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

[1]Department of Obstetrics and Gynaecology, NIHR Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK
[2]The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[3]Division of Child Health, Obstetrics and Gynaecology, School of Medicine, University of Nottingham, Nottingham, UK
[4]Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford, UK
[5]School of Healthcare Sciences, Cardiff University, Cardiff, UK
[6]Faculty of Biology, Medicine and Health, School of Medical Sciences, University of Manchester, Manchester, UK
[7]Reproductive and Vascular Biology Group, The Medical School, Newcastle University, Newcastle upon Tyne, UK
[8]Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, UK
[9]Health Economics Group, Norwich Medical School, University of East Anglia, Norwich, UK

*Corresponding author gcss2@cam.ac.uk

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

Scientific summary

Late pregnancy ultrasound screening to predict adverse outcomes

Health Technology Assessment 2021; Vol. 25: No. 15
DOI: 10.3310/hta25150

NIHR Journals Library www.journalslibrary.nihr.ac.uk

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 $\geq$ 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 predesigned 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 hierarchal 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 > 4,000g 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 a only a 44.19% probability that this conclusion is correct (i.e. there is a 55.81% probability that this conclusion is incorrect, in which case a loss will be incurred). The expected loss associated with this decision uncertainty is £31.56 per pregnancy. Equivalently, this is the expected gain if uncertainty were to be eliminated (expected value of perfect information). 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.

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.

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