
N/A, not applicable.					
a B = beta distribution.					
b Quality assessment. High – good-quality, directly relevant evidence (e.g. directly relevant population, well-conducted RCT for relative effects, or cohort for baseline effects). Medium – directly relevant evidence but poorer-quality source (e.g. retrospective cohort for relative treatment effect). Low – lack of direct evidence/informed by expert opinion. Direct – source provides required parameter. Indirect – source provides related parameter used as background evidence to inform expert opinion. Note that the same source may be used in different contexts; therefore, this results in a different relevance rating to inform different parameters.					

TABLE 11 Model inputs for probabilities

Parameter	Mean (95% CI)	Distribution summary ^a	Node	Source	Quality of evidence ^b
Mode of delivery					
EmCS delivery AGA and Exp Mgt	20.70% (19.4% to 22.06%)	~B(735, 2813)	C1	Wastlund <i>et al.</i> ¹¹	High
RR EmCS delivery SGA and Exp Mgt [FN] vs. C1	1.9 (1.4 to 2.5)	~LN(0.642, 0.14)	S_C2	Monier <i>et al.</i> ²²	Medium
RR EMCS induced, SGA [TP] vs. C1	2.9 (1.8 to 4.7)	~LN(1.065, 0.246)	S_C3	Monier <i>et al.</i> ²²	Low
RR EMCS induced, AGA, [FP SGA] vs. C1	0.84 (0.76 to 0.93)	~LN(-0.174, 0.052)	C4	Grobman <i>et al.</i> ¹⁵⁴	High
OR of EmCS delivery LGA and Exp Mgt [FN] vs. C1	1.792 (0.718 to 4.471)	~LN(0.583, 0.466)	L_C2	Blackwell <i>et al.</i> ¹⁴⁶	Medium
OR of EmCS delivery LGA and Induce [TP] vs. L_C2	0.92 (0.85 to 0.99)	~LN(-0.083, 0.037)	L_C3	Middleton <i>et al.</i> ¹⁶	Low
EmCS delivery breech and Exp Mgt [FN]	57.69% (38.67% to 75.62%)	~B(15, 11)	B_C2	Leung <i>et al.</i> ¹⁶¹	Medium
EmCS delivery breech, ECV success, remain cephalic	27.27% (6.69% to 55.64%)	~B(3, 8)	B_C3a	Wastlund <i>et al.</i> ¹¹	High
EmCS delivery breech, ECV success, revert breech	57.69% (38.67% to 75.62%)	~B(15, 11)	B_C3b	Leung <i>et al.</i> ¹⁶¹	Medium
Vaginal delivery breech, ECV fail, revert cephalic	52.38% (31.51% to 72.80%)	~D(11, 1, 9)	B_C3c	Wastlund <i>et al.</i> ¹¹	High
ELCS delivery breech, ECV fail, revert cephalic	4.76% (0.13% to 16.84%)	-	B_C3c	Wastlund <i>et al.</i> ¹¹	
EmCS delivery breech, ECV fail, revert cephalic	42.86% (23.07% to 63.97%)	-	B_C3c	Wastlund <i>et al.</i> ¹¹	
Vaginal delivery breech, ECV fail, remain breech	0% (0% to 0%)	~D(0, 54, 18)	B_C3d	Wastlund <i>et al.</i> ¹¹	High
ELCS delivery breech, ECV fail, remain breech	75% (64.47% to 84.22%)	-	B_C3d	Wastlund <i>et al.</i> ¹¹	
EmCS delivery breech, ECV fail, remain breech	25% (15.78% to 35.53%)	-	B_C3d	Wastlund <i>et al.</i> ¹¹	
Vaginal delivery breech, no ECV, revert cephalic	52.38% (31.51% to 72.80%)	~D(11, 1, 9)	B_C3e	Wastlund <i>et al.</i> ¹¹	High
ELCS delivery breech, no ECV, revert cephalic	4.76% (0.13% to 16.84%)	-	B_C3e	Wastlund <i>et al.</i> ¹¹	
EmCS delivery breech, no ECV, revert cephalic	42.86% (23.07% to 63.97%)	-	B_C3e	Wastlund <i>et al.</i> ¹¹	
Vaginal delivery breech, no ECV, remain breech	0% (0% to 0%)	~D(0, 52, 20)	B_C3f	Wastlund <i>et al.</i> ¹¹	High
ELCS delivery breech, no ECV, remain breech	72.22% (61.38% to 81.88%)	-	B_C3f	Wastlund <i>et al.</i> ¹¹	
EmCS delivery breech, no ECV, remain breech	27.77% (18.12% to 38.62%)	-	B_C3f	Wastlund <i>et al.</i> ¹¹	

continued

TABLE 11 Model inputs for probabilities (continued)

Parameter	Mean (95% CI)	Distribution summary ^a	Node	Source	Quality of evidence ^b
External cephalic version					
ECV attempted	47.46% (40.16% to 54.81%)	~B(84, 93)	B_ECV	Wastlund <i>et al.</i> ¹¹	High
ECV not attempted, spontaneous reversion to cephalic	22.58% (14.72% to 31.56%)	~B(21, 72)	B_noECV_rc	Wastlund <i>et al.</i> ¹¹	High
Probability ECV successful	14.29% (7.70% to 22.48%)	~B(12, 72)	B_ECVs	Wastlund <i>et al.</i> ¹¹	High
Probability of reverting to breech post successful ECV	8.33% (0.23% to 28.49%)	~B(1, 11)	B_ECVs_rb	Wastlund <i>et al.</i> ¹¹	High
Probability of spontaneous reversion to cephalic post ECV failure	2.31% (0.48% to 5.49%)	~B(3, 127)	B_ECVf_rc	Ben-Meir <i>et al.</i> ¹⁶²	High
Outcomes for LGA model					
Respiratory morbidity, baseline	0.32% (0.20% to 0.46%)	~B(22, 6933)	-	Morrison <i>et al.</i> ¹⁶³	High
Shoulder dystocia, baseline	0.63% (0.60% to 0.66%)	~B(1686, 265542)	-	Ouzounian <i>et al.</i> ¹⁶⁴	Medium
Other acidosis, baseline	0.68% (0.22% to 1.40%)	~B(5, 726)	-	Middleton <i>et al.</i> ¹⁶	High
Perinatal mortality, baseline	0.155% (0.145% to 0.165%)	~B(984, 634412)	-	Moraitis <i>et al.</i> ⁵⁴	Medium
RR respiratory morbidity, LGA vs. AGA [FN and ExpMan LGA policy]	0.75 (0.5125 to 0.9875)	~U(0.5, 1)	L_D2a	Expert opinion	Low
OR shoulder dystocia, LGA vs. AGA [FN and ExpMan LGA policy]	7.18 (2.06 to 25.00)	~LN(1.971, 0.637)	L_D2a	Rossi <i>et al.</i> ¹⁶⁵	High
OR other acidosis, LGA vs. AGA [FN and ExpMan LGA policy]	2.88 (1.34 to 6.22)	~LN(1.058, 0.393)	L_D2a	Rossi <i>et al.</i> ¹⁶⁵	Medium
OR perinatal mortality, LGA vs. AGA [FN and ExpMan LGA policy]	1.77 (0.30 to 10.34)	~LN(0.571, 0.901)	L_D2a	Rossi <i>et al.</i> ¹⁶⁵	Medium
OR respiratory morbidity, LGA vs. AGA, EMCS [FN and ExpMan LGA policy]	5.33 (3.50 to 7.40)	~LN(1.674, 0.167)	L_D2c	Morrison <i>et al.</i> ¹⁶³	High
P shoulder dystocia, LGA, EMCS [FN and ExpMan LGA policy]	0 (0 to 0)	N/A	L_D2c	Assumption	High
OR other acidosis, LGA, EMCS [FN and ExpMan LGA policy]	1.867 (1.217 to 2.865)	~LN(0.625, 0.218)	L_D2c	Chongsuvivatwong <i>et al.</i> ¹⁶⁶	Medium
OR perinatal mortality, LGA, EMCS [FN and ExpMan LGA policy]	1.781 (1.266 to 2.505)	~LN(0.577, 0.174)	L_D2c	Chongsuvivatwong <i>et al.</i> ¹⁶⁶	Medium
OR respiratory morbidity, LGA, IOL, vaginal delivery [TP]	0.54 (0.373 to 0.783)	~LN(-0.616, 0.19)	L_D3a	Gibson <i>et al.</i> ¹⁶⁷	Medium

TABLE 11 Model inputs for probabilities (continued)

Parameter	Mean (95% CI)	Distribution summary ^a	Node	Source	Quality of evidence ^b
RR shoulder dystocia, LGA, IOL, vaginal delivery [TP]	0.6 (0.37 to 0.98)	~LN(-0.511, 0.25)	L_D3a	Boulvain <i>et al.</i> ¹⁰¹	Medium
RR acidosis, LGA, IOL, vaginal delivery [TP]	1.66 (0.61 to 4.55)	~LN(0.507, 0.514)	L_D3a	Middleton <i>et al.</i> ¹⁶	Medium
RR perinatal mortality, LGA, IOL, vaginal delivery [TP]	0.33 (0.14 to 0.78)	~LN(-1.109, 0.439)	L_D3a	Middleton <i>et al.</i> ¹⁶	Medium
OR respiratory morbidity, LGA, IOL, EMCS [TP]	0.54 (0.373 to 0.783)	~LN(-0.616, 0.19)	L_D3c	Gibson <i>et al.</i> ¹⁶⁷	Medium
P shoulder dystocia, LGA, IOL, EMCS [TP]	0 (0 to 0)	N/A	L_D3c	Assumption	High
RR acidosis, LGA, IOL, EMCS [TP]	1.66 (0.61 to 4.55)	~LN(0.507, 0.514)	L_D3c	Middleton <i>et al.</i> ¹⁶	Medium
RR perinatal mortality, LGA, IOL, EMCS [TP]	0.33 (0.14 to 0.78)	~LN(-1.109, 0.439)	L_D3c	Middleton <i>et al.</i> ¹⁶	Medium
Risk of acidosis shoulder dystocia	0.07 (0.0630 to 0.1112)	~B(36, 478)	L_E1	MacKenzie <i>et al.</i> ¹⁶⁸	Low
Risk of BPI shoulder dystocia	0.0856 (0.0496 to 0.0936)	~B(44, 470)	L_E1	^c MacKenzie <i>et al.</i> ¹⁶⁸	Low
Risk of permanent BPI	0.055 (0.024 to 0.098)	~B(8, 137)	L_F1	^c Sandmire <i>et al.</i> ¹⁶⁹	Medium
Neonatal morbidity					
Risk of moderate neonatal morbidity (AGA) [FP]	5.62% (0.0488% to 0.0641%)	~B(198, 3325)	D1	The POP study ^{c,d}	High
Risk of severe neonatal morbidity (AGA) [FP]	0.62% (0.0039% to 0.0091%)	~B(22, 3501)	D1	The POP study ^{c,d}	High
Risk of perinatal death (AGA) [FP]	0.155% (0.145% to 0.165%)	~B(984, 634412)	D1	Moraitis <i>et al.</i> ⁵⁴	Medium
OR moderate neonatal morbidity (SGA vs. AGA, ExpMan)	2.48 (1.75 to 3.51)	~LN(0.91, 0.18)	S_D2	The POP Study ^{c,d}	High
OR severe neonatal morbidity (SGA vs. AGA, ExpMan)	1.88 (0.65 to 5.50)	~LN(0.63, 0.55)	S_D2	The POP Study ^{c,d}	High
OR perinatal death (SGA vs. AGA, ExpMan)	4.39 (3.84 to 5.03)	~LN(1.48, 0.07)	S_D2	Moraitis <i>et al.</i> ⁵⁴	High
RR moderate morbidity induce SGA vs. not inducing SGA [TP]	0.7 (0.50 to 0.98)	~LN(-0.357, 0.172)	S_D3	Middleton <i>et al.</i> ¹⁶	Low
RR severe morbidity induce SGA vs. not inducing SGA [TP]	0.7 (0.50 to 0.98)	~LN(-0.357, 0.172)	S_D3	Middleton <i>et al.</i> ¹⁶	Low
RR perinatal death induce SGA vs. not inducing SGA [TP]	0.33 (0.11 to 0.96)	~LN(-1.109, 0.553)	S_D3	Middleton <i>et al.</i> ¹⁶	Low

continued

TABLE 11 Model inputs for probabilities (continued)

Parameter	Mean (95% CI)	Distribution summary ^a	Node	Source	Quality of evidence ^b
OR of moderate neonatal morbidity if induce AGA [FP SGA or LGA]	1.92 (1.71 to 2.15)	~LN(0.652, 0.058)	D4	Stock <i>et al.</i> ¹⁶⁰	High
OR of severe neonatal morbidity if induce AGA [FP SGA or LGA]	1.92 (1.71 to 2.15)	~LN(0.652, 0.058)	D4	Stock <i>et al.</i> ¹⁶⁰	High
OR of perinatal death if induce AGA [FP SGA or LGA]	0.15 (0.03 to 0.68)	~LN(-1.897, 0.771)	D4	Stock <i>et al.</i> ¹⁶⁰	High
OR of moderate neonatal morbidity vaginal breech vs. vaginal cephalic delivery	6.70 (5.9 to 7.6)	~LN(1.902, 0.064)	B_D2a	Thorngren-Jerneck <i>et al.</i> ¹⁷⁰	High
OR of severe neonatal morbidity vaginal breech vs. vaginal cephalic delivery	6.70 (5.9 to 7.6)	~LN(1.902, 0.064)	B_D2a	Thorngren-Jerneck <i>et al.</i> ¹⁷⁰	High
OR of perinatal death vaginal breech vs. vaginal cephalic delivery	6.68 (2.75 to 16.22)	~LN(1.899, 0.453)	B_D2a	Moraitis <i>et al.</i> ⁵⁴	High
RR of moderate morbidity ELCS vs. vaginal breech delivery	0.43 (0.12 to 1.47)	~LN(-0.844, 0.627)	B_D2b	Hofmeyr <i>et al.</i> ¹⁴	High
RR of severe morbidity ELCS vs. vaginal breech delivery	0.11 (0.01 to 0.87)	~LN(-2.207, 1.055)	B_D2b	Hofmeyr <i>et al.</i> ¹⁴	High
RR of perinatal death ELCS vs. vaginal breech delivery	0.29 (0.1 to 0.86)	~LN(-1.238, 0.555)	B_D2b	Hofmeyr <i>et al.</i> ¹⁴	High
OR of moderate morbidity EMCS vs. vaginal breech delivery	0.533 (0.192 to 1.482)	~LN(-0.629, 0.522)	B_D2c	^c Pasupathy <i>et al.</i> ¹⁷¹	Medium
OR of severe morbidity EMCS vs. vaginal breech delivery	0.533 (0.192 to 1.482)	~LN(-0.629, 0.522)	B_D2c	^c Pasupathy <i>et al.</i> ¹⁷¹	Medium
OR of perinatal death EMCS vs. vaginal breech delivery	0.533 (0.192 to 1.482)	~LN(-0.629, 0.522)	B_D2c	^c Pasupathy <i>et al.</i> ¹⁷¹	Medium
Risk of long-term outcomes from neonatal morbidity					
Risk of SEN no neonatal morbidity	0.0474 (0.0467 to 0.0480)	~B(18736, 376891)	E1	MacKay <i>et al.</i> ¹⁷²	High
Risk of neurological morbidity no neonatal morbidity	0.0008 (0.0007 to 0.0008)	~B(906, 1193647)	E1	Persson <i>et al.</i> ¹⁷³	High
Risk of neonatal/infant mortality no neonatal morbidity	0.002 (0.0020 to 0.0021)	~B(2074, 1011289)	E1	Iliodromiti <i>et al.</i> ¹⁷⁴	High
OR of SEN moderate neonatal morbidity	1.55 (1.43 to 1.67)	~LN(0.438, 0.038)	E2	MacKay <i>et al.</i> ¹⁷²	High

TABLE 11 Model inputs for probabilities (continued)

Parameter	Mean (95% CI)	Distribution summary ^a	Node	Source	Quality of evidence ^b
RR of neurological morbidity moderate neonatal morbidity	10.4 (7.8 to 13.9)	~LN(2.34, 0.149)	E2	Persson <i>et al.</i> ¹⁷³	High
RR of neonatal/infant mortality moderate morbidity	12.82 (9.33 to 17.61)	~LN(2.551, 0.162)	E2	Iliodromiti <i>et al.</i> ¹⁷⁴	High
OR of SEN severe neonatal morbidity	1.66 (1.46 to 1.88)	~LN(0.507, 0.063)	E3	MacKay <i>et al.</i> ¹⁷²	High
RR of neurological morbidity severe morbidity	145.5 (104.0 to 204.1)	~LN(4.98, 0.173)	E3	Persson <i>et al.</i> ¹⁷³	High
RR of neonatal/infant mortality severe morbidity	60.61 (48.17 to 76.26)	~LN(4.104, 0.117)	E3	Iliodromiti <i>et al.</i> ¹⁷⁴	High

ELCS, elective caesarean section; EMCS, emergency caesarean section; ExpMan, expectant management; FN, false negative; FP, false positive; N/A, not applicable; OR, odds ratio; RR, relative risk; TP, true positive.

a Distributions: B = beta; D = Dirichlet; LN = log-normal; and U = uniform.

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

c Parameter estimates were based on data from the source, rather than directly from the source. Details are provided in *Appendix 6, Tables 25–30*.

d Alexandros A Moraitis, Ilianna Armata, Ulla Sovio, Peter Brocklehurst, Alexander EP Heazell, Jim G Thornton, Stephen C Robson, Aris Papageorgiou and Gordon CS Smith, University of Cambridge, 2021.

TABLE 12 Model inputs for costs and related probabilities

Parameter	Mean cost (95% CI)	Distribution summary ^a	Node	Source	Quality of evidence ^b
Ultrasound scan	£107.06 (£70.98 to £134.92)	~G(4.9604, 22.8062)	A	^c National Schedule of Reference Costs, 2016–17 – NHS Trusts and NHS Foundation Trusts ¹⁷⁵	High
Positioning scan only	£48.71 (£8.96 to £88.46)	~U(6.87, 90.55)	A	Expert opinion	N/A
Proportion scanned with ultrasound (selective screening)	0.3499 (0.3349 to 0.3650)	~B(1351, 2510)	A	Sovio <i>et al.</i> ⁸	High
IOL (difference vs. normal delivery)	£125 (–£1343 to £1594)	~N(125.3, 749.2)	B1, B2	Vijgen <i>et al.</i> ¹⁷⁶	Medium
Cost of vaginal (cephalic) delivery	£1834 (£1750 to £2236)	~G(7.2606, 252.5824)	C1 – C4	^c National Schedule of Reference Costs, 2016–17 – NHS Trusts and NHS Foundation Trusts ¹⁷⁵	High
Relative cost difference (vaginal breech vs. cephalic delivery)	1.1633 (1.0982 to 1.2284)	~N(1.1633, 0.0332)	B_C3b, B_C3d, B_C3f, B_C2	Palencia <i>et al.</i> ¹⁷⁷	Medium

continued

TABLE 12 Model inputs for costs and related probabilities (continued)

Parameter	Mean cost (95% CI)	Distribution summary ^a	Node	Source	Quality of evidence ^b
Cost of ECV	£292.30 (£287.50 to £297.1)	~U(287.22, 297.38)	B_ECV	^c James <i>et al.</i> ¹⁷⁸	Medium
Cost of emergency caesarean section	£4688 (£3816 to £5443)	~G(14.7329, 318.1354)	C1 – C4	^c National Schedule of Reference Costs, 2016–17 – NHS Trusts and NHS Foundation Trusts ¹⁷⁵	High
Cost of elective caesarean section	£3412 (£2680 to £4038)	~G(11.1212, 307.0169)	C1 – C4	^c National Schedule of Reference Costs, 2016–17 – NHS Trusts and NHS Foundation Trusts ¹⁷⁵	High
Cost of SCBU admission	£1064 (£487 to £1862)	~G(9.0371, 117.7307)	D1 – D4	^c National Schedule of Reference Costs, 2016–17 – NHS Trusts and NHS Foundation Trusts ¹⁷⁵	High
Cost of NHDU admission	£1346 (£807 to £2020)	~G(18.7696, 71.7047)	D1 – D4	^c National Schedule of Reference Costs, 2016–17 – NHS Trusts and NHS Foundation Trusts ¹⁷⁵	High
Cost of NICU admission	£2590 (£1280 to £4352)	~G(10.7403, 241.0768)	D1 – D4	^c National Schedule of Reference Costs, 2016–17 – NHS Trusts and NHS Foundation Trusts ¹⁷⁵	High
Proportion of neonates admitted to SCBU	74% (65% to 82%)	~D(74, 7, 19)	D1 – D4	Alfirevic <i>et al.</i> ¹⁷⁹	Medium
Proportion of neonates admitted to NHDU	7% (3% to 13%)	–	D1 – D4	Alfirevic <i>et al.</i> ¹⁷⁹	
Proportion of neonates admitted to NICU	19% (12% to 27%)	–	D1 – D4	Alfirevic <i>et al.</i> ¹⁷⁹	
Probability of admission to care no neonatal morbidity	0.074 (0.066 to 0.082)	~B(292, 3659)	D1 – D4	Sovio <i>et al.</i> ⁸	High
Odds ratio of admission to care moderate neonatal morbidity	11.29 (5.90 to 21.60)	~LN(2.424, 0.331)	D1 – D4	Sovio <i>et al.</i> ⁸	High
Probability of admission to care severe neonatal morbidity	1 (1 to 1)	N/A	D1 – D4	Assumption	N/A
Short-term cost of acidosis/anoxia	£3240 (£806 to £7328)	~G(3.6143, 895.6169)	L_E1, L_D2a	Own estimation ^c	Low
Short-term cost of respiratory morbidity	£2011 (£993 to £3381)	~G(10.7125, 187.6316)	L_D2a, L_D3a	Own estimation ^c	Low
Cost of transient BPI	£2066 (£1033 to £4132)	~LN(7.6334, 0.3536)	L_F1	Culligan <i>et al.</i> ¹⁸⁰	Medium
Cost of permanent BPI	£14,134 (£7068 to £28,264)	~LN(9.5563, 0.03536)	L_F1	^c Culligan <i>et al.</i> ¹⁸⁰	Medium
Cost of perinatal or infant mortality	£1664 (£1372 to £1956)	~U(1357, 1971)	D1 and E1 – 3	Mistry <i>et al.</i> ¹⁸¹	Medium

TABLE 12 Model inputs for costs and related probabilities (continued)

Parameter	Mean cost (95% CI)	Distribution summary ^a	Node	Source	Quality of evidence ^b
SEN (per annum)	£7428 (£4467 to £10,389)	~N(7428.1, 1511)	E1 – E3	Barrett <i>et al.</i> ¹⁸²	Medium
SNM (per annum)	£2930 (£1465 to £5859)	~LN(7.9826, 0.3536)	E1 – E3	^c Access economics ¹⁸³	Medium

N/A, not applicable; NHDU, neonatal high-dependency unit; NICU, neonatal intensive care unit; SCBU, special care baby unit.

a Distributions: B = beta; D = Dirichlet; G = gamma; LN = log-normal; N = normal; and U = uniform.

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

c Parameter estimates were based on data from the source, rather than directly from the source. Details are provided in *Appendix 6*.

Costs

The price year used in the analysis is 2016/17. The majority of costs were sourced from the English national schedule of reference costs.¹⁷⁵ The national schedule of reference costs reports different costs depending on how the service was delivered (e.g. elective inpatient, non-elective inpatient, outpatient procedures). We used costs from total Healthcare Resource Groups (i.e. weighted by each category by the number of yearly activities), except for cases in which only one or a few categories made logical sense. In all categories in the schedule costs were reported as mean and interquartile range. To obtain parameter estimates of costs, we fitted a gamma distribution using these data points. Where multiple cost categories were used, we first calculated a weighted average of the mean and interquartile range by the number of yearly activities in each category before fitting the gamma distribution.

Where no directly applicable cost could be identified from the reference schedule, we first attempted to obtain resource use from literature, and assign costs to this using the reference costs. When insufficient data on resource usage were available, we adopted the costs directly from the literature. Costs reported in currencies other than Great British pounds or in 2016/17 prices were converted to Great British pounds at the exchange rate of the year that the source was published and inflated to 2016/17 prices using the Hospital & Community Health Services (HCHS) index.¹⁸⁴ Where no credible estimates could be identified from the literature, we estimated the costs ourselves, assigning a wide credibility interval to represent the uncertainty. Full details on the derivation of all cost parameters are presented in *Appendix 6*.

All costs presented in Great British pounds and updated to the cost-year of 2016–17 using the Hospital & Community Health Services Index:¹⁸⁴ quality of life

We estimated age-specific quality of life for healthy neonates using EuroQol data for a general UK population.¹⁸⁵ Age-specific health state utilities were multiplied by age-specific survival,¹⁸⁶ the discounted sum over the time horizon of the model yielding the expected QALYs gained for an otherwise healthy neonate. Per definition, the quality of life following mortality is zero, and we made the simplifying assumption that all deaths during a particular year of life occurred on the first day of the year. In the absence of suitable evidence of how SEN affect quality of life, we assumed for our base-case scenario that SEN would affect costs only. In the case of SNM, we adjusted the baseline quality of life with a relative decrease following the methodology of Leigh *et al.*,¹⁸⁷ using cerebral palsy (CP) as a proxy for SNM. Full details on the derivation of quality-of-life parameters are presented in *Appendix 6*.

Analysis

The model was analysed via Monte Carlo simulation, capturing the overall uncertainty in cost-effectiveness as a function of the uncertainty of the input parameters. Health outcomes were from the fetal perspective only and ultimately presented as QALYs. Cost-effectiveness was explored through incremental cost-effectiveness ratios (ICERs) and net monetary benefits (NMBs), using a WTP threshold of £20,000 per QALY. All costs and QALYs were discounted by 3.5% per annum.¹⁸⁸ All costs were from a third-party (payer) perspective (i.e. NHS England plus SEN costs) and the reference case time horizon was 20 years (varied in sensitivity analysis).

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

Cost-effectiveness: reference case

For each of the six discrete strategies, we present mean and 95% credibility intervals for cost and QALYs gained, net benefit at a WTP of £20,000 per QALY, and incremental net monetary benefit (INMB) relative to the assumed status quo (selective scanning with IOL for macrosomia or SGA, offer of ECV for breech). The option with the highest expected NMB was identified as the most cost-effective. Decision uncertainty was expressed as the probability that each decision would be cost-effective at the reference case threshold (i.e. £20,000/QALY). The cost-effectiveness acceptability curve plots decision uncertainty as a function of WTP per QALY (see *Figure 17*).

Cost-effectiveness: sensitivity and scenario analyses

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

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

In addition, because of concerns over the validity of input data, we also explore the difference in:

- the risk of acidosis and respiratory morbidity associated with vaginal delivery of a LGA infant (vs. AGA)
- the odds ratio of perinatal death resulting from delivery by emergency caesarean section of a breech infant (vs. vaginal delivery)
- the relative risk of an emergency caesarean section from IOL for a SGA infant (vs. expectant management of an AGA infant)
- the relative risk of SEN as a result of inducing labour (vs. expectant management), and the impact that IOL has on health-related quality of life, and the sensitivity of ultrasound scanning at detecting SGA.

Value-of-information analysis

Uncertainty in cost-effectiveness results (i.e. decision uncertainty) was used to conduct a VOI analysis.¹⁸⁹ Decision uncertainty arises from parameter uncertainty. The EVPI is the expected value of eliminating all decision uncertainty, which by definition implies eliminating all parameter uncertainty. This therefore provides an upper bound for the value of all research into the decision question. The EVPPI is the expected value of eliminating uncertainty in a single parameter or group of parameters. The EVSI is the expected value of a study of sample size n . The EVSI of a study of size n less the cost of conducting it provides a measure of the expected return on investment in that research project [expected net gain of sampling (ENGS)].¹⁹⁰⁻¹⁹² An EVPPI above the plausible cost of a research project is a necessary condition for future research to be economically viable. A positive ENGS is the sufficient condition. The efficient sample size of a study is that which maximises the ENGS.

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

We report the per-patient (i.e. per mother/infant dyad) and population EVPI at a WTP of £20,000 per QALY. We then report the per-patient and population EVPPI for each parameter individually, calculated using the Sheffield Accelerated Value-of-information (SAVI) tool.¹⁵⁹ Parameters with a positive EVPPI were grouped into those that could logically be collected in one research study, and the EVPPI for that group of parameters was calculated (also with the SAVI tool¹⁵⁹). The EVSI for any parameters or groups of parameters is then calculated using the method of Heath *et al.*¹⁹³ Population values are presented as a 'conservative' estimate, assuming that the information is of value only to singleton nulliparous pregnancies (i.e. using the 1,689,663 beneficiary population) and a broader estimate that assumes the information is of value to all pregnancies in England (5,477,940 population).

Results

Stability testing

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

Cost-effectiveness results

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

TABLE 13 Results from stability testing

Simulations	Computation time (seconds)	Mean coefficient of variation (%)
10	0.10	24.68
100	0.09	7.73
1000	0.33	2.53
10,000	2.75	0.56
100,000	29.56	< 0.01

TABLE 14 Cost-effectiveness results (per woman scanned)

Screening and management	Cost (£), mean (95% credibility interval)	QALYs, mean (95% credibility interval)	NB £20,000, mean (95% credibility interval)	INB £20,000, mean (95% credibility interval)	P_CE £20,000 (%)
Selective ultrasound and induction	6090 (4420 to 7890)	13.640 (13.441 to 13.841)	266,719 (262,333 to 271,079)	0 (0 to 0)	0.65
Selective ultrasound and expectant	6091 (4424 to 7889)	13.639 (13.439 to 13.839)	266,682 (262,297 to 271,040)	-37.09 (-124.7 to 35.24)	0.22
Universal ultrasound for presentation and induction ^a	6101 (4443 to 7887)	13.645 (13.446 to 13.846)	266,806 (262,426 to 271,154)	87.36 (4.88 to 205.68)	44.19
Universal ultrasound for presentation and expectant	6102 (4446 to 7887)	13.644 (13.444 to 13.844)	266,769 (262,389 to 271,120)	50.29 (-68.06 to 186.43)	15.63
Universal ultrasound for size and expectant	6178 (4508 to 7972)	13.646 (13.446 to 13.846)	266,734 (262,351 to 271,099)	14.47 (-133.98 to 173.31)	0.51
Universal ultrasound and induction	6180 (4498 to 7983)	13.648 (13.448 to 13.849)	266,779 (262,386 to 271,147)	60.24 (-151.43 to 281.7)	38.81

INB, incremental net benefit relative to current practice (selective ultrasound and IOL); NB, net benefit; P_CE, probability of being the most cost-effective strategy.

^a Strategy with the highest expected net benefit.

Note

Management refers to management strategy when LGA is suspected; all babies that are of suspected SGA are assumed induced.

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

One-way and scenario analyses

Cost-effectiveness conclusions were sensitive only to the time horizon, the cost of an ultrasound scan for fetal presentation only, the background risk of stillbirth, moderate and severe perinatal complications, and the risk of SEN associated with IOL.¹⁸⁹

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

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

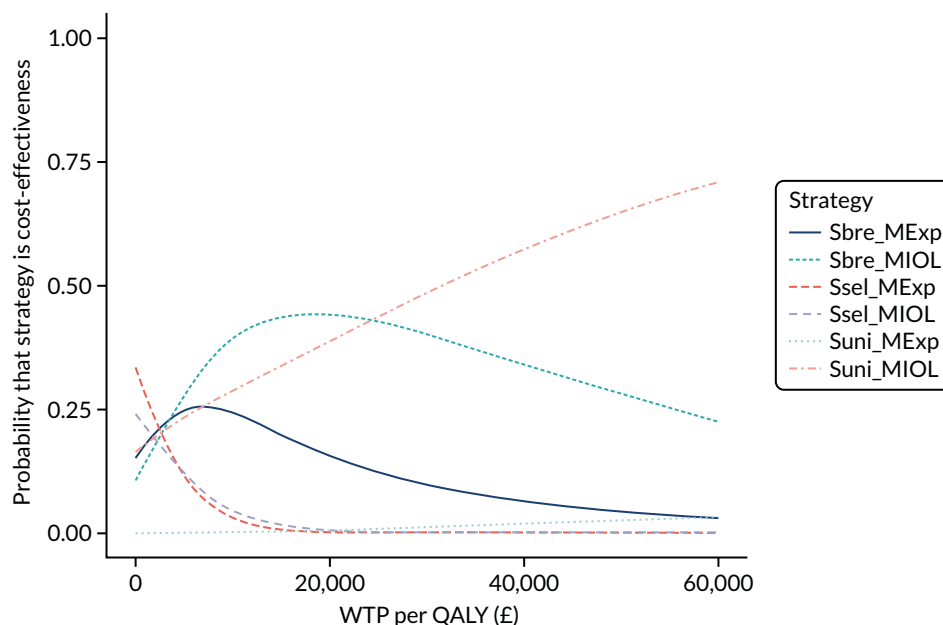


FIGURE 17 Cost-effectiveness acceptability curve for the chance that each strategy will be the most cost-effective as a function of WTP for an additional QALY. Mexp, expectant management; MIOL, IOL; Sbre, universal ultrasound for fetal presentation only; Ssel, selective ultrasound; Suni, universal ultrasound for fetal biometry plus presentation.

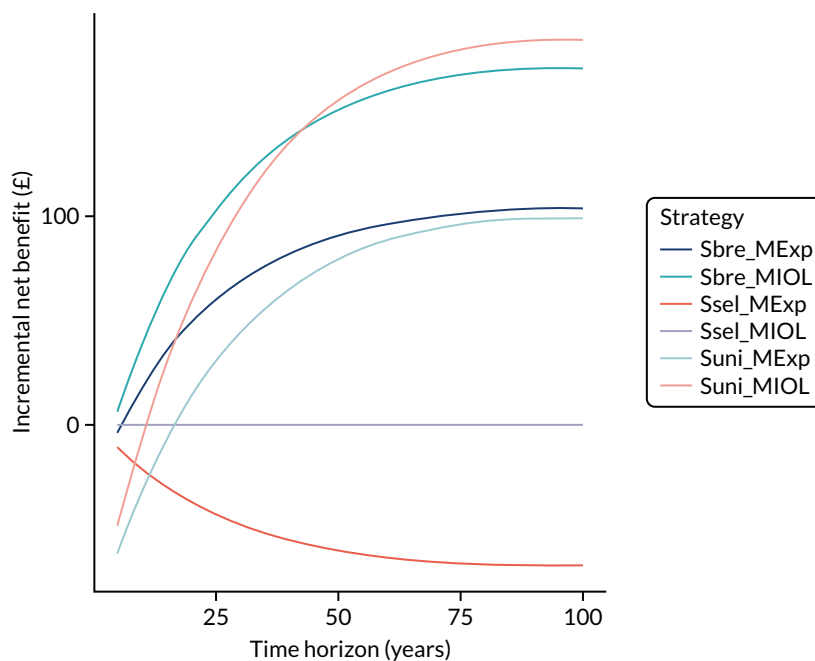


FIGURE 18 One-way sensitivity analysis of model time horizon. MExp, expectant management; MIOL, IOL; Sbre, universal ultrasound for fetal presentation only; Ssel, selective ultrasound; Suni, universal ultrasound for fetal biometry plus presentation.

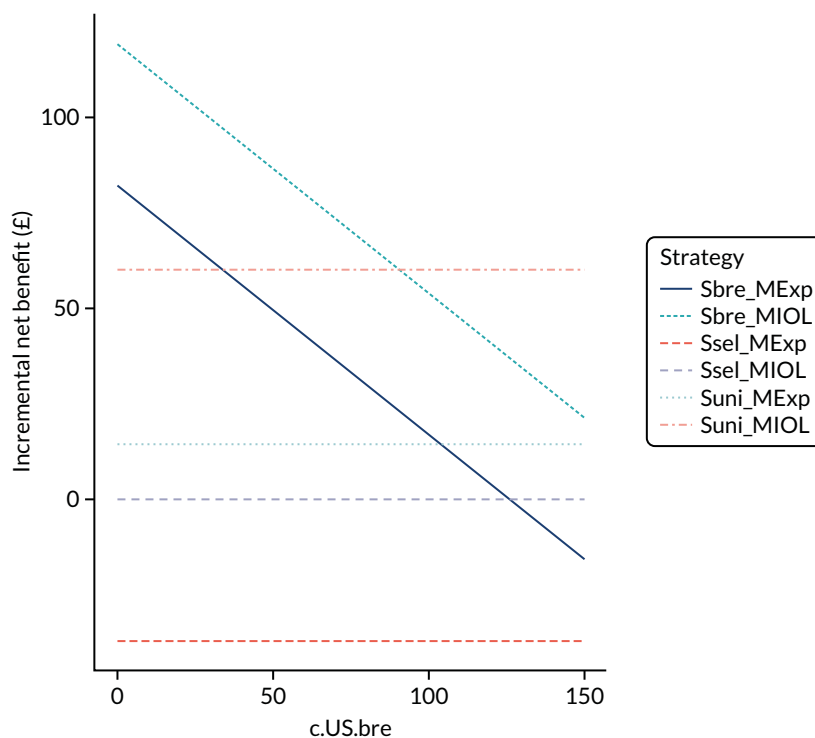


FIGURE 19 One-way sensitivity analysis of the cost of a scan for fetal presentation only. MExp, expectant management; MIOL, IOL; Sbre, universal ultrasound for fetal presentation only; Ssel, selective ultrasound; Suni, universal ultrasound for fetal biometry plus presentation.

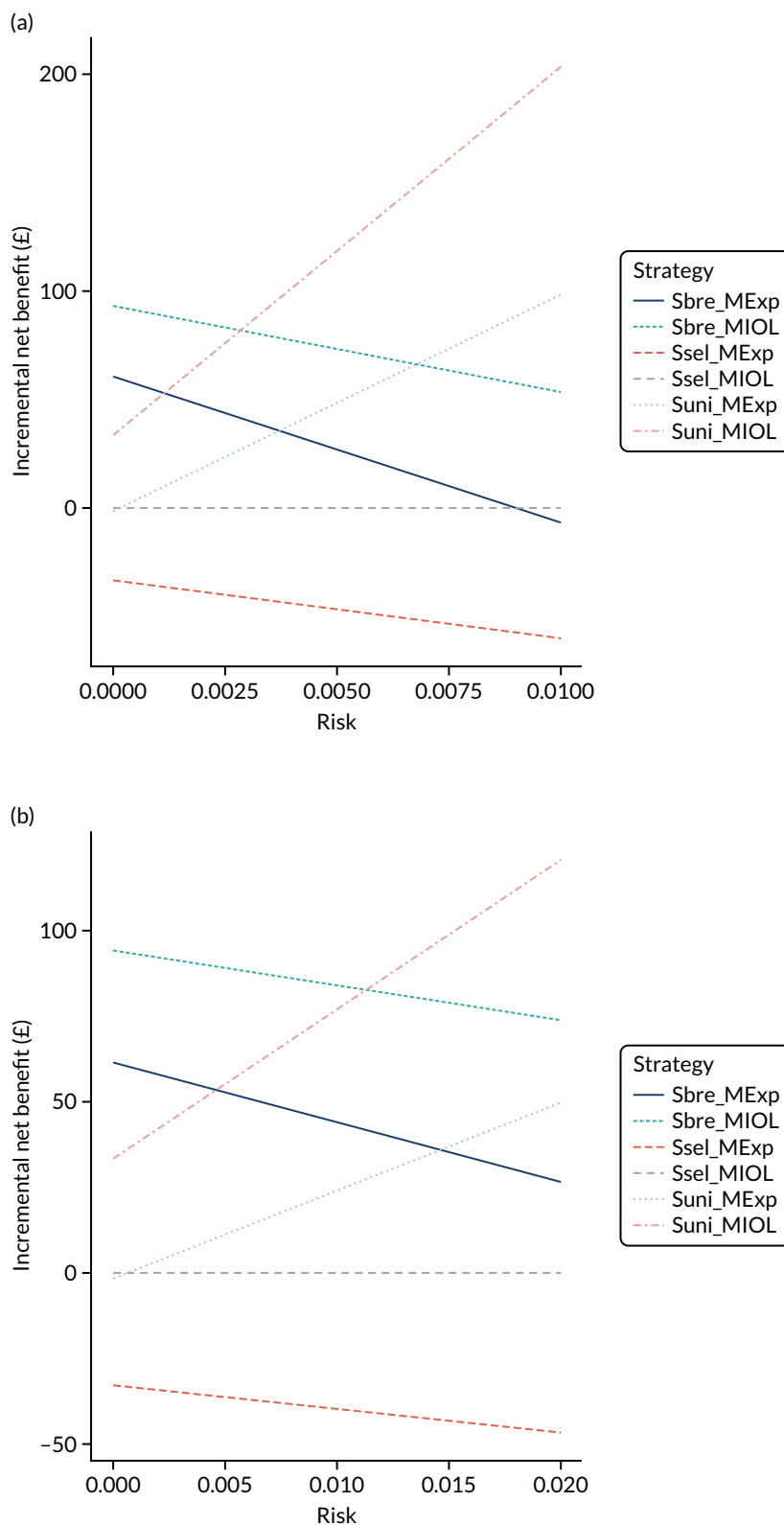


FIGURE 20 One-way sensitivity analysis of baseline risk of (a) perinatal mortality; (b) severe morbidity; and (c) moderate morbidity. MExp, expectant management; MIOL, IOL; Sbre, universal ultrasound for fetal presentation only; Ssel, selective ultrasound; Suni, universal ultrasound for fetal biometry plus presentation. (continued)

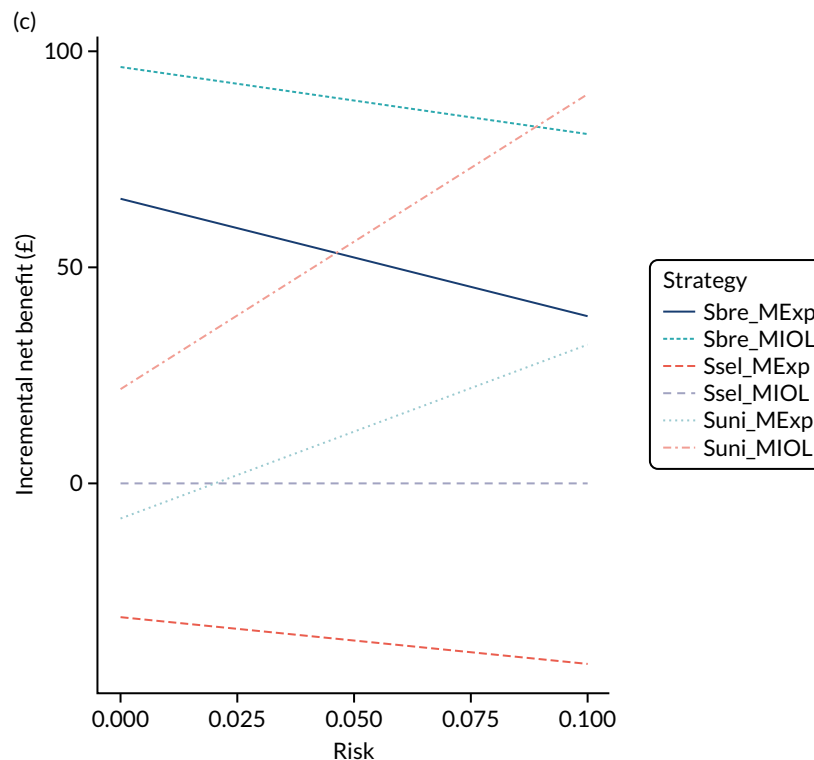


FIGURE 20 One-way sensitivity analysis of baseline risk of (a) perinatal mortality; (b) severe morbidity; and (c) moderate morbidity. MExp, expectant management; MIOL, IOL; Sbre, universal ultrasound for fetal presentation only; Ssel, selective ultrasound; Suni, universal ultrasound for fetal biometry plus presentation.

Our base-case analysis assumed a linear progression through the model whereby long-term outcomes were dependent on perinatal outcomes, which were dependent on mode of delivery alone [vaginal vs. caesarean section (emergency or elective)]. However, there is evidence to suggest that IOL may increase the risk of SEN in later life.¹⁷² We therefore explored the impact on the results via a one-way sensitivity analysis. We found that our results remained the same as long as the relative risk of SEN as a result of IOL is between approximately 0.95 and 1.3 and the estimated risk at 38 weeks' gestation was within this range.¹⁷² Below this risk, the most cost-effective strategy is to perform universal screening for both presentation and EFW, and to induce labour when SGA or LGA is suspected. Above this risk, then while the recommended scan remains a presentation-only scan, the most cost-effective intervention for suspected SGA or LGA is expectant management (i.e. IOL ceases to be the appropriate intervention; *Figure 21*). Given this, although not captured in our formal VOI analysis (because of structural assumptions), it may be worthwhile exploring the impact that inducing labour has on long-term risk of SEN in future research.

Figure 18 shows the expected INMB for different strategies compared with current practice (selective ultrasound with IOL for suspected LGA) as a function of the model's time horizon (years). Calculations are based on a WTP (i.e. valuation of one additional QALY) of £20,000.

Figure 19 shows the expected INMB for different strategies compared with current practice (selective ultrasound with IOL for suspected LGA) as a function of the cost of an ultrasound for fetal presentation only. Calculations are based on a WTP (i.e. valuation of one additional QALY) of £20,000.

Figure 20 shows the expected INMB for different strategies compared with current practice (selective ultrasound with IOL for suspected LGA) as a function of the baseline risk of perinatal mortality (see *Figure 20a*), severe neonatal morbidity (see *Figure 20b*) and moderate neonatal morbidity (see *Figure 20c*). Calculations are based on a WTP (i.e. valuation of one additional QALY) of £20,000.

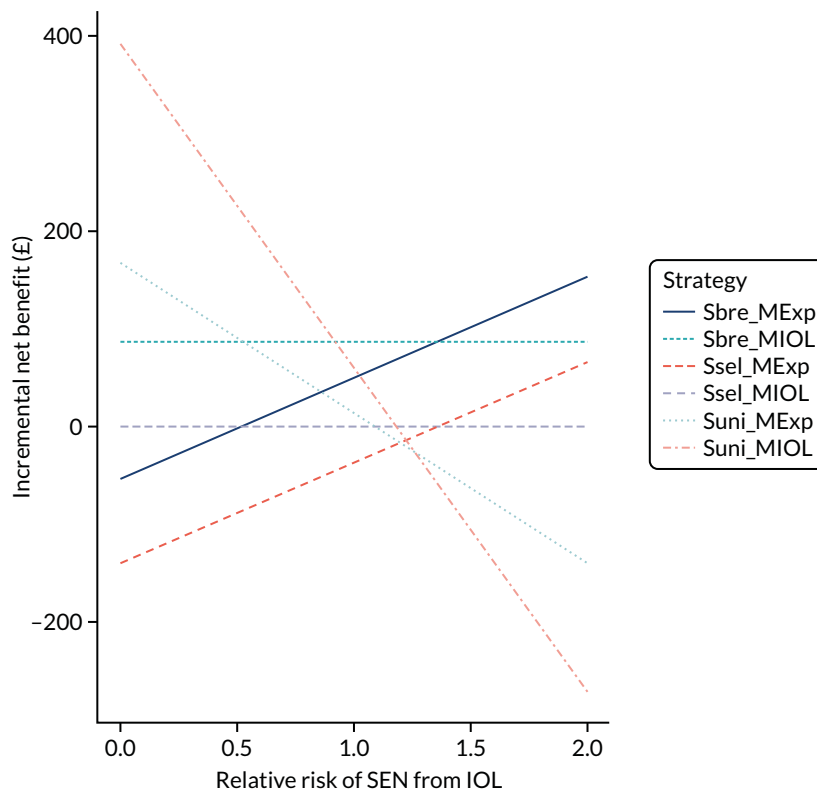


FIGURE 21 One-way sensitivity analysis on relative risk of SEN from IOL. MExp, expectant management; MIOL, IOL; Sbre, universal ultrasound for fetal presentation only; Ssel, selective ultrasound; Suni, universal ultrasound for fetal biometry plus presentation.

Figure 21 shows the expected INMB for different strategies compared with current practice (selective ultrasound with IOL for suspected LGA) as a function of the relative risk of SEN if labour is induced early (compared with expectant management). Calculations are based on a WTP (i.e. valuation of one additional QALY) of £20,000.

Value-of-information analysis

Expected value of perfect information

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

Expected value of perfect parameter information and expected value of sample information

Table 15 shows the parameters with an EVPPi exceeding £100,000 under the broader assumption that any future study will be of value to all births in England, not just low-risk singleton pregnancies. The most valuable parameter is difference in cost of delivery from IOL, accounting for 84% of the EVPI. Except for this cost, no other parameters individually account for > 1% of the total EVPI. The other parameters with the greatest contribution to EVSI are the relative risk (LGA vs. AGA) of acidosis from a vaginal delivery following IOL, the odds ratio of perinatal death (LGA vs. AGA) from an infant delivered vaginally without IOL, the relative risk (SGA vs. AGA) of emergency caesarean section following IOL and the odds ratio (SGA vs. AGA) of severe neonatal morbidity under expectant management.

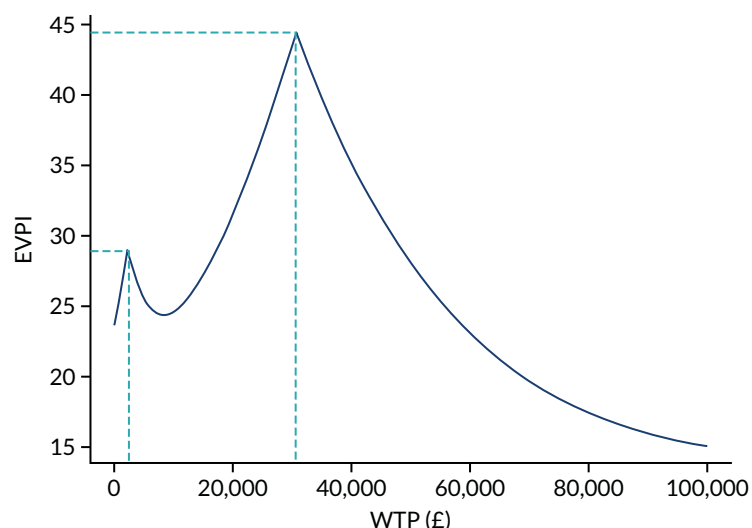


FIGURE 22 Per-patient EVPI as a function of the WTP for an additional QALY. EVPI is presented per person. WTP refers to monetary valuation of an additional QALY (£).

TABLE 15 The expected value of partial perfect information for individual parameters and groups of parameters

Parameter	Per-patient EVPPI (£)	Standard error	Percentage of EVPI	pEVPPI (£)	pEVPPI (£) ^a
Cost of delivery from IOL	26.51	0.07	84	44,790,000	145,200,000
RR for acidosis in macrosomic fetuses if induced early	0.27	0.04	1	456,000	1,478,000
OR for mortality if fetus is macrosomic	0.26	0.03	1	438,900	1,423,000
Group	0.72	0.07	2	1,215,199	3,939,513
RR for emergency caesarean section among SGA fetuses following early labour induction	0.06	0.01	0	99,290	321,900
OR for severe neonatal morbidity if fetus is SGA	0.03	0.01	0	48,740	158,000
Group	0.26	0.04	1	443,104	1,436,484

OR, odds ratio; RR, relative risk.

^a Assuming study results are applicable to all births in England.

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

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

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

The EVSI of the costing study suggests that there is scope for the study to yield a positive return on investment. For example, a two-arm study with 1000 patients in each arm has an EVSI to England of £11.3M (or £97.2M if this information is of value to all pregnancies in England, not just to low-risk nulliparous singleton pregnancies). If such a study was to cost £1M, then it would yield a net return on investment of at least £10.3M (Figure 23).

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

Expected value of perfect parameter information under alternative scenarios

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

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

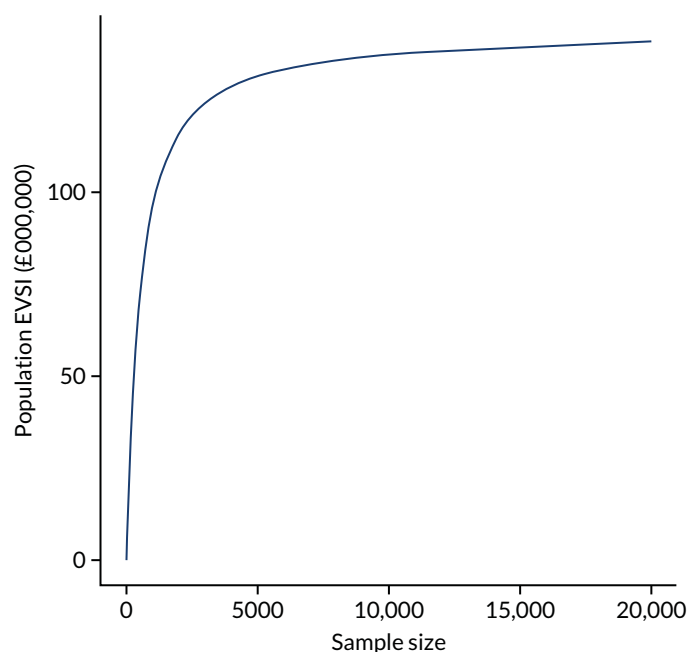


FIGURE 23 Population EVSI for a study on the cost of IOL.

The cost of IOL (specifically, the net difference in total cost between pregnancies that were induced early and that of expectant management) had the highest EVPPI in our base-case scenario, and hence the greatest expected benefit from future research. In the base-case scenario, the cost was £125 (95% CI –£1343 to £1594); more details are presented in *Appendix 6*. To test how sensitive the EVPPI was to the exact input values used, we simulated two alternative scenarios: (1) where the standard error of the mean was reduced by 50% and (2) where costs were instead obtained from the 35/39 trial,¹⁹⁴ where the cost difference was –£236 (95% CI –£646 to £174);¹⁹⁴ see *Appendix 6* for details. When the standard error was reduced by 50%, the EVPPI fell by \approx 80%. When costs were obtained from the 35/39 trial, the EVPPI was £6.3M for the beneficial population (i.e. nulliparous women).

Discussion

Main findings

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

Third-trimester scans for fetal size should take place only where clinically indicated. We estimate that the added benefits of including estimation of fetal weight in the scan may not justify the added cost; more health would be lost elsewhere than would be gained from the added knowledge and subsequent management from these scans. When LGA is suspected following ultrasound, early IOL is the preferred management irrespective of whether screening is offered routinely or following clinical indication.

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

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

The net difference in cost between an induced delivery and expectant management was the parameter that had the largest impact on decision uncertainty in the base-case scenario, and hence this is the parameter that should be prioritised in future research. It should be noted that this does not relate simply to the cost of a procedure to induce delivery; included in this definition is uncertainty about the timing of induction, and the impact on, for example, antenatal appointments, as well as the cost of the delivery itself. A study of 'reasonable size' to reduce uncertainty in this parameter is likely to yield a positive return on investment. For example, the EVSI of a study with 1000 women in each arm is worth in excess of £11M. If this was to be delivered at a cost of £1M, it would yield a > 10-fold return on investment. Alternative scenarios found that the value of future research may be less than for the base-case scenario. Nonetheless, although the exact value of future research is hard to determine, the net cost of labour induction appears influential on which screening strategy is the most cost-effective.

Of note is that studies on the outcomes for SGA or LGA fetuses are unlikely to yield a positive return on investment based on the model.

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

The results described above relate to a WTP threshold of £20,000 per QALY. At a threshold of £30,600 per QALY (just above the upper threshold of NICE's stated acceptable range of £20,000–30,000¹⁸⁸), universal scanning becomes the most cost-effective option. Furthermore, our one-way sensitivity analyses suggest that there is scope for universal scanning to be cost-effective under other assumptions. For example, the most cost-effective option remains a breech-only scan as long as the time horizon of the analysis is below 45 years only. The ideal time horizon for an economic evaluation should be sufficient to capture all relevant differences in cost and outcomes.¹⁸⁸ In many cases this implies a lifetime horizon;¹⁹⁵ however, our base-case analysis was limited to 20 years. This represents a compromise between the desire for a long time horizon and the inherent uncertainties in extrapolating relatively short-term data into long-term outcomes. We therefore acknowledge the possibility that universal ultrasound scanning may be cost-effective in the long run, but we would urge caution in any recommendation of such.

Finally, all else being equal, presentation-only scan is the most cost-effective option provided that it can be accomplished for < £90 per scan. This is a higher price than we estimated in our previous work, which estimated a maximum cost-effective price of a presentation scan of approximately £20.¹¹ This difference is due to the more detailed modelling in this analysis; where the previous analysis based QALY gains on mortalities averted and a set life expectancy, this analysis included the impact that morbidity has on costs and quality of life, and incorporates explicit survival functions.

Strengths and limitations

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

A key limitation of this study is that only fetal outcomes were considered, excluding the outcomes of the mother. Maternal outcomes may also be significant. Furthermore, the well-being of mother and child are sometimes at odds with each other, and clinical decisions frequently involve a trade off between the two. Incorporating maternal outcomes into the analysis, therefore, could have an impact on both the cost-effectiveness of the different strategies (in either direction) and our VOI analyses guiding where future research could be prioritised. However, as per our original protocol, maternal health consequences were not incorporated in this study. The primary justification for this is the lack of sufficiently reliable evidence of how screening outcomes may affect maternal quality of life.

We have previously emphasised the need for further research in this area, particularly surrounding long-term maternal consequences from mode of delivery,^{11,155} and repeat that call here.

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

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

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

In this analysis, we modelled the risk of long-term outcomes, such as SEN, as a function of neonatal morbidity. This means that clinical interventions that can alleviate neonatal morbidity are also expected to alleviate the risk of SEN. Similarly, interventions that do not affect neonatal morbidity will have no impact on the risk of SEN. However, this may not accurately capture how interventions affect the risk of SEN. This model structure has been adopted because of data limitations and to avoid overestimating the effect of intervention.

There is some evidence that the risk of SEN increases with early IOL, and the perceived risk of this is often influential in the clinical decision of whether or not to induce labour early. Our model structure captures long-term effects on SEN from early IOL if it is mediated through neonatal morbidity. However, if there is a direct link between gestational age at delivery and the risk of SEN that is not mediated through neonatal morbidity, this is uncaptured in the model. One-way sensitivity analyses exploring this suggest that our results hold as long as the risk of SEN associated with IOL (vs. expectant management) is below approximately 1.34. Above this, the recommendation for a presentation-only scan holds, but inducing labour for LGA is no longer recommended. If it is plausible that the increased risk of SEN associated with IOL exceeds 34%, then it may be worthwhile exploring this in future research. However, observational data indicate that delivery at 38 weeks' gestation is associated with < 34% increase in risk.¹⁷²

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

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

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

Comparison with other studies

A previous review of studies of universal ultrasound assessment during late pregnancy found no clear benefit of universal ultrasound.²¹ In this study, we have found that universal ultrasound may be associated with better clinical outcomes. Whether or not universal screening is cost-effective, however, depends on the features included in such a scan. Our analysis shows that universal ultrasound for fetal size is unlikely to be cost-effective, unless the valuation of additional health is higher than that recommended by current UK guidelines.¹⁸⁸ By contrast, universal ultrasound for fetal presentation alone is likely to be cost-effective, although uncertainty persists over whether or not fetal presentation can be assessed sufficiently cheaply using ultrasound to make such a screening policy feasible.

Furthermore, the findings also align with our cost-effectiveness analyses of universal ultrasound for individual complications only. When exploring the cost-effectiveness of universal ultrasound for breech presentation only, we found that whether or not such a screening programme could be cost-effective largely depended on the price at which fetal presentation could be detected.¹¹ It seemed unlikely that screening for SGA or LGA only would be cost-effective, but we highlighted that the effectiveness of labour induction was uncertain and may warrant further research. This joint analysis confirms these findings, and has allowed us to point more specifically towards those parameters for which further research may have a meaningful impact on the decision problem.

Implementation considerations

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

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

The consequences of future research are likely to go beyond the perspective employed in this analysis. First, our analysis focused on nulliparous women with singleton pregnancies, but, for many parameters, reducing uncertainty would be helpful to women regardless of parity. To address this, we provided two population values of information: one based on nulliparous singleton pregnancies and the other based on all pregnancies. Second, the scope of our study was limited to England, but many findings are likely to be just as applicable to the rest of the UK, and indeed to other high-income countries as well. If the VOI analyses are considered applicable to the entire UK, the EVPI, EVPPI and EVSI figures should be multiplied by approximately 25% to reflect this (England accounts for approximately 80% of the UK population). Third, the economic perspective of this study was NHS England and education services only, but many consequences would go beyond this. For instance, it has been estimated that the majority of the costs associated with stillbirth and CP are indirect (e.g. from decreased productivity, extra monitoring for subsequent pregnancies and mourning^{181,183,196}). When considering such perspectives, both the attractiveness of universal ultrasound and the value of future research are likely to increase.

Conclusions

The remit of this work was to advise the National Institute for Health Research on the current body of evidence regarding the cost-effectiveness of late-pregnancy ultrasound screening and specifically whether or not there is value in commissioning further research in the area and, if so, what this research should focus on.

Our results suggest that universal ultrasound for fetal presentation only may be both clinically and economically justified, but implementation research is needed before it is adopted into routine care. Specifically, this must explore whether or not a scan can be conducted by a midwife during a routine antenatal visit. Universal ultrasound including estimation of fetal weight is of borderline cost-effectiveness and is sensitive to certain assumptions. Our formal VOI analysis suggests that future research should be focused on the net cost of IOL compared with expectant management.

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

Aims

The aims of this section were to:

1. assess pregnant women's knowledge about the current antenatal care pathway for low-risk pregnancies
2. assess pregnant women's understanding of the potential benefits and drawbacks of third-trimester screening
3. estimate pregnant women's willingness to participate in a future randomised clinical trial, examine which trial design they would prefer to participate in, and calculate the expected recruitment rate.

Methods

To evaluate both the quantitative and the qualitative aspects of the above aims, we conducted a survey and ran focus groups. For each aim we collaborated with the National Institute for Health Research Cambridge Biomedical Research Centre Communications and patient and public involvement (PPI) department of Cambridge University Hospitals NHS Foundation Trust (CUHFT). Amanda Stranks, the head of the PPI department of CUHFT, had an active role in the writing and testing of the survey as well as the design, recruitment and running of the focus groups, as explained below.

The objective of the survey was to meet the requirements of aims 1 and 3 by involving a large and representative number of women. We planned to recruit low-risk nulliparous women after their ultrasound scan at 12 or 20 weeks' gestation, given that the scans at these points confirm a viable pregnancy. We excluded any high-risk pregnancies with either maternal or fetal pathology. The questionnaire was approved by all of the collaborators of the study and tested by the PPI office in CUHFT to ensure that it was understood by the women. We received feedback from five anonymous individuals and modified our form accordingly. We include the final version of the questionnaire in *Appendix 8*. In brief, this questionnaire had three parts. The first two questions were about the woman's knowledge of current antenatal care and her willingness to have an additional ultrasound scan in the third trimester. The second part included three questions about potential participation in a future randomised controlled trial. We discussed two possible trial designs. The first study (study A) would randomise low-risk women to have a scan at 36 weeks' gestation or not (the latter being current standard of care). The ultrasound results would be revealed to their clinical care team and their management would be affected accordingly. In the second study (study B) all women would have an ultrasound at 36 weeks' gestation. If there was a major problem (e.g. breech presentation or very small amount of fluid around the infant), the result would be revealed to the care team. In all other cases the result would be blinded to the women and the clinicians. Finally, we included some questions on women's demographics, such as age, ethnicity and education, to ensure that the sample was diverse. All of the replies were anonymised.

The second part of this section was running groups in which we could discuss the qualitative aspects of all the above aims. We planned to recruit women who had recently delivered (within the last 2 years), and discuss in detail the benefits and potential risks of third-trimester screening. To advertise the focus groups, we used the mailing list of the PPI office, personal contact by midwives, and social

media including Facebook (www.facebook.com; Facebook, Inc., Menlo Park, CA, USA), Twitter (www.twitter.com; Twitter, Inc., San Francisco, CA, USA) and WhatsApp (Facebook, Inc., Menlo Park, CA, USA) to address groups of mothers in the broader area of Cambridge. The focus group discussion was designed by Alexandros A Moraitis, Gordon CS Smith and Amanda Stranks.

Results

Survey

We collected 100 replies from pregnant women attending for their routine dating or anomaly scan at the Rosie Hospital, Cambridge. We present the results in *Table 16*. The respondents were diverse in age group, ethnicity and education level. The majority (85%) were aware that women with low-risk

TABLE 16 Results of the survey of low-risk pregnant women (*n* = 100)

Question	Answer	Number of responses
1. Were you aware that women whose pregnancies are straightforward are NOT routinely scanned after 20 weeks' gestation?	Yes	85
	No	15
2. 'I would like to have the option of a scan at around 36 weeks as part of my routine NHS care'	Agree/strongly agree	84
	Neither agree nor disagree	13
	Disagree/strongly disagree	3
3. I would be likely to agree to take part in study A	Agree/strongly agree	76
	Neither agree nor disagree	17
	Disagree/strongly disagree	7
4. I would be likely to agree to take part in study B	Agree/strongly agree	66
	Neither agree nor disagree	18
	Disagree/strongly disagree	16
5. If you are happy to participate in one of the above research projects which one would you prefer?	Study A	10
	Study B	23
	Both	32
	N/A - missing	35
Maternal age (years)	< 30	38
	≥ 30	60
	Missing	2
Ethnicity	White British	40
	Other British	20
	Other European	17
	Asian/African	8
	Missing	15
Age stopped education (years)	< 22	53
	≥ 22	39
	Missing	8

pregnancies are not offered routine ultrasound in the third trimester and 84% said that they would like to have a routine third-trimester scan. Regarding participation in a future clinical trial, 76% agree or strongly agree that they would participate in study A and 66% in study B. When asked which study they would prefer to participate in, out of the 65 women who answered this question, 10 (15.4%) preferred study A, 23 (35.4%) preferred study B, and 32 (49.2%) would be happy to participate in either study.

Focus group

Eight women showed an initial interest in participating in our focus groups. Owing to difficulties with child care, four of the women could not participate in a focus group on any of multiple suggested dates. We managed to run one focus group with four participants. The focus group was run by Alexandros A Moraitis and Amanda Stranks (PPI lead in CUHFT). The participant characteristics are as follows:

- Participant A had one previous delivery at low risk. She had measured slightly small on symphysis-fundal height (2 cm below AGA) but had no extra scans. Normal uncomplicated delivery of 2.49-kg infant at 40 weeks' gestation. Her motivation for participation was to find out whether or not she needed a third scan. She also mentioned that her husband is French and as in France all pregnant women have a third-trimester scan she wanted to know why this is not the policy in the UK.
- Participant B had two previous deliveries (now 4- and 2-year-old), both of which were low risk. The first infant was born in the birth centre, for the second she had IOL for post dates. Both deliveries were uncomplicated. Her motivation for participation was that four of her friends had had stillbirths at term in the last few years, which she found very stressful as she was planning a third pregnancy.
- Participant C had one previous delivery, which was initially high risk due to low BMI, and she had growth scans at 32 and 36 weeks' gestation (both normal). She was then discharged to midwifery care and delivered in the midwifery unit without complications. Her motivation for participating was finding out whether or not she had needed all these scans as it had been difficult to attend the appointments because of work.
- Participant D had one previous delivery, initially low risk. Owing to low pregnancy associated plasma protein-A (PAPP-A) she was closely monitored during pregnancy. She had IOL at 37 weeks' gestation because of suspected FGR. She delivered vaginally a 2.1-kg infant (2nd centile), who stayed in the NICU for 3 days. Her motivation for participation was to find out whether or not this might have been missed had the PAPP-A not been marginally abnormal in the first trimester.

We initially discussed the women's opinions on the current screening schedule and whether or not they would want an additional ultrasound scan in the third trimester. Two participants (A and B) thought that two scans are not enough and that there is a long period after 20 weeks' gestation during which they do not know about the fetus's well-being. They both believed that an additional scan would make them feel more reassured. One participant (C) considered herself low risk (despite her low BMI) and had found it difficult to attend the additional scans that she was offered. Finally, the fourth participant thought that the schedule was about right and she wanted to have more evidence that the additional scans would be beneficial before these were introduced.

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

Finally, we discussed participation in a future trial. All women said that they would be happy to participate in a future trial. When we specifically discussed the two potential study designs they all preferred study B (screening all women and randomising to blind, or not blind, the result) because they would be reassured about the infant's presentation and that a diagnosis of a severe problem would be revealed. The main comment about blinding were that we had to make it clear which conditions would be revealed and which would not. In addition, they wanted us to explain clearly that we were not withholding information from them but simply collecting more of it, and that they would receive the normal standard of care if they were randomised to the control group. When we discussed the timing of consent, all of the women stated that they would be happy to be approached in the first or second trimester. However, they would prefer to have a second discussion about randomisation at 36 weeks' gestation because they felt that they would have forgotten the details of the consent form at 12 or 20 weeks' gestation and they would prefer to have a longer conversation at that point.

Discussion and conclusions

We were able to collect both quantitative and qualitative data about the opinions of women on third-trimester ultrasound screening. We found that there was a clear interest in having an additional ultrasound scan in the third trimester, which was also confirmed in the focus group by all but one participant. This also confirms the previously published finding by the Stillbirth Priority Setting Partnership,¹⁹⁷ which included responses from > 300 parents and 700 professionals and concluded that the question of whether or not a third-trimester ultrasound scan can reduce the risk of stillbirth was one of the most important research priorities. We also found that the majority of women would be happy to participate in a future randomised controlled trial and we would expect a recruitment rate of at least two out of three women, which is similar to the recruitment rate of the POP study in which the ultrasound result was blinded to the women and the clinicians. In total, 66% of women who replied to our questionnaire, and all of the focus group participants, would be happy with the blinding of the ultrasound result if there was no severe problem, something that we would have to define clearly.

Reflections/clinical perspective

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

Chapter 13 Designing a randomised controlled trial of screening and intervention

Implications of the health economic analysis

The economic analysis demonstrated that although, on average, the most cost-effective approach was to screen all nulliparous pregnant women with a presentation-only scan, this had only a 44% probability of being true, and a scan that included fetal biometry had a $\approx 39\%$ chance of being the most cost-effective. Moreover, if the time scale was increased, it became likely that such a scan in late pregnancy would be the most cost-effective approach. These observations indicate that implementing such a scan could be considered. However, one of the major obstacles to implementing such a policy is that there is no direct evidence from a RCT that this screening and intervention is clinically effective. The Cochrane review of universal late-pregnancy ultrasound failed to show any benefit of this to the mother or infant.²¹ However, as discussed in the introduction, this review has a number of methodological issues and it is more accurate to state that it does not provide a definite answer the question of whether or not universal late-pregnancy ultrasound reduces the risk of perinatal death.

Interestingly, the VOI analysis highlighted reducing uncertainty about the costs of IOL. Given the above, this may be regarded as somewhat counterintuitive. However, the parameters used in the VOI analysis in relation to the screening performance of ultrasound and the effect of intervention were known with a degree of precision that meant that reducing their uncertainty was not the most cost-effective research question. For example, the ability of ultrasound to predict SGA, the relationship between SGA birthweight and the risk of stillbirth, and the ability of IOL to reduce the risk of stillbirth are all known quite precisely and are based on high-quality data. Consequently, even though there is no direct evidence to indicate that universal late-pregnancy ultrasound would reduce the risk of stillbirth, the model estimates quite a high chance that it is the most cost-effective approach and does not highlight reducing the uncertainty in these parameters in the VOI analysis. By contrast, previous health economic analyses of IOL have generated quite wide CIs,^{176,194} and hence the model has identified that reducing this uncertainty is the key question.

Case for considering a randomised controlled trial of screening and intervention

In this chapter we consider the practicalities of designing a RCT of screening and intervention using fetal biometry in nulliparous women at 36 weeks' gestation. We have done this because, even though the parameters in the modelling were reasonably certain, these parameters were calculated from a range of different study designs (i.e. we did not perform the VOI analysis based on the uncertainty of parameters calculated from a large RCT of late pregnancy screening and intervention in nulliparous women). Rather, we performed the analysis using parameters from a range of observational studies and a range of studies of interventions in women who were deemed to be high risk for other reasons. The concern in this case is external validity. The parameters may be reasonably certain in relation to the setting where they were derived but there is an unquantifiable uncertainty in relation to how well they inform our research question. The obvious way to address this would be to perform a study in the setting of interest. Such a study could be the definitive study or it could be a pilot or a proof-of-principle study. The former might be a trial of screening compared with not screening, with perinatal death as the primary outcome. The latter might exploit alternative study designs and use of proxies. Hence, there are a number of important considerations to take into account when designing a RCT of screening and intervention using universal ultrasound, and we will consider each of these in turn.

Candidate primary outcomes

In relation to the primary outcome of a RCT, we believe that the strongest case can be made for perinatal death. First, losing an infant at term is clearly a devastating outcome for a family. In the absence of a lethal anomaly, preventing death would lead to an entire life gained which, from a health-care and health economic perspective, is a gain of unique magnitude. Second, the main intervention available is earlier delivery. There is strong evidence that IOL is effective in reducing the risk of perinatal death. Over two-thirds of perinatal deaths at term are antepartum stillbirths⁵⁴ (i.e. intrauterine fetal death prior to the onset of labour). Self-evidently, antepartum stillbirth cannot occur after an infant has been delivered.¹⁷ Delivery at or after 38–39 weeks' gestation carries the same risk of intrapartum stillbirth and neonatal death as delivery at a later week of gestation.^{17,198} These epidemiological observations underlie the 67% reduction in the risk of perinatal death associated with IOL at term.¹⁶

Proxies

The main problem with a primary outcome of perinatal death is that the outcome is uncommon, and this will result in major issues of statistical power. Indicators of perinatal morbidity would be an alternative outcome to perinatal death. First, as the same factors might be involved in death and morbidity, the latter could be used as proxies of the former. Second, perinatal morbidity is of importance in its own right. For example, birth asphyxia is one of the major determinants of the burden of litigation in the health service as a result of devastating effects on the later health of the child, such as CP. There is evidence to support the use of a single indicator in both roles. An Apgar score of < 4 at 5 minutes was associated with a relative risk of early neonatal death of ≈ 360 ¹⁷⁴ and a relative risk of CP of > 400.¹⁷³ Hence, a primary outcome based on perinatal morbidity, such as an Apgar score of < 4, could be clinically important, both as a proxy of death and as a determinant of long-term outcome. Morbidity could be a more pragmatic outcome as rates of severe morbidity are much greater than the risks of death, and hence it may be easier to design a trial with morbidity as the primary outcome.

Subgroups

A further refinement to the primary outcome is to study subgroups of the given event that were actually associated with the infant being born SGA or LGA. It is self-evident that screening for SGA or LGA will primarily have an impact on outcomes related to fetal growth disorder. Many adverse perinatal outcomes, both lethal and non-lethal, are unrelated to fetal growth abnormalities. Consequently, if a screening study of fetal biometry has a primary outcome that includes infants in the full range of birthweight, most of the primary outcomes in both arms of the trial will be unrelated to fetal growth disorder, which is not preventable by screening for fetal growth disorder and intervention. This means that the potential for screening to have an impact on the rate of death is limited and extremely large sample sizes would be required. For example, around one-third of perinatal deaths at term are related to being SGA or LGA.⁵⁴ The background rate of perinatal death at term is ≈ 2 per 1000. Even if a screening test was perfect (i.e. detected all cases of growth disorder), and even if the intervention was perfect (i.e. prevented all such deaths), a power calculation still indicates that > 100,000 women would have to be recruited to the trial. However, if the primary outcome was perinatal death of a SGA or LGA infant, the sample size would be $\approx 22,000$ (note that this is used to illustrate the point that it is not a practical proposition, as the screening and intervention characteristics were assumed to be perfect). An analogy might be a trial of breast cancer screening. Screening reduces deaths related to breast cancer but does not reduce all-cause mortality.¹⁹⁹ This is likely to be explained by the fact that no study could be sufficiently powered to detect an effect of screening for breast cancer on all-cause mortality

because most deaths are due to other causes. Consequently, one approach to addressing the problems of statistical power in trials of screening using fetal biometry would be to define primary outcomes related to fetal growth abnormalities. An insistence on evidence that shows a reduction in all-cause perinatal death would simply remove the possibility of screening and intervention being implemented, which could lead to avoidable harm that could have been prevented in a cost-effective way.

Early delivery and iatrogenic harm

Routine induction at term had less dramatic effects on the risk of neonatal morbidity, with a 12% reduction in the risk of NICU admission and a 30% reduction in the risk of a low Apgar score. Moreover, these effects may be lost or even reversed in the context of early-term IOL. Most trials in the Cochrane review of term induction were of pregnancies at 41 weeks' gestation and beyond.¹⁶ As post-term pregnancy is associated with an increased risk of neonatal morbidity, preventing this outcome should improve immediate neonatal outcomes as well as preventing stillbirth. In the context of IOL at < 39 weeks' gestation, epidemiological data indicate that the intervention may actually increase neonatal morbidity.¹⁶⁰ The potential for earlier intervention to cause harm is increasingly recognised. The Awareness of fetal movements and care package to reduce fetal mortality (AFFIRM) study²⁰⁰ reported a stepped-wedge RCT of a programme to inform women about reduced fetal movements and to standardise intervention. Although it did not show a significant reduction in stillbirth, the intervention was associated with increased risks of neonatal morbidity.²⁰⁰ This trial has some parallels with the current question. Despite the fact that women were selected on the basis of having a risk factor (i.e. reduced fetal movements, which is associated with stillbirth), it still failed to demonstrate a reduction in stillbirth rates, and the intervention was associated with increased rates of intervention and adverse outcomes. The result of the trial underlines two key issues: (1) the need for better predictors of adverse outcome and (2) the potential for intervention to cause harm.

Current status of screening tests

Unfortunately, the results of our systematic reviews of diagnostic effectiveness and a Cochrane DTA review²³ failed to identify any ultrasonic marker that was clearly predictive of the risk of stillbirth in the context of scanning women in late pregnancy using ultrasound. Moreover, if we regard neonatal morbidity as a proxy of stillbirth, again, tests performed very poorly. Finally, actual birthweight in the < 3rd percentile was associated with a 0.9–1% risk of perinatal death at term compared with a background risk of just over 0.2%.⁵⁴ Hence, even knowing that the actual birthweight was < 3rd percentile would be associated with a positive LR of between 4 and 5. In the POP study, of 562 women whose scan indicates that their infant was SGA, only 12% of women delivered an infant with a birthweight in the < 3rd percentile; a further 23% delivered an infant \geq 3rd and < 10th percentile but about two-thirds of the women delivered an infant \geq 10th percentile. Hence, on the basis of the association between the EFW and the actual birthweight, and their relationship between the actual birthweight and the risk of stillbirth, it is highly unlikely that detecting a SGA infant is strongly predictive of the risk of stillbirth. Given the lack of information, we model outcomes with variable incidence and assess different screening test values to establish what characteristics would be required of a test to make a trial of screening and intervention feasible.

Possible trial designs

Broadly speaking, there are two main approaches to trial design (*Figure 24*).³² First (hereinafter referred to as screen vs. no screen), women might be randomised (1) to be screened, with the offer of

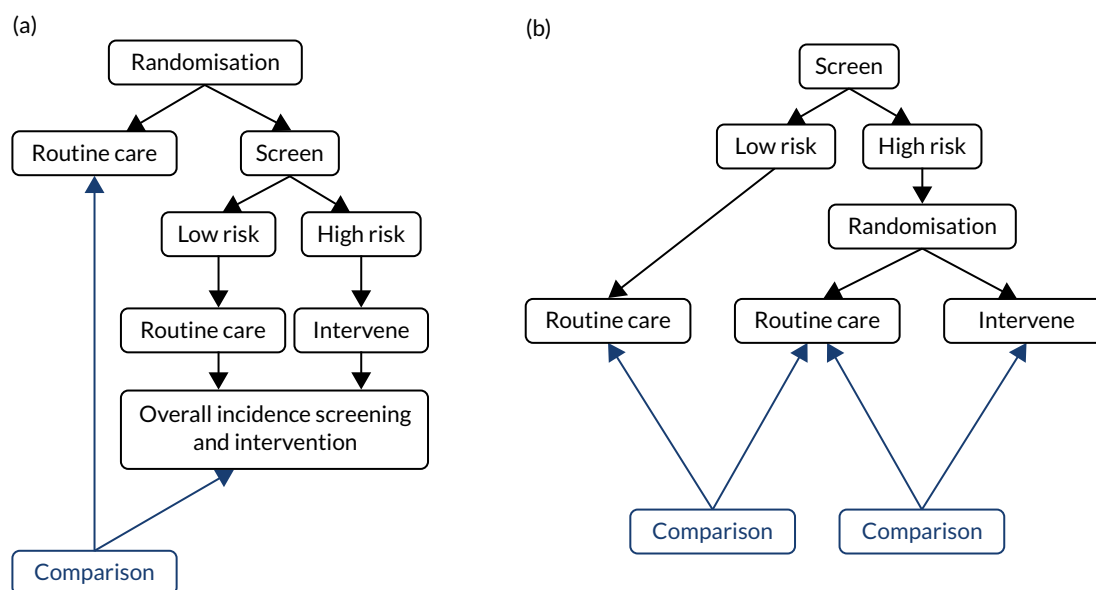


FIGURE 24 Flow charts of possible trial designs: (a) screen vs. no screen; and (b) screen all.

intervention if they screen positive, or (2) to receive routine care, which currently requires scanning only if there is a conventional clinical indication. The result of this trial design is a simple comparison between the two groups. In the event of a negative result, it is impossible to determine whether the result was because the screening test did not work or because the intervention did not mitigate the higher risks in screen-positive women. The second approach is to screen the whole population and randomise high-risk women to an intervention or to routine care (masking the result in the latter group), hereafter referred to as 'screen all'. The advantages of the second approach are that the number of women who need to be recruited is substantially fewer and that the same trial can assess both the diagnostic effectiveness of the screening test and the clinical effectiveness of the intervention. The two approaches are illustrated in *Figure 24*.

Acceptability of the 'screen-all' approach

When discussing the possibility of randomising women with a high-risk screening result, some of the co-applicants expressed concerns. Interestingly, however, when we surveyed pregnant women, they actually preferred a study design that involved all participants being scanned. In the focus group, women tended to be more concerned about being offered interventions. The observations underline the different perspectives of pregnant women and professionals. We envisaged that women who are recruited to a 'screen all' approach would have some information revealed irrespective of their randomisation status. For example, we do not feel that it would be practical or ethical not to reveal the presentation of the infant as cephalic or non-cephalic. Hence, this would probably be revealed in a 'screen all' trial design. In the POP study, although scans were blinded, breech presentation was revealed. Subsequent interviews with participants were highly positive about this element of the study where the infant was breech [Dacey 2015; www.repository.cam.ac.uk/handle/1810/280595 (accessed June 2019)]. However, a drawback of this approach is that a 'screen all' design, which reveals breech presentation, would not capture the health benefits of detecting breech presentation. Other features that should be considered in revealing the result are the presence of previously undiagnosed major congenital anomalies and placenta praevia. In the POP study, there was no cases of placenta praevia, but two patients had major anomalies diagnosed where revealing the result optimised care and, in one case (unilateral hydrothorax with severe mediastinal shift), is likely to have prevented intrauterine fetal demise.

Power calculations

To determine the feasibility of a RCT we performed power calculations using the two different study designs represented above. The sample size calculations are presented in *Table 17*. All power calculations have been performed for a p -value < 0.05 (two-sided) with 90% power to detect the effect. We selected a range of possible primary outcomes: perinatal death, severe neonatal morbidity, any neonatal morbidity and delivery of a SGA infant with complications. In relation to perinatal death, we found no adequately powered studies of the diagnostic effectiveness of ultrasound to predict this outcome and the Cochrane DTA review²³ of SGA also found no data in relation to this question. Therefore, we modelled a series of possible screening performances, varying the screen-positive rate and positive LR. In relation to morbidity, we used two studies reporting data from the POP study, from *The Lancet*⁸ and *The Lancet Child & Adolescent Health*.¹⁴⁹ As described above, the POP study was one of only two studies (Perinatal Ireland Genesis study being the other) that performed blinded ultrasound scanning in late gestation in nulliparous women. Unfortunately, the Genesis study did not report the association between SGA and morbidity, and the only publication in relation to LGA is in abstract form

TABLE 17 Sample size calculations for different outcomes, screening tests and trial designs

Screening test	SPR (%)	PPV (%)	Sample size (n)			Reference
			Screen vs. no screen	Screen all, randomise high risk		
				Number needed to screen	Number of high-risk women	
Perinatal death (background = 0.2%)						
LR+ = 2	10	0.4	1,488,448	234,740	23,474	
LR+ = 3	10	0.6	644,156	156,260	15,626	
LR+ = 5	10	1.0	219,382	93,460	9346	
LR+ = 2	5	0.4	6,110,172	469,480	23,474	
LR+ = 3	5	0.6	2,680,882	312,520	15,626	
LR+ = 5	5	1.0	940,096	186,920	9346	
LR+ = 10	5	2.0	219,382	92,760	4638	
Any neonatal morbidity^a						
EFW < 10th	14	10.3	36,910	6014	842	Sovio <i>et al.</i> ⁸
EFW < 10th + ACGV	4.3	15.7	172,522	12,279	528	Sovio <i>et al.</i> ⁸
Severe neonatal morbidity^a						
EFW < 10th	14	1.07	422,336	63,743	8924	Sovio <i>et al.</i> ⁸
EFW < 10th + ACGV	4.3	2.33	965,714	93,256	4010	Sovio <i>et al.</i> ⁸
Complicated SGA^b						
EFW < 10th	14	7.5	13,920	8457	1184	Gaccioli <i>et al.</i> ¹⁴⁹
EFW < 10th + ACGV	4.3	11.2	73,538	17,860	768	Gaccioli <i>et al.</i> ¹⁴⁹
Delphi ^c	11.3	8.5	16,952	9168	1036	Gaccioli <i>et al.</i> ¹⁴⁹
ACGV, abdominal circumference growth velocity in the lowest decile (see Sovio <i>et al.</i> ⁸); LR+, positive likelihood ratio; SPR, screen-positive rate.						
a Neonatal morbidity and severe neonatal morbidity are defined in Sovio <i>et al.</i> ⁸						
b Complicated SGA is defined in Gaccioli <i>et al.</i> ¹⁴⁹ (In brief: delivery of an infant with a birthweight < 10th percentile where either the mother had a diagnosis of pre-eclampsia or the infant experienced neonatal morbidity.)						
c Fulfilled definition of late FGR using criteria of Gordjin <i>et al.</i> ¹⁴² (except MCA Doppler not included).						

only and addresses shoulder dystocia. The two POP study publications^{8,149} address the relationship between SGA, SGA combined with reduced growth velocity (which was the best-performing predictor of morbidity from a range of candidate predictors of FGR) and the Delphi consensus definition of late FGR.

In all of these calculations we assumed that the intervention would reduce the risk of the given event by 50%. Given the lack of data, a range of figures could be considered. We used this figure as we felt that it was conservative in relation to perinatal death. It could be argued, based on the discussion above, that it is optimistic in relation to neonatal morbidity. However, by concentrating the outcome of morbidity on infants that are actually SGA, it is plausible that the combined effect of making the diagnosis and intervening could substantially reduce the rate of adverse events. It should be borne in mind that in the relevant RCT, DIGITAT,⁹⁹ randomisation occurred after ultrasound scanning led to suspicion of SGA. Hence, the group randomised to expectant management would still have received enhanced monitoring and high-risk care during labour as the infant was known to be SGA. By contrast, routine care in a trial of screening means that neither antenatal nor intrapartum care is tailored to the suspected SGA status of the fetus.

Implications of sample size calculations

We present the data on sample size calculations but we are not recommending a specific trial design. It is also possible that a trial may be considered where the combination of screening parameters, intervention effect and outcome are not listed in *Table 17*. The exact design of the trial would depend on the resources available and the research question. We do, however, discuss some of the issues that may motivate a choice.

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

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

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

Overall conclusions

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

Consultation with the National Screening Committee

We sent the scientific summary of the project and *Chapter 13* to the UK National Screening Committee (NSC) Evidence Lead, who has worked for the UK NSC for > 15 years. The UK NSC would be happy to contribute to any further HTA discussions where this is useful. Following preliminary discussion, the applicants plan to submit a proposal to the UK NSC to suggest that it recommends a screening programme for breech presentation near term. Their evidence review process is outlined on its website [www.gov.uk/government/publications/uk-nsc-evidence-review-process (accessed June 2019)].

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

- If the uncertainty around the costs of IOL were reduced, how likely is it that the NSC would recommend screening for fetal size near term based on a model that lacked direct evidence from a RCT that involved screening? For example, if the currently funded HTA trial around IOL for suspected fetal macrosomia confirms improved outcomes with intervention, would the combination of the diagnostic effectiveness of ultrasound as a screening test for LGA and the clinical effectiveness of IOL as an intervention in LGA be regarded as acceptable evidence for screening? The issue of interpretation is that screened women are likely to have lower prior odds of complications than women identified as having a LGA fetus through a clinically indicated scan. Hence, extrapolation of the results of the trial may involve an assumption that is untrue.

OVERALL CONCLUSIONS AND ASSESSMENT OF EVIDENCE REQUIRED

- If direct evidence of a beneficial effect of screening from a RCT was required, would this have to come from a 'screen versus no screen' trial or would evidence from a 'screen all' trial suffice?
- What outcomes would be acceptable? Specifically –
 - Would screening be recommended on the basis of an effect on proxies?
 - Would screening be considered on the basis of an effect on a subgroup, for example, subgroups of neonatal morbidity or mortality confined to infants who were actually small or large at birth?
 - Would screening be considered on the basis of an effect on a composite outcome?

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

- RCTs based on intervention from screen-positive women would provide much stronger support for a programme than evidence derived from RCTs of high-risk women (i.e. those not identified through screening the general population).
- Data from a 'screen versus no screen' study would be preferred to those from a 'screen all' design. However, if there were absolute methodological obstacles to 'screen versus no screen', one approach would be to show proof of principle with a 'screen all' study, consider other studies to address any shortfall arising from this design and other criteria, and then perform a stepped-wedge RCT trial when implementing the new test.

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

Acknowledgements

Steering group members

David Cromwell, Gianluca Baio, Kathryn Cook, Elizabeth Duff, Neil Marlow, Tracey Mills and Dharmintra Pasupathy.

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

For their roles as second reviewers on one of the systematic reviews, we would also like to thank Tom Bainton (umbilical artery), Norman Shreeve (LGA), Illianna Armata (borderline AFI) and Dexter Hayes (cerebroplacental ratio).

Contributions of authors

Gordon CS Smith (<https://orcid.org/0000-0003-2124-0997>) (Professor, Head of Department of Obstetrics and Gynaecology) conceived the project and contributed to protocol development, management of the project, planning of the systematic reviews, conceptualisation of the economic models, clinical interpretation of findings and writing of the report.

Alexandros A Moraitis (<https://orcid.org/0000-0003-4634-1129>) (Research Associate, Obstetrics and Gynaecology) performed the systematic reviews of clinical effectiveness, drafted and edited the final report, designed questionnaires for determining which conditions the systematic reviews should focus on, commissioned the focus group, and contributed to the identification of data and conceptualisation of the economic models.

David Wastlund (<https://orcid.org/0000-0002-5074-4740>) (Research Assistant, Health Economics) conceptualised and programmed the economic models, identified and estimated data for the economic analyses, performed the cost-effectiveness and VOI analyses, performed the systematic review on AFI, and drafted and edited the final report.

Jim G Thornton (<https://orcid.org/0000-0001-9764-6876>) (Professor, Obstetrics and Gynaecology) provided input on which systematic reviews to undertake and the design of future research, helped design the questionnaire for identifying topics for systematic reviews, and commented on drafts of the systematic reviews, the economic analyses and the final report.

Aris Papageorghiou (<https://orcid.org/0000-0001-8143-2232>) (Professor, Obstetrics and Gynaecology) contributed to designing the methods for systematic reviews, provided input on the design of future research, and commented on drafts of the report.

Julia Sanders (<https://orcid.org/0000-0001-5712-9989>) (Professor, Clinical Nursing and Midwifery) provided input on which systematic reviews to undertake, and edited drafts of the economic analyses and final report.

Alexander EP Heazell (<https://orcid.org/0000-0002-4303-7845>) (Professor, Obstetrics) helped design the methods for the systematic reviews, provided input on which clinical areas the systematic reviews should target and edited the final report.

ACKNOWLEDGEMENTS

Stephen C Robson (<https://orcid.org/0000-0001-7897-7987>) (Professor, Fetal Medicine) reviewed chapters on systematic reviews, contributed to the design of PPI and edited the final report.

Ulla Sovio (<https://orcid.org/0000-0002-0799-1105>) (Senior Research Associate, Applied Medical Statistics) contributed to the statistical analysis of the systematic reviews, reviewed and commented on drafts of the systematic reviews and the economic analyses, contributed to data analysis for the economic models, and edited the final report.

Peter Brocklehurst (<https://orcid.org/0000-0002-9950-6751>) (Professor of Women's Health) provided input on which systematic reviews to undertake and the design of future research, helped design the questionnaire for identifying topics for systematic reviews, and commented on drafts of the systematic reviews, the economic analyses and the final report.

Edward CF Wilson (<https://orcid.org/0000-0002-8369-1577>) (Senior Lecturer, Health Economics) designed and programmed the models for the economic and VOI analysis, designed methods for the quantification of data for the economic analysis, and drafted and edited the final report.

Publications

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

Wastlund D, Moraitis AA, Thornton JG, Sanders J, White IR, Brocklehurst P, *et al.* The cost-effectiveness of universal late-pregnancy screening for macrosomia in nulliparous women: a decision-analysis. *BJOG* 2019;**126**:1243–50.

Moraitis AA, Bainton T, Sovio U, Brocklehurst P, Heazell AEP, Thornton JG, *et al.* The fetal umbilical artery Doppler as a tool for universal third trimester screening: a systematic review and meta-analysis of diagnostic test accuracy. *Placenta* 2021; in press.

Wilson ECF, Wastlund D, Moraitis AA, Smith GCS. Late pregnancy ultrasound to screen for and manage potential birth complications in nulliparous women: a cost-effectiveness and value of information analysis. *Value Health* 2021; in press.

Data-sharing statement

All available data are contained within the report. All requests for access to study data should be made to the corresponding author.

Patient data

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety, and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that it is stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: <https://understandingpatientdata.org.uk/data-citation>.

References

1. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017;**390**:1151–210. [https://doi.org/10.1016/s0140-6736\(17\)32152-9](https://doi.org/10.1016/s0140-6736(17)32152-9)
2. NICE. *NICE CG62: Antenatal Care for Uncomplicated Pregnancies*. London; 2008.
3. Public Health England. *NHS Fetal Anomaly Screening Programme (FASP)*. URL: www.gov.uk/topic/population-screening-programmes/fetal-anomaly (accessed June 2019).
4. American College of Obstetricians and Gynecologists. Practice Bulletin No. 175: Ultrasound in Pregnancy. *Obstet Gynecol* 2016;**128**:e241–e56. <https://doi.org/10.1097/aog.0000000000001815>
5. Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the use of head, body, and femur measurements – a prospective study. *Am J Obstet Gynecol* 1985;**151**:333–7. [https://doi.org/10.1016/0002-9378\(85\)90298-4](https://doi.org/10.1016/0002-9378(85)90298-4)
6. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology* 1991;**181**:129–33. <https://doi.org/10.1148/radiology.181.1.1887021>
7. Kiserud T, Piaggio G, Carroli G, Widmer M, Carvalho J, Neerup Jensen L, *et al*. The World Health Organization Fetal Growth Charts: a multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight. *PLOS Med* 2017;**14**:e1002220. <https://doi.org/10.1371/journal.pmed.1002220>
8. Sovio U, White IR, Dacey A, Pasupathy D, Smith GCS. Screening for fetal growth restriction with universal third trimester ultrasonography in nulliparous women in the Pregnancy Outcome Prediction (POP) study: a prospective cohort study. *Lancet* 2015;**386**:2089–97. [https://doi.org/10.1016/S0140-6736\(15\)00131-2](https://doi.org/10.1016/S0140-6736(15)00131-2)
9. Hoffman C, Galan HL. Assessing the ‘at-risk’ fetus: Doppler ultrasound. *Curr Opin Obstet Gynecol* 2009;**21**:161–6. <https://doi.org/10.1097/GCO.0b013e3283292468>
10. Krebs C, Macara LM, Leiser R, Bowman AW, Greer IA, Kingdom JC. Intrauterine growth restriction with absent end-diastolic flow velocity in the umbilical artery is associated with maldevelopment of the placental terminal villous tree. *Am J Obstet Gynecol* 1996;**175**:1534–42. [https://doi.org/10.1016/S0002-9378\(96\)70103-5](https://doi.org/10.1016/S0002-9378(96)70103-5)
11. Wastlund D, Moraitis AA, Dacey A, Sovio U, Wilson ECF, Smith GCS. Screening for breech presentation using universal late-pregnancy ultrasonography: a prospective cohort study and cost effectiveness analysis. *PLOS Med* 2019;**16**:e1002778. <https://doi.org/10.1371/journal.pmed.1002778>
12. Thorpe-Beeston JG, Banfield PJ, Saunders NJ. Outcome of breech delivery at term. *BMJ* 1992;**305**:746–7. <https://doi.org/10.1136/bmj.305.6856.746>
13. Impey LWM, Murphy DJ, Griffiths M, Penna LK on behalf of the Royal College of Obstetricians and Gynaecologists. External cephalic version and reducing the incidence of term breech presentation. *BJOG* 2017;**124**:e178–e92. <https://doi.org/10.1111/1471-0528.14466>
14. Hofmeyr GJ, Hannah M, Lawrie TA. Planned caesarean section for term breech delivery. *Cochrane Database Syst Rev* 2015;**7**:CD000166. <https://doi.org/10.1002/14651858.CD000166.pub2>

15. Smith GC, Fretts RC. Stillbirth. *Lancet* 2007;**370**:1715–25. [https://doi.org/10.1016/s0140-6736\(07\)61723-1](https://doi.org/10.1016/s0140-6736(07)61723-1)
16. Middleton P, Shepherd E, Crowther CA. Induction of labour for improving birth outcomes for women at or beyond term. *Cochrane Database Syst Rev* 2018;**5**:CD004945. <https://doi.org/10.1002/14651858.CD004945.pub4>
17. Smith GC. Life-table analysis of the risk of perinatal death at term and post term in singleton pregnancies. *Am J Obstet Gynecol* 2001;**184**:489–96. <https://doi.org/10.1067/mob.2001.109735>
18. Smith GCS. The risk of perinatal death at term. *BJOG* 2019;**126**:1258. <https://doi.org/10.1111/1471-0528.15827>
19. Smith GCS. Should we implement universal screening with late pregnancy ultrasound to prevent stillbirth? *BJOG* 2018;**125**:101–3. <https://doi.org/10.1111/1471-0528.14782>
20. Alfirevic Z, Stampalija T, Dowswell T. Fetal and umbilical Doppler ultrasound in high-risk pregnancies. *Cochrane Database Syst Rev* 2017;**6**:CD007529. <https://doi.org/10.1002/14651858.CD007529.pub4>
21. Bricker L, Medley N, Pratt JJ. Routine ultrasound in late pregnancy (after 24 weeks' gestation). *Cochrane Database Syst Rev* 2015;**6**:CD001451. <https://doi.org/10.1002/14651858.CD001451.pub4>
22. Monier I, Blondel B, Ego A, Kaminiski M, Goffinet F, Zeitlin J. Poor effectiveness of antenatal detection of fetal growth restriction and consequences for obstetric management and neonatal outcomes: a French national study. *BJOG* 2015;**122**:518–27. <https://doi.org/10.1111/1471-0528.13148>
23. Heazell AE, Hayes DJ, Whitworth M, Takwoingi Y, Bayliss SE, Davenport C. Biochemical tests of placental function versus ultrasound assessment of fetal size for stillbirth and small-for-gestational-age infants. *Cochrane Database Syst Rev* 2019;**5**:CD012245. <https://doi.org/10.1002/14651858.CD012245.pub2>
24. Brouwers L, van der Meiden-van Roest AJ, Savelkoul C, Vogelvang TE, Lely AT, Franx A, van Rijn BB. Recurrence of pre-eclampsia and the risk of future hypertension and cardiovascular disease: a systematic review and meta-analysis. *BJOG* 2018;**125**:1642–54. <https://doi.org/10.1111/1471-0528.15394>
25. Phillips C, Velji Z, Hanly C, Metcalfe A. Risk of recurrent spontaneous preterm birth: a systematic review and meta-analysis. *BMJ Open* 2017;**7**:e015402. <https://doi.org/10.1136/bmjopen-2016-015402>
26. Lamont K, Scott NW, Jones GT, Bhattacharya S. Risk of recurrent stillbirth: systematic review and meta-analysis. *BMJ* 2015;**350**:h3080. <https://doi.org/10.1136/bmj.h3080>
27. Kinzler WL, Kaminsky L. Fetal growth restriction and subsequent pregnancy risks. *Semin Perinatol* 2007;**31**:126–34. <https://doi.org/10.1053/j.semperi.2007.03.004>
28. Brocklehurst P, Hardy P, Hollowell J, Linsell L, Macfarlane A, McCourt C, *et al*. Perinatal and maternal outcomes by planned place of birth for healthy women with low risk pregnancies: the Birthplace in England national prospective cohort study. *BMJ* 2011;**343**:d7400. <https://doi.org/10.1136/bmj.d7400>
29. The Stillbirth Collaborative Research Network Writing Group. Association between stillbirth and risk factors known at pregnancy confirmation. *JAMA* 2011;**306**:2469–79. <https://doi.org/10.1001/jama.2011.1798>

30. Haas DM, Parker CB, Wing DA, Parry S, Grobman WA, Mercer BM, *et al.* A description of the methods of the Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b). *Am J Obstet Gynecol* 2015;**212**:539.e1–539.e24. <https://doi.org/10.1016/j.ajog.2015.01.019>
31. Revicki DA, Lenderking WR. Methods and issues associated with the use of quality-adjusted life-years. *Expert Rev Pharmacoecon Outcomes Res* 2012;**12**:105–14. <https://doi.org/10.1586/erp.11.100>
32. Smith GC. Researching new methods of screening for adverse pregnancy outcome: lessons from pre-eclampsia. *PLOS Med* 2012;**9**:e1001274. <https://doi.org/10.1371/journal.pmed.1001274>
33. Morris RK, Meller CH, Tamblyn J, Malin GM, Riley RD, Kilby MD, *et al.* Association and prediction of amniotic fluid measurements for adverse pregnancy outcome: systematic review and meta-analysis. *BJOG* 2014;**121**:686–99. <https://doi.org/10.1111/1471-0528.12589>
34. Alfirevic Z, Stampalija T, Medley N. Fetal and umbilical Doppler ultrasound in normal pregnancy. *Cochrane Database Syst Rev* 2015;**4**:CD001450. <https://doi.org/10.1002/14651858.CD001450.pub4>
35. Moraitis AA, Bainton T, Sovio U, Brocklehurst P, Heazell AEP, Thornton JG, *et al.* The fetal umbilical artery Doppler as a tool for universal third trimester screening: a systematic review and meta-analysis of diagnostic test accuracy. *Placenta* 2021; in press.
36. Pasupathy D, Dacey A, Cook E, Charnock-Jones DS, White IR, Smith GC. Study protocol. A prospective cohort study of unselected primiparous women: the pregnancy outcome prediction study. *BMC Pregnancy Childbirth* 2008;**8**:51. <https://doi.org/10.1186/1471-2393-8-51>
37. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529–36. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>
38. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;**20**:2865–84. <https://doi.org/10.1002/sim.942>
39. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;**58**:982–90. <https://doi.org/10.1016/j.jclinepi.2005.02.022>
40. Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;**323**:157–62. <https://doi.org/10.1136/bmj.323.7305.157>
41. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;**58**:882–93. <https://doi.org/10.1016/j.jclinepi.2005.01.016>
42. Akolekar R, Ciobanu A, Zingler E, Syngelaki A, Nicolaides KH. Routine assessment of cerebroplacental ratio at 35–37 weeks' gestation in the prediction of adverse perinatal outcome. *Am J Obstet Gynecol* 2019;**221**:65.e1–65.e18.
43. Bolz N, Kalache KD, Proquitt H, Slowinski T, Hartung JP, Henrich W, *et al.* Value of Doppler sonography near term: can umbilical and uterine artery indices in low-risk pregnancies predict perinatal outcome? *J Perinat Med* 2013;**41**:165–70. <https://doi.org/10.1515/jpm-2012-0042>
44. Cooley SM, Donnelly JC, Walsh T, MacMahon C, Gillan J, Geary MP. The impact of umbilical and uterine artery Doppler indices on antenatal course, labor and delivery in a low-risk primigravid population. *J Perinat Med* 2011;**39**:143–9. <https://doi.org/10.1515/JPM.2010.130>

45. Filmar G, Panagopoulos G, Minior V, Barnhard Y, Divon MY. Elevated umbilical artery systolic/diastolic ratio in the absence of fetal growth restriction. *Arch Gynecol Obstet* 2013;**288**:279–85. <https://doi.org/10.1007/s00404-013-2764-5>
46. Fischer RL, Kuhlman KA, Depp R, Wapner RJ. Doppler evaluation of umbilical and uterine-arcuate arteries in the postdates pregnancy. *Obstet Gynecol* 1991;**78**:363–8.
47. 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. *Eur J Obstet Gynecol Reprod Biol* 1997;**71**:11–19. [https://doi.org/10.1016/S0301-2115\(96\)02606-1](https://doi.org/10.1016/S0301-2115(96)02606-1)
48. Hanretty KP, Primrose MH, Neilson JP, Whittle MJ. Pregnancy screening by Doppler uteroplacental and umbilical artery waveforms. *Br J Obstet Gynaecol* 1989;**96**:1163–7. <https://doi.org/10.1111/j.1471-0528.1989.tb03191.x>
49. Schulman H, Winter D, Farmakides G, Ducey J, Guzman E, Coury A, Penny B. Pregnancy surveillance with Doppler velocimetry of uterine and umbilical arteries. *Am J Obstet Gynecol* 1989;**160**:192–6. [https://doi.org/10.1016/0002-9378\(89\)90118-X](https://doi.org/10.1016/0002-9378(89)90118-X)
50. Sijmons EA, Reuwer PJ, van Beek E, Bruinse HW. The validity of screening for small-for-gestational-age and low-weight-for-length infants by Doppler ultrasound. *Br J Obstet Gynaecol* 1989;**96**:557–61. <https://doi.org/10.1111/j.1471-0528.1989.tb03255.x>
51. Valino N, Giunta G, Gallo DM, Akolekar R, Nicolaides KH. Biophysical and biochemical markers at 30–34 weeks' gestation in the prediction of adverse perinatal outcome. *Ultrasound Obstet Gynecol* 2016;**47**:194–202. <https://doi.org/10.1002/uog.14928>
52. Valino N, Giunta G, Gallo DM, Akolekar R, Nicolaides KH. Biophysical and biochemical markers at 35–37 weeks' gestation in the prediction of adverse perinatal outcome. *Ultrasound Obstet Gynecol* 2016;**47**:203–9. <https://doi.org/10.1002/uog.15663>
53. Weiner Z, Reichler A, Zlozover M, Mendelson A, Thaler I. The value of Doppler ultrasonography in prolonged pregnancies. *Eur J Obstet Gynecol Reprod Biol* 1993;**48**:93–7. [https://doi.org/10.1016/0028-2243\(93\)90246-9](https://doi.org/10.1016/0028-2243(93)90246-9)
54. Moraitis AA, Wood AM, Fleming M, Smith GC. Birth weight percentile and the risk of term perinatal death. *Obstet Gynecol* 2014;**124**:274–83. <https://doi.org/10.1097/aog.0000000000000388>
55. Flenady V, Wojcieszek AM, Middleton P, Ellwood D, Erwich JJ, Coory M, et al. Stillbirths: recall to action in high-income countries. *Lancet* 2016;**387**:691–702. [https://doi.org/10.1016/S0140-6736\(15\)01020-X](https://doi.org/10.1016/S0140-6736(15)01020-X)
56. Giussani DA. The fetal brain sparing response to hypoxia: physiological mechanisms. *J Physiol* 2016;**594**:1215–30. <https://doi.org/10.1113/JP271099>
57. Akolekar R, Syngelaki A, Gallo DM, Poon LC, Nicolaides KH. Umbilical and fetal middle cerebral artery Doppler at 35–37 weeks' gestation in the prediction of adverse perinatal outcome. *Ultrasound Obstet Gynecol* 2015;**46**:82–92. <https://doi.org/10.1002/uog.14842>
58. Bakalis S, Akolekar R, Gallo DM, Poon LC, Nicolaides KH. Umbilical and fetal middle cerebral artery Doppler at 30–34 weeks' gestation in the prediction of adverse perinatal outcome. *Ultrasound Obstet Gynecol* 2015;**45**:409–20. <https://doi.org/10.1002/uog.14822>
59. Bligh LN, Alsolai AA, Greer RM, Kumar S. Cerebroplacental ratio thresholds measured within 2-weeks before birth and risk of Cesarean section for intrapartum fetal compromise and adverse neonatal outcome. *Ultrasound Obstet Gynecol* 2018;**52**:340–6. <https://doi.org/10.1002/uog.17542>

60. Bligh LN, Al Solai A, Greer RM, Kumar S. Diagnostic performance of cerebroplacental ratio thresholds at term for prediction of low birthweight and adverse intrapartum and neonatal outcomes in a term, low-risk population. *Fetal Diagn Ther* 2018;**43**:191–8. <https://doi.org/10.1159/000477932>
61. Flatley C, Kumar S. Is the fetal cerebroplacental ratio better than the estimated fetal weight in predicting adverse perinatal outcomes in a low risk cohort? *J Matern Fetal Neonatal Med* 2019;**32**:2380–6. <https://doi.org/10.1080/14767058.2018.1438394>
62. Khalil AA, Morales-Roselló J, Morlando M, Hannan H, Bhide A, Papageorghiou A, et al. Is fetal cerebroplacental ratio an independent predictor of intrapartum fetal compromise and neonatal unit admission? *Am J Obstet Gynecol* 2015;**213**:54.e1–54.e10. <https://doi.org/10.1016/j.ajog.2014.10.024>
63. Maged AM, Abdelhafez A, Al Mostafa W, Elsherbiny W. Fetal middle cerebral and umbilical artery Doppler after 40 weeks gestational age. *J Matern Fetal Neonatal Med* 2014;**27**:1880–5. <https://doi.org/10.3109/14767058.2014.892068>
64. Monaghan C, Binder J, Thilaganathan B, Morales-Roselló J, Khalil A. Perinatal loss at term: the role of uteroplacental and fetal doppler assessment. *Ultrasound Obstet Gynecol* 2018;**52**:72–7. <https://doi.org/10.1002/uog.17500>
65. Morales-Roselló J, Khalil A, Morlando M, Papageorghiou A, Bhide A, Thilaganathan B. Changes in fetal Doppler indices as a marker of failure to reach growth potential at term. *Ultrasound Obstet Gynecol* 2014;**43**:303–10. <https://doi.org/10.1002/uog.13319>
66. Prior T, Mullins E, Bennett P, Kumar S. Prediction of intrapartum fetal compromise using the cerebroumbilical ratio: a prospective observational study. *Am J Obstet Gynecol* 2013;**208**:124.e1–6. <https://doi.org/10.1016/j.ajog.2012.11.016>
67. Prior T, Paramasivam G, Bennett P, Kumar S. Are fetuses that fail to achieve their growth potential at increased risk of intrapartum compromise? *Ultrasound Obstet Gynecol* 2015;**46**:460–4. <https://doi.org/10.1002/uog.14758>
68. Rial-Crestelo M, Martinez-Portilla RJ, Cancemi A, Caradeux J, Fernandez L, Peguero A, et al. Added value of cerebro-placental ratio and uterine artery Doppler at routine third trimester screening as a predictor of SGA and FGR in non-selected pregnancies. *J Matern Fetal Neonatal Med* 2019;**32**:2554–60. <https://doi.org/10.1080/14767058.2018.1441281>
69. Sabdia S, Greer RM, Prior T, Kumar S. Predicting intrapartum fetal compromise using the fetal cerebro-umbilical ratio. *Placenta* 2015;**36**:594–8. <https://doi.org/10.1016/j.placenta.2015.01.200>
70. Stumpfe FM, Kehl S, Pretscher J, Baier F, Bayer CM, Schwenke E, et al. Correlation of short-term variation and Doppler parameters with adverse perinatal outcome in low-risk fetuses at term. *Arch Gynecol Obstet* 2019;**299**:411–20. <https://doi.org/10.1007/s00404-018-4978-z>
71. Twomey S, Flatley C, Kumar S. The association between a low cerebro-umbilical ratio at 30–34 weeks gestation, increased intrapartum operative intervention and adverse perinatal outcomes. *Eur J Obstet Gynecol Reprod Biol* 2016;**203**:89–93. <https://doi.org/10.1016/j.ejogrb.2016.05.036>
72. Dunn L, Sherrell H, Kumar S. Review: Systematic review of the utility of the fetal cerebroplacental ratio measured at term for the prediction of adverse perinatal outcome. *Placenta* 2017;**54**:68–75. <https://doi.org/10.1016/j.placenta.2017.02.006>
73. Phelan JP, Ahn MO, Smith CV, Rutherford SE, Anderson E. Amniotic fluid index measurements during pregnancy. *J Reprod Med* 1987;**32**:601–4.

74. Ashwal E, Hirsch L, Melamed N, Aviram A, Wiznitzer A, Yogev Y. The association between isolated oligohydramnios at term and pregnancy outcome. *Arch Gynecol Obstet* 2014;**290**:875–81. <https://doi.org/10.1007/s00404-014-3292-7>
75. Ghosh G, Marsál K, Gudmundsson S. Amniotic fluid index in low-risk pregnancy as an admission test to the labor ward. *Acta Obstet Gynecol Scand* 2002;**81**:852–5. <https://doi.org/10.1034/j.1600-0412.2002.810909.x>
76. Hassan AA. The role of amniotic fluid index in the management of postdate pregnancy. *J Coll Physicians Surg Pak* 2005;**15**:85–8. <https://doi.org/02.2005/jcpsp.8588>
77. Hsieh TT, Hung TH, Chen KC, Hsieh CC, Lo LM, Chiu TH. Perinatal outcome of oligohydramnios without associated premature rupture of membranes and fetal anomalies. *Gynecol Obstet Invest* 1998;**45**:232–6. <https://doi.org/10.1159/000009974>
78. Locatelli A, Vergani P, Toso L, Verderio M, Pezzullo JC, Ghidini A. Perinatal outcome associated with oligohydramnios in uncomplicated term pregnancies. *Arch Gynecol Obstet* 2004;**269**:130–3. <https://doi.org/10.1007/s00404-003-0525-6>
79. Megha B, Indu C. Correlation of amniotic fluid index with perinatal outcome. *J Obstet Gynecol India* 2014;**64**:32–5. <https://doi.org/10.1007/s13224-013-0467-2>
80. Melamed N, Pardo J, Milstein R, Chen R, Hod M, Yogev Y. Perinatal outcome in pregnancies complicated by isolated oligohydramnios diagnosed before 37 weeks of gestation. *Am J Obstet Gynecol* 2011;**205**:241.e1–6. <https://doi.org/10.1016/j.ajog.2011.06.013>
81. Morris JM, Thompson K, Smithey J, Gaffney G, Cooke I, Chamberlain P, et al. The usefulness of ultrasound assessment of amniotic fluid in predicting adverse outcome in prolonged pregnancy: a prospective blinded observational study. *BJOG* 2003;**110**:989–94. <https://doi.org/10.1111/j.1471-0528.2003.02417.x>
82. Myles TD, Santolaya-Forgas J. Normal ultrasonic evaluation of amniotic fluid in low-risk patients at term. *J Reprod Med* 2002;**47**:621–4.
83. Naveiro-Fuentes M, Prieto AP, Ruiz RS, Badillo MPC, Ventoso FM, Vallejo JLG. Perinatal outcomes with isolated oligohydramnios at term pregnancy. *J Perinat Med* 2016;**44**:793–8. <https://doi.org/10.1515/jpm-2015-0198>
84. Quiñones JN, Odibo AO, Stringer M, Rochon ML, Macones GA. Determining a threshold for amniotic fluid as a predictor of perinatal outcome at term. *J Matern Fetal Neonatal Med* 2012;**25**:1319–23. <https://doi.org/10.3109/14767058.2011.632453>
85. Rainford M, Adair R, Scialli AR, Ghidini A, Spong CY. Amniotic fluid index in the uncomplicated term pregnancy. Prediction of outcome. *J Reprod Med* 2001;**46**:589–92.
86. Shanks A, Tuuli M, Schaecher C, Odibo AO, Rampersad R. Assessing the optimal definition of oligohydramnios associated with adverse neonatal outcomes. *J Ultrasound Med* 2011;**30**:303–7. <https://doi.org/10.7863/jum.2011.30.3.303>
87. Zhang J, Troendle J, Meikle S, Klebanoff MA, Rayburn WF. Isolated oligohydramnios is not associated with adverse perinatal outcomes. *BJOG* 2004;**111**:220–5. <https://doi.org/10.1111/j.1471-0528.2004.00060.x>
88. Moraitis AA, Armata I, Sovio U, Smith GC. *Borderline low amniotic fluid index and adverse pregnancy outcome at term: prospective cohort study and meta-analysis of diagnostic test accuracy*. Abstracts of the British Maternal & Fetal Medicine Society (BMFMS) 21st Annual Conference, 28–29 March 2019, Edinburgh, UK, abstract no. EP.317.

89. Asgharnia M, Faraji R, Salamat F, Ashrafkhani B, Dalil Heirati SF, Naimian S. Perinatal outcomes of pregnancies with borderline versus normal amniotic fluid index. *Iran J Reprod Med* 2013;**11**:705–10.
90. Banks EH, Miller DA. Perinatal risks associated with borderline amniotic fluid index. *Am J Obstet Gynecol* 1999;**180**:1461–3. [https://doi.org/10.1016/S0002-9378\(99\)70037-2](https://doi.org/10.1016/S0002-9378(99)70037-2)
91. Choi SR. Borderline amniotic fluid index and perinatal outcomes in the uncomplicated term pregnancy. *J Matern Fetal Neonatal Med* 2016;**29**:457–60. <https://doi.org/10.3109/14767058.2015.1004051>
92. Gumus II, Koktener A, Turhan NO. Perinatal outcomes of pregnancies with borderline amniotic fluid index. *Arch Gynecol Obstet* 2007;**276**:17–19. <https://doi.org/10.1007/s00404-006-0309-x>
93. Jamal A, Kazemi M, Marsoosi V, Eslamian L. Adverse perinatal outcomes in borderline amniotic fluid index. *Int J Reprod Biomed* 2016;**14**:705–8. <https://doi.org/10.29252/ijrm.14.11.705>
94. Kwon JY, Kwon HS, Kim YH, Park YW. Abnormal Doppler velocimetry is related to adverse perinatal outcome for borderline amniotic fluid index during third trimester. *J Obstet Gynaecol Res* 2006;**32**:545–9. <https://doi.org/10.1111/j.1447-0756.2006.00459.x>
95. Petrozella LN, Dashe JS, McIntire DD, Leveno KJ. Clinical significance of borderline amniotic fluid index and oligohydramnios in preterm pregnancy. *Obstet Gynecol* 2011;**117**:338–42. <https://doi.org/10.1097/AOG.0b013e3182056766>
96. Rutherford SE, Phelan JP, Smith CV, Jacobs N. The four-quadrant assessment of amniotic fluid volume: an adjunct to antepartum fetal heart rate testing. *Obstet Gynecol* 1987;**70**:353–6.
97. Sahin E, Madendag Y, Tayyar AT, Sahin ME, Col Madendag I, Acmaz G, *et al.* Perinatal outcomes in uncomplicated late preterm pregnancies with borderline oligohydramnios. *J Matern Fetal Neonatal Med* 2018;**31**:3085–8. <https://doi.org/10.1080/14767058.2017.1364722>
98. Wood SL, Newton JM, Wang L, Lesser K. Borderline amniotic fluid index and its relation to fetal intolerance of labor: a 2-center retrospective cohort study. *J Ultrasound Med* 2014;**33**:705–11. <https://doi.org/10.7863/ultra.33.4.705>
99. Boers KE, Vijgen SM, Bijlenga D, van der Post JA, Bekedam DJ, Kwee A, *et al.* Induction versus expectant monitoring for intrauterine growth restriction at term: randomised equivalence trial (DIGITAT). *BMJ* 2010;**341**:c7087. <https://doi.org/10.1136/bmj.c7087>
100. Campbell S, Wilkin D. Ultrasonic measurement of fetal abdomen circumference in the estimation of fetal weight. *Br J Obstet Gynaecol* 1975;**82**:689–97. <https://doi.org/10.1111/j.1471-0528.1975.tb00708.x>
101. Boulvain M, Senat MV, Perrotin F, Winer N, Beucher G, Subtil D, *et al.* Induction of labour versus expectant management for large-for-date fetuses: a randomised controlled trial. *Lancet* 2015;**385**:2600–5. [https://doi.org/10.1016/S0140-6736\(14\)61904-8](https://doi.org/10.1016/S0140-6736(14)61904-8)
102. Aviram A, Yogev Y, Ashwal E, Hirsch L, Hadar E, Gabbay-Benziv R. Prediction of large for gestational age by various sonographic fetal weight estimation formulas-which should we use? *J Perinatol* 2017;**37**:513–7. <https://doi.org/10.1038/jp.2017.5>
103. Balsyte D, Schäffer L, Burkhardt T, Wisser J, Kurmanavicius J. Sonographic prediction of macrosomia cannot be improved by combination with pregnancy-specific characteristics. *Ultrasound Obstet Gynecol* 2009;**33**:453–8. <https://doi.org/10.1002/uog.6282>
104. Benacerraf BR, Gelman R, Frigoletto FD. Sonographically estimated fetal weights: accuracy and limitation. *Am J Obstet Gynecol* 1988;**159**:1118–21. [https://doi.org/10.1016/0002-9378\(88\)90425-5](https://doi.org/10.1016/0002-9378(88)90425-5)

REFERENCES

105. Ben-Haroush A, Yogev Y, Hod M, Bar J. Predictive value of a single early fetal weight estimate in normal pregnancies. *Eur J Obstet Gynecol Reprod Biol* 2007;**130**:187–92. <https://doi.org/10.1016/j.ejogrb.2006.04.018>
106. Ben-Haroush A, Melamed N, Mashiach R, Meizner I, Yogev Y. Use of the amniotic fluid index combined with estimated fetal weight within 10 days of delivery for prediction of macrosomia at birth. *J Ultrasound Med* 2008;**27**:1029–32. <https://doi.org/10.7863/jum.2008.27.7.1029>
107. Benson CB, Coughlin BF, Doubilet PM. Amniotic fluid volume in large-for-gestational-age fetuses of nondiabetic mothers. *J Ultrasound Med* 1991;**10**:149–51. <https://doi.org/10.7863/jum.1991.10.3.149>
108. Burkhardt T, Schmidt M, Kurmanavicius J, Zimmermann R, Schaffer L. Evaluation of fetal anthropometric measures to predict the risk for shoulder dystocia. *Ultrasound Obstet Gynecol* 2014;**43**:77–82. <https://doi.org/10.1002/uog.12560>
109. Chauhan SP, Parker D, Shields D, Sanderson M, Cole JH, Scardo JA. Sonographic estimate of birth weight among high-risk patients: feasibility and factors influencing accuracy. *Am J Obstet Gynecol* 2006;**195**:601–6. <https://doi.org/10.1016/j.ajog.2006.04.012>
110. Chervenak JL, Divon MY, Hirsch J, Girz BA, Langer O. Macrosomia in the postdate pregnancy: is routine ultrasonographic screening indicated? *Am J Obstet Gynecol* 1989;**161**:753–6. [https://doi.org/10.1016/0002-9378\(89\)90395-5](https://doi.org/10.1016/0002-9378(89)90395-5)
111. Cohen JM, Hutcheon JA, Kramer MS, Joseph KS, Abenhaim H, Platt RW. Influence of ultrasound-to-delivery interval and maternal–fetal characteristics on validity of estimated fetal weight. *Ultrasound Obstet Gynecol* 2010;**35**:434–41. <https://doi.org/10.1002/uog.7506>
112. Crimmins S, Mo C, Nassar Y, Kopelman JN, Turan OM. Polyhydramnios or excessive fetal growth are markers for abnormal perinatal outcome in euglycemic pregnancies. *Am J Perinatol* 2018;**35**:140–5. <https://doi.org/10.1055/s-0037-1606186>
113. Cromi A, Ghezzi F, Di Naro E, Siesto G, Bergamini V, Raio L. Large cross-sectional area of the umbilical cord as a predictor of fetal macrosomia. *Ultrasound Obstet Gynecol* 2007;**30**:861–6. <https://doi.org/10.1002/uog.5183>
114. De Reu PAOM, Smits LJ, Oosterbaan HP, Nijhuis JG. Value of a single early third trimester fetal biometry for the prediction of birth weight deviations in a low risk population. *J Perinat Med* 2008;**36**:324–9. <https://doi.org/10.1515/JPM.2008.057>
115. Freire DM, Cecatti JG, Paiva CS. [Correlation between estimated fetal weight by ultrasound and neonatal weight.] *Rev Bras Ginecol Obstet* 2010;**32**:4–10. <https://doi.org/10.1590/S0100-72032010000100002>
116. Galvin DM, Burke N, Burke G, Breathnach F, McAuliffe F, Morrison J, et al. 94: Accuracy of prenatal detection of macrosomia > 4,000g and outcomes in the absence of intervention: results of the prospective multicenter genesis study. *Am J Obstet Gynecol* 2017;**216**:S68. <https://doi.org/10.1016/j.ajog.2016.11.983>
117. Gilby JR, Williams MC, Spellacy WN. Fetal abdominal circumference measurements of 35 and 38 cm as predictors of macrosomia. A risk factor for shoulder dystocia. *J Reprod Med* 2000;**45**:936–8. <https://doi.org/10.1097/00006254-200106000-00007>
118. Hasenoehrl G, Pohlhammer A, Gruber R, Staudach A, Steiner H. Fetal weight estimation by 2D and 3D ultrasound: comparison of six formulas. *Ultraschall Med* 2009;**30**:585–90. <https://doi.org/10.1055/s-0028-1109185>
119. Hendrix NW, Grady CS, Chauhan SP. Clinical vs. sonographic estimate of birth weight in term parturients. A randomized clinical trial. *J Reprod Med* 2000;**45**:317–22.

120. Henrichs C, Magann EF, Brantley KL, Crews JH, Sanderson M, Chauhan SP. Detecting fetal macrosomia with abdominal circumference alone. *J Reprod Med* 2003;**48**:339–42.
121. Humphries J, Reynolds D, Bell-Scarborough L, Lynn N, Scardo JA, Chauhan SP. Sonographic estimate of birth weight: relative accuracy of sonographers versus maternal-fetal medicine specialists. *J Matern Fetal Neonatal Med* 2002;**11**:108–12. <https://doi.org/10.1080/jmf.11.2.108.112>
122. Kayem G, Grangé G, Bréart G, Goffinet F. Comparison of fundal height measurement and sonographically measured fetal abdominal circumference in the prediction of high and low birth weight at term. *Ultrasound Obstet Gynecol* 2009;**34**:566–71. <https://doi.org/10.1002/uog.6378>
123. Kehl S, Brade J, Schmidt U, Berlit S, Bohlmann MK, Sütterlin M, et al. Role of fetal abdominal circumference as a prognostic parameter of perinatal complications. *Arch Gynecol Obstet* 2011;**284**:1345–9. <https://doi.org/10.1007/s00404-011-1888-8>
124. Levine AB, Lockwood CJ, Brown B, Lapinski R, Berkowitz RL. Sonographic diagnosis of the large for gestational age fetus at term: does it make a difference? *Obstet Gynecol* 1992;**79**:55–8.
125. Melamed N, Yogev Y, Meizner I, Mashlach R, Pardo J, Ben-Haroush A. Prediction of fetal macrosomia: effect of sonographic fetal weight-estimation model and threshold used. *Ultrasound Obstet Gynecol* 2011;**38**:74–81. <https://doi.org/10.1002/uog.8930>
126. Miller JM, Korndorffer FA, Gabert HA. Fetal weight estimates in late pregnancy with emphasis on macrosomia. *J Clin Ultrasound* 1986;**14**:437–42. <https://doi.org/10.1002/jcu.1870140606>
127. Miller JM, Brown HL, Khawli OF, Pastorek JG, Gabert HA. Ultrasonographic identification of the macrosomic fetus. *Am J Obstet Gynecol* 1988;**159**:1110–14. [https://doi.org/10.1016/0002-9378\(88\)90423-1](https://doi.org/10.1016/0002-9378(88)90423-1)
128. Nahum GG, Pham KQ, McHugh JP. Ultrasonic prediction of term birth weight in Hispanic women. Accuracy in an outpatient clinic. *J Reprod Med* 2003;**48**:13–22.
129. Nahum GG, Stanislaw H. A computerized method for accurately predicting fetal macrosomia up to 11 weeks before delivery. *Eur J Obstet Gynecol Reprod Biol* 2007;**133**:148–56. <https://doi.org/10.1016/j.ejogrb.2006.08.011>
130. Nicod AC, Hohlfeld P, Vial Y. [Performance of ultrasound estimation of fetal weight in fetuses weighing ≤ 2000 g and more than 4000 g.] *Rev Med Suisse* 2012;**8**:2022–4, 2026–7.
131. O'Reilly-Green CP, Divon MY. Receiver operating characteristic curves of sonographic estimated fetal weight for prediction of macrosomia in prolonged pregnancies. *Ultrasound Obstet Gynecol* 1997;**9**:403–8. <https://doi.org/10.1046/j.1469-0705.1997.09060403.x>
132. Pates JA, McIntire DD, Casey BM, Leveno KJ. Predicting macrosomia. *J Ultrasound Med* 2008;**27**:39–43. <https://doi.org/10.7863/jum.2008.27.1.39>
133. Peregrine E, O'Brien P, Jauniaux E. Clinical and ultrasound estimation of birth weight prior to induction of labor at term. *Ultrasound Obstet Gynecol* 2007;**29**:304–9. <https://doi.org/10.1002/uog.3949>
134. Pollack RN, Hauer-Pollack G, Divon MY. Macrosomia in postdates pregnancies: the accuracy of routine ultrasonographic screening. *Am J Obstet Gynecol* 1992;**167**:7–11. [https://doi.org/10.1016/S0002-9378\(11\)91615-9](https://doi.org/10.1016/S0002-9378(11)91615-9)
135. Rossavik IK, Joslin GL. Macrosomatia and ultrasonography: what is the problem? *South Med J* 1993;**86**:1129–32. <https://doi.org/10.1097/00007611-199310000-00010>
136. Sapir A, Khayyat I, Drukker L, Rabinowitz R, Samueloff A, Sela HY. 365: Ultrasound predication of shoulder dystocia in low risk term singleton deliveries. *Am J Obstet Gynecol* 2017;**216**(Suppl. 1):S221. <https://doi.org/10.1016/j.ajog.2016.11.623>

137. Smith GC, Smith MF, McNay MB, Fleming JE. The relation between fetal abdominal circumference and birthweight: findings in 3512 pregnancies. *Br J Obstet Gynaecol* 1997;**104**:186–90. <https://doi.org/10.1111/j.1471-0528.1997.tb11042.x>
138. Sovio U, Moraitis AA, Wong HS, Smith GCS. Universal vs selective ultrasonography to screen for large-for-gestational-age infants and associated morbidity. *Ultrasound Obstet Gynecol* 2018;**51**:783–91. <https://doi.org/10.1002/uog.17491>
139. Sritippayawan S, Anansakunwat W, Suthantikorn C. The accuracy of gestation-adjusted projection method in estimating birth weight by sonographic fetal measurements in the third trimester. *J Med Assoc Thai* 2007;**90**:1058–67.
140. Sylvestre G, Divon MY, Onyeije C, Fisher M. Diagnosis of macrosomia in the postdates population: combining sonographic estimates of fetal weight with glucose challenge testing. *J Matern Fetal Med* 2000;**9**:287–90. [https://doi.org/10.1002/1520-6661\(200009/10\)9:5<287::AID-MFM6>3.0.CO;2-1](https://doi.org/10.1002/1520-6661(200009/10)9:5<287::AID-MFM6>3.0.CO;2-1)
141. Weiner Z, Ben-Shlomo I, Beck-Fruchter R, Goldberg Y, Shalev E. Clinical and ultrasonographic weight estimation in large for gestational age fetus. *Eur J Obstet Gynecol Reprod Biol* 2002;**105**:20–4. [https://doi.org/10.1016/S0301-2115\(02\)00140-9](https://doi.org/10.1016/S0301-2115(02)00140-9)
142. Weiner E, Fainstein N, Mizrahi Y, Elyashiv O, Mevorach-Zussman N, Bar J, et al. 410: Comparison between three methods for the detection of macrosomia and growth restriction in patients presenting in active labor—a prospective study. *Am J Obstet Gynecol* 2016;**214**:S225–S6. <https://doi.org/10.1016/j.ajog.2015.10.451>
143. Shepard MJ, Richards VA, Berkowitz RL, Warsof SL, Hobbins JC. An evaluation of two equations for predicting fetal weight by ultrasound. *Am J Obstet Gynecol* 1982;**142**:47–54. [https://doi.org/10.1016/s0002-9378\(16\)32283-9](https://doi.org/10.1016/s0002-9378(16)32283-9)
144. Ouzounian JG. Shoulder dystocia: incidence and risk factors. *Clin Obstet Gynecol* 2016;**59**:791–4. <https://doi.org/10.1097/GRF.0000000000000227>
145. Little SE, Edlow AG, Thomas AM, Smith NA. Estimated fetal weight by ultrasound: a modifiable risk factor for cesarean delivery? *Am J Obstet Gynecol* 2012;**207**:309.e1–309.e6. <https://doi.org/10.1016/j.ajog.2012.06.065>
146. Blackwell SC, Refuerzo J, Chadha R, Carreno CA. Overestimation of fetal weight by ultrasound: does it influence the likelihood of cesarean delivery for labor arrest? *Am J Obstet Gynecol* 2009;**200**:340.e1–340.e3. <https://doi.org/10.1016/j.ajog.2008.12.043>
147. Parry S, Severs CP, Sehdev HM, Macones GA, White LM, Morgan MA. Ultrasonographic prediction of fetal macrosomia. Association with cesarean delivery. *J Reprod Med* 2000;**45**:17–22.
148. Gordijn SJ, Beune IM, Thilaganathan B, Papageorgiou A, Baschat AA, Baker PN, et al. Consensus definition of fetal growth restriction: a Delphi procedure. *Ultrasound Obstet Gynecol* 2016;**48**:333–9. <https://doi.org/10.1002/uog.15884>
149. Gaccioli F, Sovio U, Cook E, Hund M, Charnock-Jones DS, Smith GCS. Screening for fetal growth restriction using ultrasound and the sFLT1/PIGF ratio in nulliparous women: a prospective cohort study. *Lancet Child Adolesc Health* 2018;**2**:569–81. [https://doi.org/10.1016/S2352-4642\(18\)30129-9](https://doi.org/10.1016/S2352-4642(18)30129-9)
150. Bond DM, Gordon A, Hyett J, de Vries B, Carberry AE, Morris J. Planned early delivery versus expectant management of the term suspected compromised baby for improving outcomes. *Cochrane Database Syst Rev* 2015;**11**:CD009433. <https://doi.org/10.1002/14651858.CD009433.pub2>

151. Boulvain M, Irion O, Dowswell T, Thornton JG. Induction of labour at or near term for suspected fetal macrosomia. *Cochrane Database Syst Rev* 2016;5:CD000938. <https://doi.org/10.1002/14651858.CD000938.pub2>
152. Royal College of Obstetricians and Gynaecologists (RCOG). *The Investigation and Management of the Small-for-Gestational-Age Fetus, Investigation and Management (Green-top Guideline No. 31)*. London: RCOG; 2013.
153. NHS England. *Saving Babies' Lives Version Two: A Care Bundle for Reducing Perinatal Mortality*. URL: www.england.nhs.uk/wp-content/uploads/2019/03/Saving-Babies-Lives-Care-Bundle-Version-Two-Updated-Final-Version.pdf (accessed February 2021).
154. Grobman WA, Rice MM, Reddy UM, Tita ATN, Silver RM, Mallett G, et al. Labor induction versus expectant management in low-risk nulliparous women. *N Engl J Med* 2018;**379**:513–23. <https://doi.org/10.1056/NEJMoa1800566>
155. Wastlund D, Moraitis AA, Thornton JG, Sanders J, White IR, Brocklehurst P, et al. The cost-effectiveness of universal late-pregnancy screening for macrosomia in nulliparous women: a decision analysis. *BJOG* 2019;**126**:1243–50. <https://doi.org/10.1111/1471-0528.15809>
156. Impey LWM, Murphy DJ, Griffiths M, Penna LK on behalf of the Royal College of Obstetricians and Gynaecologists. Management of breech presentation. *BJOG* 2017;**124**:e151–e77. <https://doi.org/10.1111/1471-0528.14465>
157. Rouse DJ, Owen J, Goldenberg RL, Cliver SP. The effectiveness and costs of elective cesarean delivery for fetal macrosomia diagnosed by ultrasound. *JAMA* 1996;**276**:1480–6. <https://doi.org/10.1001/jama.1996.03540180036030>
158. Wickham H, Henry L. *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*. 2019. URL: <https://tidyr.tidyverse.org/> (accessed May 2019).
159. Strong M. SAVI: SAVI Sheffield Accelerated Value of Information. 2015. URL: <https://www.sciencedirect.com/science/article/pii/S1098301515048354?via%3Dihub> (accessed May 2019).
160. Stock SJ, Ferguson E, Duffy A, Ford I, Chalmers J, Norman JE. Outcomes of elective induction of labour compared with expectant management: population based study. *BMJ* 2012;**344**:e2838. <https://doi.org/10.1136/bmj.e2838>
161. Leung WC, Pun TC, Wong WM. Undiagnosed breech revisited. *Br J Obstet Gynaecol* 1999;**106**:638–41. <https://doi.org/10.1111/j.1471-0528.1999.tb08360.x>
162. Ben-Meir A, Elram T, Tsafirir A, Elchalal U, Ezra Y. The incidence of spontaneous version after failed external cephalic version. *Am J Obstet Gynecol* 2007;**196**:157.e1–3. <https://doi.org/10.1016/j.ajog.2006.10.889>
163. Morrison JJ, Rennie JM, Milton PJ. Neonatal respiratory morbidity and mode of delivery at term: influence of timing of elective caesarean section. *Br J Obstet Gynaecol* 1995;**102**:101–6. <https://doi.org/10.1111/j.1471-0528.1995.tb09060.x>
164. Ouzounian JG, Gherman RB. Shoulder dystocia: are historic risk factors reliable predictors? *Am J Obstet Gynecol* 2005;**192**:1933–5; discussion 5–8. <https://doi.org/10.1016/j.ajog.2005.02.054>
165. Rossi AC, Mullin P, Prefumo F. Prevention, management, and outcomes of macrosomia: a systematic review of literature and meta-analysis. *Obstet Gynecol Surv* 2013;**68**:702–9. <https://doi.org/10.1097/01.ogx.0000435370.74455.a8>
166. Chongsuvivatwong V, Bachtiar H, Chowdhury ME, Fernando S, Suwanrath C, Kor-Anantakul O, et al. Maternal and fetal mortality and complications associated with cesarean section deliveries in teaching hospitals in Asia. *J Obstet Gynaecol Res* 2010;**36**:45–51. <https://doi.org/10.1111/j.1447-0756.2009.01100.x>

167. Gibson KS, Waters TP, Bailit JL. Maternal and neonatal outcomes in electively induced low-risk term pregnancies. *Am J Obstet Gynecol* 2014;**211**:249.e1–249.e16. <https://doi.org/10.1016/j.ajog.2014.03.016>
168. MacKenzie IZ, Shah M, Lean K, Dutton S, Newdick H, Tucker DE. Management of shoulder dystocia: trends in incidence and maternal and neonatal morbidity. *Obstet Gynecol* 2007;**110**:1059–68. <https://doi.org/10.1097/01.AOG.0000287615.35425.5c>
169. Sandmire HF, DeMott RK. The Green Bay cesarean section study. IV. The physician factor as a determinant of cesarean birth rates for the large fetus. *Am J Obstet Gynecol* 1996;**174**:1557–64. [https://doi.org/10.1016/s0002-9378\(96\)70606-3](https://doi.org/10.1016/s0002-9378(96)70606-3)
170. Thorngren-Jerneck K, Herbst A. Low 5-minute Apgar score: a population-based register study of 1 million term births. *Obstet Gynecol* 2001;**98**:65–70. <https://doi.org/10.1097/00006250-200107000-00012>
171. Pasupathy D, Wood AM, Pell JP, Fleming M, Smith GC. Time trend in the risk of delivery-related perinatal and neonatal death associated with breech presentation at term. *Int J Epidemiol* 2009;**38**:490–8. <https://doi.org/10.1093/ije/dyn225>
172. MacKay DF, Smith GC, Dobbie R, Pell JP. Gestational age at delivery and special educational need: retrospective cohort study of 407,503 schoolchildren. *PLOS Med* 2010;**7**:e1000289. <https://doi.org/10.1371/journal.pmed.1000289>
173. Persson M, Razaz N, Tedroff K, Joseph KS, Cnattingius S. Five and 10 minute Apgar scores and risks of cerebral palsy and epilepsy: population based cohort study in Sweden. *BMJ* 2018;**360**:k207. <https://doi.org/10.1136/bmj.k207>
174. Iliodromiti S, Mackay DF, Smith GC, Pell JP, Nelson SM. Apgar score and the risk of cause-specific infant mortality: a population-based cohort study. *Lancet* 2014;**384**:1749–55. [https://doi.org/10.1016/S0140-6736\(14\)61135-1](https://doi.org/10.1016/S0140-6736(14)61135-1)
175. NHS Improvement. *National Schedule of Reference Costs, 2016–17 – NHS Trusts and NHS Foundation Trusts*. London: NHS; 2017.
176. Vijgen SM, Boers KE, Opmeer BC, Bijlenga D, Bekedam DJ, Bloemenkamp KW, *et al*. Economic analysis comparing induction of labour and expectant management for intrauterine growth restriction at term (DIGITAT trial). *Eur J Obstet Gynecol Reprod Biol* 2013;**170**:358–63. <https://doi.org/10.1016/j.ejogrb.2013.07.017>
177. Palencia R, Gafni A, Hannah ME, Ross S, Willan AR, Hewson S, *et al*. The costs of planned cesarean versus planned vaginal birth in the Term Breech Trial. *CMAJ* 2006;**174**:1109–13. <https://doi.org/10.1503/cmaj.050796>
178. James M, Hunt K, Burr R, Johanson R. A decision analytical cost analysis of offering ECV in a UK district general hospital. *BMC Health Serv Res* 2001;**1**:6. <https://doi.org/10.1186/1472-6963-1-6>
179. Alfirevic Z, Keeney E, Dowswell T, Welton NJ, Medley N, Dias S, *et al*. Which method is best for the induction of labour? A systematic review, network meta-analysis and cost-effectiveness analysis. *Health Technol Assess* 2016;**20**(65). <https://doi.org/10.3310/hta20650>
180. Culligan PJ, Myers JA, Goldberg RP, Blackwell L, Gohmann SF, Abell TD. Elective cesarean section to prevent anal incontinence and brachial plexus injuries associated with macrosomia – a decision analysis. *Int Urogynecol J Pelvic Floor Dysfunct* 2005;**16**:19–28. <https://doi.org/10.1007/s00192-004-1203-3>
181. Mistry H, Heazell AE, Vincent O, Roberts T. A structured review and exploration of the healthcare costs associated with stillbirth and a subsequent pregnancy in England and Wales. *BMC Pregnancy Childbirth* 2013;**13**:236. <https://doi.org/10.1186/1471-2393-13-236>

182. Barrett B, Mosweu I, Jones CR, Charman T, Baird G, Simonoff E, *et al.* Comparing service use and costs among adolescents with autism spectrum disorders, special needs and typical development. *Autism* 2015;**19**:562–9. <https://doi.org/10.1177/1362361314536626>
183. Access Economics. *The Economic Impact of Cerebral Palsy in Australia in 2007*. Sydney, NSW: Access Economics Pty limited; 2008.
184. Curtis L, Burns A. *Unit Costs of Health and Social Care 2017*. Canterbury: PSSRU, University of Kent; 2017.
185. Szende A, Janssen B, Cabasés J. *Self-Reported Population Health: An International Perspective Based on EQ-5D*. Dordrecht: Springer Netherlands; 2014. <https://doi.org/10.1007/978-94-007-7596-1>
186. Office for National Statistics. *National Life Tables, United Kingdom, 1980–82 to 2014–16*. Newport: Office for National Statistics; 2017.
187. Leigh S, Granby P, Turner M, Wieteska S, Haycox A, Collins B. The incidence and implications of cerebral palsy following potentially avoidable obstetric complications: a preliminary burden of disease study. *BJOG* 2014;**121**:1720–8. <https://doi.org/10.1111/1471-0528.12897>
188. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal 2013*. London: NICE; 2013.
189. Wilson ECF, Wastlund D, Moraitis AA, Smith GCS. Late pregnancy ultrasound to screen for and manage potential birth complications in nulliparous women: a cost-effectiveness and value of information analysis. *Value Health* 2021; in press.
190. Wilson EC. A practical guide to value-of-information analysis. *PharmacoEconomics* 2015;**33**:105–21. <https://doi.org/10.1007/s40273-014-0219-x>
191. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999;**18**:341–64. [https://doi.org/10.1016/S0167-6296\(98\)00039-3](https://doi.org/10.1016/S0167-6296(98)00039-3)
192. Pratt J, Raiffa H, Schlaifer R. *Introduction to Statistical Decision Theory*. Cambridge, MA: Massachusetts Institute of Technology; 1995.
193. Heath A, Baio G. Calculating the expected value of sample information using efficient nested Monte Carlo: a tutorial. *Value Health* 2018;**21**:1299–304. <https://doi.org/10.1016/j.jval.2018.05.004>
194. Walker KF, Dritsaki M, Bugg G, Macpherson M, McCormick C, Grace N, *et al.* Labour induction near term for women aged 35 or over: an economic evaluation. *BJOG* 2017;**124**:929–34. <https://doi.org/10.1111/1471-0528.14557>
195. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M, ISPOR-SMDM Modeling Good Research Practices Task Force. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force – 2. *Value Health* 2012;**15**:804–11. <https://doi.org/10.1016/j.jval.2012.06.016>
196. Heazell AEP, Siassakos D, Blencowe H, Burden C, Bhutta ZA, Cacciatore J, *et al.* Stillbirths: economic and psychosocial consequences. *Lancet* 2016;**387**:604–16. [https://doi.org/10.1016/S0140-6736\(15\)00836-3](https://doi.org/10.1016/S0140-6736(15)00836-3)
197. Heazell AE, Whitworth MK, Whitcombe J, Glover SW, Bevan C, Brewin J, *et al.* Research priorities for stillbirth: process overview and results from UK Stillbirth Priority Setting Partnership. *Ultrasound Obstet Gynecol* 2015;**46**:641–7. <https://doi.org/10.1002/uog.15738>
198. Eskes M, Ensing S, Groenendaal F, Abu-Hanna A, Ravelli A. The risk of intrapartum/neonatal mortality and morbidity following birth at 37 weeks of gestation: a nationwide cohort study. *BJOG* 2019;**126**:1252–7. <https://doi.org/10.1111/1471-0528.15748>

REFERENCES

199. Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of breast cancer screening: systematic review and meta-analysis to update the 2009 U.S. Preventive Services Task Force recommendation. *Ann Intern Med* 2016;**164**:244–55. <https://doi.org/10.7326/M15-0969>
200. Norman JE, Heazell AEP, Rodriguez A, Weir CJ, Stock SJE, Calderwood CJ, *et al.* Awareness of fetal movements and care package to reduce fetal mortality (AFFIRM): a stepped wedge, cluster-randomised trial. *Lancet* 2018;**392**:1629–38. [https://doi.org/10.1016/s0140-6736\(18\)31543-5](https://doi.org/10.1016/s0140-6736(18)31543-5)
201. NHS Digital. *NHS Maternity Statistics, England 2016–17*. Leeds: NHS Digital; 2017.
202. American College of Obstetricians and Gynecologists. Practice bulletin no. 173: fetal macrosomia. *Obstet Gynecol* 2016;**128**:e195–e209. <https://doi.org/10.1097/aog.0000000000001767>
203. Cullinney KA, Parry GK, Brown J, Crowther CA. Regimens of fetal surveillance of suspected large-for-gestational-age fetuses for improving health outcomes. *Cochrane Database Syst Rev* 2016;**4**:CD011739. <https://doi.org/10.1002/14651858.CD011739.pub2>
204. Benner JS, Morrison MR, Karnes EK, Kocot SL, McClellan M. An evaluation of recent federal spending on comparative effectiveness research: priorities, gaps, and next steps. *Health Aff* 2010;**29**:1768–76. <https://doi.org/10.1377/hlthaff.2010.0687>
205. Royal College of Obstetricians and Gynaecologists (RCOG). *Green-top Guideline No. 42: Shoulder Dystocia*. London: RCOG; 2012.
206. Hannah ME, Hannah WJ, Hewson SA, Hodnett ED, Saigal S, Willan AR. Planned caesarean section versus planned vaginal birth for breech presentation at term: a randomised multicentre trial. *Lancet* 2000;**356**:1375–83. [https://doi.org/10.1016/s0140-6736\(00\)02840-3](https://doi.org/10.1016/s0140-6736(00)02840-3)
207. Gherman RB, Ouzounian JG, Miller DA, Kwok L, Goodwin TM. Spontaneous vaginal delivery: a risk factor for Erb's palsy? *Am J Obstet Gynecology* 1998;**178**:423–7. [https://doi.org/10.1016/s0002-9378\(98\)70413-2](https://doi.org/10.1016/s0002-9378(98)70413-2)
208. NHS Digital. *NHS Staff Earnings Estimates to September 2017 – Provisional Statistics*. Leeds: NHS Digital; 2017.
209. NHS Purchasing and Supply Agency. *Cost-effectiveness of Ultrasound Elastography in the Assessment of Liver Fibrosis*. Chester: NHS Purchasing and Supply Agency; 2009.
210. Curtis L. *Unit Costs of Health and Social Care 2008*. Canterbury: PSSRU, University of Kent; 2008.
211. Malloy MH, Freeman DH. Respiratory distress syndrome mortality in the United States, 1987 to 1995. *J Perinatol* 2000;**20**:414–20. <https://doi.org/10.1038/sj.jp.7200420>
212. Curtis L, Burns A. *Unit Costs of Health and Social Care 2016*. Canterbury: PSSRU, University of Kent; 2016.
213. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol* 1997;**39**:214–23. <https://doi.org/10.1111/j.1469-8749.1997.tb07414.x>
214. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199–208. [https://doi.org/10.1016/0168-8510\(90\)90421-9](https://doi.org/10.1016/0168-8510(90)90421-9)
215. Young NL, Rochon TG, McCormick A, Law M, Wedge JH, Fehlings D. The health and quality of life outcomes among youth and young adults with cerebral palsy. *Arch Phys Med Rehabil* 2010;**91**:143–8. <https://doi.org/10.1016/j.apmr.2009.08.152>

Appendix 1 Supporting data for the systematic review of the diagnostic effectiveness of universal ultrasonic screening using late pregnancy umbilical artery Doppler flow velocimetry in the prediction of adverse perinatal outcome

MEDLINE and EMBASE

Date range searched: inception to 19 March 2019.

Search strategy

1. exp pregnant woman/
2. exp pregnancy/
3. pregnan*.mp.
4. exp prenatal diagnosis/
5. exp fetus echography/
6. exp Doppler ultrasonography/
7. arterial doppler.mp.
8. doppler velocimetry.mp.
9. doppler ultraso*.mp.
10. umbilical arter*.mp.
11. 1 or 2 or 3
12. 4 or 5 or 6
13. 7 or 8 or 9 or 10
14. 11 and 12
15. 13 and 14.

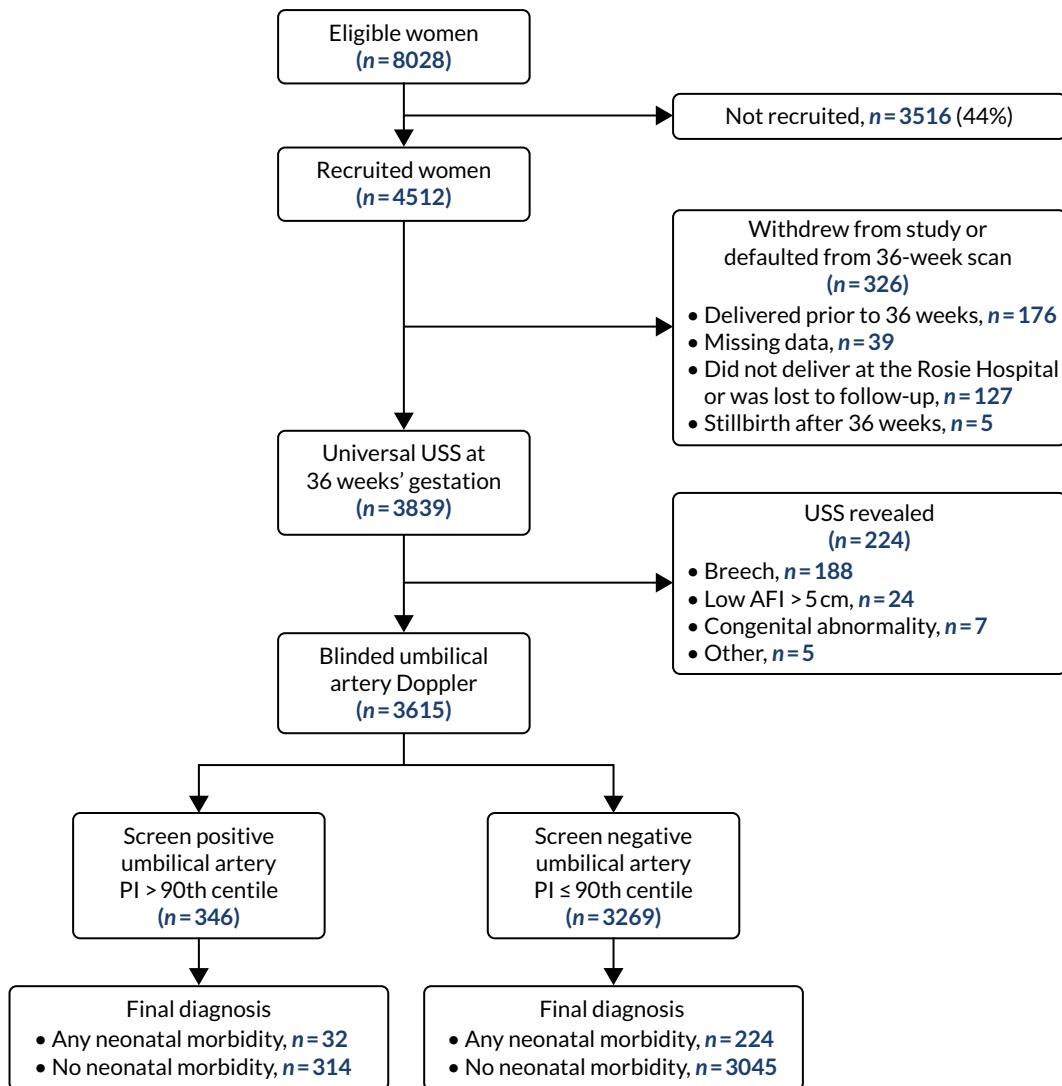


FIGURE 25 The POP study inclusion flow chart. USS, ultrasound scan.

TABLE 18 Maternal characteristics and birth outcomes of POP study

Characteristic	UA PI > 90th centile (N = 346)	UA PI < 90th centile (N = 3269)	p-value	Overall baseline characteristics (N = 3615)
Maternal characteristic				
Age (years), median (IQR)	29.7 (26.2–32.7)	30.3 (26.8–33.3)	0.05	30.2 (26.7–33.3)
Deprivation quartile, n (%)				
1 (lowest)	97 (28.0)	784 (24.0)	0.14	881 (24.4)
2	73 (21.1)	776 (23.7)		849 (23.5)
3	92 (26.6)	773 (23.7)		865 (23.9)
4 (highest)	71 (20.5)	799 (24.4)		870 (24.1)
Missing	13 (3.7)	137 (4.2)		150 (4.2)

TABLE 18 Maternal characteristics and birth outcomes of POP study (continued)

Characteristic	UA PI > 90th centile (N = 346)	UA PI < 90th centile (N = 3269)	p-value	Overall baseline characteristics (N = 3615)
White ethnicity, n (%)	324 (93.6)	3036 (92.9)	0.53	3360 (93.0)
Missing	6 (1.7)	56 (1.7)		62 (1.7)
Married, n (%)	229 (66.2)	2238 (68.5)	0.39	2467 (68.2)
Smoker, n (%)	24 (6.9)	152 (4.7)	0.06	176 (4.9)
Any alcohol consumption, n (%)	13 (3.8)	155 (4.7)	0.40	168 (4.7)
Missing	0 (0)	1 (0)		1 (0)
BMI (kg/m ²), median (IQR)	24.3 (21.7–28.1)	24.0 (21.8–27.2)	0.44	24.0 (21.8–27.3)
One or more previous miscarriage(s), n (%)	34 (9.8)	331 (10.1)	0.86	365 (10.1)
Chronic hypertension, n (%)	25 (7.3)	161 (4.9)	0.06	186 (5.1)
Pre-eclampsia, n (%)	29 (8.4)	204 (6.2)	0.12	233 (6.5)
Missing	0 (0)	2 (0.1)		2 (0.1)
DM, n (%)				
Type 1 or type 2	2 (0.6)	10 (0.3)	0.14	12 (0.3)
Gestational	20 (5.8)	124 (3.8)		144 (4.0)
Birth outcome				
Birthweight (g), median (IQR)	3263 (2970–3560)	3470 (3170–3770)	< 0.001	3445 (3150–3750)
Gestational age (weeks), median (IQR)	40.4 (39.3–41.1)	40.4 (39.4–41.3)	0.74	40.4 (39.4–41.3)
< 37	3 (0.9)	34 (1.0)	0.19 ^a	37 (1.0)
37	22 (6.4)	133 (4.1)		155 (4.3)
38	35 (10.1)	360 (11.0)		395 (10.9)
39	71 (20.5)	641 (19.6)		712 (19.7)
40	92 (26.6)	1001 (30.6)		1093 (30.2)
41	102 (29.5)	909 (27.8)		1011 (30.0)
≥ 42	21 (6.1)	191 (5.8)		212 (5.9)
IOL, n (%)	125 (36.1)	1081 (33.1)	0.25	1206 (33.4)
Mode of delivery, n (%)				
Spontaneous vaginal	178 (51.5)	1662 (50.8)	0.20	1840 (50.9)
Assisted vaginal	86 (24.9)	821 (25.1)		907 (25.1)
Intrapartum caesarean	54 (15.6)	601 (18.4)		655 (18.1)
Pre-labour caesarean	27 (7.8)	176 (5.4)		203 (5.6)
Missing	1 (0.3)	9 (0.3)		10 (0.3)
DM, diabetes mellitus; IQR, interquartile range. a p-value for trend.				

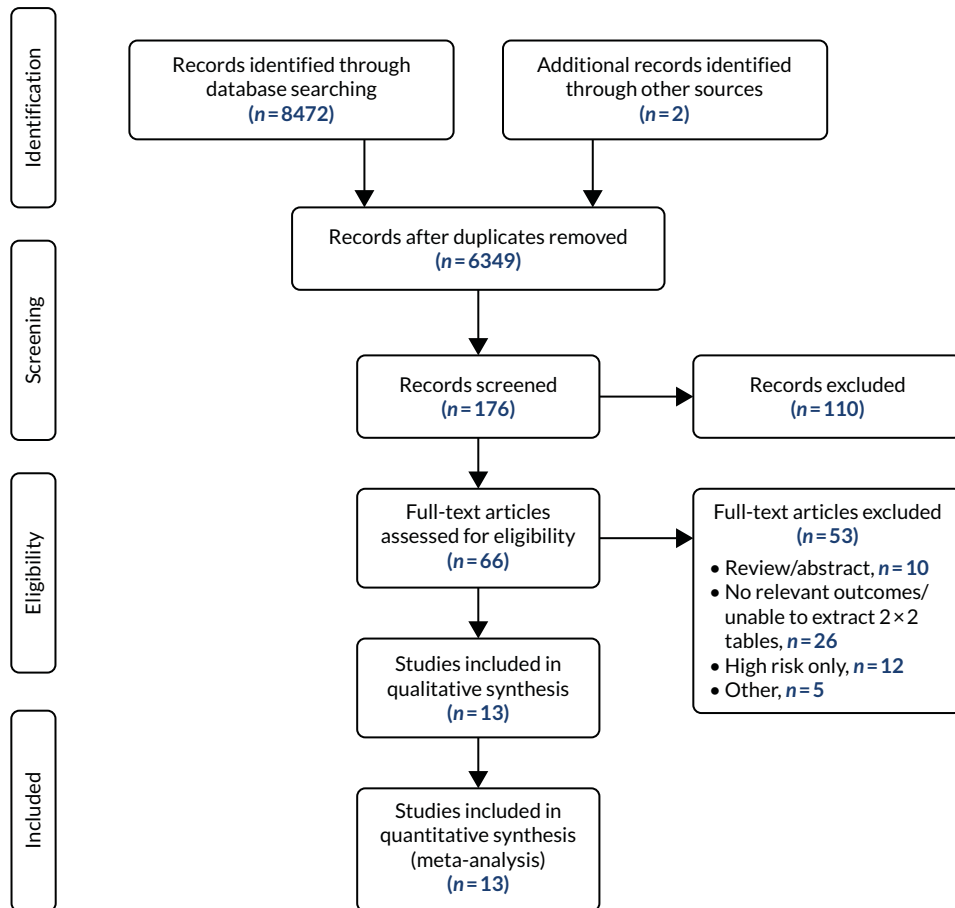


FIGURE 26 Literature search PRISMA flow diagram for the systematic review on umbilical artery Doppler.

	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Akolekar 2019 ⁴²	+	-	-	+	+	+	+
Bolz 2013 ⁴³	+	+	+	+	+	+	+
Cooley 2011 ⁴⁴	+	+	+	+	+	+	+
Filmar 2013 ⁴⁵	?	+	-	+	-	+	+
Fischer 1991 ⁴⁶	?	+	+	-	+	+	+
Goffinet 1997 ⁴⁷	+	+	-	+	+	+	+
Hanretty 1989 ⁴⁸	+	+	+	+	+	+	+
Moraitis 2021 ³⁵	+	+	+	+	+	+	+
Schulman 1989 ⁴⁹	?	+	-	+	?	+	+
Sijmons 1989 ⁵⁰	+	+	+	+	+	+	+
Valino 2016 ⁵¹	+	+	-	+	+	+	+
Valino 2016 ⁵²	+	+	-	+	+	+	+
Weiner 1993 ⁵³	+	+	-	-	+	+	+

- High
? Unclear
+ Low

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

TABLE 19 Characteristics of studies included in the meta-analysis

Study (first author and year of publication)	Type of study; setting	Number of fetuses and selection (all singleton, non-anomalous unless otherwise stated)	Index test	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Akolekar 2019 ⁴²	Prospective cohort; two NHS hospitals, UK Between March 2014 and September 2018 (potential overlap with Valino <i>et al.</i> studies ^{51,52})	<i>n</i> = 47,211 Universal, > 36 weeks' gestation	PI > 90th centile Not blinded	Between 35 ⁺⁶ and 37 ⁺⁶ weeks' gestation	Adverse perinatal outcome (composite of stillbirth, neonatal deaths and HIE grade 2 or 3), perinatal hypoxia (cord arterial pH of < 7.0, 5-minute Apgar score of < 7, NICU admission), caesarean section for fetal compromise, SGA < 3rd centile	Median gestational age at delivery 40.0 (39.0–40.9) weeks	Nulliparous: 45.4% for those with no adverse outcome, 58.5% for those with adverse outcome
Bolz 2013 ⁴³	Prospective cohort; single hospital, Germany	<i>n</i> = 514 Low risk, term, cephalic only Excluded maternal disease, SGA, RFM	PI > 1.2 Blinded umbilical artery Doppler	Within 1 week from delivery Mean gestational age 39 ⁺² weeks	Neonatal acidosis (cord arterial pH of < 7.10)	Mean gestational age: 40 ⁺¹ weeks	Nulliparity: not reported IOL: not reported
Cooley 2011 ⁴⁴	Prospective cohort; single hospital, Ireland	<i>n</i> = 810 Mixed risk, nulliparous only. Only included Caucasians aged 18–40 years	PI > 95th centile Umbilical artery blinded but EFW not blinded	Around 36 weeks' gestation (not specified)	Emergency caesarean section, PIH, pre-eclampsia, preterm delivery (< 37 weeks' gestation), SGA < 10th centile, SGA < 3rd centile, 5-minute Apgar score of < 7, cord arterial pH of < 7.10, NICU admission, stillbirth	Not reported	Nulliparity: all IOL: 22.4%
Filmar 2013 ⁴⁵	Retrospective cohort; single hospital, New York, NY, USA	<i>n</i> = 251 Mixed risk, EFW > 10th centile	S/D ratio > 90th centile (persistent), not blinded	Mean gestational age 35.3 weeks for abnormal umbilical artery group. Mean gestational age 34.4 weeks for control group	NICU admission, 5-minute Apgar score of < 7	Median gestational age: 37 weeks for the abnormal umbilical artery group, 39 weeks for the control group	Nulliparity: not reported IOL: not reported

continued

TABLE 19 Characteristics of studies included in the meta-analysis (continued)

Study (first author and year of publication)	Type of study; setting	Number of fetuses and selection (all singleton, non-anomalous unless otherwise stated)	Index test	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Fischer 1991 ⁴⁶	Prospective cohort; single hospital, PA, USA	<i>n</i> = 75 Low risk, post dates > 41 weeks' gestation. Excluded maternal disease, suspected IUGR	S/D ratio > 3.0 S/D ratio > 2.4 Blinded umbilical artery Doppler	Mean interval from scan to delivery: 2 days	Composite perinatal outcome: 1. non-reassuring intrapartum fetal heart rate 2. umbilical artery pH of < 7.15 or a venous pH of < 7.2 3. 5-minute Apgar score of < 7 4. meconium-stained liquor 5. NICU admission 6. birthweight < 10th centile	Mean gestational age: at delivery 292.2 days	Nulliparity: 57% IOL: not reported
Goffinet 1996 ⁴⁷	Prospective cohort; 17 hospitals, France	<i>n</i> = 1903 Low risk, excluded maternal disease, suspected IUGR	RI > 90th centile Not blinded	Between 28 and 34 weeks' gestation	PIH, pre-eclampsia, intervention for fetal distress, 5-minute Apgar score of < 7, NICU admission, birthweight < 3rd centile, birthweight 3–10th centile	Mean gestational age: 39.2 weeks for those with an abnormal umbilical artery, 39.4 weeks for those with a normal umbilical artery	Nulliparous: 43.0% for those with an abnormal umbilical artery, 45.3% for those with normal umbilical aretery
Hanretty 1989 ⁴⁸	Prospective cohort; single hospital, Glasgow, UK	<i>n</i> = 395 Universal	S/D ratio > 95th centile Blinded umbilical artery Doppler	34–36 weeks' gestation	PIH, SGA < 5th centile, 5-minute Apgar score of < 6, NICU admission	Mean gestational age: 38.9 weeks for those with an abnormal umbilical artery, 39.5 weeks for those with a normal umbilical artery	Nulliparity: not reported IOL: not reported
Moraitis 2021 ³⁵	Prospective cohort; single hospital, Cambridge, UK	<i>n</i> = 3615 Universal, nulliparous only, > 36 weeks' gestation	PI > 90th centile Blinded	Mean 36 weeks' gestation	NICU admission, metabolic acidosis, 5-minute Apgar score < 7, composite neonatal morbidity (one or more of the above), composite severe neonatal morbidity, SGA < 10th centile, SGA < 3rd centile	40.4 (39.3–41.1) weeks' gestation	Nulliparity: all IOL: 36.1% for those with an abnormal umbilical artery Doppler, 33.1% for those with a normal umbilical artery Doppler

Study (first author and year of publication)	Type of study; setting	Number of fetuses and selection (all singleton, non-anomalous unless otherwise stated)	Index test	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Schulman 1989 ⁴⁹	Prospective cohort; single hospital, NY, USA	n = 255 Mixed	S/D ratio > 3 Not blinded	Around 30 weeks' gestation	SGA < 15th centile	Not reported	Nulliparous: not reported IOL: not reported
Sijmons 1989 ⁵⁰	Prospective cohort; single hospital, the Netherlands	n = 368 Mixed (randomly selected)	PI > 95th centile Blinded umbilical artery Doppler	At 28 and 34 weeks' gestation	SGA < 10th centile, SGA < 3rd centile	Not reported	Nulliparous: not reported IOL: not reported
Valino 2016 ⁵¹	Retrospective cohort; three NHS hospitals, south-east England, UK	n = 8262	PI > 95th centile	30 ⁺⁰ -34 ⁺⁶ weeks' gestation	Term pre-eclampsia, term SGA < 10th centile, stillbirth, caesarean section for fetal distress, cord arterial pH of < 7.0, 5-minute Apgar score of < 7, NICU admission	Mean 40.0 weeks' gestation	Nulliparous: 49.2% IOL: 15.5%
	May 2011-August 2014	Universal	PI > 90th centile Not blinded	Mean 32.2 weeks' gestation			
Valino 2016 ⁵²	Retrospective cohort; two NHS hospitals, south-east England, UK	n = 3953	PI > 95th centile	35 ⁺⁰ -37 ⁺⁶ weeks' gestation	Pre-eclampsia, SGA < 10th centile, caesarean section for fetal distress, cord arterial pH of < 7.0, 5-minute Apgar score of < 7, NICU admission	Mean 40.0 weeks' gestation	Nulliparous: 49.7% IOL: 19.1%
	February 2014-December 2014 (potential overlap with above)	Universal	Not blinded	Mean 36.1 weeks' gestation			
Weiner 1993 ⁵³	Prospective cohort; single hospital, Israel	n = 142 Low risk, term only afert 41 weeks' gestation	RI > 95th centile Not blinded	After 41 weeks' gestation	Composite adverse outcome: 1. 5-minute Apgar score of < 7 2. NICU admission 3. Caesarean section for fetal distress, SGA < 5th centile	Mean 41.8 weeks' gestation	Nulliparous: n = 43 IOL: not reported

IUGR, intrauterine growth restriction; PIH, pregnancy-induced hypertension; S/D ratio, systolic/diastolic ratio.

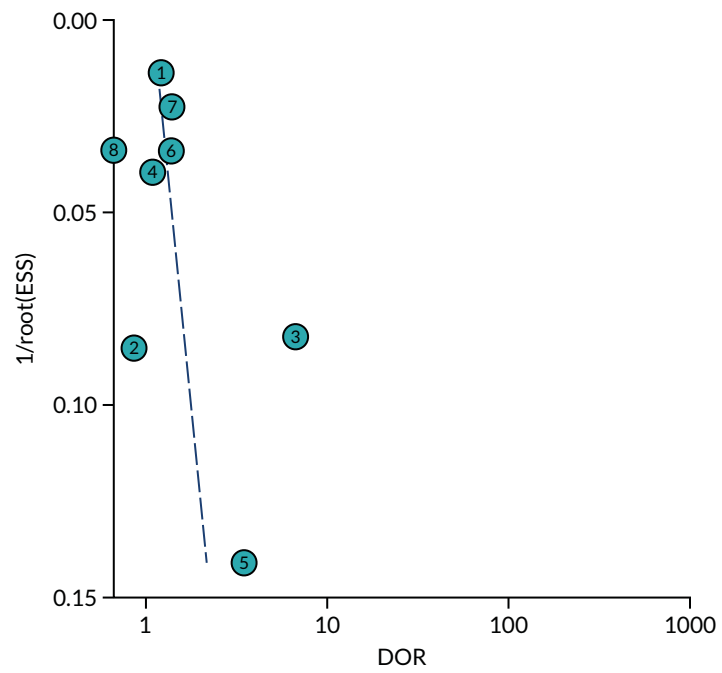


FIGURE 28 Deeks' funnel plot for publication bias for umbilical artery Doppler for the prediction of neonatal unit admission. Deeks' funnel plot asymmetry test, $p = 0.52$. ESS, effective sample size.

Appendix 2 Supporting data for the systematic review of the diagnostic effectiveness of universal ultrasonic screening using late pregnancy cerebroplacental ratio in the prediction of adverse perinatal outcome

MEDLINE and EMBASE

Date range searched: inception to 30 May 2019.

Search strategy

1. exp pregnant woman/
2. exp pregnancy/
3. pregnan*.mp.
4. exp fetus echography/
5. exp prenatal diagnosis/
6. exp Doppler ultrasonography/
7. exp fetus monitoring/
8. ultraso*.mp.
9. exp middle cerebral artery/
10. middle cerebral artery.mp.
11. uteroplacental.mp.
12. utero-placental.mp.
13. cerebroplacental.mp.
14. cerebro-placental.mp.
15. cerebroumbilical.mp.
16. cerebro-umbilical.mp.
17. fetal brain doppler.mp.
18. fetal cerebral doppler.mp.
19. 1 or 2 or 3
20. 4 or 5 or 6 or 7 or 8
21. 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18
22. 19 and 20
23. 21 and 22.

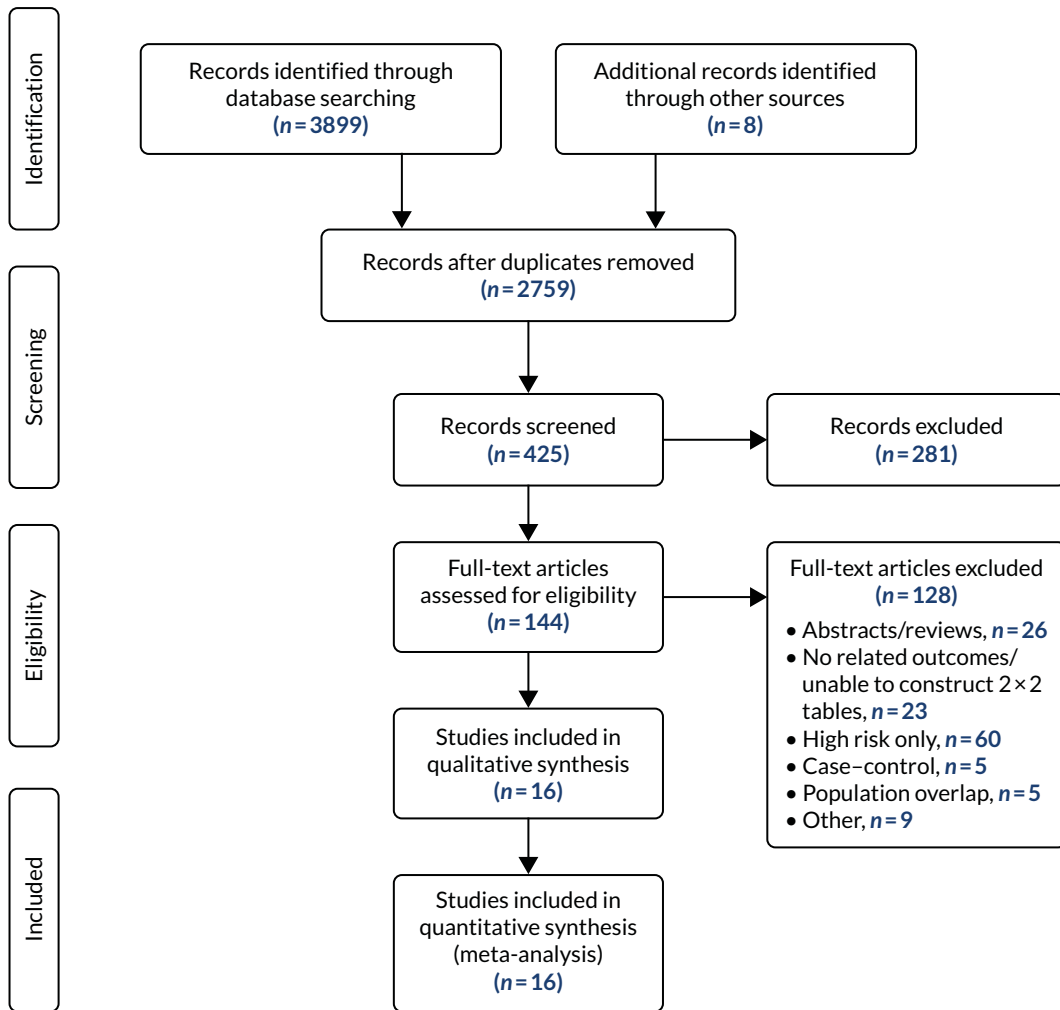


FIGURE 29 Literature search PRISMA flow diagram for the systematic review on CPRs.

	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Akolekar 2015 ⁵⁷	+	+	-	+	+	+	+
Akolekar 2019 ⁴²	+	+	-	+	+	+	+
Bakalis 2015 ⁵⁸	+	+	-	-	+	+	+
Bligh 2018 ⁵⁹	+	+	+	+	+	+	+
Bligh 2018 ⁶⁰	+	+	+	+	+	+	+
Flatley 2019 ⁶¹	+	+	-	+	+	+	+
Khalil 2015 ⁶²	+	+	-	+	+	+	+
Maged 2014 ⁶³	?	+	-	+	?	+	+
Monaghan 2018 ⁶⁴	+	+	-	+	+	+	+
Morales-Roselló 2014 ⁶⁵	+	+	-	+	+	+	+
Prior 2013 ⁶⁶	+	+	+	-	+	+	+
Prior 2015 ⁶⁷	+	+	+	-	+	+	+
Rial-Crestelo 2019 ⁶⁸	+	+	+	-	+	+	+
Sabdía 2015 ⁶⁹	+	+	-	+	+	+	+
Stumpfe 2019 ⁷⁰	+	+	-	-	+	+	+
Twomey 2016 ⁷¹	+	+	-	-	+	+	+

- High
? Unclear
+ Low

FIGURE 30 Risk of bias and applicability concerns using the QUADAS-2 tool for the studies included in the meta-analysis of CPRs.

TABLE 20 Characteristics of studies included in the meta-analysis of CPRs to predict adverse pregnancy outcome

Study (first author and year of publication)	Type of study; setting	Number of fetuses and selection (all singleton, non-anomalous unless otherwise stated)	Index test CPR = MCA PI/umbilical artery PI (unless otherwise stated)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Akolekar 2015 ⁵⁷	Prospective cohort; two NHS hospitals (King's College London and Medway Maritime Hospital), UK	n = 6038	CPR < 5th centile	35 ⁺⁰ to 37 ⁺⁶ weeks	Cord arterial pH of < 7.0, 5-minute Apgar score of < 7, NICU admission	Median 39.9 (IQR 39.0–40.7) weeks	Nulliparous: 49.8%
	Between February 2014 and December 2014	Universal screening	Not blinded	Median 36.1 (IQR 36.0–36.6) weeks			IOL: 20% overall
Akolekar 2019 ⁴²	Prospective cohort; two NHS hospitals (King's College London and Medway Maritime Hospital), UK	n = 47,211	CPR < 10th centile	Between 35 ⁺⁰ and 37 ⁺⁶ weeks	Adverse perinatal outcome (composite of stillbirths, neonatal deaths and HIE grade 2 or 3), perinatal hypoxia (composite of cord arterial pH of < 7.0 and venous < 7.1, 5-minute Apgar score of < 7, NICU admission for > 24 hours), caesarean section for fetal compromise, SGA < 3rd centile	Median gestational age at delivery 40.0 (39.0–40.9) weeks	Nulliparous: 45.4% for those with no adverse outcome, 58.5% for those with adverse outcome
	Between March 2014 and September 2018; significant population overlap with the 2015 Akolekar <i>et al.</i> study ⁵⁷	Universal screening	Not blinded				IOL: not reported
Bakalis 2015 ⁵⁸	Prospective cohort; three NHS hospitals (King's College London, University College London, Medway Maritime Hospital), UK	n = 30,780	CPR < 5th centile	30 ⁺⁰ to 34 ⁺⁶ weeks, mean 32.3 (IQR 32.0–32.9) weeks	Stillbirth, emergency caesarean section for fetal distress, cord arterial pH of < 7.0, cord venous pH of 7.1, 5-minute Apgar score of < 7, NNU admission, NICU admission	Median 40 (IQR 39.0–40.9) weeks	Nulliparous: 50.2%
	Between May 2011 and August 2014; likely to be population overlap with Akolekar <i>et al.</i> studies ^{42,57}	Universal screening	Not blinded				Further analysed in SGA vs. AGA and delivery < 2 weeks from scan vs. > 2 weeks from scan IOL: 14.5% overall

Study (first author and year of publication)	Type of study; setting	Number of fetuses and selection (all singleton, non-anomalous unless otherwise stated)	Index test CPR = MCA PI/umbilical artery PI (unless otherwise stated)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Bligh 2018 ⁵⁹	Prospective cohort; single hospital, Brisbane, QLD, Australia (May 2014–August 2016)	n = 437 Low risk Uncomplicated, term only	CPR < 10th centile Blinded	From 36 ⁺¹ weeks' gestation Within 2 weeks of delivery	Caesarean section for fetal distress. Composite adverse neonatal outcome (cord arterial pH of < 7.10, 5-minute Apgar score of < 7 or NICU admission)	Median 40 weeks (IQR 39.3–40.9 weeks)	Nulliparous: 87.4% IOL: not reported
Bligh 2018 ⁶⁰	Prospective cohort; single hospital, Brisbane, QLD, Australia (May 2014–August 2016)	n = 437 Low risk Uncomplicated, term only	CPR < 10th centile CPR < 5th centile Blinded	From 36 weeks' gestation Within 2 weeks of delivery	SGA < 10th centile SGA < 5th centile	Median 40 weeks (IQR 39.3–40.9 weeks)	Nulliparous: 87.4% IOL: not reported
Flatley 2019 ⁶¹	Retrospective cohort; single hospital, Brisbane, QLD, Australia (2010–15) (likely to be some population overlap with Bligh <i>et al.</i> ^{59,60})	n = 2425 Mixed risk Excluded preterm delivery < 37 weeks' gestation, maternal hypertension and diabetes mellitus	CPR < 10th centile Not blinded	Between 36 and 38 weeks' gestation	Cord arterial pH of < 7.00, 5-minute Apgar score of ≤ 3, NICU admission, perinatal death. Composite of all of the above (SCNO) caesarean section for fetal distress. SGA < 10th centile, SGA < 5th centile	Term only, 54.5% of those with an abnormal CPR delivered < 39 weeks, 36.4% of those with a normal CPR	Nulliparous: 65.4% of those with an abnormal CPR, 48.0% of those with a normal CPR IOL: 46.4% for those with an abnormal CPR, 39.5% for those with a normal CPR
Khalil 2015 ⁶²	Retrospective cohort; one tertiary NHS hospital (St George's), UK (2000–13)	n = 9772 Low risk Term only. For the analysis of operative delivery for fetal distress, the patients who had elective caesarean section were excluded	CPR < 0.6765 MoM Not blinded	Within 2 weeks of delivery Median 40.4 weeks for those admitted to NNU, 40.4 weeks for those not admitted to NNU	NNU admission Operative delivery of fetal distress, (including instrumental delivery and caesarean section)	Median 41.1 weeks for both those admitted and those not admitted to NNU	Nulliparous: 65.2% of those admitted to NNU, 54.6% for those not admitted to NNU IOL: 44.1% for NNU, 39.4% for no NNU

continued

TABLE 20 Characteristics of studies included in the meta-analysis of CPRs to predict adverse pregnancy outcome (continued)

Study (first author and year of publication)	Type of study; setting	Number of fetuses and selection (all singleton, non-anomalous unless otherwise stated)	Index test CPR = MCA PI/umbilical artery PI (unless otherwise stated)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Maged 2014 ⁶³	Prospective cohort; single hospital, Cairo, Egypt	n = 100 Low risk Included those delivered between 40 and 42 weeks' gestation Excluded PPROM, APH, patients in labour and maternal HTN/DM	CPR < 1.05 Not blinded	37.8 weeks' gestation for those with adverse outcome, 39.5 weeks' gestation for those with normal outcome	Caesarean section for fetal distress Composite adverse pregnancy outcome defined as one or more of caesarean section for fetal distress, 5-minute Apgar score of < 7, MAS, NICU admission	283.1 days for those with adverse outcome, 281.7 days for those with normal outcome	Nulliparous: not reported IOL: not reported
Monaghan 2018 ⁶⁴	Retrospective cohort; single NHS hospital (St George's), UK January 2008–June 2016 (likely to be population overlap with Khalil <i>et al.</i> ⁶²)	n = 7013 Mixed risk (had ultrasound scan based on NHS indications) Only included those delivered after 36 weeks' gestation	CPR < 10th centile CPR < 5th centile Not blinded	36.4 weeks for all live births, 37 weeks for perinatal deaths	Perinatal death	Median: 40.1 weeks' gestation for all live births, 39 weeks' gestation for perinatal deaths	Nulliparous: not reported IOL: not reported
Morales-Roselló 2014 ⁶⁵	Retrospective cohort; single NHS hospital (St George's), UK, 2002–12 (likely to be population overlap with Khalil <i>et al.</i> ⁶² and Monaghan <i>et al.</i> ⁶⁴)	n = 11,576 Mixed risk Term only with ultrasound scan within 14 days of delivery	CPR < 0.6765 MoM Not blinded	Mean 40.1 ± 1.5 weeks	SGA < 10th centile	Mean 40.8 ± 1.3 weeks	Nulliparous: not reported IOL: not reported

Study (first author and year of publication)	Type of study; setting	Number of fetuses and selection (all singleton, non-anomalous unless otherwise stated)	Index test CPR = MCA PI/umbilical artery PI (unless otherwise stated)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Prior 2013 ⁶⁶	Prospective cohort; single NHS hospital (Queen Charlotte's and Chelsea), UK. (March 2011–March 2014)	n = 400 Low risk Term only. Recruited before active labour. Excluded pre-eclampsia, FGR, intrauterine infection	CPR < 10th centile Blinded	Mean 40 weeks' gestation + 2 days (range 37 ⁺⁰ –42 ⁺¹ weeks)	Caesarean section for fetal compromise, 5-minute Apgar score of < 7, cord arterial pH of < 7.20, NNU admission	Within 72 hours from scan	Nulliparous: 65.5% IOL: not reported
Prior 2015 ⁶⁷	Prospective cohort; single tertiary NHS hospital (Chelsea), UK. (likely to be population overlap with the Prior <i>et al.</i> study ⁶⁶)	n = 775 Low risk Term only. Recruited before active labour or IOL (for post dates or social). Excluded SGA/FGR, PIH/pre-eclampsia, PPRM	CPR < 0.6765 MoM Blinded	Median 41 weeks' gestation (range 37–42 weeks)	Caesarean section for fetal distress, 5-minute Apgar score of < 7, cord arterial pH of < 7.20, NNU admission	Within 72 hours from scan	Nulliparous: 80.8% IOL: not reported
Rial-Crestelo 2019 ⁶⁸	Prospective cohort; single hospital, Barcelona, Spain. January 2013–December 2016	n = 1030 Universal screening	CPR < 10th centile Doppler blinded for those with EFW > 10th centile	Between 32 ⁺⁰ and 34 ⁺⁶ weeks, mean 33 weeks	SGA < 10th centile	Mean 40 weeks' gestation	Nulliparous: 70% of those born SGA, 54% of those not born SGA IOL: not reported
Sabdia 2015 ⁶⁹	Retrospective cohort; single hospital, Brisbane, QLD, Australia (June 1998–November 2013)	n = 1381 Mixed risk Included cephalic with umbilical artery PI < 95th centile	CPR < 10th centile (1.20) Not blinded	Between 35 and 37 weeks' gestation	Operative delivery for fetal distress (caesarean section or instrumental), 5-minute Apgar score of < 7, NICU admission	Median gestational age 36 weeks for those with an abnormal CPR, 38 weeks for those with a normal CPR	Nulliparous: 53.9% of those with an abnormal CPR, 40.4% of those with a normal CPR IOL: not reported

continued

TABLE 20 Characteristics of studies included in the meta-analysis of CPRs to predict adverse pregnancy outcome (continued)

Study (first author and year of publication)	Type of study; setting	Number of fetuses and selection (all singleton, non-anomalous unless otherwise stated)	Index test CPR = MCA PI/umbilical artery PI (unless otherwise stated)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Stumpfe 2019 ⁷⁰	Retrospective cohort; single tertiary centre, Germany (January 2016–April 2017)	<i>n</i> = 1008 Low risk Term only, excluded those in labour, elective caesarean section, EFW < 10th centile	CPR < 0.6765 MoM Not blinded	Term, within 72 hours of delivery	Caesarean section for fetal distress, 5-minute Apgar score of < 7, cord arterial pH of < 7.10	Term (not further specified)	Nulliparous: not specified IOL: 42.4% overall
Twomey 2016 ⁷¹	Retrospective cohort; single hospital, Brisbane, QLD, Australia (January 2007–December 2013) (population overlap with Sabdia <i>et al.</i> ⁶⁹)	<i>n</i> = 1224 Mixed risk Excluded women who had elective caesarean section	CPR < 1 Not blinded	30–34 weeks, median 32.1 weeks	Caesarean section for fetal compromise, cord arterial pH of < 7.0, 5-minute Apgar score of ≤ 3, NNU admission, SGA < 10th centile, SGA < 5th centile	Mean gestational age 32 weeks for those with a CPR < 1, 37 weeks for those with a CPR > 1	Nulliparous: 43.2% IOL: not reported

HIE, hypoxic–ischaemic encephalopathy; IQR, interquartile range; MoM, multiples of median; USS, ultrasound scan.

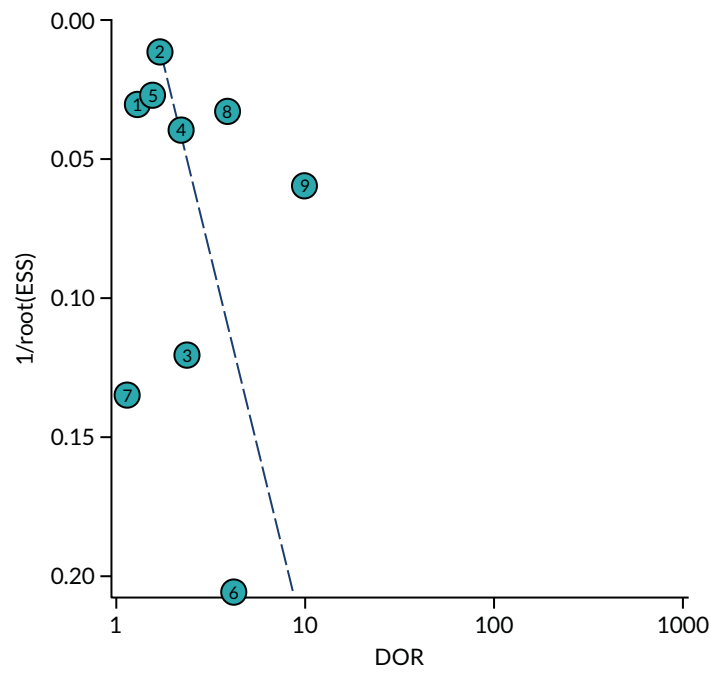


FIGURE 31 Deeks' funnel plot for publication bias for CPRs for the prediction of neonatal unit admission. Deeks' funnel plot asymmetry test: $p = 0.28$. ESS, effective sample size.

Appendix 3 Supporting data for the systematic review of the diagnostic effectiveness of universal ultrasonic screening using severe oligohydramnios in the prediction of adverse perinatal outcome

MEDLINE and EMBASE

Date range searched: 1 January 2011 to 5 June 2019.

Search strategy

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

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

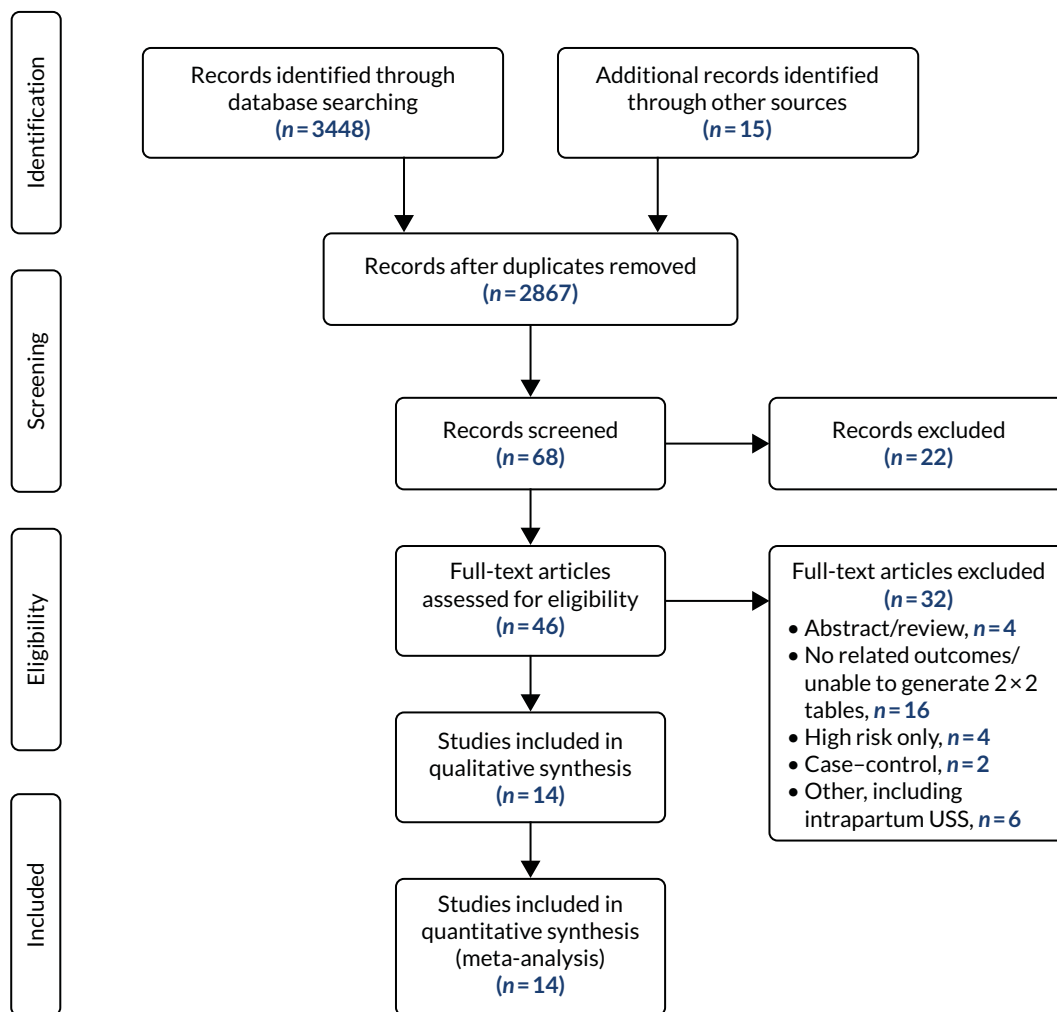


FIGURE 32 The PRISMA flow diagram for the systematic review of severe oligohydramnios. USS, ultrasound scan.

	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Ashwal 2014 ⁷⁴	+	+	-	+	+	+	+
Ghosh 2002 ⁷⁵	+	+	-	-	+	+	+
Hassan 2005 ⁷⁶	+	+	-	+	-	+	+
Hsieh 1998 ⁷⁷	+	+	-	?	+	+	+
Locatelli 2004 ⁷⁸	+	+	-	+	+	+	+
Megha 2014 ⁷⁹	?	+	-	+	+	+	+
Melamed 2011 ⁸⁰	+	+	-	+	+	+	+
Morris 2003 ⁸¹	+	+	+	+	+	+	+
Myles 2002 ⁸²	+	+	-	?	+	+	+
Naveiro-Fuentes 2016 ⁸³	+	+	-	+	+	+	+
Quiñones 2012 ⁸⁴	+	+	+	+	+	+	+
Rainford 2001 ⁸⁵	+	+	-	-	+	+	+
Shanks 2011 ⁸⁶	?	+	-	+	+	+	+
Zhang 2004 ⁸⁷	+	+	-	+	+	+	+

- High
? Unclear
+ Low

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

TABLE 21 Characteristics of studies included in the meta-analysis of severe oligohydramnios

Study (first author and year of publication)	Type of study; setting	Number of fetuses and selection (all singleton, non-anomalous unless otherwise stated)	Index test	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Ashwal 2014 ⁷⁴	Retrospective cohort; single university hospital, Israel	<i>n</i> = 23,267 Low risk	AFI < 5 cm	Within 1 week from delivery	Caesarean section for fetal distress, operative vaginal delivery for fetal distress, 5-minute Apgar score of < 7, umbilical artery pH of < 7.10, NICU admission, need for intubation, MAS or HIE. Also stillbirth, neonatal death, IVH, meconium amniotic fluid (not MAS)	39 ⁺⁸ ± 1.1 weeks for isolated oligohydramnios; 39.3 ± 1.1 weeks for normal AFI	Nulliparous: <i>n</i> = 442 (44.8%) for isolated oligohydramnios, <i>n</i> = 6848 (30.7%) for normal AFI IOL: <i>n</i> = 273 (27.7%) for oligohydramnios, <i>n</i> = 824 (3.7%) for normal
Ghosh 2002 ⁷⁵	Prospective cohort; single hospital, Sweden	<i>n</i> = 333 Low risk	AFI < 5 cm Not blinded	In early labour or before IOL	Operative delivery for fetal distress, caesarean section for fetal distress, 5-minute Apgar score of < 7, cord arterial pH of < 7.10, NICU admission	Mean gestational age 283 days for those with AFI < 5 cm, 280 days for those with AFI > 5 cm	Nulliparous: 26/49 of those with AFI < 5 cm, 134 for those with AFI > 5 cm
Hassan 2005 ⁷⁶	Cross-sectional; single hospital, Pakistan	<i>n</i> = 260 Low risk	AFI < 6 cm	After 41 ⁺⁰ weeks	Neonatal death, caesarean section, meconium-stained amniotic fluid	After 41 ⁺⁰ weeks	Nulliparous: 34% of those with low AFI, 19.7% of those with normal AFI IOL: not specified
Hsieh 1998 ⁷⁷	Retrospective cohort; single hospital, Taiwan (Province of China)	<i>n</i> = 27,506 Universal Excluded those with AFI > 24 cm, PPROM	AFI < 5 cm Not blinded	Not specified	Stillbirth, SGA < 10th centile, 5-minute Apgar score of < 7, NICU admission, neonatal death	Not specified	Nulliparous: not specified IOL: not specified

Study (first author and year of publication)	Type of study; setting	Number of fetuses and selection (all singleton, non-anomalous unless otherwise stated)	Index test	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Locatelli 2004 ⁷⁸	Prospective cohort; single hospital, Italy	<i>n</i> = 3049 Universal Routine scan at 40 weeks' gestation Excluded those with PPRM and those with other indications for ultrasound scan	AFI < 5 cm Not blinded	40 weeks' gestation	Meconium-stained amniotic fluid, caesarean section for fetal distress, SGA < 10th centile, Apgar score of < 7, cord arterial pH of < 7.0	40 ⁺⁰ -41 ⁺⁶ weeks' gestation	Nulliparous: 72% for those with low AFI, 58% for those with normal AFI IOL: 83% for those with low AFI, 25% for those with normal AFI
Megha 2013 ⁷⁹	Prospective cohort; single centre, India	<i>n</i> = 200 Mixed Selection not specified	AFI < 5 cm Blinded	34-41 weeks' gestation Within 7 days of delivery	Caesarean section for fetal distress, meconium-stained fluid, 5-minute Apgar score of < 7, cord arterial pH of < 7.10. Admission to NICU for > 48 hours	Not specified. 56% of those with low AFI delivered < 37 weeks' gestation vs. 34.3% of those with normal AFI	Nulliparous: 68% of those with low AFI, 58.9% of those with normal AFI IOL: 72% of those with low AFI, 51% of those with normal AFI
Melamed 2011 ⁸⁰	Matched cohort (3: 1); single hospital, Israel	<i>n</i> = 432 Low risk Excluded pregnancies with pre-eclampsia/DM/GDM, EFW < 10th centile, abnormal umbilical artery Doppler, and PROM	AFI < 5 cm Not blinded	Gestational age at initial ultrasound scan: 33.9 weeks for low AFI, 33.9 weeks for normal AFI Gestational age at last scan not reported	Caesarean section for fetal distress, meconium-stained fluid, preterm delivery (< 37 weeks' gestation), admission to NICU	37.3 ± 1.6 weeks for cases, 39.1 ± 1.8 weeks for controls	Nulliparous: 62 (57.4%) of cases, 186 (57.4%) of controls IOL: 54 (50%) of cases, 31 (9.6%) of controls

continued

TABLE 21 Characteristics of studies included in the meta-analysis of severe oligohydramnios (continued)

Study (first author and year of publication)	Type of study; setting	Number of fetuses and selection (all singleton, non-anomalous unless otherwise stated)	Index test	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Morris 2003 ⁸¹	Prospective cohort; single hospital, Oxford, UK	<i>n</i> = 1584 Low risk Term only (> 40 weeks' gestation). Excluded non-vertex and those with clinically required ultrasound	AFI < 5 cm SDP < 2 cm Not blinded	At or after 40 weeks' gestation (59% at 40 weeks)	Caesarean section for fetal distress, NICU admission, 5-minute Apgar score of < 7	At or after 40 weeks' gestation (615 at 41 weeks' gestation)	Nulliparous: 778 (49.1%) IOL: 643 (40.6%)
Myles 2002 ⁸²	Prospective cohort; single hospital, FL, USA	<i>n</i> = 266 Low risk Term only. Excluded non-vertex, SROM, polyhydramnios, and any pregnancies with fetal or maternal complications	AFI < 5 cm SDP < 2.5 cm Not blinded	Between 37 ⁺⁰ and 41 ⁺⁶ weeks (not specified)	Caesarean section for fetal distress, NICU admission, meconium-stained amniotic fluid	Not specified	Nulliparous: not specified IOL: not specified
Naveiro-Fuentes 2016 ⁸³	Retrospective cohort; single hospital, Spain	<i>n</i> = 27,708 Low risk Term only. Routine antenatal scan at 39 weeks' gestation. Excluded pregnancies with maternal or fetal pathology including suspected IUGR	AFI < 5 cm Not blinded	39 weeks' gestation	Caesarean section for fetal distress, instrumental delivery for fetal distress, meconium-stained fluid, SGA (< 10th centile), 5-minute Apgar score of < 7, admission to NICU, umbilical artery pH of < 7.10	279 ± 7.3 days for those with oligohydramnios, 278.2 ± 7.5 days for normal	Nulliparous: 65.1% of those with low AFI IOL: not reported

Study (first author and year of publication)	Type of study; setting	Number of fetuses and selection (all singleton, non-anomalous unless otherwise stated)	Index test	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comments
Quiñones 2012 ⁸⁴	Prospective cohort; two centres, PA, USA	<i>n</i> = 308 Low risk Between 37 and 40 weeks' gestation, excluded pregnancies with maternal or obstetric complications (including suspected FGR)	AFI < 5 cm AFI < 8 cm AFI < 10 cm SDP < 2cm	37–40 weeks' gestation (mean 38.1 ± 0.9 weeks' gestation)	Fetal vulnerability index, which is defined as one or more of the following: 5-minute Apgar score of < 3, umbilical cord pH of < 7.0, intrapartum fetal death, neonatal seizures, intubation in the absence of meconium, or NICU admission for > 24 hours	Mean gestational age 39.9 ± 0.8 weeks	Nulliparous: 50%
Rainford 2001 ⁸⁵	Retrospective cohort; single hospital, USA	<i>n</i> = 232 Low risk Term only. Excluded those with any maternal or fetal complications	AFI < 5 cm Not blinded	Within 4 days of delivery	Operative delivery for fetal distress, NICU admission, 5-minute Apgar score of < 7, meconium-stained amniotic fluid	Mean gestational age 40.1 weeks for those with oligohydramnios, 40.9 weeks for normal AFI	Nulliparous: 17% for low AFI, 20% for normal AFI IOL: 98% of those with low AFI, 51% of those with normal AFI
Shanks 2011 ⁸⁶	Retrospective cohort; single centre, USA	<i>n</i> = 17,877 Mixed risk Selection criteria not specified	AFI < 5 cm AFI < 5th centile Not blinded	Mean 34.38 ± 3.04 weeks' gestation	NICU admission	Mean 38.27 ± 2.86 weeks' gestation	Nulliparous: <i>n</i> = 7069 (39.5%)
Zhang 2004 ⁸⁷	Clinical trial (ultrasound scan screening vs. no screening). For this study data used by the screening group	<i>n</i> = 6657 in the low-risk group. All women had two research scans at 15–22 and 31–35 weeks' gestation. Excluded multiple pregnancies and those with any maternal or fetal conditions	AFI < 5 cm Not blinded	31–35 weeks' gestation	Caesarean section for fetal distress, 5-minute Apgar score of < 7, NICU admission, perinatal mortality	Mean gestational age 39.6 weeks for those with oligohydramnios, 39.8 weeks' gestation for those with normal AFI	Nulliparous: 53% of those with oligohydramnios, 45% of normal AFI IOL: not specified

DM, diabetes mellitus; GDM, gestational diabetes mellitus; HIE, hypoxic-ischaemic encephalopathy; IUGR, intrauterine growth restriction; IVH, intraventricular haemorrhage; MAS, meconium aspiration syndrome; PPROM, preterm premature rupture of membranes; SROM, spontaneous rupture of membranes; USS, ultrasound scan.

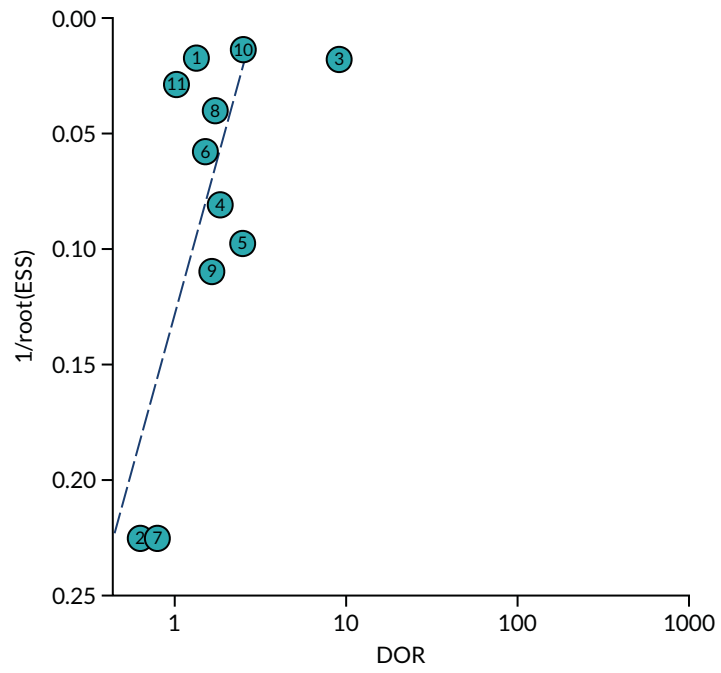


FIGURE 34 Deeks' funnel plot for publication bias for severe oligohydramnios for the prediction of neonatal unit admission. Deeks' funnel plot asymmetry test: $p = 0.54$. ESS, effective sample size.

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

MEDLINE and EMBASE

Date range searched: inception to 18 June 2019.

Search strategy

1. exp Pregnant Women/
2. exp pregnancy/
3. pregnan\$.mp.
4. exp oligohydramnios/
5. oligohydramnio\$.mp.
6. exp Amniotic Fluid/
7. amniotic fluid index.mp.
8. AFI.mp.
9. liquor volume.mp.
10. ow.mp.
11. borderline.mp.
12. decreased.mp.
13. perinatal.mp.
14. peripartum.mp.
15. fetal.mp.
16. 1 or 2 or 3
17. 4 or 5 or 6 or 7 or 8 or 9
18. 13 or 14 or 15
19. 16 and 17 and 18
20. 10 or 11 or 12
21. 19 and 20.

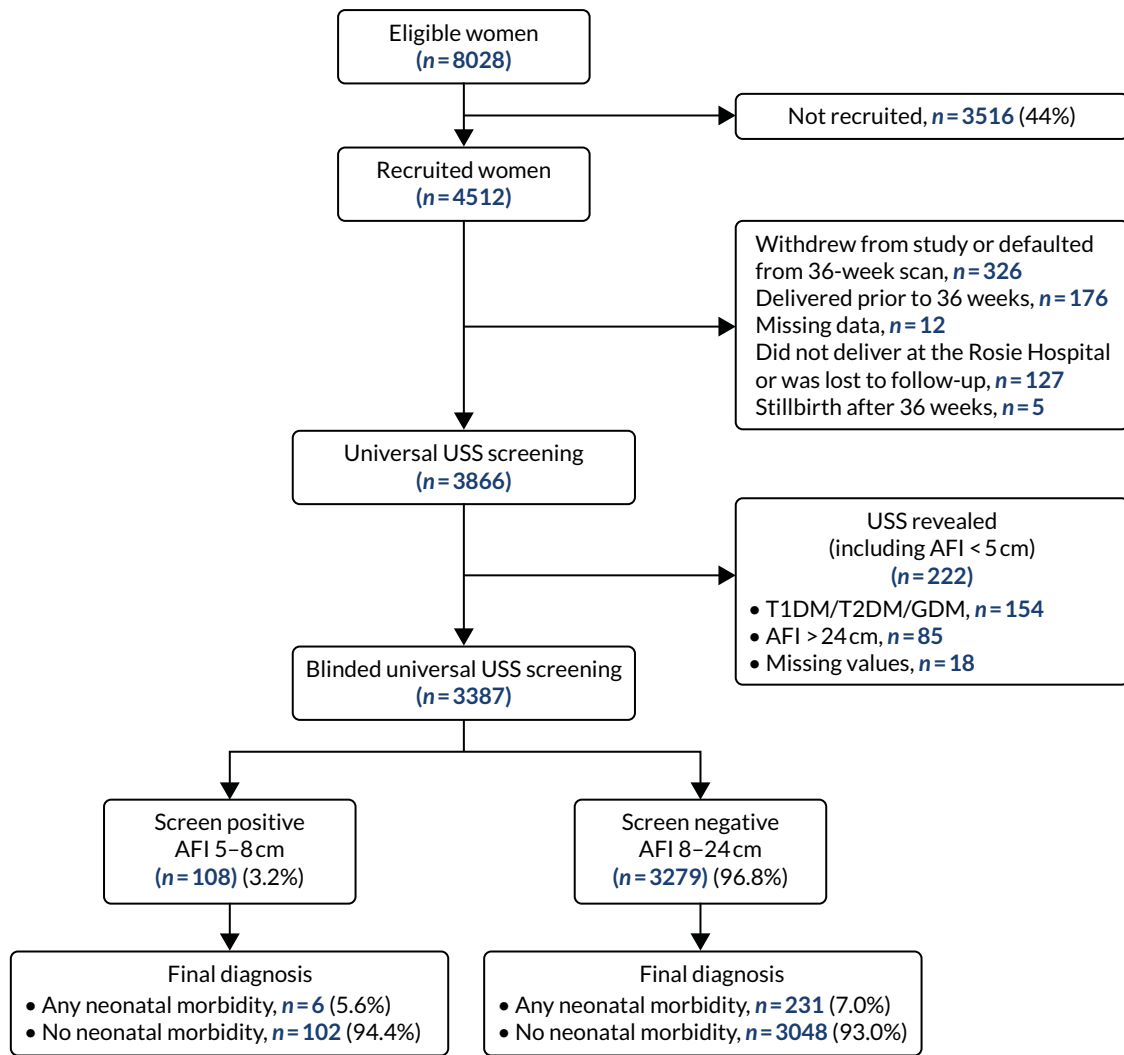


FIGURE 35 The POP study inclusion flow chart.

TABLE 22 Patient characteristics and birth outcomes of POP study

Characteristic	Borderline AFI 5–8 cm (n = 108)	Normal AFI 8–24 cm (N = 3279)	p-value	Overall baseline characteristics (N = 3387)
Maternal characteristic				
Age (years), median (IQR)	30.1 (26.7–33.2)	30.3 (26.2–33.7)	0.60	30.1 (26.7–33.2)
Deprivation quartile, n (%)				
1 (lowest)	29 (26.9)	808 (24.6)	0.53	837 (24.7)
2	28 (25.9)	769 (23.5)		797 (23.5)
3	23 (21.3)	776 (23.7)		799 (23.6)
4 (highest)	25 (23.2)	783 (23.9)		808 (23.9)
Missing	3 (2.8)	143 (4.4)		146 (4.3)
White ethnicity, n (%)	96 (88.9)	3052 (93.1)	0.16	3148 (92.9)
Missing	3 (2.8)	54 (1.7)		57 (1.7)
Married, n (%)	81 (75.0)	2222 (67.8)	0.11	2303 (68.0)
Smoker	3 (2.8)	164 (5.0)	0.29	167 (4.9)
Any alcohol consumption	1 (0.9)	154 (4.7)	0.06	155 (4.6)
Missing	0 (0.0)	1(0.0)		1 (0.0)
BMI (kg/m ²), median (IQR)	23.4 (21.6–26.5)	23.9 (21.8–27.1)	0.19	23.9 (21.8–27.0)
One or more previous miscarriage(s), n (%)	8 (7.4)	327 (10.0)	0.38	335 (9.9)
Chronic hypertension, n (%)	4 (3.7)	164 (5.0)	0.54	
Pre-eclampsia, n (%)	9 (8.3)	201 (6.1)	0.35	210 (6.2)
Missing	0 (0)	2 (0.1)		2 (0.1)
Birth outcome				
Birthweight (g), median (IQR)	3260 (3005–3520)	3460 (3150–3770)	< 0.001	3450 (3150–3760)
Gestational age (weeks), median (IQR)	40.0 (38.8–40.9)	40.4 (39.6–41.3)	< 0.001	40.4 (39.6–41.3)
IOL, n (%)	41 (38.0)	1016 (31.0)	0.12	1057 (31.2)
Mode of delivery, n (%)				
Spontaneous vaginal	70 (64.8)	1685 (51.4)	0.04	1755 (51.8)
Assisted vaginal	19 (17.6)	832 (25.4)		851 (25.1)
Intrapartum caesarean	13 (12.0)	596 (18.2)		609 (18.0)
Pre-labour caesarean	6 (5.6)	157 (4.8)		163 (4.8)
Missing	0 (0.0)	9 (0.3)		9 (0.3)
IQR, interquartile range.				

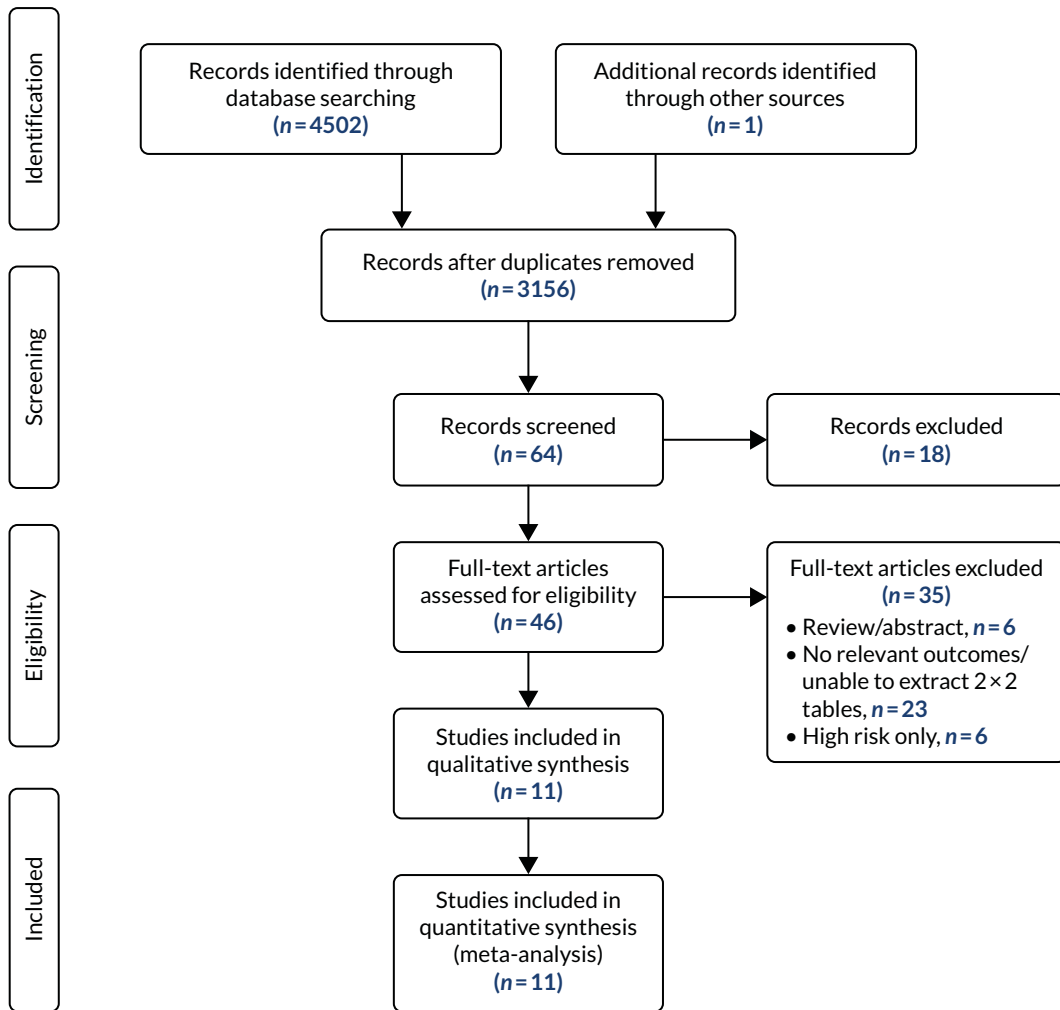


FIGURE 36 The PRISMA flow diagram for the systematic review of borderline oligohydramnios.

	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Asgharnia 2013 ⁸⁹	?	+	-	?	+	+	+
Banks 1999 ⁹⁰	?	+	-	+	+	+	+
Choi 2016 ⁹¹	+	+	-	+	+	+	+
Gumus 2007 ⁹²	+	+	-	?	+	+	+
Jamal 2016 ⁹³	-	+	-	+	-	+	+
Kwon 2006 ⁹⁴	+	+	-	?	+	+	+
Moraitis 2021 ³⁵	+	+	+	+	+	+	+
Petrozella 2011 ⁹⁵	+	+	-	+	+	+	+
Rutherford 1987 ⁹⁶	+	+	-	?	+	+	+
Sahin 2018 ⁹⁷	+	+	-	+	+	+	+
Wood 2014 ⁹⁸	+	+	-	?	+	+	+

- High
? Unclear
+ Low

FIGURE 37 Risk-of-bias and applicability concerns for included studies in systematic review of borderline oligohydramnios using the QUADAS-2 tool.

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

Study (first author and year of publication)	Type of study; setting	Population and selection (singletons only unless otherwise specified)	Index test	Gestational age at ultrasound	Reference standard	Gestational age at delivery (mean unless otherwise specified)	Other comments
Asgharnia 2013 ⁸⁹	Retrospective cohort; single hospital, Islamic Republic of Iran	<i>n</i> = 235 Mixed risk Pregnancies > 28 weeks. Excluded PPRM, uterine anomalies, vaginal bleeding	5 < AFI < 10 cm Not blinded	> 28 weeks' gestation (mean gestational age not reported)	RDS, 5-minute Apgar score of < 7, NICU, IUGR, SGA < 10th centile	Mean gestational age not reported Preterm: BAFI 40.4% Normal AFI 14.9%	Nulliparous: BAFI 68.1%, normal AFI 58.2% IOL: BAFI 22.3%, normal AFI 10.6%
Banks, 1999 ⁹⁰	Retrospective cohort; single hospital, USA	<i>n</i> = 214 Mixed risk Pregnancies with antepartum testing within 1 week of delivery	5 cm < AFI < 10 cm Not blinded	Not reported	Intrapartum fetal distress, meconium-stained amniotic fluid, SGA < 10th centile	Not reported	Nulliparous: not reported IOL: not reported
Choi 2016 ⁹¹	Retrospective cohort; single hospital, the Republic of Korea	<i>n</i> = 721 Low risk Uncomplicated, term pregnancies only Excluded SROM, elective caesarean section, breech presentation, pre-eclampsia, and other maternal disease	5.1 ≤ AFI ≤ 8.0 cm	Within 1 week of delivery	Meconium-stained amniotic fluid, caesarean section for fetal distress, 5-minute Apgar score of < 7, NICU admission, SGA < 10th centile	BAFI: 39.2 weeks Normal AFI: 39.4 weeks	Nulliparous: BAFI 66.1%, normal AFI 57.3% IOL: BAFI 60.7%, normal AFI 27.4%
Gumus, 2007 ⁹²	Retrospective cohort; single hospital, Turkey	<i>n</i> = 367 Mixed risk Excluded PROM, uterine anomalies, vaginal bleeding	5 cm < AFI < 10 cm	Not reported	Intrapartum fetal distress, meconium-stained amniotic fluid, SGA < 10th centile), NICU admission, RDS	BAFI 37.7 weeks for normal AFI 38.3 weeks Preterm: BAFI 18.9%, normal AFI 9.7%	IOL: BAFI 73.3% Normal AFI 54.5%

Study (first author and year of publication)	Type of study; setting	Population and selection (singletons only unless otherwise specified)	Index test	Gestational age at ultrasound	Reference standard	Gestational age at delivery (mean unless otherwise specified)	Other comments
Jamal 2016 ⁹³	Matched cohort (matched 1 : 1); single hospital, Islamic Republic of Iran	n = 128 Mixed risk Term only. Excluded PPROM, anomalies, maternal medical diseases, contraindications for vaginal delivery	5.1 ≤ AFI ≤ 8.0	37–40 weeks' gestation Within 1 week of delivery	Meconium-stained amniotic fluid, 5-minute Apgar score of < 7, umbilical artery pH of < 7.0, NICU admission, SGA < 10th centile	BAFI (median): 37 ⁺⁵ weeks Normal AFI: 38 ⁺⁶ weeks	Nulliparous: not reported IOL: not reported
Kwon 2006 ⁹⁴	Retrospective cohort; single hospital, the Republic of Korea	n = 3740 Mixed risk Excluded fetal malformations, SROM pre-eclampsia, chromosomal anomalies, AFI > 25 cm	5.1 ≤ AFI ≤ 8.0	Within 2 weeks of delivery	Perinatal death, NICU admission, caesarean section for fetal distress, 5-minute Apgar score of < 7, SGA < 10th centile	BAFI: 36.3 weeks' gestation Normal AFI: 38.0 weeks' gestation	Nulliparous: not reported IOL: not reported
The POP Study ^a	Prospective cohort; single centre; Cambridge, UK	n = 3387 Nulliparous only Universal screening	5 cm < AFI < 8 cm Blinded	36 weeks' gestation	NICU admission, metabolic acidosis, 5-minute Apgar score of < 7, composite morbidity (all above), composite severe morbidity		Nulliparous only
Petrozella, 2011 ⁹⁵	Retrospective cohort; regional hospitals, USA	n = 27,601 Mixed risk Those who received USS between 24 and 34 weeks' gestation Excluded AFI > 24 cm, SROM	5 cm < AFI < 8 cm	24 ⁺⁰ to 33 ⁺⁶ weeks' gestation Mean gestational age 29.2 weeks	Caesarean section for fetal distress, SGA < 10th centile, SGA < 3rd centile, neonatal death	BAFI 37.1 weeks' gestation Normal AFI 39.2 weeks' gestation Preterm: BAFI 37%, normal AFI 8%	Nulliparous: not reported IOL: not reported

continued

TABLE 23 Characteristics of studies included in the meta-analysis of borderline oligohydramnios (continued)

Study (first author and year of publication)	Type of study; setting	Population and selection (singletons only unless otherwise specified)	Index test	Gestational age at ultrasound	Reference standard	Gestational age at delivery (mean unless otherwise specified)	Other comments
Rutherford, 1987 ⁹⁶	Retrospective cohort; single hospital, USA	<i>n</i> = 286 Mixed risk Those who had antepartum surveillance Excluded PPROM	5 cm < AFI < 8 cm	Not reported	Meconium, caesarean section for fetal distress, 5-minute Apgar score of < 7	Not reported	Nulliparous: not reported IOL: not reported
Sahin, 2018 ⁹⁷	Prospective (matched 1 : 3); single hospital, Turkey	<i>n</i> = 430 Low risk Excluded maternal disease, IUGR chromosomal/fetal abnormalities, SROM, abnormal Doppler	5 cm < AFI ≤ 8 cm	Between 34 ⁺⁰ and 36 ⁺⁶ weeks' gestation Mean 35,4 weeks' gestation	5-minute Apgar score of < 7, caesarean section for fetal distress, RDS, meconium-stained amniotic fluid, meconium aspiration syndrome, NICU, neonatal death	BAFI: 37.5 weeks Normal AFI: 38.6 weeks Preterm: BAFI 15.9%, normal AFI 8.4%	Nulliparous: not reported IOL: BAFI 34.6%, normal AFI 23.8%
Wood 2014 ⁹⁸	Retrospective cohort (matched 1 : 3); two hospitals, USA	<i>n</i> = 739 Low risk Exclusion criteria: AFI ≤ 5 cm, PPROM, pre-eclampsia	5 cm < AFI ≤ 10 cm	Not reported	Caesarean section for fetal distress, SGA, meconium-stained amniotic fluid, 5-minute Apgar score of < 7, NICU admission, preterm delivery	BAFI: 38.3 weeks Normal AFI: 38.9 weeks	Nulliparous: not reported IOL: not reported

IUGR, intrauterine growth restriction; PPROM, preterm premature rupture of membranes; SROM, spontaneous rupture of membranes; USS, ultrasound scan.

a Alexandros A Moraitis, Ilianna Armata, Ulla Sovio, Peter Brocklehurst, Alexander EP Heazell, Jim G Thornton, Stephen C Robson, Aris Papageorghiou and Gordon CS Smith, University of Cambridge, 2021.

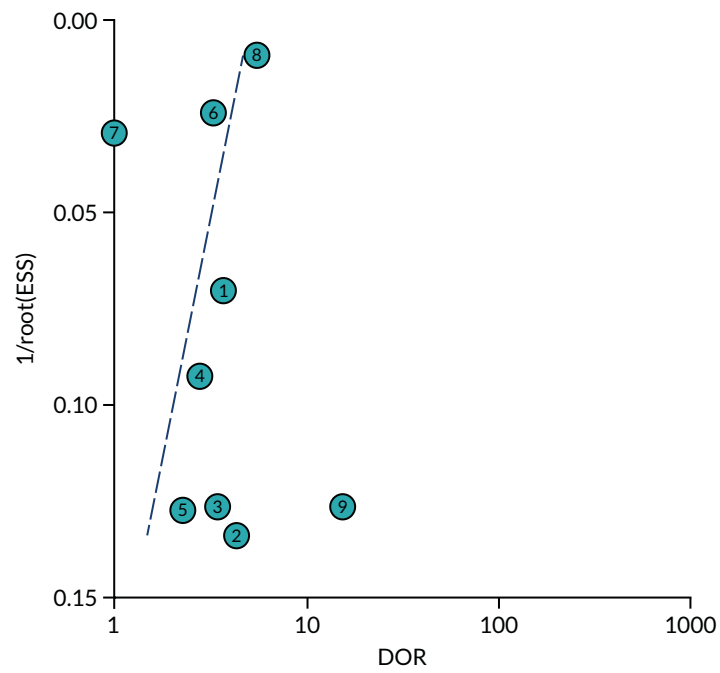


FIGURE 38 Deeks' funnel plot for publication bias for borderline oligohydramnios for the prediction of SGA < 10th centile. Deeks' funnel plot asymmetry test: $p = 0.33$. ESS, effective sample size.

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

MEDLINE and EMBASE

Date range searched: inception to 22 October 2018.

Search strategy

1. exp fetus echography/
2. ultrasonography, prenatal.mp.
3. exp ultrasound/
4. ultraso*.mp.
5. sonograph*.mp.
6. exp biometry/
7. USS.mp.
8. estimated fetal weight.mp.
9. EFW.mp.
10. abdominal circumference.mp.
11. AC.mp.
12. exp macrosomia/
13. macrosomi*.mp.
14. exp fetus weight/
15. fetal weight.mp.
16. exp birth weight/
17. birthweight.mp.
18. large for gestational age.mp.
19. LGA.mp.
20. large fetus.mp.
21. exp brachial plexus injury/or brachial plexus injury.mp.
22. exp shoulder dystocia/or shoulder dystocia.mp.
23. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
24. 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22
25. 23 and 24
26. exp pregnancy/
27. 25 and 26.

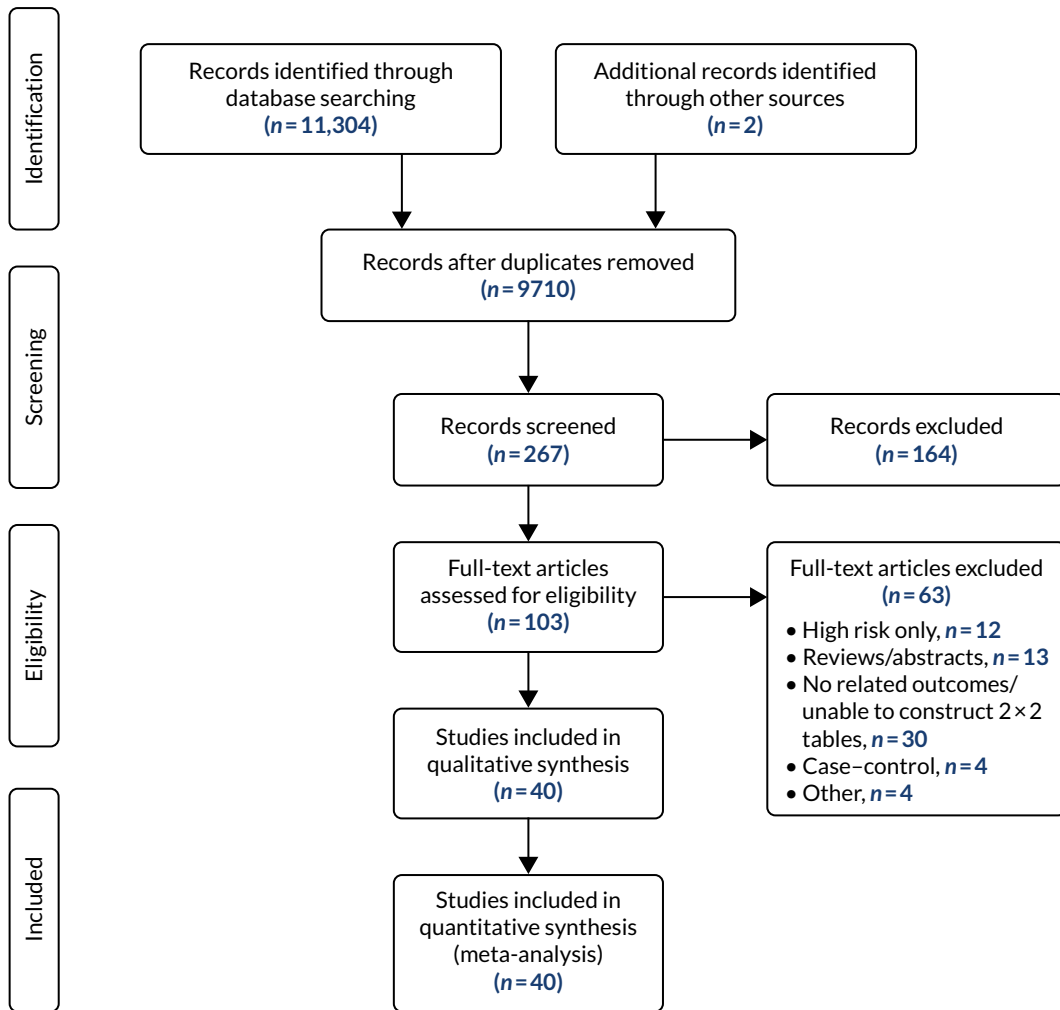


FIGURE 39 The PRISMA flow diagram for the systematic review of macrosomia.

	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Aviram 2017 ¹⁰²	+	+	-	+	+	+	+
Balsyte 2009 ¹⁰³	+	+	-	+	+	+	+
Benecerraf 1988 ¹⁰⁴	+	+	-	?	+	+	+
Ben-Haroush 2007 ¹⁰⁵	+	+	-	-	+	+	+
Ben-Haroush 2008 ¹⁰⁶	+	+	-	-	+	+	+
Benson 1991 ¹⁰⁷	?	+	-	?	+	+	+
Burkhardt 2014 ¹⁰⁸	+	+	-	+	+	+	+
Chauhan 2006 ¹⁰⁹	+	+	-	+	+	+	+
Chervenak 1989 ¹¹⁰	+	+	-	+	-	+	+
Cohen 2010 ¹¹¹	-	+	-	-	+	+	+
Crimmins 2018 ¹¹²	+	+	-	+	+	+	+
Cromi 2007 ¹¹³	+	+	-	+	+	+	+
De Reu 2008 ¹¹⁴	+	+	-	-	+	+	+
Freire 2010 ¹¹⁵	+	+	-	+	+	+	+
Galvin 2017 ¹¹⁶	+	+	+	+	+	+	+
Gilby 2000 ¹¹⁷	+	+	-	+	+	+	+
Hasenoehrl 2009 ¹¹⁸	+	+	-	+	+	+	+
Hendrix 2000 ¹¹⁹	+	+	-	+	+	+	+
Henrichs 2003 ¹²⁰	+	+	-	+	+	+	+
Humpries 2002 ¹²¹	+	+	-	+	+	+	+
Kayem 2009 ¹²²	+	+	-	+	+	+	+
Kehl 2011 ¹²³	+	+	-	-	+	+	+
Levine 1992 ¹²⁴	+	+	-	+	+	+	+
Melamed 2011 ¹²⁵	+	+	-	-	+	+	+
Miller 1986 ¹²⁶	+	+	-	+	+	+	+
Miller 1988 ¹²⁷	+	+	-	+	+	+	+
Nahum 2003 ¹²⁸	+	+	-	+	+	+	+
Nahum 2007 ¹²⁹	+	+	-	+	+	+	+
Nicod 2012 ¹³⁰	+	+	-	+	+	+	+
O'Reilly-Green 1997 ¹³¹	+	+	-	+	+	+	+
Pates 2008 ¹³²	+	+	-	+	+	+	+
Peregrine 2007 ¹³³	+	+	-	-	+	+	+
Pollack 1992 ¹³⁴	+	+	-	+	-	+	+
Rossavik 1993 ¹³⁵	+	+	-	+	+	+	+
Sapir 2017 ¹³⁶	+	+	-	+	+	+	+
Smith 1997 ¹³⁷	+	+	-	+	+	+	+
Sovio 2018 ¹³⁸	+	+	+	+	+	+	+
Sritippayawan 2007 ¹³⁹	+	+	-	+	+	+	+
Sylvestre 2000 ¹⁴⁰	+	+	-	+	-	+	+
Weiner 2002 ¹⁴¹	+	+	-	-	+	+	+



FIGURE 40 Risk-of-bias applicability concerns for included studies for systematic review of macrosomia.

TABLE 24 Characteristics of studies included in the meta-analysis of macrosomia

Study (first author and year of publication)	Type of study; setting	Number of total fetuses (LGA fetuses), risk, and selection (all singleton, non-anomalous unless otherwise stated)	Index test (blinding)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comment (inclusion of T1DM, T2DM and GDM)
Aviram 2017 ¹⁰²	Retrospective cohort; single hospital, Israel	<i>n</i> = 7996 (1618) Risk: mixed Selection: mixed risk, term only. Excluded SGA deliveries, intrapartum and SROM	EFW (20 formulas) Hadlock (AC/FL/BPD) Hadlock (AC/FL/HC) Hadlock (AC/FL/BPD/HC) Hadlock (AC/FL) Hadlock (AC/BPD) Shepard (AC/BPD) Threshold: > 90th centile Blinded: no	Within 1 week from delivery	BW > 90th centile	Mean for LGA group: 39.4 weeks' gestation, mean for AGA group: 38.3 weeks' gestation	DM/GDM: included (21% for LGA, 14% for AGA)
Balsyte 2009 ¹⁰³	Retrospective cohort; single hospital, Switzerland	<i>n</i> = 1062 (135) Risk: mixed Selection: term only	EFW Hadlock (AC/FL/HC) Threshold: > 4000 g Blinded: no	Within 1 week from delivery	BW > 4000 g	Mean 39.3 weeks' gestation	DM/GDM: not reported
Benecerraf 1988 ¹⁰⁴	Retrospective cohort; single hospital, Boston, MA, USA	<i>n</i> = 1301 (324) Risk: mixed Selection: included all pregnancies apart from breech and multiples	EFW (Birnholz) Threshold: > 4000 g, > 3800 g Blinded: no	Within 1 week from delivery	BW > 4000 g	Not specified	DM/GDM: included

Study (first author and year of publication)	Type of study; setting	Number of total fetuses (LGA fetuses), risk, and selection (all singleton, non-anomalous unless otherwise stated)	Index test (blinding)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comment (inclusion of T1DM, T2DM and GDM)
Ben-Haroush 2007 ¹⁰⁵	Prospective cohort; single hospital, Israel	<i>n</i> = 259 (23) Risk: universal Selection: routine scan. Included SGA. Excluded hypertensives and diabetics	EFW Hadlock (AC/FL/BPD) Threshold: > 90th centile Blinded: no	Mean 32 weeks' gestation	BW > 4000 g	Mean 39 weeks' gestation	DM/GDM: excluded
Ben-Haroush 2008 ¹⁰⁶	Retrospective cohort; single hospital, Israel	<i>n</i> = 1925 (140) Risk: mixed Selection: term only	EFW Hadlock (AC/FL) EFW + AFI Threshold: EFW > 4000 g, AFI > 95 mm (60th centile) Blinded: no	Interval from ultrasound scan to delivery 2.5 days	BW > 4000 g	Mean for LGA 40 weeks' gestation, mean for normal BW 39.4 weeks' gestation	DM/GDM: excluded
Benson 1991 ¹⁰⁷	Retrospective cohort; Boston, MA, USA	<i>n</i> = 412 (32) Risk: mixed Selection: not specified. Excluded diabetics	EFW Hadlock (AC/FL/BPD) Threshold: > 90th centile Blinded: no	Within 1 week from delivery	BW > 90th centile	Not specified	DM/GDM: excluded
Burkhardt 2014 ¹⁰⁸	Retrospective cohort; single hospital, Zurich, Switzerland	<i>n</i> = 12,794 Risk: mixed Selection: all term, with vertex presentation with scan with 7 days	EFW, AC Hadlock (AC/FL/BPD) Threshold: > 4000 g, > 4500 g and > 35 cm, > 39 cm Blinded: no	Within 1 week from delivery	Shoulder dystocia	281 days for shoulder dystocia, 278 days for no shoulder dystocia	DM/GDM: 7.5% for those with shoulder dystocia, 2.7% for those without shoulder dystocia

continued

TABLE 24 Characteristics of studies included in the meta-analysis of macrosomia (continued)

Study (first author and year of publication)	Type of study; setting	Number of total fetuses (LGA fetuses), risk, and selection (all singleton, non-anomalous unless otherwise stated)	Index test (blinding)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comment (inclusion of T1DM, T2DM and GDM)
Chauhan 2006 ¹⁰⁹	Retrospective cohort; single hospital, Houston, TX, USA	<i>n</i> = 1954 (119) Risk: mixed Selection: pregnancies undergoing fetal surveillance. Included SGA, hypertensives (22%) and SROM (5%)	EFW Hadlock (AC/FL/BPD) Threshold: > 90th centile Blinded: no	Within 4 weeks from delivery; 64% within 7 days from delivery	BW > 90th centile	34% preterm	DM/GDM: included (13%)
Chervenak 1989 ¹¹⁰	Prospective cohort; single hospital, New Jersey, USA	<i>n</i> = 317 (81) Risk: low Selection: uncomplicated pregnancies after 41 weeks' gestation	EFW Hadlock AC/BPD or AC/FL if BPD not available Threshold: > 4000 g Blinded: not clear	> 41 weeks' gestation	BW > 4000 g	Mean 42 ± 0.6 weeks	DM/GDM: excluded
Cohen 2010 ¹¹¹	Retrospective cohort; single hospital, Montréal, QC, Canada	<i>n</i> = 1099 (105) Risk: mixed Selection: only included pregnancies with ultrasound scan on the same or next day as delivery	EFW Hadlock (AC/FL/BPD/HC) Threshold: > 90th centile Blinded: no	On the same day as or next day of delivery	BW > 4000 g	Mean 275.2 days	DM/GDM: included (11.6%)
Crimmins 2018 ¹¹²	Retrospective cohort; single hospital, Baltimore, MD, USA	<i>n</i> = 945 (40) Risk: mixed Selection: all pregnancies > 34 weeks' gestation with normal oGCT	AFG defined as EFW > 90th centile (Hadlock-AC/FL/BPD) or AC > 95th centile Polyhydramnios > 25 cm Threshold: as above Blinded: no	> 34 weeks' gestation	BW > 4000 g Shoulder dystocia NICU admission	Not specified	DM/GDM: excluded

Study (first author and year of publication)	Type of study; setting	Number of total fetuses (LGA fetuses), risk, and selection (all singleton, non-anomalous unless otherwise stated)	Index test (blinding)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comment (inclusion of T1DM, T2DM and GDM)
Cromi 2007 ¹¹³	Retrospective cohort; two hospitals, Switzerland	<i>n</i> = 1026 (53) Risk: mixed Selection: all singletons > 34 weeks' gestation with ultrasound scan within 4 weeks of delivery. Excluded SROM	EFW, AC Hadlock (AC/FL/BPD) Threshold: > 95th centile Blinded: no	Within 4 weeks of delivery Mean 37.3 weeks' gestation	BW > 4000 g, BW > 4500 g	> 34 weeks' gestation; mean 39.2 weeks' gestation	DM/GDM: included (8.8%)
De Reu 2008 ¹¹⁴	Retrospective cohort; single hospital, the Netherlands	<i>n</i> = 3449 (285) Risk: universal Selection: women with no risk factors or pathology. Did not exclude SGA	AC Threshold: > 75th/90th/95th centile Blinded: no	Between 27 and 33 weeks' gestation	BW > 90th centile, BW > 95th centile	Mean 278.7 days	DM/GDM: excluded
Freire 2010 ¹¹⁵ (article in Portuguese)	Retrospective cohort; two hospitals, Brazil	<i>n</i> = 114 (8) Risk: mixed Selection: those with ultrasound scan within 7 days of delivery	EFW Hadlock (AC/FL/BPD/HC) Threshold: > 90th centile Blinded: no	Within 7 days of delivery	BW > 90th centile	15.6% preterm, 84.4% at term	DM/GDM: not reported
Galvin 2017 ¹¹⁶ (GENESIS study) (abstract only)	Prospective cohort; large multicentre study, Ireland	<i>n</i> = 2336 Risk: low Selection: term, uncomplicated, cephalic only	EFW (not specified) Threshold: 4000 g Blinded: yes	Between 39 ⁺⁰ and 40 ⁺⁶ weeks' gestation	Shoulder dystocia NICU admission	Not specified	DM/GDM: excluded

continued

TABLE 24 Characteristics of studies included in the meta-analysis of macrosomia (continued)

Study (first author and year of publication)	Type of study; setting	Number of total fetuses (LGA fetuses), risk, and selection (all singleton, non-anomalous unless otherwise stated)	Index test (blinding)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comment (inclusion of T1DM, T2DM and GDM)
Gilby 2000 ¹¹⁷	Retrospective cohort; single hospital, FL, USA	<i>n</i> = 1996 (318) Risk: mixed Selection: all singleton > 36 weeks' gestation with ultrasound scan within 1 week from delivery	AC Threshold: > 35 cm, > 38 cm Blinded: no	Within 1 week from delivery	BW > 4500 g	> 36 weeks' gestation, mean not reported	DM/GDM: not reported
Hasenoehrl 2006 ¹¹⁸	Prospective cohort; single hospital, Austria	<i>n</i> = 200 (33) Risk: low Selection: included those with ultrasound scan within 1 week. Excluded only fetal anomaly	EFW (Schild) Threshold: > 4000 g Blinded: no	Mean 39.2 weeks' gestation	BW > 4000 g	Mean interval 2.0 days	DM/GDM: not reported
Hendrix 2000 ¹¹⁹	Prospective (RCT); GA, USA	<i>n</i> = 367 (39) Risk: low Selection: term only	EFW Hadlock AC/BPD Threshold: > 4000 g Blinded: no	> 37 weeks' gestation	BW > 4000 g	Mean 39.1 weeks' gestation	DM/GDM: not reported
Henricks 2003 ¹²⁰	Prospective cohort; SC, USA	<i>n</i> = 256 (21) Risk: universal Selection: term only	AC Threshold: > 35 cm Blinded: no	> 37 weeks' gestation	BW > 4000 g	Mean 39.1 weeks' gestation	DM/GDM: not reported
Humphries 2002 ¹²¹	Retrospective cohort; SC, USA	<i>n</i> = 238 (29) Risk: mixed Selection: term only, with ultrasound scan within 2 weeks	EFW Combs (AC/FL/FL) Threshold: > 4000 g Blinded: no	Within 2 weeks of delivery	BW > 4000 g	> 37 weeks' gestation	DM/GDM: not reported

Study (first author and year of publication)	Type of study; setting	Number of total fetuses (LGA fetuses), risk, and selection (all singleton, non-anomalous unless otherwise stated)	Index test (blinding)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comment (inclusion of T1DM, T2DM and GDM)
Kayem 2009 ¹²²	Prospective cohort; multiple hospitals, France and Belgium	<i>n</i> = 1689 (124) Risk: low Selection: as part of a prospective cohort for breech. Term only, with ultrasound scan within 10 days of delivery	AC Threshold: > 36.3 cm Blinded: no	Within 10 days of delivery	BW > 4000 g	Median 39 weeks' gestation	DM/GDM: not reported
Kehl 2011 ¹²³	Prospective cohort; single hospital, Germany	<i>n</i> = 258 (30) Risk: universal Selection: term only with vertex presentation and ultrasound scan within 3 days of delivery	AC Threshold: > 36 cm Blinded: no	Within 3 days of delivery	BW > 4000 g	40 ⁺⁵ weeks' gestation for AC > 36 cm 39 ⁺⁶ weeks' gestation for AC < 36 cm	DM/GDM: not reported
Levine 1992 ¹²⁴	Retrospective cohort; single hospital, New York, NY, USA	<i>n</i> = 406 (68) Risk: mixed Selection: term only. Included pregnancies with diabetes (22%) and previous caesarean section (20%)	EFW Hadlock (AC/FL/HC) Threshold: > 90th centile Blinded: no	5–10 days before delivery	BW > 90th centile	Mean 39.4 weeks	DM/GDM: included (22%)

continued

TABLE 24 Characteristics of studies included in the meta-analysis of macrosomia (continued)

Study (first author and year of publication)	Type of study; setting	Number of total fetuses (LGA fetuses), risk, and selection (all singleton, non-anomalous unless otherwise stated)	Index test (blinding)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comment (inclusion of T1DM, T2DM and GDM)
Melamed 2011 ¹²⁵	Retrospective cohort; single hospital, Israel	<i>n</i> = 4765 (431) Risk: mixed Selection: all deliveries with ultrasound scan within 3 days of delivery. DM/GDM and SROM excluded	EFW (multiple) and AC Hadlock (AC/FL/BPD) Hadlock (AC/FL/HC) Hadlock (AC/FL/BPD/HC) Hadlock (AC/FL) Shepard (AC/BPD) Threshold: > 4000 g, > 36 cm Blinded: no	Within 3 days of delivery	BW > 4000 g	Mean 38.1 weeks	DM/GDM: excluded
Miller 1986 ¹²⁶	Retrospective cohort; single hospital, LO, USA	<i>n</i> = 150 (28) Risk: mixed Selection: term only, included diabetes, pre-eclampsia, prior caesarean section. Excluded SGA	EFW Hadlock (AC/FL) Shepard (AC/BPD) Threshold: > 4000 g Blinded: no	Within 7 days of delivery	BW > 4000 g	Term (mean gestational age not reported)	DM/GDM: included
Miller 1988 ¹²⁷	Retrospective cohort; single hospital, LO, USA	<i>n</i> = 382 (58) Risk: mixed Selection: term only, excluded SROM	EFW and AC Hadlock (AC/FL/BPD) Threshold: EFW > 4100 g, AC > 36.4 cm Blinded: no	Within 7 days of delivery Mean gestational age 275.8 days	BW > 4000 g	Mean gestational age 279.1 days	DM/GDM: not reported

Study (first author and year of publication)	Type of study; setting	Number of total fetuses (LGA fetuses), risk, and selection (all singleton, non-anomalous unless otherwise stated)	Index test (blinding)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comment (inclusion of T1DM, T2DM and GDM)
Nahum 2003 ¹²⁸	Retrospective cohort; single hospital, CA, USA	<i>n</i> = 74 (12) Risk: mixed Selection: only included Hispanic ethnicity, term only	EFW (11 formulas) Hadlock (AC/FL/BPD) Hadlock (AC/FL/HC) Hadlock (AC/FL/BPD/HC) Hadlock (AC/BPD) Shepard (AC/BPD) Threshold: > 4000 g Blinded: no	Within 3 weeks of delivery	BW > 4000 g	Term (mean gestational age not reported)	DM/GDM: included (23.0%)
Nahum 2007 ¹²⁹	Retrospective cohort; single hospital, CA, USA	<i>n</i> = 98 (16) Risk: low risk Selection: term only. Excluded medical complications (pre-eclampsia, DM)	EFW Hadlock (AC/FL/BPD) Hadlock (AC/BPD) Hadlock (AC/FL) Threshold: > 4000 g Blinded: no	Within 3 weeks of delivery	BW > 4000 g	Term (mean gestational age not reported)	DM/GDM: excluded
Nicod 2012 ¹³⁰ (article in French)	Retrospective cohort; single hospital, Switzerland	<i>n</i> = 708 (141) Risk: mixed risk Selection: pregnancies with ultrasound scan within 7 days of delivery	EFW Hadlock (AC/FL/BPD/HC) Hadlock (AC/FL) Threshold: > 4000 g Blinded: no	Within 7 days of delivery	BW > 4000 g	Not reported	DM/GDM: not reported

continued

TABLE 24 Characteristics of studies included in the meta-analysis of macrosomia (continued)

Study (first author and year of publication)	Type of study; setting	Number of total fetuses (LGA fetuses), risk, and selection (all singleton, non-anomalous unless otherwise stated)	Index test (blinding)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comment (inclusion of T1DM, T2DM and GDM)
O'Reilly-Green 1997 ¹³¹	Retrospective cohort; single hospital; New York, NY, USA	<i>n</i> = 445 (107) Risk: low Selection: prolonged pregnancies defined as gestational age > 40 ⁺⁴ weeks	EFW Hadlock (AC/FL/BPD) Threshold: > 4000 g, > 4500 g Blinded: no	Within 3 weeks of delivery	BW > 4000 g, BW > 4500 g	Gestational age > 40 ⁺⁴ weeks' gestation	DM/GDM: excluded
Pates 2007 ¹³²	Retrospective cohort; single hospital, TX, USA	<i>n</i> = 3115 (239) Risk: mixed Selection: those with clinically indicated ultrasound scan within 7 days of delivery	EFW and AFI Hadlock (AC/FL/BPD/HC) Threshold: > 4000 g, AFI > 20 cm (95th centile) Blinded: no	Within 7 days of delivery	BW > 4000 g	Not reported	DM/GDM: included (11%)
Peregrine 2007 ¹³³	Prospective cohort; single hospital, London, UK	<i>n</i> = 262 (48) Risk: mixed Selection: pregnancies with gestational age > 35 ⁺⁶ weeks undergoing IOL. Excluded those with IUD or antepartum haemorrhage	EFW Hadlock (AC/FL) Shepard (AC/BPD) Threshold: > 4000 g Blinded: yes	Exactly before IOL	BW > 4000 g	Median gestational age 41 weeks' gestation	DM/GDM: not reported
Pollack 1992 ¹³⁴	Retrospective cohort; single hospital, New York, NY, USA	<i>n</i> = 519 (119) Risk: mixed Selection: postdate pregnancies > 41 weeks' gestation	EFW Hadlock (AC/FL) Threshold: > 4000 g, > 4500 g Blinded: no	Within 7 days of delivery	BW > 4000 g	> 41 weeks' gestation	DM/GDM: not reported

Study (first author and year of publication)	Type of study; setting	Number of total fetuses (LGA fetuses), risk, and selection (all singleton, non-anomalous unless otherwise stated)	Index test (blinding)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comment (inclusion of T1DM, T2DM and GDM)
Rossavik 1993 ¹³⁵	Retrospective cohort; single hospital, OK, USA	n = 498 (36) Risk: mixed Selection: infants with ultrasound scan within 2 weeks of delivery (if gestational age > 38 weeks) or within 1 week of delivery (if gestational age < 38 weeks)	EFW Hadlock (AC/FL/HC) Threshold: > 4000 g Blinded: no	Within 2 weeks of delivery (if gestational age > 38 weeks) or within 1 week of delivery (if gestational age < 38 weeks)	BW > 4000 g	Not reported	DM/GDM: not reported
Sapir 2017 ¹³⁶ (abstract only)	Retrospective cohort; single hospital, Israel	n = 6214 Risk: mixed Selection: term only; no GDM with scan within 7 days of delivery	EFW, AC Threshold: > 4000 g, > 4500 g, AC > 39 cm Blinded: no	Within 1 week of delivery	Shoulder dystocia	Term (not specified)	DM/GDM: excluded
Smith 1997 ¹³⁷	Retrospective cohort; single hospital, Glasgow, UK	n = 1213 (16) Risk: mixed Selection: non-diabetic pregnancies with ultrasound scan within 7 days of delivery	EFW and AC Hadlock (AC/FL) Threshold: > 4000 g, > 4500 g, AC > 36 cm, AC > 38 cm Blinded: no	Within 7 days of delivery	BW > 4500 g	Not reported	DM/GDM: excluded

continued

TABLE 24 Characteristics of studies included in the meta-analysis of macrosomia (continued)

Study (first author and year of publication)	Type of study; setting	Number of total fetuses (LGA fetuses), risk, and selection (all singleton, non-anomalous unless otherwise stated)	Index test (blinding)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comment (inclusion of T1DM, T2DM and GDM)
Sovio 2018 ¹³⁸	Prospective cohort; single hospital, Cambridge, UK	<i>n</i> = 3866 (177) Risk: universal Selection: unselected number of nulliparous women who delivered after 36 weeks' gestation	EFW, ACGV Hadlock (AC/FL/BPD/HC) Threshold: > 90th centile (population/customised) Blinded: yes	Regular research scan at 36 weeks' gestation (median 36.4 weeks' gestation)	BW > 90th centile, BW > 97th centile BW > 4000 g, BW > 4500 g, shoulder dystocia, neonatal morbidity (composite of metabolic acidosis, 5-minute Apgar score of < 7, NICU admission), severe neonatal morbidity	Median 40.4 weeks' gestation	DM/GDM: included (4.3%)
Sritippayawan 2007 ¹³⁹	Prospective cohort; single hospital, Thailand	<i>n</i> = 328 (3) Risk: low Selection: pregnancies > 34 weeks' gestation. Excluded IUFD, any medical complication	EFW Hadlock (AC/FL/BPD/HC) Threshold: > 4000 g Blinded: no	> 34 weeks' gestation, mean interval 16.9 days from delivery	BW > 4000 g	Mean gestational age 39.4 weeks	DM/GDM: excluded

Study (first author and year of publication)	Type of study; setting	Number of total fetuses (LGA fetuses), risk, and selection (all singleton, non-anomalous unless otherwise stated)	Index test (blinding)	Gestational age at ultrasound	Reference standard	Gestational age at delivery	Other comment (inclusion of T1DM, T2DM and GDM)
Sylvestre 2000 ¹⁴⁰	Retrospective cohort; single hospital, New York, NY, USA	<i>n</i> = 656 (147) Risk: low Selection: postdate pregnancies only (> 41 weeks' gestation)	EFW (Hadlock or Shepard/not specified) Threshold: > 4000 g Blinded: no	> 41 weeks' gestation	BW > 4000 g	41.3 weeks' gestation	DM/GDM: not reported
Weiner 2002 ¹⁴¹	Prospective cohort; single centre, Israel	<i>n</i> = 315 (134) Risk: mixed Selection: offered routine clinical screening to all women at term. Those with suspected EFW > 3700 g had ultrasound scan. Only included those with ultrasound scan with 3 days of delivery	EFW Shepard (AC/BPD) Threshold: > 4000 g Blinded: no	Ultrasound scan with 3 days of delivery	BW > 4000 g BW > 4500 g Shoulder dystocia	40.1 weeks' gestation for both groups	DM/GDM: included (9.2%)

BPD, biparietal diameter; BW, birthweight; DM, diabetes mellitus; FL, femur length; GDM, gestational diabetes mellitus; HC, head circumference; IUF, intrauterine fetal death; oGCT, oral glucose challenge test; SROM, spontaneous rupture of membranes; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus; USS, ultrasound scan.

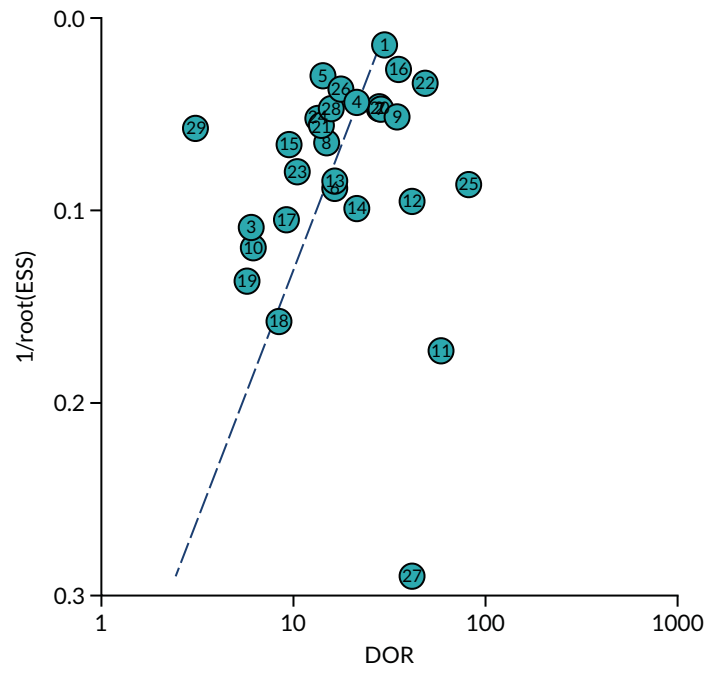


FIGURE 41 Deeks' funnel plot for publication bias for the prediction of LGA (birthweight > 4000 g or > 90th centile). Deeks' funnel plot asymmetry test: $p = 0.02$. ESS, effective sample size.

Appendix 6 Derivation of input parameters for economic simulation model

Beneficial population

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

The NHS Maternity Statistics²⁰¹ states that there were 636,401 births in England in the financial year 2016–17. Of these, 91.8% were at ≥ 37 weeks' gestation, 33.6% of which were to nulliparous women.²⁰¹ The statistics do not disaggregate by reason for elective caesarean section (specifically, whether or not because of suspected breech position). Therefore, this means that there were:

$$636,401 \times 0.918 \times 0.336 = 196,297 \quad (1)$$

deliveries in England per annum that met our population definition.

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

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

Probabilities

Prevalence of small for gestational age fetuses, large for gestational age fetuses and breech presentation: nodes A1 and A2

LGA and SGA are defined as a birthweight in the highest and lowest decile of the distribution, respectively.^{202,203} The prevalence of each in the population is therefore 10%.

The prevalence of breech presentation at the third-trimester scan is estimated at 4.6%, based on the POP study, a large prospective cohort study conducted in Cambridge, UK.¹¹

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

Estimates of the sensitivity and specificity of ultrasound scanning were based on the POP study.^{8,11,138} Note that, because of the structure of the model, these figures are not the true sensitivity and specificity of the tests per se, but the probability of detection if everyone is screened ('universal screening') compared with the probability of detection with selective screening. The estimates are thus the actual sensitivity and specificities multiplied by the proportion of the population screened. Note that we assume that the sensitivity and specificity of a positioning scan are 100%, as this is an extremely simple procedure requiring solely the identification of the skull and spinal column to determine orientation of the fetus.

Interventions for breech presentation: nodes B_ECV, B_ECVs, B_noECV, B_ECVs_rC and B_ECVf_RC

Data on the proportion of mothers accepting ECV, the success rate and the reversion rates were extracted from the POP study.⁸ These methods and results have been published separately.¹¹

Delivery modes for true negative (appropriate for gestational age infants): node C1

An otherwise healthy infant (i.e. true negative for SGA, LGA and breech presentation, node C1) can be delivered by emergency caesarean section or vaginally.

A study of 14,100 singleton liveborn and stillborn infants in French maternity units in 2010 found that approximately 19.4% (2504/12881) of non-SGA infants were delivered by emergency caesarean section.²² The POP study found that 19.9% (735/3689) of non-breech position infants were delivered via emergency caesarean section.¹¹ A 2018 Cochrane systematic review¹⁶ of IOL compared with expectant management in women at or beyond term found an 18.42% (1056/5734) caesarean section rate in the expectant management arm (see analysis 1.13¹⁶).

The most relevant population to this analysis is the POP study.¹¹ Of the 3689 deliveries, 141 were by elective caesarean section. Our defined population excludes elective caesarean sections for indications other than breech presentation; therefore, we assume that 20.7% (735/3548) of AGA deliveries result in emergency caesarean section (95% CI 19.4% to 22.06%), with 79.3% of AGA babies being delivered vaginally.

We chose to use data from the POP study¹¹ (a prospective cohort study) for the risk of emergency caesarean section, rather than those from Monier *et al.*²² (a population-based setting), because the study design of the former made the validity of the numbers easier to verify. Compared with a network meta-analysis, relying on a single study risks potentially overestimating uncertainty; however, because of time constraints, conducting a network meta-analysis was unfeasible.

Delivery modes for false negatives for small for gestational age fetuses and large for gestational age fetuses: nodes S_C2, L_C2

If an infant is SGA and this is not spotted (i.e. is a false negative, node S_C2), the relative risk of emergency caesarean section is taken from the French cohort study, which reported an adjusted relative risk of 'caesarean after onset of labour' (assumed to meet the definition of emergency caesarean section) in low-risk pregnancies of 1.9 (95% CI 1.4 to 2.5; table 3, Monier *et al.*,²² figures reported to only one decimal place).

If a baby is LGA and this is not spotted (i.e. is a false negative, node L_C2), the odds ratio of emergency caesarean section compared with that for an AGA infant is assumed to be 1.792 (95% CI 0.718 to 4.471). This probability was obtained from a retrospective analysis carried out in the USA in 2005 that included 241 nulliparous women whose pregnancies were induced and who were delivered at term.¹⁴⁶ Breech position, stillbirth and pregnancies with other abnormalities were excluded. All women underwent estimation of fetal weight with ultrasound prior to labour. In total, 23 out of 241 (9.5%) overestimated the EFW by $\geq 15\%$. Caesarean section delivery rates for labour arrest (assumed to be emergency caesarean section) were 34.8% in the overestimation group and 13.3% in the no-overestimation group. This equates to 8 out of 23 and 29 out of 218 in each group, respectively, yielding an odds ratio of 1.792 with a standard error of the log of the odds ratio of 0.466.

Delivery modes for true positives for small for gestational age fetuses and large for gestational age fetuses: nodes S_C3, L_C3

The relative risk of 'caesarean after onset of labour' (assumed to meet the definition of emergency caesarean section) in true-positive SGA infants following induction compared with true-negative infants (i.e. AGA infants) is assumed to be 2.9 (node S_C3). This may be an overestimate as according to the data source²² this is the relative risk of emergency caesarean section for true-positive SGA fetuses, whether or not labour was induced, and only 27.1% (36/133) were induced at < 39 weeks' gestation.

We could not identify data on how early IOL would affect the risk of emergency caesarean section among true-positive LGA pregnancies. For this reason, we used data from Middleton *et al.*,¹⁶ implicitly assuming the same relative risk reduction for LGA pregnancies as for non-LGA pregnancies. The relative risk for induced versus non-induced LGA pregnancies was 0.92 (95% CI 0.85 to 0.99) and was modelled using log-normal distribution (mean -0.08, standard error 0.037).

If the policy for LGA infants is expectant management (node L_C2), then the emergency caesarean section rate is assumed the same as for a false-negative diagnosis.

Delivery modes for false positives for small for gestational age and macrosomia: nodes S_C4, L_C4, L_C1

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

A prospective RCT ($n = 6106$) of IOL at 39 weeks' gestation in low-risk nulliparous women yielded a relative risk of (emergency) caesarean section of 0.84 (95% CI 0.76, 0.93) associated with induction.¹⁵⁴ Note that the Monier *et al.*²² study described above reported a relative risk of emergency caesarean section in false positives for SGA of 1.0 (95% CI 0.5 to 2.2). However, as a RCT is generally considered at a lower risk of bias than an observational study, we opted for the RCT results¹⁵⁴ and applied these to nodes S_C4 and L_C4, representing the probabilities of emergency caesarean section following IOL for false-positive diagnoses of SGA and LGA, respectively.

Where the selected management strategy for LGA is expectant management, the risk of emergency caesarean section after a false-positive diagnosis (node L_C1) is logically assumed to be the same as that for an AGA infant (node C1).

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

If an infant is breech and is a false negative for this (i.e. undetected breech, node B_C2), we assume that the probability of an emergency caesarean section is 57.7% (95% CI 38.67% to 75.62%). No comparative data were identified for the risk of emergency caesarean section with unidentified breech compared with that with cephalic presentation. However, a retrospective cohort study of the case notes of 131 women in Hong Kong in 1997 found that, of those with undiagnosed breech at labour, and excluding those in whom ECV was subsequently attempted, 11 (42.3%) had a vaginal breech delivery and 15 (57.7%) had a caesarean section (table 2, Leung *et al.*¹⁶¹). Caesarean sections are labelled as the sum of elective and emergencies, but, given that these were undiagnosed until labour, we have interpreted these as all emergency caesarean section.

Nodes B_C3a to B_C3f represent delivery modes with and without ECV, taking into account success or failure as well as spontaneous reversion (to either breech or cephalic presentation). All estimates are obtained from the POP study¹¹ except for node B_C3b, representing delivery modes where ECV was successful but the infant subsequently reverted to the breech position, because of a lack of relevant observations in the POP study data. We assumed the same distribution as per a false-negative diagnosis of breech (57.69% probability of emergency caesarean section, node B_C2).¹⁶¹ Note that we assume this to be an independent probability with the same parameters as node B_C2, rather than taking the exact same value, to reflect that this is a different outcome measure from B_C2, but with the same likelihood.

Perinatal morbidity: true negative (appropriate for gestational age infants) – node D1

Node D1 represents the baseline risk of neonatal morbidity as a result of expectant management of an otherwise healthy, non-SGA infant, taken from the POP study (Table 25) (see Table 11 and systematic review⁵⁴). Outcomes include no, moderate and severe neonatal morbidity, and perinatal death. Moderate neonatal morbidity was defined as meeting one or more of the following criteria: a

5-minute Apgar score of < 7, delivery with metabolic acidosis (defined as cord blood pH of < 7.1 and base deficit > 10 mmol/l), or admission to the neonatal unit at term (defined as admission < 48 hours after birth at ≥ 37 weeks' gestation and discharge ≥ 48 hours after admission). Severe neonatal morbidity was defined as hypoxic-ischaemic encephalopathy, use of inotropes, need for mechanical ventilation, or severe metabolic acidosis (defined as a cord blood pH of < 7.0 and base deficit of > 12 mmol/l).

The RCOG *Green-top Guideline No. 20a*¹³ states that a 0.1% risk of perinatal mortality is associated with a planned cephalic vaginal delivery. However, this figure includes all stillbirths and neonatal deaths. The relevant figure for the purpose of our model comprises intrapartum stillbirths and neonatal deaths only; deaths prior to this are assumed unrelated to orientation or size of the fetus, and thus do not affect the results of the incremental analysis. To estimate the risk of stillbirth and perinatal mortality, we used observational data from Moraitis *et al.*,⁵⁴ because delivery before 37 weeks' gestation was an exclusion criterion of the study. For baseline risk, we used mortality for spontaneous vaginal and assisted vaginal deliveries only. In the study, spontaneous and assisted vaginal deliveries accounted for 88.07% and 59.48% of antepartum stillbirths and delivery-related perinatal mortality, respectively. Data from the POP study showed the risk of stillbirth/perinatal mortality as a function of birthweight. Using these data, we estimated that the total number of stillbirths and perinatal mortality for spontaneous and vaginal deliveries would have been 809.66 and 455.54, respectively, if all infants had been AGA. Multiplying these numbers with the corresponding proportions of deaths resulting from spontaneous and instrumental vaginal deliveries, we estimated that the total mortality for these categories would have been 984 cases ($n = 635,396$). Modelling this using a beta distribution, the baseline risk (i.e. for AGA pregnancies delivered vaginally) was 0.155% (95% CI 0.145% to 0.165%).

The probabilities of no, moderate or severe morbidity and perinatal death would ideally be modelled as a Dirichlet distribution. However, as these statistics are taken from different sources, they are modelled as independent beta distributions. This may overestimate the uncertainty in morbidity risk. Furthermore, we assume that the risk of neonatal morbidity in an AGA infant is independent of delivery mode. A priori, an emergency caesarean section is expected to be associated with a higher risk of perinatal morbidity. However, the relevant population is infants who are not breech, SGA or LGA, but who are delivered by emergency caesarean section for other reasons. After factoring out these indications for emergency caesarean section, the assumption may not be so unreasonable.

Perinatal morbidity: false-negative small for gestational age infants – node S_D2

The same sources (POP study and Moraitis *et al.*⁵⁴) for node D1 report the odds of adverse outcome in SGA infants (i.e. in the bottom decile of the distribution). The odds ratio of moderate and severe morbidity and stillbirth for SGA compared with AGA infants in the absence of intervention (i.e. induction) is 2.48, 1.88 and 4.89, respectively (node S_D2). Again, we assume that the risk of neonatal morbidity in SGA infants is solely a function of the infant's size and not of the mode of delivery.

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

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

Perinatal morbidity: false-negative large for gestational age infants – nodes L_D2a and L_D2c

Baselines

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

The baseline probability of respiratory morbidity was extracted from a study of the influence of timing of elective caesarean section on respiratory morbidity, conducted in Cambridge, UK.²⁰⁴ All deliveries between 1985 and 1993 at the centre ($n = 33,289$) were included in the analysis and all cases of respiratory distress syndrome or transient tachypnoea necessitating admission to neonatal intensive care were recorded. Of the entire sample, 6955 deliveries occurred at term (39^{+0} to 39^{+6} weeks' gestation) and were delivered vaginally. Among these babies, 22 had respiratory morbidity, reported as 0.32% (95% CI 0.18% to 0.45%). Assigning a beta distribution to these figures yields a similar (but slightly different) 95% CI of 0.20% to 0.46%. This was used as the baseline risk (i.e. the risk for AGA infants).

The baseline probability of shoulder dystocia was based on figures quoted in the RCOG guidelines for the management of shoulder dystocia.²⁰⁵ This reported incidences in the literature of between 0.58% and 0.70%. The best-quality study informing the estimate was a retrospective analysis by Ouzounian *et al.*¹⁶⁴ This reported 1686 cases of shoulder dystocia among 267,228 vaginal births, yielding an incidence of 0.63% (95% CI 0.60% to 0.66%).¹⁶⁴

The baseline probability of other acidosis (i.e. not secondary to shoulder dystocia) was based on a Cochrane systematic review comparing induction with expectant management.¹⁶ Analysis 1.4 of the review reported incidence of birth asphyxia, with 5 out of 731 pregnancies in the expectant management arm, yielding a base probability of 0.68%.¹⁶

The baseline risk of perinatal morbidity was assumed to be the same as described above (node D1), that is an estimated risk of 0.155% (95% CI 0.145% to 165%), based on our own estimations using data from Moraitis *et al.*⁵⁴ As this baseline risk was not specific to fetal size, we used the same baseline risk for SGA and LGA fetuses and distinguished their risk using their odds ratios instead.

To estimate the baseline risk of perinatal death, we used observational data from Moraitis *et al.*,⁵⁴ because delivery before 37 weeks' gestation was an exclusion criterion of the study. For baseline risk, we used mortality for spontaneous vaginal and assisted vaginal deliveries only. In the study, spontaneous and assisted vaginal deliveries accounted for 88.07% and 59.48% of antepartum stillbirths and delivery-related perinatal mortality, respectively. Data from the POP study showed the risk of stillbirth and perinatal mortality as a function of birthweight. Using these data, we estimated that the total number of perinatal deaths for spontaneous and vaginal deliveries would have been 809.66 and 455.54, respectively, if all infants had been AGA. Multiplying these numbers by the proportion of deaths resulting from spontaneous and instrumental vaginal deliveries, we estimated that the total mortality for these categories would have been 984 cases ($n = 635,396$). Modelling this using a beta distribution, the baseline risk (i.e. for AGA pregnancies delivered vaginally) was 0.155% (95% CI 0.145% to 165%).

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

Undetected large for gestational age infant (false negative), vaginal delivery (L_D2a)

No data were available on the relative risk or odds ratio of respiratory morbidity for undetected LGA with a vaginal delivery (node L_D2a). Expert opinion estimated that these infants were at either the same or a lower risk of respiratory morbidity than AGA infants. We therefore used a point estimate relative risk of 0.75, and assigned a uniform distribution between 0.5 and 1. Note that as relative risks are more intuitive than odds ratios from an elicitation point of view, we report this as a relative risk rather than an odds ratio.

The odds ratio of shoulder dystocia in a LGA infant delivered vaginally (vs. an AGA infant) is assumed to be 7.18 (95% CI 2.06 to 25.00). This is based on a systematic review reporting the incidence of shoulder dystocia in all infants with a birthweight ≥ 4000 g (table 2 of Rossi *et al.*¹⁶⁵). Two source studies were meta-analysed with a random-effects model. Importantly, these data are not disaggregated by delivery method. However, it is reasonable to assume that caesarean section eliminates the risk of shoulder dystocia and, therefore, this represents the odds ratio of LGA infants delivered vaginally.

The same table in the review¹⁶⁵ also reported the odds ratio of asphyxia in a LGA infant (vs. an AGA infant) of 2.88 (95% CI 1.34 to 6.22). We assume that this meets our definition of 'other acidosis' and apply the figures accordingly, but with the caveat that this is not disaggregated by delivery mode and so may overestimate the risk (e.g. asphyxia may be the reason for an emergency caesarean section).

The same table in the review¹⁶⁵ also reported the odds ratio of perinatal death in a LGA infant (vs. an AGA infant) of 1.77 (95% CI 0.30 to 10.34). We apply this to our definition of perinatal mortality, again noting that this is not disaggregated by delivery mode. The rarity of the outcome is also reflected in the wide CI, implying a high degree of uncertainty.

Undetected large for gestational age infant (false negative), emergency caesarean section (node L_D2c)

The relative risk of respiratory morbidity for a macrosomic infant delivered via emergency caesarean section compared with an AGA infant (Table 26) delivered vaginally was taken from the Cambridge cohort²⁰⁴ described in *Baselines* (table 2 of Morrison *et al.*¹⁶³). As stated above, this study was not specific to LGA infants, but the risk of respiratory morbidity is most plausibly associated with intervention to speed delivery rather than the presence of LGA. The source table reports the odds ratio of respiratory morbidity with 'caesarean section labour' (assumed to meet the definition of emergency caesarean section) at 39⁺⁰ to 39⁺⁶ weeks' gestation as 3.2 (95% CI 1.4 to 7.4) relative to the baseline of vaginal delivery at 40⁺⁰ to 40⁺⁶ weeks' gestation. Rebasings relative to vaginal delivery at 39⁺⁰ to 39⁺⁶ weeks' gestation yields an odds ratio of 1.674 (95% CI 1.253 to 2.001).

The relative risk of shoulder dystocia for emergency caesarean section was assumed to be zero.

The relative risk of other acidosis for a LGA infant delivered via emergency caesarean section compared with an AGA infant (Table 27) was taken from Chongsuvivatwong *et al.*¹⁶⁶ (as for elective caesarean section described above, and thus the same caveats are attached).

Finally, the relative risk of perinatal mortality for a LGA infant delivered via emergency caesarean section compared with that for an AGA infant was taken from the same source¹⁶⁶ (Table 28).

TABLE 26 Risk of respiratory morbidity from emergency caesarean section

	Odds ratio	95% CI
Caesarean section labour	0.6	0.4 to 1
Vaginal	3.2	1.4 to 7.4
Rebased	5.33	3.5 to 7.4
ln	1.674	1.253 to 2.001
SE	0.167	

SE, standard error.

TABLE 27 Risk of acidosis from emergency caesarean section

Mode of delivery	n	Severe acidosis rate/1000	95% CI	Implied n from raw numbers
Vaginal	12,591	4.3	3.2 to 5.6	54
Emergency caesarean section	4328	8	5.5 to 11.1	35
Mode of delivery	Asphyxia	No acidosis	Total	
Vaginal	54	12,537	12,591	
Emergency caesarean section	35	4293	4328	
Total	89	16,830		
Ratio		95% CI		
OR	1.867	1.217 to 2.865		
lnOR	0.625			
SE(lnOR)	0.218			

OR, odds ratio; SE, standard error.

TABLE 28 Risk of perinatal mortality from emergency caesarean section

Mode of delivery	n	Deaths per 1000 deliveries	95% CI	Implied n from raw numbers
Vaginal	12,591	7	5.6 to 8.6	88
Emergency caesarean section	4328	12.4	9.3 to 16.2	54
Mode of delivery	Dead	Alive	Total	
Vaginal	88	12,503	12,591	
Emergency caesarean section	54	4274	4328	
Total	142	16,777		
Ratio		95% CI		
OR	1.781	1.266 to 2.505		
lnOR	0.577			
SE(lnOR)	0.174			

OR, odds ratio; SE, standard error.

Perinatal morbidity, true-positive small for gestational age infants: induction of labour – node S_D3

If a SGA infant is induced, we assume that the relative risk is 0.7 for moderate and severe morbidity and 0.33 for perinatal death (node S_D3). These data are based on a systematic review of IOL compared with expectant management in low-risk women at or beyond term (approximately 10,000 observations; odds ratios not reported).¹⁶ Critically, this is not the treatment effect for SGA infants, for which we were unable to identify any data, and the relative risk for moderate and severe morbidity was based on data reporting a 5-minute Apgar score of < 7. However, the central estimates of relative risks (0.7 and 0.33, respectively) were considered plausible by clinical experts (GCSS and AAM), and the CIs represented plausible summaries of their epistemic uncertainty.

Perinatal morbidity, true-positive large for gestational age infants: expectant management and induction of labour – nodes L_D3a and L_D3c

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

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

Respiratory complications

A retrospective cross-sectional study of maternal and neonatal outcomes in induced low-risk term pregnancies ($n = 131,243$) reported neonatal complications by week of delivery comparing IOL with expectant management.¹⁶⁷ The adjusted odds ratio of respiratory complications at week 39 is reported as 0.540 (95% CI 0.373 to 0.783; see table 4¹⁶⁷). This was used as odds relative to an AGA infant, whether delivered vaginally or by emergency caesarean section (L_D3a and L_D3c respectively). Of note is that these data are not LGA specific.

Shoulder dystocia

A Cochrane systematic review¹⁰¹ of IOL compared with expectant management for suspected fetal macrosomia estimated a relative risk of shoulder dystocia of 0.6 (95% CI 0.37 to 0.98) (analysis 1.3 of Boulvain *et al.*¹⁰¹). We therefore applied this relative risk, noting that the baseline comparator is MGT_LGA_TP >> L_C2 or MGT_LGA_TA >> L_C3. That is:

$$P(\text{dystocia} \mid \text{vaginal delivery at node L_D3a}) = P(\text{dystocia} \mid \text{vaginal delivery at node L_D2a}) \times \text{RR}, \quad (2)$$

and:

$$P(\text{dystocia} \mid \text{EmCS at node L_D3c}) = P(\text{dystocia} \mid \text{EmCS at node L_D2c}) \times \text{RR}. \quad (3)$$

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

Acidosis

The Boulvain *et al.*¹⁰¹ Cochrane review did not report the incidence of acidosis or asphyxia. Therefore, we sourced data from the Middleton *et al.*¹⁶ Cochrane review, which compared IOL with expectant management in all pregnancies at term. Analysis 1.4¹⁶ reported a relative risk of birth asphyxia of 1.66 (95% CI 0.61 to 4.55). We used this to represent the relative risk of 'other acidosis'.

Perinatal mortality

The Cochrane systematic review¹⁰¹ of IOL compared with expectant management for suspected fetal macrosomia observed zero events in the included studies. We therefore used the Middleton *et al.*¹⁶ Cochrane review, Analysis 1.1,¹⁶ reporting a relative risk of 0.33 (95% CI 0.14 to 0.78) compared with AGA infants that received expectant management.

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

Perinatal morbidity, false-positive small for gestational age or large for gestational age infants: induction of labour – node D4

Following an incorrect diagnosis of SGA or following an incorrect diagnosis of LGA under the IOL policy, an AGA infant will be induced. Evidence suggests that this reduces the risk of stillbirth, but with the consequence of increasing perinatal complications; a retrospective database analysis of induction compared with expectant management at 37 weeks' gestation found an odds ratio of 0.15 (95% CI 0.03 to 0.68) for perinatal death and 1.92 (95% CI 1.71 to 2.15) for admission to a neonatal unit or special care baby unit.¹⁶⁰ We assumed admission to these specialist units was a proxy for moderate and severe complications, so we applied these odds ratios to the baseline risks.

Perinatal morbidity: false-positive large for gestational age infants – expectant management

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

Perinatal morbidity: breech – false negative and true positive (B_D2a – B_D2c)

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

Vaginal breech delivery (B_D2a): perinatal death

The RCOG *Green-top Guideline No. 20a*¹³ states that vaginal delivery in the breech position is associated with a risk of perinatal mortality of 2 in 1000, but 0.5 in 1000 with elective caesarean section, compared with a 1.0 in 1000 risk for a cephalic vaginal delivery. This is based largely on a Cochrane systematic review of planned caesarean section for term breech delivery,¹⁴ the largest contributor to which was the Term Breech Trial (TBT).²⁰⁶

As described in *Perinatal morbidity: true negative (appropriate for gestational age infants) – node D1*, the risk of perinatal mortality of 1.0 in 1000 includes all deaths around the time of delivery. However, our figure of interest is solely intrapartum stillbirth and neonatal death (the implicit assumption is that prepartum deaths are due to causes other than breech, LGA or SGA). A retrospective cohort study of all term singleton births in delivery units in Scotland between 1992 and 2008 ($n = 784,576$) found a mortality rate of 0.04% (234/537,745) associated with cephalic vaginal deliveries.⁵⁴ The same study reported a mortality rate of 0.29% (5/1719) associated with breech vaginal deliveries, yielding an odds ratio of 6.68 (95% CI 2.75 to 16.22).

Vaginal breech delivery (B_D2a): moderate and severe morbidity

We estimate the relative risk of moderate and severe morbidity associated with breech vaginal delivery compared with cephalic vaginal delivery to be 6.7 (95% CI 5.9 to 7.6). This is based on a large

retrospective cohort analysis of the Swedish Medical Birth Registry from 1988 to 1997 reporting the odds ratio of a 5-minute Apgar score of < 7 .¹⁷⁰ We assume that the odds ratios are identical for moderate and severe morbidity. This may be a reasonable assumption: the odds ratio for perinatal death calculated above is 6.68, extremely close to the 6.7 reported here.

Elective caesarean section delivery (B_D2b): perinatal death

A Cochrane systematic review of elective caesarean section compared with vaginal delivery for term breech delivery (Hofmeyr *et al.*,¹⁴ analysis 1.3) found an overall global relative risk of perinatal death of 0.29 (95% CI 0.10 to 0.86).

Elective caesarean section delivery (B_D2b): moderate and severe morbidity

The same review¹⁴ reported a relative risk of a 5-minute Apgar score of < 7 of 0.43 (95% CI 0.12 to 1.47), and of a 5-minute Apgar score of < 4 of 0.11 (95% CI 0.01 to 0.87) (analyses 1.4 and 1.5,¹⁴ respectively). We therefore use this as the relative risk of moderate and severe perinatal morbidity, respectively, associated with elective caesarean section compared with planned vaginal breech delivery.

Emergency caesarean section delivery (B_D2c): perinatal death

A study of 32,776 breech presentations in Scotland between 1985 and 2004¹⁷¹ found 9018 emergency caesarean section deliveries (4108 pre labour and 4910 post labour), of which 14 led to perinatal or neonatal death (0.16%). As stated above, the Moraitis review⁵⁴ reported a mortality rate of 0.29% (5/1719) associated with breech vaginal deliveries. This yields an odds ratio of 0.533 (95% CI 0.192 to 1.482). As this odds ratio is based on combining data from different sources, we explore this parameter in greater detail in a one-way sensitivity analysis.

Emergency caesarean section delivery (B_D2c): moderate and severe morbidity

In the absence of evidence on the effect of emergency caesarean section compared with vaginal breech delivery for the risk of moderate and severe neonatal morbidity, we assumed that the odds ratio would be the same as the odds ratio of perinatal death, that is 0.533 (95% CI 0.192 to 1.482).

Long term outcomes following no, moderate and severe perinatal morbidity (appropriate for gestational age infants, small for gestational age infants and breech presentation): nodes E1–E3

Long-term outcomes were no complications, SEN, SNM, and neonatal/infant death. The risks of each were assumed to be dependent solely on level of perinatal morbidity (where perinatal morbidity is a function of abnormality and delivery management).

A large retrospective cohort study of school children reported the risk of SEN by 5-minute Apgar score, *inter alia*.¹⁷² In total, 4.7% [= 18,736/(18,736 + 376,891)] of children with a 5-minute Apgar score at birth of 8–10 had SEN. We used this as the risk of SEN for children with no neonatal complications (node E1). The same study also reported odds ratio for 5-minute Apgar scores of 4–7 and 0–3, which were used as the increase in risk for moderate and severe neonatal morbidities (nodes E2 and E3).

We used CP as a proxy for SNM. A large retrospective cohort study of births in Sweden analysed the risk of CP by 5-minute Apgar score.¹⁷³ We calculated the baseline risk of CP as the sum of the number of children with CP with a 5-minute Apgar score of ≥ 7 divided by the total number of children with a 5-minute Apgar score of ≥ 7 [= (69 + 163 + 674)/(27,664 + 129,096 + 1,037,793) = 0.08%, node E1]. The study also reported adjusted hazard ratios by individual Apgar score, rather than grouped categorisations (< 4 , 4 to < 7 and ≥ 7). A weighted geometric mean hazard ratio (and 95% CI) was calculated for each group as per *Table 29*, and divided by the weighted 7–10 results. We interpreted the hazard ratio as the relative risk. These are different, but related concepts; the former takes account of time, whereas the latter assumes that all events happen simultaneously. Given the simple structure of our model, and the relative rarity of CP, we felt that this was a sufficient approximation.

TABLE 29 Baseline risk of CP by 5-minute Apgar score

5-minute Apgar score	By single score			Grouped		
	Number of children	Number of children with CP	Adjusted hazard ratio (95% CI)	Number of children	Number of children with CP	Adjusted hazard ratio (95% CI)
0	136	13	277.7 (154.4 to 499.5)	1447	130	145.5 (104 to 204.1)
1	215	23	238.2 (153 to 371)			
2	388	29	124 (83.8 to 183.4)			
3	708	65	148.3 (112.8 to 195)			
4	1097	53	75.9 (56.4 to 102)	17,470	185	10.4 (7.8 to 13.9)
5	1830	39	32.6 (23.4 to 45.6)			
6	4259	42	15.4 (11.2 to 21.2)			
7	10,284	51	6.9 (5.1 to 9.4)			
8	27,664	69	3.8 (3 to 4.9)	1,194,553	906	1 (reference)
9	129,096	163	1.9 (1.6 to 2.2)			
10	1,037,793	674	1 (reference)			

This table was produced using data from figure 1 in Persson *et al.*¹⁷³

Infant mortality data were extracted from routine Scottish data from 1992 to 2010.¹⁷⁴ A total of 1,013,363 neonates had a normal 5-minute Apgar score at birth (defined as ≥ 7) (see *Table 29*). There were 628 neonatal (birth to 28 days) and 1446 infant deaths (29 days to 1 year), a total of 0.2%. This was assumed to form the baseline risk of neonatal/infant mortality (node E1). Adjusted relative risks of neonatal and infant mortality were reported in the appendix of the paper.¹⁷⁴ To generate an overall relative risk over 12 months, a weighted geometric mean (and 95% CIs) of the risks reported by Iliodromiti *et al.*¹⁷⁴ for neonatal and infant mortality was calculated, with weights of 1 and 12 for neonatal and infant mortality, respectively (representing the relative length of the time periods; *Table 30*). Relative risks for Apgar scores of 4–6 and 0–3 were used for moderate and severe neonatal morbidity, respectively (nodes E2 and E3).

Long-term outcomes following large for gestational age infants at birth: nodes L_E1, L_F1, L_G

In our model, LGA infants are at risk of no perinatal complications, respiratory morbidity, shoulder dystocia, other acidosis or perinatal mortality. LGA infants developing shoulder dystocia are at risk of no long-term complications, BPI or acidosis. BPI can be transient or permanent. Acidosis can lead to no long-term complications, SEN, SNM or perinatal mortality. The RCOG *Green-top Guideline No. 42*²⁰⁵ states that ‘fewer than 10% resulting in permanent [injuries]’, based on findings from Gherman *et al.*²⁰⁷

TABLE 30 Relative risk of CP by 5-minute Apgar score

5-minute Apgar score	Model weight for neonates (months)	Adjusted RR (95% CI)	Model weight for infants (months)	Adjusted RR (95% CI)	Pooled adjusted RR (95% CI)
0–3	1/13	188.4 (141.7 to 250.5)	12/13	55.14 (44.03 to 69.06)	60.61 (48.17 to 76.26)
4–6	1/13	34.16 (23.41 to 49.86)	12/13	11.81 (8.64 to 16.15)	12.82 (9.33 to 17.61)
7–10	1/13	1 (reference)	12/13	1 (reference)	1 (reference)

RR, relative risk.

These figures in turn rely on the study by Sandmire *et al.*¹⁶⁹ In total, in 8 out of 145 cases BPI injuries were permanent. We modelled this using a beta distribution, yielding a risk of permanent BPI of 5.5% (95% CI 2.4% to 9.8%).

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

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

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

Transient BPI leads to a background risk of long-term complications, SEN, SNM and neonatal mortality (node E1).

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

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

Costs

Costs of ultrasound scan for fetal size

We obtained the cost of an ultrasound scan for fetal size (and presentation) from the *National Schedule of Reference Costs, 2016–17*.¹⁷⁵ We used data for 'Ante-Natal Standard Ultrasound scan (NZ21Z)', as reported for outpatient procedures. The reference costs contained the mean as well as lower and upper interquartile range for costs, listed by every type of service provider. We calculated a weighted average for the mean/interquartile ranges based on the reported numbers of activities over the year for each provider. We then fitted a gamma distribution to the weighted mean/interquartile range, obtaining the parameters $\alpha = 4.6904$ and $\beta = 22.8062$, and yielding a total cost of £107.06 per scan (95% CI £70.89 to £134.92).

Cost of ultrasound scan for fetal presentation only

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

Midwife-led screening in primary care setting

We theorised that an ultrasound scan for fetal presentation alone could be provided by a midwife in conjunction with a standard antenatal visit in primary care. Although NHS reference costs are provided for 'Ante-Natal Standard Ultrasound scan (NZ21Z)',¹⁷⁵ these scans frequently involve an assessment of fetal anatomy and/or biometry and, because these require much more time and training to assess than fetal presentation alone, we deemed that it was inappropriate to use this cost as an estimate for the cost of an ultrasound scan for fetal presentation alone.

Following the methodology of Wastlund *et al.*,¹¹ we estimated the cost of ultrasound scanning for fetal presentation as a function of the midwife's time, the equipment cost and the cost of the room/facilities where the scan would take place.

We obtained the cost of the midwife's time from the *Unit Costs of Health and Social Care 2017*.¹⁸⁴ We used the total hourly cost for band 5 nurses, £36; this was consistent with the costs reported for midwives in *NHS Staff Earnings Estimates to September 2017 – Provisional Statistics*.²⁰⁸ In addition to the scan itself, time would be needed to make the woman feel comfortable with the process, and to document the results of the scan; therefore, we estimated that the average scan would require 5–10 minutes in total. In the absence of data on how much it would cost to provide ultrasound equipment and sufficient training, we estimated that this could be provided for a total cost between £1000 and £20,000. We assumed that the average machine would be operated 400–3000 times annually over the 5-year time horizon. We assumed that room costs would be between £4500 and £6000 annually²⁰⁹ and that rooms would be in operation 1573 hours per year.¹⁸⁴

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

Sonographer-led ultrasound in designated setting

If implementing ultrasound assessment in primary care (as part of a standard antenatal visit) would not be possible, the most feasible alternative would be to perform the scan at a designated ultrasonography unit. A scan for fetal presentation alone is much swifter and technically less complicated than the type of scan typically performed as part of a standard antenatal visit. For this reason, we did not consider 'Ante-Natal Standard Ultrasounds Scan (NZ21Z)' in the NHS reference costs¹⁷⁵ to be a suitable cost estimate. Instead, we used the data for 'Ultrasound Scan with duration of less than 20 minutes, without Contrast (RD40Z)' from the reference costs¹⁷⁵ for diagnostic imaging. The national schedule of reference costs report costs as mean (£52) and interquartile range (£37–60) only. To capture the uncertainty of this cost appropriately, we fitted a gamma distribution to the mean and interquartile range. The resulting parameter estimation was a gamma distribution with $\alpha = 9.2207$ and $\beta = 5.6395$. This resulted in a total cost of ultrasound scan for fetal presentation of £52.00 (95% CI £24.05 to £90.55) per scan.

Cost for base-case scenario

Because there is genuine uncertainty about the feasibility of providing midwife-led ultrasound screening for fetal presentation only, quantifying the reasonable cost for this parameter was problematic. For the base-case scenario, we used a uniform distribution of costs, ranging between the lower end of the 95% CI if midwife-led screening was possible (£6.87) and the upper end of the 95% CI for sonographer-led screening (£90.55). This way, all plausible costs of ultrasound screening for fetal presentation alone were incorporated into the sensitivity and VOI analysis.

Cost per mode of delivery

We obtained data on costs for different modes of deliveries from the national schedule of reference costs.¹⁷⁵ For a (cephalic) vaginal delivery, we used data for a normal delivery without epidural or assistance. For all modes of deliveries, the reference costs were presented for different levels of complications (CC scores), and we calculated a weighted average cost for all levels. The reference costs reports the mean as well as the lower and upper interquartile range for costs, listed by types of clinical setting (e.g. elective inpatient, non-elective inpatient, outpatient procedures). We calculated a weighted average for the mean/interquartile ranges based on the reported numbers of activities over the year for each setting. For each of the three modes of deliveries (cephalic vaginal, planned caesarean section and emergency caesarean section), we fitted a gamma distribution to the resulting weighted mean/interquartile range. For vaginal delivery, this yielded the parameters $\alpha = 7.2606$ and

beta = 252.5824, with a total cost of £1834.47 (95% CI £1750.43 to £2236.05). The corresponding values for planned caesarean section were alpha = 11.1212 and beta = 307.0169, with a total cost of £3411.93 (95% CI £2679.80 to £4038.29). For emergency caesarean section the values were alpha = 14.7329 and beta = 318.1354, for a total cost of £4688.27 (95% CI £3816.15 to £5443.02).

As the *National Schedule of Reference Costs, 2016–17*¹⁷⁵ does not list separate costs for vaginal breech deliveries, we made the simplifying assumption that these costs would have the same ratio to the costs of elective caesarean section as reported by Palencia *et al.*¹⁷⁷ For that study, the costs were CA\$7255 and CA\$8440 for elective caesarean section and vaginal breech delivery, respectively, with a mean cost difference of CA\$1185 (95% CI CA\$719 to CA\$1663). We fitted a normal distribution (mean 1.1633, standard deviation 0.0332) to calculate the relative cost increase from vaginal breech delivery to elective caesarean section. This yielded a relative cost increase of 1.1633 (95% CI 1.0982 to 1.2284). To obtain the cost of vaginal breech delivery for our model, we multiplied the cost of elective caesarean section (as calculated above from the *National Schedule of Reference Costs, 2016–17*¹⁷⁵) by the relative cost increase from vaginal breech delivery.

Cost of external cephalic version

We obtained the cost of ECV from the cost analysis of offering ECV in the UK reported by James *et al.*¹⁷⁸ The authors provided two different estimates of costs, using low (£186.70) and high (£193.30) staff costs. To convert to 2017's price level, we used the HCHS inflation index: compared with baseline, the index was £302.30 for year 2017,¹⁸⁴ and £196.50 for year 2001.²¹⁰ The resulting cost per ECV was £287.20 and £297.40 for low and high staff costs, respectively. We interpreted this as the feasible range that costs could assume, and let the model sample from this interval using a uniform distribution.

Cost of neonatal unit admission

To capture the cost of admission to neonatal care following delivery, we used cost data from the *National Schedule of Reference Costs, 2016–17*.¹⁷⁵ We divided neonatal critical care into three levels: 'intensive care', 'high-dependency' and 'special care'. For intensive and high-dependency care we used currency codes XA01Z and XA02Z, respectively, and for special care we used a weighted average of currency codes XA03Z to XA05Z. We assumed that the proportion of admittance to each level of neonatal care and length of stay was the same as the one reported by Alfirevic *et al.*¹⁷⁹ This meant that 19%, 7% and 74% of admitted neonates went to intensive, high-dependency and special care, respectively, and that the length of stay was 2, 1.5 and 2 days, respectively. To capture the uncertainty in the cost of care, we fitted a gamma distribution based on the mean and interquartile values, as reported in the reference costs.¹⁷⁵

To estimate the number of neonates admitted to neonatal care as a function of neonatal morbidity at delivery, we reanalysed data from the POP study.⁸ We used 5-minute Apgar score as a proxy for neonatal morbidity at delivery: a 5-minute Apgar score of ≥ 7 , 4–6, and 0–3 was equivalent to no, moderate and severe neonatal morbidity, respectively. This meant that the risk of admittance was 7.4% (95% CI 6.6% to 8.2%) with no morbidity and 47.4% (95% CI 31.9% to 63.1%) with moderate morbidity; we modelled this using the beta distribution. For severe morbidity, we instead made the simplifying assumption that all neonates with severe morbidity would be admitted to a neonatal unit because of the small sample number of infants with severe neonatal morbidity in the POP study. In the absence of evidence as to how the level of neonatal morbidity at birth affects the chance of ending up in each tier of neonatal care, we assumed that the proportions were constant, and that the level of neonatal morbidity affected the level of overall admittance only.

Cost from respiratory morbidity

Morrison *et al.*¹⁶³ reported the incidence and length of stay at hospital for respiratory morbidity. A total of 28% of the morbidities consisted of respiratory distress syndrome and the rest of transient tachypnoea of the newborn. The average stay at the NICU was 4 days for respiratory distress syndrome and 0.6 days of transient tachypnoea of the newborn. The NHS cost of NICU admission is £1295 per day (interquartile range £1015–1541).¹⁷⁵ Given this, the average cost for a case of

respiratory distress syndrome is £5180 (interquartile range £4060–6164), and the cost for transient tachypnoea of the newborn is £777 (interquartile range £609–925). Assuming that respiratory distress syndrome and transient tachypnoea of the newborn make up 28% and 72% of respiratory morbidities, respectively, the average cost of a case of respiratory morbidity would be £2010 (interquartile range £1575–2392). Owing to the very low mortality rate from respiratory distress among infants born at term, we made the simplifying assumption that respiratory distress could lead to NICU admission, but would otherwise have no consequences.²¹¹ To capture the uncertainty of the cost of respiratory morbidity in one parameter, we fitted a gamma distribution based on the mean and interquartile range. The resulting distributions had parameters $\alpha = 10.7125$ and $\beta = 187.6316$, yielding a total cost of £2011 (95% CI £993 to £3381).

Cost of acidosis without long-term consequences

In the absence of data on the costs associated with short-term acidosis (i.e. acidosis that requires neonatal treatment but resolves without any other health consequences), we made the simplifying assumption that treatment would be required at the NICU for 1–4 days, with equal probabilities. To obtain per-day costs, we fitted a gamma distribution for the unit cost of NICU care using cost data from the *National Schedule of Reference Costs, 2016–17*,¹⁷⁵ based on mean and interquartile range. Combining the time and per-day costs, we obtained a total cost distribution. To be able to capture total cost uncertainty in a single parameter, we fitted a gamma distribution to the total cost. The resulting parameter ($\alpha = 3.6143$ and $\beta = 895.6169$) had a total cost of £3240 (95% CI £806 to 7328).

Cost of transient and permanent brachial plexus injury

To estimate the costs associated with BPI, we assumed the same resource use as that reported by Culligan *et al.*¹⁸⁰ Transient BPI costs included a hospital consultation by a specialist, weekly physical therapy for 4 months and one needle electromyography test. Permanent BPI costs included the costs from transient BPI but with weekly physical therapy for 3 years instead, plus one outpatient visit to a specialist, and magnetic resonance imaging of the shoulder.¹⁸⁰ We obtained costs for the specialist consultations and weekly physiotherapy treatments from the *Unit Costs of Health and Social Care 2016*;²¹² these were £199 and £87, respectively. The costs for electromyography and magnetic resonance imaging were taken from the *National Schedule of Reference Costs, 2016–17* (AA33D and RD01C);¹⁷⁵ these were £269.20 and £106.59, respectively. All costs were updated to the price year 2016–17 using the HCHS index.¹⁸⁴ We assumed that all costs except the cost of physiotherapy were incurred in the first year of life and discounted accordingly; the discount rate was 3.5% as recommended by NICE.¹⁸⁸ The total discounted costs from transient and permanent BPI were £2066 and £14,133, respectively.

To account for uncertainty, Culligan *et al.*¹⁸⁰ expanded their cost estimate into a plausible range of costs, which ranged between 50% and 200% of the point estimate. However, directly incorporating this plausible range into our own estimation (after adjusting for cost differences) by using uniform distribution would have been inappropriate, as this would have overestimated costs. Instead, we interpreted the plausible range as a 95% CI for total costs, and then fitted a log-normal distribution to the appropriate mean and 95% CI range. This way, the lower and upper 95% CI were still 50% and 200% of the point estimate, respectively, but in this case following a log-normal distribution. For transient BPI, the resulting distribution had a logged standard error of 0.3536, and the total costs were £2066 (95% CI £1033 to £4132). The corresponding figures for permanent BPI were a logged standard error of 0.3536, and a total cost of £14,133 (95% CI £7067 to £28,264).

Cost of perinatal death

We used the cost of stillbirth as a proxy for the cost of perinatal death. The direct costs of stillbirth were obtained from Mistry *et al.*¹⁸¹ The authors estimated that the costs would be between £1242 (core investigation and counselling only) and £1804 depending on the clinical scenario surrounding the stillbirth and what tests were needed. The authors chose not to present a most plausible estimate within this, but instead just reported these costs as the full range of costs for stillbirth. For this reason, we interpreted these costs as the upper and lower boundaries that the cost of perinatal death could

reasonably assume. We updated these costs to the price year of 2016–17 (the original source used price year 2010) using the HCHS index,¹⁸⁴ and used a uniform distribution.

Cost of special educational needs

We obtained the cost of SEN from Barrett *et al.*,¹⁸² using the difference in costs between SEN and typically developing groups. The cost difference was £6315 (95% CI £3798 to £8832). These costs were estimated for the cost year of 2007–8; hence, we inflated these to the value of price year 2016–17 using the HCHS index,¹⁸⁴ resulting in a cost difference of £7428 (95% CI £4467 to £10,389). This cost was applied annually for years 6–17 of life (the typical school years) and discounted using a discount rate of 3.5%, as recommended by NICE.¹⁸⁸

The cost of severe neurological morbidity

We used CP as a proxy for SNM. In the absence of English cost data that are detailed enough to provide an annual cost for the relevant payer perspective, we obtained the annual cost of CP from Cerebral Palsy Australia.¹⁸³ We used total per-capita cost for the health system, as well as indirect costs (e.g. programme services, aids and home modifications), but we omitted productivity losses, dead-weight losses from financial transactions and costs for informal carers. The annual average cost per case of CP in 2005 was AU\$5362. We converted this to Great British pounds (£) using the exchange rate at 31 December 2005, and updated to the price level of 2016/17 using the HCHS index.¹⁸⁴ This gave a total annual cost of £2929.60. Because the data were derived from the nationwide population of people with CP, this average annual cost is applicable to any year of life.

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

Quality of life

Baseline long-term quality-adjusted life-years

In the absence of neonatal morbidity at birth, lifetime QALYs were calculated using survival and quality-of-life weights for a general UK population. Survival rates were obtained from the Office for National Statistics.¹⁸⁶ These were adjusted using age-specific quality-of-life data from EuroQol. The quality of life for each age group was modelled using a normal distribution with mean and standard errors as provided by EuroQol for the UK using the time trade-off method.¹⁸⁵ We finally limited the total QALYs to the model's time horizon and discounted these QALYs, using a discount rate of 3.5% as recommended by NICE.¹⁸⁸

Quality of life for brachial plexus injury

We obtained the estimated quality of life following BPI from Culligan *et al.*¹⁸⁰ These data were estimated as a plausible range by an expert panel, and the authors used a uniform distribution within the plausible range. The authors provided separate estimates for different complexity levels of BPI. We assumed that long-term BPI in the context of our model would be equivalent to either 'permanent brachial plexus injury (mild to moderate)' or 'permanent brachial plexus injury (severe) and uncomplicated delivery'. We therefore chose to consider the plausible range to stretch between 0.30 (the lower boundary for severe BPI) and 0.70 (the upper boundary for mild to moderate BPI).

Long-term health outcomes following severe neurological morbidity

To get an estimate of the long-term consequences from SNM, we constructed a model based on the work by Leigh *et al.*,¹⁸⁷ using CP as a proxy for SNM. Analogous to Leigh *et al.*,¹⁸⁷ we divided all cases of CP into five levels according to the Gross Motor Function Classification System (GMFCS), which

describes the ambulatory functionality of people with CP.²¹³ We obtained the GMFCS-specific quality of life by letting the model sample values from the gamma distribution provided by Leigh *et al.*,¹⁸⁷ then subtracting these values from 1 (highest possible quality of life) to provide utility weights. A benefit of using these quality-of-life weights was that they were derived using the EuroQol-5 Dimensions,²¹⁴ facilitating comparison with the quality of life of the general population. We let quality of life decrease over time at the same rate as Leigh *et al.*,¹⁸⁷ thereby indirectly assuming that ageing has no greater effect on quality of life for those with CP than for otherwise healthy people in the UK.

Because CP affects mortality as well as quality of life, we had to adjust the model for survival. We calculated GMFCS-specific survival rates using the average mortality rates provided by Leigh *et al.*¹⁸⁷ for each GMFCS and age group (0–10 years, 11–20 years and 21–30 years). Unlike for Leigh *et al.*,¹⁸⁷ our model was not probabilistic in regard to survival; parameter uncertainty was restricted to quality of life only. In the absence of evidence on GMFCS-specific mortality rates beyond 30 years, we made the conservative assumption that the mortality rate for those born with SNM would mimic the general population in the UK after this age.

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

Combining quality of life with survival, we obtained expected lifetime QALYs for neonates born with SNM. We finally limited the total QALYs to the model's time horizon and discounted these QALYs, using a discount rate of 3.5% as recommended by NICE.¹⁸⁸

Appendix 7 Brief summary of economic analyses of universal screening for breech presentation, large for gestational age fetuses and small for gestational age fetuses

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

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

In the following section, we present the economic analysis of universal ultrasound for three conditions: breech presentation, LGA and SGA. The economic analyses of screening for breech presentation¹¹ and LGA¹⁵⁵ have been published. It should be noted that the term 'macrosomia' was used in the publication of the LGA analysis. Although macrosomia is differentiated from LGA, the two are closely related and the definition of macrosomia in this particular analysis was the same as that of LGA.

Breech presentation

Background

Despite the relative ease with which breech presentation can be identified on ultrasound screening, the assessment of fetal presentation at term is often based on clinical examination only. Owing to limitations in this approach, many women present in labour with undiagnosed breech presentation, with increased risk of fetal morbidity and mortality. This study sought to determine the cost-effectiveness of universal ultrasound scanning for breech presentation near term (at 36 weeks' gestation) in nulliparous women.

Methods

To estimate the effects of universal ultrasound screening for breech presentation, we analysed the outcomes for women with a breech presentation in the POP study. The POP study was a prospective cohort study between 14 January 2008 and 31 July 2012, in which nulliparous women, in addition to receiving care in accordance with current clinical practice, attended a research screening ultrasound examination at 36 weeks' gestation. All cases of breech presentation were revealed to both the woman and the attending clinician. By analysing the patients' journals, we noted whether or not breech presentation had been suspected prior to the research scan.

Where breech presentation was detected, ECV was routinely offered. If the ECV was unsuccessful or not performed, the woman was offered either planned caesarean section at 39 weeks' gestation or attempted vaginal breech delivery. We noted if an ECV had been offered, accepted, performed and was successful; where it had not been performed, we noted the reason. We also analysed the mode of delivery as a function of the ECV status.

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

Results

Breech presentation was detected in 179 out of 3879 women (4.6%). In most cases ($n = 96$), there had been no prior suspicion of noncephalic presentation, indicating that up to 54.9% (95% CI 47.5% to 62.1%) of all breech presentations might have been undetected in the absence of universal ultrasound. ECV was attempted for 84 (46.9%) women and was successful for 12 (success rate: 14.3%). Overall, 19 of the 179 women delivered vaginally (10.6%), 110 delivered by elective caesarean section (61.5%) and 50 delivered by emergency caesarean section (27.9%). There were no women with undiagnosed breech presentation in labour in the cohort.

On average, 40 scans were needed per detection of a previously undiagnosed breech presentation (95% CI 33 to 49 scans). The economic analysis indicated that, compared with current practice, universal late-pregnancy ultrasound would identify around 14,826 otherwise undiagnosed breech presentations across England annually. It would also reduce emergency caesarean section and vaginal breech deliveries by 0.7 and 1.0 percentage points, respectively, around 4196 and 6061 deliveries across England annually. Universal ultrasound would also prevent 7.89 neonatal mortalities annually.

We found that a key determinant of the cost-effectiveness of universal ultrasound was the cost of the ultrasound itself. We also noted that there was a high degree of uncertainty surrounding this cost, because no NHS cost data were available for an ultrasound scan for fetal presentation only. We therefore estimated the cost thresholds at which universal ultrasound may be cost-effective. We found that universal ultrasound would be cost-effective if fetal presentation could be assessed for \leq £19.80, assuming a WTP per QALY of £20,000; for a WTP threshold of £30,000, the threshold for cost-effectiveness was £23.10. If the fetal presentation could be assessed for $<$ £12.90 per mother, universal ultrasound would be cost saving.

Conclusions

According to our estimates, universal late-pregnancy ultrasound in nulliparous women would (1) virtually eliminate undiagnosed breech presentation, (2) be expected to reduce fetal mortality in breech presentation and (3) be cost-effective if fetal presentation could be assessed for \leq £19.80 per woman.

Large for gestational age fetuses

Background

Large for gestational age fetuses (i.e. those with an EFW in the highest decile) are at increased risk of complications at delivery. This may manifest in increased neonatal morbidity and mortality, as well as maternal morbidity. Ultrasound screening can be used to diagnose LGA antenatally, but this approach is known to have low predictive value. Furthermore, there is no general agreement on how best to

manage suspected LGA. Possible interventions include scheduling an elective caesarean section or early IOL. However, uncertainty regarding the clinical effectiveness of these interventions persists and intervention may cause unnecessary harm if given without clinical need.

There is currently no national programme that couples screening for LGA with a proven, disease-modifying intervention. Currently, clinical examination of third-trimester pregnancies does not routinely include ultrasound, but women may be selected for ultrasound scanning following clinical suspicion of LGA (selective ultrasound). An alternative approach would be to prospectively scan all women for LGA (universal ultrasound) at around 36 weeks' gestation, but whether or not the benefits of such an approach would justify the increased costs and risk of harmful interventions is unclear.

Methods

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

We simulated outcomes at delivery using sources of data on probabilities, costs and health outcomes obtained from literature. Outcomes included mode of delivery, as well as respiratory morbidity, shoulder dystocia, acidosis and death of the neonate. Long-term neonatal outcomes were then modelled based on the outcomes at delivery; these included permanent BPI, severe anoxic brain damage and neonatal mortality. Maternal health outcomes were based on the mode of delivery. Probabilistic sensitivity analysis was used to capture overall uncertainty in the outcomes as a function of uncertainty in its input parameters. Overall outcomes included expected costs to NHS England and QALYs gained from each strategy. To identify the most cost-effective screening policy we calculated the expected net benefit of each screening management strategy and compared these using ICERs and cost-effectiveness acceptability curves.

Results

Compared with selective ultrasound, universal ultrasound increased by 0.0038 QALYs (95% CI 0.0012 to 0.0076 QALYs), but also increased costs by £123.50 (95% CI £99.60 to £149.90). Overall, the health gains were unable to justify the cost increase at current UK thresholds. The most cost-effective policy was selective ultrasound coupled with IOL where LGA was suspected.

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

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

Conclusions

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

Limitations of the analysis include that LGA was the only criterion evaluated for intervention. In clinical practice, the choice between interventions is typically based on other factors as well, and not all pregnancies in which the fetus is suspected to be LGA would be managed in the same way.

However, by comparing the outcomes for different interventions, our analysis estimates the value of universal ultrasound screening for LGA. Another limitation was the weak evidence base for long-term maternal outcomes following different modes of deliveries; this is something that could be the subject of future research.

Small for gestational age fetuses

Background

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

Methods

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

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

Results

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

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

Conclusions

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

Appendix 8 Questionnaire for attitudes towards universal ultrasound screening in late pregnancy



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

Background

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

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

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

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

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

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

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

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

Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree

(don't want scan)

(do want scan)

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

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

Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree
(wouldn't want to take part) (would take part)

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

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

Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree
(wouldn't want to take part) (would take part)

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

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

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

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

About you

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

Ethnicity:

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

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

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

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

*This report presents independent research funded by the National Institute for Health Research (NIHR).
The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the
Department of Health and Social Care*

Published by the NIHR Journals Library