NIHR | National Institute for Health and Care Research

Journals Library

Check for updates

Health Technology Assessment

Volume 29 • Issue 62 • December 2025

ISSN 2046-4924

Accuracy and clinical effectiveness of fetal growth monitoring strategies for the prediction of small for gestational age at birth: a systematic review and meta-analysis

Evangelos Danopoulos, Nigel Armstrong, Kevin McDermott, Jiongyu Chen, Xiaoyu Tian, Caro Noake and Marie Westwood

DOI 10.3310/AJLK7403

NIHR | National Institute for Health and Care Research

Health Technology Assessment

Check for updates

Extended Research Article

Accuracy and clinical effectiveness of fetal growth monitoring strategies for the prediction of small for gestational age at birth: a systematic review and meta-analysis

Evangelos Danopoulos,[1] Nigel Armstrong,[1] Kevin McDermott,[1] Jiongyu Chen,[1] Xiaoyu Tian,[1] Caro Noake[1] and Marie Westwood[1,2*]

[1]Kleijnen Systematic Reviews (KSR) Ltd, York, UK
[2]Clinical Epidemiology and Public Health, Bristol Medical School, University of Bristol, UK

[*]**Corresponding author** marie.westwood@bristol.ac.uk

Published December 2025
DOI: 10.3310/AJLK7403

This report should be referenced as follows:

Danopoulos E, Armstrong N, McDermott K, Chen J, Tian X, Noake C, Westwood M. Accuracy and clinical effectiveness of fetal growth monitoring strategies for the prediction of small for gestational age at birth: a systematic review and meta-analysis. *Health Technol Assess* 2025;**29**(62). https://doi.org/10.3310/AJLK7403

Health Technology Assessment

HTA TAR

ISSN 2046-4924 (Online)

Impact factor: 4

A list of Journals Library editors can be found on the NIHR Journals Library website

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 4 and is ranked 30th (out of 174 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

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.

Criteria for inclusion in the *Health Technology Assessment* journal

Manuscripts 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 article

The research reported in this issue of the journal was commissioned and funded by the Evidence Synthesis Programme on behalf of NICE as award number NIHR135862. The contractual start date was in March 2023. The draft manuscript began editorial review in November 2023 and was accepted for publication in January 2025. 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' manuscript 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 article.

This article presents independent research funded by the National Institute for Health and Care 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, 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, the HTA programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).

Abstract

Background: Smallness for gestational age has been associated with an increased risk of neonatal/fetal adverse outcomes. The Healthcare Safety Investigation Branch has issued a safety recommendation aimed at improving fetal growth monitoring strategies and reducing risk for babies.

Objectives: The objective was to summarise available evidence to inform the Healthcare Safety Investigation Branch recommendation. The review comprised four research questions on: effects of fetal growth monitoring on neonatal/parental outcomes; effects of implementing fetal growth monitoring guidelines on neonatal/parental outcomes; accuracy of fetal growth monitoring strategies for predicting smallness for gestational age neonates/fetal growth restriction and factors affecting the accuracy of fetal growth monitoring strategies.

Methods: Nineteen databases were searched from 2000 to March 2023 and were updated September 2023. Pregnant people with and without risk factors were included. Each review question had further eligibility criteria. For accuracy results, summary estimates of the sensitivity and specificity with 95% confidence intervals for the prediction of smallness for gestational age at delivery were calculated. Random-effects models were used for the meta-analysis of clinical outcomes. Further outcomes, including the results of risk of bias assessments, were summarised narratively.

Results: Fifty-eight studies (78 publications) were included in the review.

Q1 – Antenatal identification of smallness for gestational age pregnancies was associated with increased rates of intervention (two retrospective cohort studies, n = 100, 198 and 2928), but the available evidence did not support an effect on stillbirths or neonatal outcomes.

Q2 – Meta-analysis (three observational studies and one randomised controlled trial, n = 318,523) indicated that implementation of the Growth Assessment Protocol was associated with a reduction in the risk of stillbirth and risk ratio of 0.79 (95% confidence interval 0.74 to 0.84). Meta-analyses (one observational study and one randomised controlled trial, n = 11,978) indicated that Growth Assessment Protocol implementation was associated with a reduction in the risk of 5-minute Apgar score < 7, risk ratio of 0.78 (95% confidence interval 0.64 to 0.95); however, the effect estimate for neonatal intensive care unit admission was highly uncertain, 0.59 (95% confidence interval 0.02 to 20.03).

Q3 (53 studies) and Q4 (15 studies) – regarding accuracy, the highest sensitivity for both general and high-risk populations was achieved using a combination of estimated fetal weight and abdominal circumference tests, where the threshold was defined as either parameter < 10th percentile. No clear trends were observed for the type of reference charts, either for the use of general versus local reference charts (either the estimated fetal weight or birthweight) or for the use of non-sex-specific versus sex-specific birthweight reference charts (nine studies).

Limitations and conclusions: There is limited evidence linking fetal growth monitoring tests results to the changes in fetal/neonatal outcomes. There is some evidence supporting the reduction of adverse outcomes by Growth Assessment Protocol implementation. Testing during the third trimester is likely to result in more accurate prediction of smallness for gestational age at birth than earlier testing. Use of a locally derived reference chart for estimated fetal weight may result in optimised sensitivity for a given birthweight reference chart (definition of smallness for gestational age).

Future work: Large diagnostic cohort studies and comparative studies are needed to further examine whether and how fetal growth monitoring testing and implementation of guidance can affect clinical outcomes.

Study registration: This study is registered as PROSPERO CRD42023408030.

Funding: This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis programme (NIHR award ref: NIHR135862) and is published in full in *Health Technology Assessment*; Vol. 29, No. 62. See the NIHR Funding and Awards website for further award information.

Contents

List of tables

List of figures

List of abbreviations

AC	abdominal circumference
ACOG	American College of Obstetricians and Gynecologists
AGA	appropriate for gestational age
AID	autoimmune disease
APLS	antiphospholipid syndrome
BMI	body mass index
BPD	biparietal diameter
BW	birthweight
CCT	controlled clinical trial
CDSR	Cochrane Database of Systematic Reviews
CENTRAL	Cochrane Central Register of Controlled Trials
CFEF	Collège Français d'Échographie Foetale
CHD	congenital heart disease
CI	confidence interval
CNGOF	Collège National des Gynécologues et Obstétriciens Français
CPG	clinical practice guideline
CPR	cerebroplacental ratio
CRD	Centre for Reviews and Dissemination
CRL	crown–rump length
DARE	Database of Abstracts of Reviews of Effects
DGGG	Deutsche Gesellschaft für Gynäkologie und Geburtshilfe
DM	diabetes mellitus
DTA	diagnostic test accuracy
EFW	estimated fetal weight
FBV	fetal brain volume
FGP	fetal growth potential
FGR	fetal growth restriction
FH	fundal height
FL	femur length
FLuV	fetal lung volume
FLV	fetal liver volume
FMF	Fetal Medicine Foundation
FN	false negative
FP	false positive
FV	femur volume
GA	gestational age
GAP	Growth Assessment Protocol
GDM	gestational diabetes mellitus
GIN	Guidelines International Network
GROW	Gestation Related Optimal Weight
HC	head circumference
HDU	high-dependency unit
hh	hand-held
HSIB	Healthcare Safety Investigation Branch
HT	hypertension
HTA	Health Technology Assessment
HV	humerus volume
ICTRP	International Clinical Trials Registry Platform
INAHTA	International Network of Agencies for Health Technology Assessment
IQR	interquartile range
IRIS	International Randomized Study of Interferon and STI571
ISUOG	International Society of Ultrasound in Obstetrics and Gynecology
IUGR	intrauterine growth restriction
KSR	Kleijnen Systematic Reviews Ltd
LGA	large for gestational age
LILACS	Latin American and Caribbean Health Sciences Literature
LMP	last menstrual period
LV	left ventricular
MCA	middle cerebral artery
MeSH	medical subject heading
MH	Mantel–Haenszel
MRI	magnetic resonance imaging
N/A	not applicable

NC	not calculable
NICE	National Institute for Health and Care Excellence
NICHD	National Institute of Child Health and Human Development
NICU	neonatal intensive care unit
NIH	National Institutes of Health
NIHR	National Institute for Health and Care Research
NISC	Neonatal Intensive and Special Care
NR	not reported
ONS	Office for National Statistics
OR	odds ratio
PAPP-A	pregnancy-associated plasma protein A
PET	pre-eclampsia
PI	Perinatal Institute
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies-2
RCOG	Royal College of Obstetricians and Gynaecologists
RCPCH	Royal College of Paediatrics and Child Health
RCT	randomised controlled trial
RoB	risk of bias
ROC	receiver operating characteristic
RR	risk ratio
RV	renal volume
SACN	Scientific Advisory Committee on Nutrition
SBLCB	Saving Babies Lives Care Bundle
SC	standard care
SCBU	Special Care Baby Unit
SFH	symphysis fundal height
SGA	small for gestational age
SLE	systemic lupus erythematosus
st	soft tissue
TCD	transverse cerebellar diameter
TN	true negative
TP	true positive
TV	thigh volume
UAD	umbilical artery Doppler
UAV	upper arm volume
US	ultrasound
USS	ultrasound scan
UtA	uterine artery
WHO	World Health Organization

Plain language summary

An independent report on maternity safety investigations conducted during the COVID-19 pandemic identified 11 cases of stillborn infants who were smaller than would be usual for the number of weeks of pregnancy. Being smaller than would be usual for the number of weeks of pregnancy has been linked to an increased risk of stillbirth. In 8 of the 11 cases, this smallness was not detected until birth.

This finding raised concerns about the way in which babies' growth is checked during pregnancy.

This research has systematically searched the published literature to identify research studies about the methods used to check babies' growth during pregnancy. Our report summarises the evidence about how good current measurement methods are at finding when babies are smaller than would be usual for the number of weeks of pregnancy, how are the results of these measurements used (e.g. to decide that a baby needs to be delivered early) and how does this affect how well babies and mothers do after delivery.

There is a lack of evidence to link the increased detection of small babies before birth to reductions in stillbirths and improved outcomes for babies, but there is evidence that an increased detection may lead to more induction of labour and increased numbers of assisted and caesarean births. There is a lack of evidence to determine whether current United Kingdom clinical guidance has been effective in reducing stillbirths and improving outcomes for babies. Evidence suggests that ultrasound during the third trimester, to measure the baby and estimate weight, is the method likely to detect most small babies before birth.

Scientific summary

Background

Smallness for gestational age has been found to be associated with a higher risk of stillbirth. According to the latest report by the Office for National Statistics, low birthweight (BW) is one of the key risk factors for neonatal and infant mortality in England and Wales.

In the UK, pregnant people are typically screened using ultrasound (US) at around 12 weeks of gestation and approximately at the middle of pregnancy for gestational aging. After mid-pregnancy, they are offered further US monitoring, to assess fetal growth, only where clinically indicated.

The Healthcare Safety Investigation Branch (HSIB) 2021 National Learning Report, '*Intrapartum stillbirth: learning from maternity safety investigations that occurred during the COVID-19 pandemic, 1 April to 30 June 2020*', identified 11 cases of stillborn small for gestational age (SGA) infants, of whom 8 were not detected until birth, a finding which could be viewed as a further indication of the poor sensitivity of the methods currently used to detect SGA. The HSIB Report made eight safety recommendations, including Safety recommendation R/2021/148:

HSIB recommends that the Department of Health and Social Care commission a review to improve the reliability of existing assessment tools for fetal growth and fetal heart rate to minimise the risk for babies.

This systematic review has been commissioned and undertaken to summarise the available evidence to inform the fetal growth monitoring component of this recommendation.

Objectives

The overall objective of this project was to summarise the available evidence to inform the fetal growth monitoring component of the HSIB safety recommendation to '*commission a review to improve the reliability of existing assessment tools for fetal growth and fetal heart rate to minimise the risk for babies*'. The following research questions were defined to address the project objectives:

1. What are the effects, on clinical outcomes [e.g. neonatal morbidity, rates of brain injury, unplanned neonatal intensive care unit (NICU) admissions and parental morbidity] and rates of stillbirth and neonatal death, of interventions (e.g. iatrogenic delivery) which are made based on the findings of fetal growth monitoring to detect SGA/fetal growth restriction (FGR)?
    a. What are the effects of fetal growth monitoring to detect SGA/FGR on rates of pre-term iatrogenic delivery and gestational age at iatrogenic delivery?
2. What are the effects, on neonatal and parental outcomes, of implementing published guidelines for fetal growth monitoring?
3. What is the accuracy of different methods of monitoring fetal growth, for example, symphysis fundal height, US measurement of fetal size [fetal abdominal circumference (AC) or estimated fetal weight (EFW)], for predicting SGA/FGR at delivery?
4. What are the effects, on the performance (accuracy) of different methods of monitoring fetal growth, of key operational variables [e.g. timing of monitoring, type of growth reference chart used (customised or population-based), experience and training of clinical practitioner performing monitoring] and parental characteristics [e.g. body mass index (BMI)/obesity and ethnicity]?
    a. What factors affect the failure to obtain a satisfactory measurement or lack of clinical confidence in the reported result?

Methods

Assessment of clinical effectiveness

Nineteen databases, including MEDLINE and EMBASE, research registers, a conference proceedings resource and a pre-print resource were searched for relevant studies from 2000 to March 2023. The main EMBASE and MEDLINE searches were rerun in their entirety in September 2023. Search results were screened for relevance independently by two reviewers. Full-text inclusion assessment, data extraction and quality assessment were conducted by one reviewer and were checked by a second. The methodological quality of included diagnostic test accuracy (DTA) studies was assessed using Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2). The methodological quality of included randomised controlled trials (RCTs) was assessed using the revised Cochrane Risk of Bias Tool for Randomised Trials (RoB 2). The methodological quality of observational 'before-and-after' studies was assessed using a checklist, devised by the authors, for this review.

Where multiple accuracy studies assessed the same growth monitoring strategy, summary estimates of the sensitivity and specificity together with 95% confidence intervals (CIs) and prediction regions for the prediction of SGA at delivery were calculated. All meta-analyses involved fewer than four studies and estimated separate pooled estimates of sensitivity and specificity, using random-effects logistic regression. For comparative studies that evaluated the same clinical outcome(s), intervention/monitoring method and comparator, summary effect estimates with 95% CIs were calculated via random-effects models using the Mantel–Haenszel method with a Paule-Mandel estimator. Where between-study heterogeneity limited the applicability of meta-analysis, a narrative summary utilising text, tables and figures has been provided. The *Results of the assessment of clinical effectiveness* section of this report is structured by research question.

Results

Fifty-eight studies (78 publications) were included in the review.

Study quality

The methodological quality of diagnostic accuracy studies included in this review was high; in 44 of the 58 QUADAS-2 assessments, studies were rated as having low risk of bias (RoB) and low concerns regarding applicability on all domains.

Using RoB 2, all included RCTs were rated as having some concerns due to possible deviations from the intended interventions but were otherwise rated as low RoB.

For included comparative observational studies, the areas of concern were retrospective study design (three studies), inappropriate exclusion of participants and differences in selection criteria before and after the introduction of monitoring (one study).

Research question 1

Two studies (Michaeli J, Michaeli O, Rozitzky A, Grisaru-Granovsky S, Feldman N, Srebnik N. Application of prospect theory in obstetrics by evaluating mode of delivery and outcomes in neonates born small or appropriate for gestational age. *JAMA Network Open* 2022;**5**:e222177, and Nymark Hansen D, Sand Odgaard H, Uldbjerg N, Sinding M, Sorensen A. Screening for small for gestational-age fetuses. *Acta Obstet Gynecol Scand* 2019;**99**:503–9) provided data to inform research question 1. Both were retrospective cohort studies, which, in addition to providing some accuracy data for growth monitoring strategies using EFW US, also investigated the relationship between antenatal classification [fetal growth monitoring test result, SGA or appropriate for gestational age (AGA)], whether or not this classification was correct in relation to BW classification of SGA or AGA [test result category: true positive (TP), false positive (FP), false negative (FN) or true negative (TN)] and parental and neonatal clinical outcomes. Both studies reported some results indicating that both antenatal identification of SGA pregnancies (TP vs. FN) and antenatal misclassification of AGA pregnancies as SGA (FP vs. TN) were associated with increased rates of intervention (induction of labour and caesarean deliveries). Despite the apparent increased rates of intervention in SGA pregnancies that were identified antenatally, there was no evidence that the antenatal identification of SGA was associated with reduction in stillbirths and there was no clear evidence of an association with improved neonatal outcomes; the only statistically significant effect identified was a reduction in the rate of NICU admission in SGA neonates identified antenatally (TP) compared with those not identified (FN), unadjusted odds ratio (OR), calculated from the data reported in Michaeli *et al.* (2022), 0.63 (0.52, 0.76). There was no evidence that antenatal misclassification of AGA neonates as SGA (FP) had any statistically significant detrimental effects on neonatal outcomes compared to correct classification as AGA (TN).

Research question 2

Four comparative observational studies and one RCT assessed the clinical effects of implementing guidance on fetal growth monitoring and all studies were concerned with the effects of implementing the Growth Assessment Protocol (GAP). The results of meta-analyses of three studies (n = 28,911) indicated that the GAP implementation was associated with an increase in the rates of induction of labour [pooled OR was 1.16 (95% CI 1.01 to 1.34), I^2 14%], but not in the risk of caesarean birth [pooled risk ratio (RR) was 1.05 (95% CI 0.86 to 1.27), I^2 61%]. The results of meta-analysis also indicated that the GAP implementation was associated with a reduction in the risk of stillbirth [RR 0.79 (95% CI 0.74 to 0.84), four studies, I^2 0%, n = 64,9272]. Additionally, in the only included RCT, the GAP implementation was associated with a small reduction in the rate of stillbirth, difference −0.07 (95% CI −0.14 to −0.01) (adjusted for age, ethnicity, parity and stratification factor). Only two studies reported data for the adverse neonatal outcomes, NICU admission and 5-minute Apgar score < 7; pooling of the results indicated that implementation was associated with a reduction in the risk of poor 5-minute Apgar score, RR of 0.78 (95% CI 0.64 to 0.95), I^2 0%, but it had no statistically significant effect on NICU admission, 0.59 (95% CI 0.02 to 20.03), I^2 93%. However, it should be noted that the adjusted difference in the rate of low Apgar scores reported in the RCT indicated that the GAP implementation had no statistically significant effect, −0.2 (95% CI −0.4 to 0.1).

Research question 3

Single US (assessing EFW US, AC US and combinations of these parameters) during the third trimester on singleton pregnancies were the most commonly evaluated growth monitoring strategies with respect to test accuracy. The majority of studies calculated EFW using a variation of the Hadlock equations. The most frequently used threshold for all the tests was < 10th percentile. In general, between-study comparisons of test accuracy are of limited value in that DTA studies are observational studies, and hence, comparing the performance of tests based on evaluations in different studies does not take account of differences in study population or other factors that may affect test performance. For this systematic review, between-study comparisons of the performance of different growth parameters were particularly problematic (even for the selected data set described above) due to the high degree of variation in other components of the test strategy (EFW equations, test reference charts and BW reference charts). Three studies, one conducted in a general unselected population and two conducted in high-risk populations, reported the results of within-study comparisons of EFW US calculated using Hadlock 2c, AC US and OR/AND combinations of the two parameters, using test and BW thresholds of < 10th percentile. The results of these studies consistently indicated that, as might be expected, the highest sensitivity was achieved when using a combination of EFW and AC, where the test threshold was defined as either parameter < 10th percentile (the OR combination), and that a combination where the test threshold was defined as both parameters < 10th percentile (the AND combination) gave the lowest sensitivity; EFW US alone consistently had either lower or equal sensitivity to AC US alone. The results of all of three within-study comparisons reported generally high specificity values (> 90%), with little variation in specificity across test strategies (EFW US calculated using Hadlock 2c, AC US and OR/AND combinations of the two parameters).

Comparison of different testing timings between the first, second and third trimesters indicated that the third-trimester testing strategies performed better with respect to maximising the antenatal detection of SGA babies. In terms of single versus multiple/serial testing, sensitivity estimates were very low for all strategies other than where multiple testing occurred during the third trimester. Both these outcomes suggest that there is more to be gained from testing during the third trimester alone; the limited available data do not support a benefit from earlier testing.

All studies, included in this review, that assessed the accuracy of growth monitoring strategies in unselected general populations, evaluated universal testing (screening) rather than clinically indicated testing as currently recommended in the UK. One UK 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) compared the accuracy of the third-trimester EFW US when applied universally in a general population to its accuracy when applied only where clinically indicated. This study used a general reference chart and a threshold of < 10th percentile for EFW and a local UK BW chart. For BW < 10th percentile, the sensitivity was 57% for universal testing and 20% for clinically indicated testing, with corresponding specificities of 90% and 98%.

Research question 4

There were insufficient data to adequately inform the assessment of how the accuracy of individual test strategies for fetal growth monitoring may vary with maternal characteristics (e.g. BMI/obesity and ethnicity) or the type, training and experience of the clinical practitioner performing the monitoring.

Most of the available data for research question 4 were concerned with variations in the accuracy of the metrics used to monitor fetal growth (EFW US and AC US) with the reference chart used for the index test or the reference chart used for BW. Nine included studies used 18 different reference charts for the index test and 15 different BW reference charts, rendering meaningful between-study comparisons challenging. Overall, no clear trend in test strategy performance was observable across studies, either for the use of general versus local reference charts (either EFW or BW) or for the use of non-sex-specific versus sex-specific BW reference charts. One large UK study (Mathewlynn S, Impey L, Ioannou C. Detection of small- and large-for-gestational age using different combinations of prenatal and postnatal charts. *Ultrasound Obstet Gynecol* 2022;**60**:373–80) used five different BW reference charts resulting in prevalences of SGA (< 10th percentile) at birth ranging from 4.2% to 10.9%. In Mathewlynn *et al.* (2022), the use of the locally derived reference chart for EFW always resulted in the highest sensitivity and lowest specificity estimates for any given BW reference chart (definition of SGA).

Conclusions

There is a lack of evidence linking fetal growth monitoring and the results of tests used to monitor fetal growth to changes in rates of stillbirth, perinatal death or adverse neonatal clinical outcomes.

There is some evidence to suggest that implementation of the GAP care pathway in UK settings may be associated with a reduction in adverse neonatal outcomes, but evidence about the effects of implementation on stillbirth was inconsistent. It is also unclear to what extent observed effects were attributable to GAP or other contemporaneous changes to routine care. The extent to which any effects of GAP, which is a complex, multifactorial intervention, may be attributable to the antenatal detection of SGA is also unknown.

There is a lack of evidence to assess the effectiveness of implementing UK clinical guidelines with respect to reducing rates of stillbirth and adverse neonatal outcomes.

Where fetal growth monitoring is to be implemented, there is insufficient evidence to support strongly favouring any one test strategy [combination of test timing, parameter measured (including any formula used, as for estimating EFW) and threshold and reference chart] for use in fetal growth monitoring. The available evidence suggests that testing during the third trimester is likely to result in a more accurate prediction of SGA at birth than earlier testing and that (for the general pregnant population) universal third-trimester US is likely to offer an improved sensitivity compared to testing based on clinical indication. Evidence from within-study comparisons of different test parameters suggests that, for a single US examination in the third trimester, a combination of EFW OR AC < 10th percentile could offer an increased sensitivity relative to either parameter used alone. There is also some evidence to suggest that, when using EFW US to monitor fetal growth, the use of a locally derived reference chart for EFW may result in the highest sensitivity for a given BW reference chart (definition of SGA).

Study registration

This study is registered as PROSPERO CRD42023408030.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis programme (NIHR award ref: NIHR135862) and is published in full in *Health Technology Assessment*; Vol. 29, No. 62. See the NIHR Funding and Awards website for further award information.

Chapter 1 Objective

The overall objective of this project was to summarise the available evidence to inform the fetal growth monitoring component of the Healthcare Safety Investigation Branch (HSIB) safety recommendation to '*commission a review to improve the reliability of existing assessment tools for fetal growth and fetal heart rate to minimise the risk for babies*'.[1] The following research questions were defined to address the project objectives:

1. What are the effects, on clinical outcomes [e.g. on neonatal morbidity, rates of brain injury, unplanned neonatal intensive care unit (NICU) admissions and parental morbidity] and rates of stillbirth and neonatal death on clinical outcomes, of interventions (e.g. iatrogenic delivery) which are made based on the findings of fetal growth monitoring to detect small for gestational age (SGA)/fetal growth restriction (FGR)?
    a. What are the effects of fetal growth monitoring to detect SGA/FGR on rates of pre-term iatrogenic delivery and gestational age (GA) at iatrogenic delivery?
2. What are the effects, on neonatal and parental outcomes, of implementing published guidelines for fetal growth monitoring?
3. What is the accuracy of different methods of monitoring fetal growth, for example, symphysis fundal height (SFH), ultrasound (US) measurement of fetal size [fetal abdominal circumference (AC) or estimated fetal weight (EFW)], for predicting SGA/FGR at delivery?
4. What are the effects, on the performance (accuracy) of different methods of monitoring fetal growth, of key operational variables [e.g. timing of monitoring, type of growth reference chart used (customised or population-based), experience and training of clinical practitioner performing monitoring] and parental characteristics [e.g. body mass index (BMI)/obesity, ethnicity]?
    a. What factors affect the failure to obtain a satisfactory measurement or lack of clinical confidence in the reported result?

Chapter 2 Background and definition of the decision problem(s)

Rationale for monitoring fetal growth

Smallness for gestational age has been found to be associated with a higher risk of stillbirth[2] and an increased risk of perinatal mortality not attributable to congenital abnormality.[3] Furthermore, birthweight (BW) has been found to be a marker or risk factor for diseases in adult life.[4] According to the latest report by the Office for National Statistics (ONS), low BW is one of the key risk factors for neonatal and infant mortality in England and Wales.[5] The infant mortality rate of babies weighing < 2500 g was 30.2 deaths per 1000 live births in 2021, which had increased from 27.9 deaths per 1000 live births in 2020. On the other hand, the infant mortality of babies weighing ≥ 2500 was 0.8 deaths per 1000 live births, which was consistent with the 2020 rate.

Different thresholds have been used, in the research literature, to define restricted growth. According to the Royal College of Obstetricians and Gynaecologists (RCOG), SGA refers to infants with a BW less than the 10th percentile and this is the threshold, which is generally used in the UK clinical practice.[6]

The terms FGR or intrauterine growth restriction (IUGR) and SGA are not interchangeable; 50–70% of SGA fetuses have fetal growth, which is appropriate for the parental size and ethnicity (i.e. not FGR).[6] Diagnosis of a SGA fetus and FGR is important in order to guide management and direct timely delivery.[6] However, when evaluating monitoring strategies, it is important to the potential harms that may arise for increased intervention both for SGA infants and infants, who may be incorrectly categorised as SGA as a result of increased monitoring. SGA is often used as a proxy measure for FGR/IUGR, and although not synonymous, there is a correlation between them. In fact, the higher the rate of SGA, the greater the likelihood that it is a result of IUGR.[7]

Monitoring methods

Fetal growth monitoring strategies can involve the assessment of parental characteristics and clinical and US estimations of fetal growth.[8] Risk assessment protocols based on parental characteristics are usually the first step in monitoring strategies, and these are used to assess which pregnancies are most likely to benefit from further monitoring to estimate fetal size (secondary screening). Methods of estimating fetal size can be broadly divided into physical, based on the clinical palpation of fundal height (FH) in relation to anatomical landmarks (e.g. symphysis, umbilicus and xiphisternum), and internal, based on measurements of fetal size parameters via ultrasonography; Doppler ultrasonography can also be used to assess the umbilical and/or uterine artery (UtA) and/or blood, which is a risk factor for FGR.[9] The predominant growth monitoring method has historically been FH measurement, as it is a non-invasive and easy-to-perform physical test. Although palpation measurements can be subjective, creating wide interobserver differences, they are often the only alternative in the absence of US machines.[10] The widely used SFH method for SGA screening involves measuring the distance between the upper border of the pubic symphysis and the uterine fundus using a tape measure (at 24+ weeks of gestation). If the SFH measurement in centimetres corresponds to the GA ± 2 (weeks), the fetus is regarded to be growing normally. When used in isolation in developing countries, specificity to predict SGA was relatively high (0.87), but sensitivity was low (0.58), providing small informational value.[11]

The best prediction of BW is based on US estimations, but it is also more complicated and expensive.[8,12] As such, it is mainly used as a secondary screening tool. On the whole, the accuracy of US estimations has improved over the years due to technological advancements.[8] Timings of US examinations vary widely and cover all three trimesters, and they can be applied as single one-off primary screening tests, but they are usually serial, especially when growth abnormality is detected. US monitoring methods vary considerably. There are seven main components of each US screening/monitoring method.

These are:

- timing of test:
  - gestation week
  - configuration: single or serial
- fetal parameters measured: EFW, AC, biparietal diameter (BPD), cerebroplacental ratio (CPR), etc.
- test end point (EFW percentile, EFW *z*-score, AC percentile, growth deviation, etc.)
- estimation formula used to calculate the test end point
- test threshold definition
- reference chart for threshold.

The fetal parameters that are examined and measured during the US examination are commonly defined by the formula that is being used to estimate the fetal weight. These fetal parameters include AC, BPD, head circumference (HC), femur length (FL), etc. In a meta-analysis, Goto (2017)[8] captured a wide range of these different formulas and corresponding fetal parameters, which are summarised in *Table 1*. The most popular equations, the Hadlock models,[28–30] are presented separately in *Table 2*; we have assigned numbers (1a–1f and 2a–2d) to the Hadlock models which we have used, in the *Results of the assessment of clinical effectiveness* section of this report, to indicate the specific model used.

The variation between the components in the screening/monitoring methods hinders the easy comparison of effectiveness. The usefulness of different estimation parameters/formulas and the timing of the US examination were the focus of a meta-analysis that included 13 recent studies (published after 2015), which concluded that US fetal anthropometric formulas/parameters, excluding AC and FL, were useful for secondary SGA screening in pregnancies, which have been clinically assessed as at risk for SGA but unsuitable for primary screening in the general pregnant population.[12] A further meta-analysis of 21 studies found that third-trimester screening performs better when conducted near 37 weeks of gestation, while both end points of AC and EFW (< 10th percentile) have similar performance for predicting late SGA.[31]

In the UK, pregnant people are typically screened using US at around 12 weeks of gestation and approximately at the middle of pregnancy for gestational aging.[32] After mid-pregnancy, they are offered further US monitoring, to assess fetal growth, only where clinically indicated.

Growth reference charts

Population centiles have been used historically to define weight centiles; customised charts have become preferred over time, as they are considered to be more effective in identifying small babies at risk for adverse perinatal outcomes.[6] Population-based norms have been criticised for not being able to differentiate between abnormal growth and constitutionally small (or large), but otherwise healthy fetuses, leading to misclassifications.[33] Popular customisable growth charts, such as Gardosi (2004)[34] and the Gestation Related Optimal Weight (GROW) method, which are developed and offered by the Perinatal Institute (PI),[35] can be customised for a number of physiological variables that are known to affect growth, including: GA, parental height, weight and ethnicity as well as parity and fetal sex. Customisable charts can refer to specific countries (e.g. GROW) or be international. The World Health Organization (WHO) has created fetal growth charts, intended for worldwide use, for EFW based on common US biometric measurements.[4] The charts' percentiles can be customised by country, parental characteristics and fetal sex.

The superiority of customised versus population-based charts to identify SGA neonates at risk for adverse outcomes has been questioned; a meta-analysis has found no differences between the two approaches.[33]

TABLE 1 Formulas for US-based EFW calculation

Author (year)	US formulas
Campbell (1975)[13]	$\text{Log}_{10}\text{BW} = -4.564 + (0.282 \times \text{AC}) - (0.00331 \times \text{AC}^2)$
Combs (1993)[14]	$\text{BW} = (0.23718 \times \text{AC}^2 \times \text{FL}) + (0.03312 \times \text{HC}^3)$
Higginbottom (1975)[15]	$\text{BW} = 0.0816 \times \text{AC}^3$
Hsieh I (1987)[16]	$\text{Log}_{10}\text{BW} = 2.1315 + (0.056541 \times \text{AC} \times \text{BPD}) - (0.00015515 \times \text{BPD} \times \text{AC}^2) + (0.000019782 \times \text{AC}^3) + (0.052594 \times \text{BPD})$
Hsieh II (1987)[16]	$\text{Log}_{10}\text{BW} = 2.7193 + (0.0094962 \times \text{AC} \times \text{BPD}) - (0.1432 \times \text{FL}) - (0.00076742 \times \text{AC} \times \text{BPD}^2) + (0.001745 \times \text{FL} \times \text{BPD}^2)$
Jordaan I (1983)[17]	$\text{Log}_{10}\text{BW} = 0.6328 + (0.1881 \times \text{AC}) - (0.0043 \times \text{AC}^2) + (0.000036239 \times \text{AC}^3)$
Jordaan II (1983)[17]	$\text{Log}_{10}\text{BW} = -1.1683 + (0.0377 \times \text{AC}) + (0.0950 \times \text{BPD}) - (0.0015 \times \text{BPD} \times \text{AC})$
Jordaan III (1983)[17]	$\text{Log}_{10}\text{BW} = 0.9119 + (0.0488 \times \text{HC}) + (0.0824 \times \text{AC}) - (0.001599 \times \text{HC} \times \text{AC})$
Jordaan IV (1983)[17]	$\text{Log}_{10}\text{BW} = 2.3231 + (0.02904 \times \text{AC}) + (0.0079 \times \text{HC}) - (0.0058 \times \text{BPD})$
Ott (1985)[18]	$\text{Log}_{10}\text{BW} = -2.0661 + (0.04355 \times \text{HC}) + (0.05394 \times \text{AC}) - (0.0008582 \times \text{HC} \times \text{AC}) + (1.2592 \times \text{FL/AC})$
Sabbagha (1989)[19]	LGA: $\text{BW} = 5426.9 - (94.98 \times \text{SUM}) + (0.54262 \times \text{SUM}^2)$ AGA: $\text{BW} = -55.3 - (16.35 \times \text{SUM}) + (0.25838 \times \text{SUM}^2)$ SGA: $\text{BW} = 1849.4 - (47.13 \times \text{SUM}) + (0.37721 \times \text{SUM}^2)$
Salomon (2007)[20]	$\text{Log10(EFW)} = 1.326 + 0.0107\ \text{HC} + 0.0438\ \text{AC} + 0.158\ \text{FL} - 0.00326\text{AC} \times \text{FL}$
Shinozuka (1987)[21]	$\text{BW} = (0.23966 \times \text{AC}^2 \times \text{FL}) + (1.6230 \times \text{BPD}^3)$
Shepard (1982)[22]	$\text{Log}_{10}\text{BW} = -1.7492 + (0.166 \times \text{BPD}) + (0.46 \times \text{AC}) - (0.002646 \times \text{AC} \times \text{BPD})$
Stirnemann (2017) (INTERGROWTH-21st)[23]	$\ln(\text{EFW}) = 5.08482 - 54.06633 \times (\text{AC}/100)^3 - 95.80076 \times (\text{AC}/100)^3 \times \log(\text{AC}/100) + 3.136370 \times (\text{HC}/100)$
Vintzileos (1987)[24]	$\text{Log}_{10}\text{BW} = 1.879 + (0.084 \times \text{BPD}) + (0.026 \times \text{AC})$
Warsof I (1977)[25]	$\text{Log}_{10}\text{BW} = -1.8367 + (0.092 \times \text{AC}) - (0.000019 \times \text{AC}^3)$
Warsof II (1977)[25]	$\text{Log}_{10}\text{BW} = -1.599 + (0.144 \times \text{BPD}) + (0.032 \times \text{AC}) - (0.000111 \times \text{BPD}^2 \times \text{AC})$
Warsof I (1986)[26]	$\text{LnBW} = 4.6914 + (0.00151 \times \text{FL}^2) - (0.000019 \times \text{FL}^3)$
Warsof II (1986)[26]	$\text{LnBW} = 2.792 + (0.108 \times \text{FL}) + (0.0036 \times \text{AC}^2) - (0.0027 \times \text{FL} \times \text{AC})$
Woo I (1985)[27]	$\text{Log}_{10}\text{BW} = 0.59 + (0.08 \times \text{AC}) + (0.28 \times \text{FL}) - (0.00716 \times \text{AC} \times \text{FL})$
Woo II (1985)[27]	$\text{Log}_{10}\text{BW} = 1.63 + (0.16 \times \text{BPD}) + (0.00111 \times \text{AC}^2) - (0.0000859 \times \text{BPD} \times \text{AC}^2)$
Woo III (1985)[27]	$\text{Log}_{10}\text{BW} = 1.54 + (0.15 \times \text{BPD}) + (0.00111 \times \text{AC}^2) - (0.0000764 \times \text{BPD} \times \text{AC}^2) + (0.05 \times \text{FL}) - (0.000992 \times \text{FL} \times \text{AC})$

AGA, appropriate for GA; LGA, large for GA; SUM, GA (week) + [2 × AC (cm)] + HC (cm) + FL (cm).

Note
AC in cm; BPD in cm; BW in g; FL in cm; HC in cm.

Clinical guidelines on fetal growth monitoring

The RCOG Green-top Guideline No. 31[6] provides evidence-based recommendations regarding the investigation and management of the SGA fetus. The guideline defines SGA as an EFW or AC < 10th percentile and severe SGA as an EFW or AC < the 3rd percentile. The recommendation is that all pregnant people should be assessed for risk factors for a SGA fetus/neonate. Methods for predicting the likelihood of SGA fetus/neonates in the first and second trimesters include medical and obstetric history and examination, parental serum screening and UtA Doppler. When risk factors are identified, further surveillance is required. The recommended methods of screening for the SGA fetus/neonate during the second and third trimesters are abdominal palpation and measurement of SFH (including customised charts).

TABLE 2 Hadlock formulas for US-based EFW calculation

Author (year)	Model	Parameters	Regression equation formula
Hadlock (1984)[28]	Model 1a	AC	Log_{10} BW = 2.695 + 0.253 (AC) – 0.00275 $(AC)^2$
Hadlock (1984)[28]	Model 1b	AC, BPD	Log_{10} BW = 1.1134 + 0.05845 (AC) – 0.000604 $(AC)^2$ + 0.007365 $(BPD)^2$ + 0.000595 (BPD × AC) + 0.1694 (BPD)
Hadlock (1984)[28]	Model 1c	AC, HC	Log_{10} BW = 1.182 + 0.0273 (HC) + 0.07057 (AC) – 0.00063 $(AC)^2$ – 0.0002184 (HC × AC)
Hadlock (1984)[28]	Model 1d	AC, FL	Log_{10} BW = 1.3598 + 0.051 (AC) + 0.1844 (FL) – 0.0037 (AC × FL)
Hadlock (1984)[28]	Model 1e	BPD, AC, FL	Log_{10} BW = 1.4787 – 0.003343 (AC × FL) + 0.001837 $(BPD)^2$ + 0.0458 (AC) + 0.158 (FL)
Hadlock (1984)[28]	Model 1f[a]	HC, AC, FL	Log_{10} BW = 1.5662 – 0.0108 (HC) + 0.0468 (AC) + 0.171 (FL) + 0.00034 $(HC)^2$ – 0.003685 (AC × FL)
Hadlock (1984)[28]	Model 1g	BPD, AC, HC, FL	Log_{10} BW = 1.5115 + 0.0436 (AC) + 0.1517 (FL) – 0.00321 (AC × FL) + 0.0006923 (BPD × HC)
Hadlock (1985)[29]	Model 2a	AC, FL	Log_{10} weight = 1.304 + 0.05281 (AC) + 0.1938 (FL) – 0.004 (AC × FL)
Hadlock (1985)[29]	Model 2b	BPD, AC, FL	Log_{10} weight = 1.335 – 0.0034 (AC × FL) + 0.0316 (BPD) + 0.0457 (AC) + 0.1623 (FL)
Hadlock (1985)[29]	Model 2c[b]	HC, AC, FL	Log_{10} weight = 1.326 – 0.00326 (AC × FL) + 0.0107 (HC) + 0.0438 (AC) + 0.158 (FL)
Hadlock (1985)[29]	Model 2d[c]	BPD, HC, AC, FL	Log_{10} weight = 1.3596 – 0.00386 (AC × FL) + 0.0064 (HC) + 0.00061 (BPD × AC) + 0.0424 (AC) + 0.174 (FL)

a Model 1f was proposed as the optimal equation by Hadlock (1984).[28]
b Model 2c was proposed as the optimal equation by Hadlock (1985).[29]
c Model 2d was proposed as the optimal equation by Hadlock (1991).[30]

The recommended method for the diagnosis of a SGA fetus is the measurement of fetal AC or EFW < 10th percentile, preferably using a customised fetal weight reference.

The National Institute for Health and Care Excellence (NICE) guideline NG201[36] on fetal growth monitoring suggests the use of the above guidance by RCOG or the NHS Saving Babies' Lives Care Bundle (SBLCB) version 2.[37] Risk assessment for FGR should be offered at the first antenatal appointment and again in the second trimester followed by SFH measurements at each antenatal appointment after 24 + 0 weeks. If these measurements indicate SGA, then US scan (USS) should be offered.

The care bundle for reducing perinatal mortality, SBLCB version 2, was released by the NHS in 2019[37] and was superseded by version 3[38] in 2023. It is comprised of five evidence-based and/or best practice elements of care, one of which covers fetal growth: risk assessment, surveillance and management.[38] Version 2 makes a distinction between SGA and FGR, stating that, although SGA is defined as EFW < 10th percentile, fetuses < 3rd centile are far more likely to be FGR. Risk assessment is used to identify people with an increased FGR risk so that they can be put into a surveillance pathway. According to version 3, women at low risk for FGR should be monitored using FH measurements before 28 + 6 weeks of gestation, while women at high risk should have a UtA Doppler assessment between 18 + 0 and 23 + 6 weeks of gestation and undergo US surveillance of fetal growth at 3–4 weekly intervals until delivery.[38]

The PI provides the customised growth chart software GROW in addition to the Growth Assessment Protocol (GAP) service specification, which details the agreement between the individual trust or health board and the PI.[39] GAP is a risk assessment-based approach, while its pathways include SFH and EFW measurements plotted serially on customised charts. Furthermore, GAP is based on training and accreditation of all staff as well as continuous audit, reporting and benchmarking of performance. The current GAP care pathway, shown in *Figure 1*, has been updated to align with the SBLCB version 3.[40]

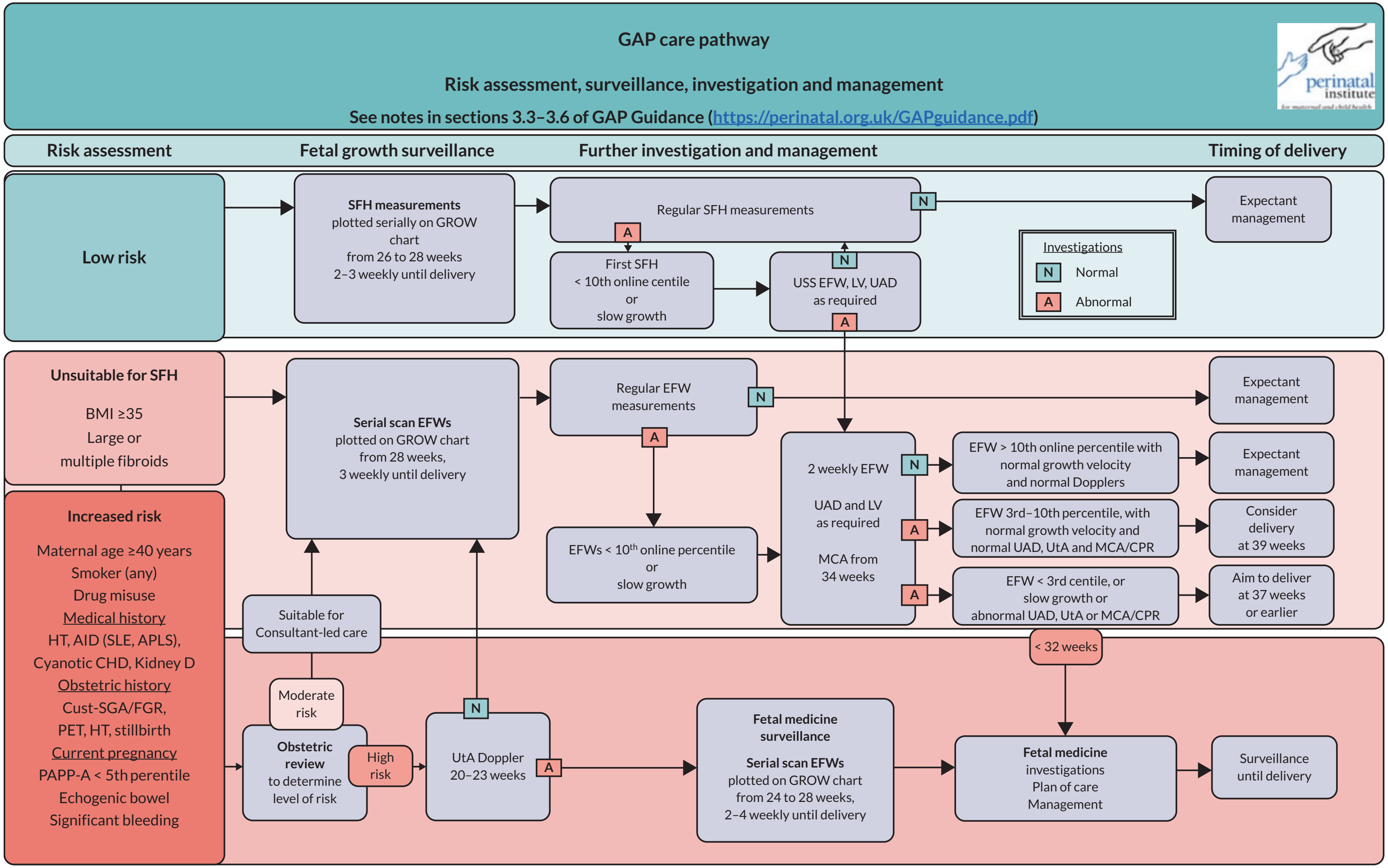


FIGURE 1 The GAP care pathway AID, autoimmune disease; APLS, antiphospholipid syndrome; CHD, congenital heart disease; D, disease; HT, hypertension; LV, left ventricular; MCA, middle cerebral artery; PAPP-A, pregnancy-associated plasma protein A; PET, pre-eclampsia; SLE, systemic lupus erythematosus; UAD, umbilical artery Doppler. Reproduced with permission from Perinatal Institute.[40]

The International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) has recently published practice guidelines for the diagnosis and management of SGA fetus and FGR.[41] They also endorse the definition of SGA as EFW or AC below the 10th percentile of a given reference range, but note that other thresholds exist, such as the 5th and 3rd percentiles or a *z*-score of –2. The ISUOG's approach is different as they have separate monitoring and managing recommendations for SGA fetuses and FGR (early-onset and late-onset FGR). Similarly to the NHS SBLCB, they point out that the key difference between SGA and FGR is that a fetus may be SGA but not at an increased risk of adverse perinatal outcome, while a non-SGA fetus may very well be FGR and at risk. Regarding FGR, they do not consider that fetal size alone is a sufficient measure (unless AC or EFW is below the third percentile) and propose monitoring fetal growth velocity as a more effective strategy. Possible FGR should be considered when there is a drop in fetal growth velocity (drop in AC or EFW) of > 2 quartiles or > 50 percentiles (e.g. from 70th percentile to or below 20th percentile).[41] In order to further distinguish between SGA and FGR, Doppler velocimetry of the uteroplacental and fetoplacental circulations can be used.

Organisations around the world have clinical practice guidelines (CPGs) providing definitions for SGA and FGR as well as recommendations for screening, diagnosis and management. The Society of Obstetricians and Gynaecologists of Canada define SGA as a fetus whose EFW is < 10th percentile on US, but do not specify their proposed growth charts.[42] They differentiate the diagnosis of SGA and pathologic growth abnormalities, stating one does not imply the other. On the other hand, the American College of Obstetricians and Gynecologists (ACOG) use the term SGA for newborns whose BW is < 10th percentile for GA and FGR for fetuses with an EFW or AC < 10th percentile for GA.[43] In Europe, both the German Society of Gynecology and Obstetrics [Deutsche Gesellschaft für Gynäkologie und Geburtshilfe (DGGG)][44] and the French College of Gynaecologists and Obstetricians [Collège National des Gynécologues et Obstétriciens Français (CNGOF)][45] use the term SGA for either in utero EFW or BW < 10th percentile and propose the use of growth curves adjusted for parental characteristics. They make a distinction to FGR/IUGR as a condition that usually corresponds with SGA, but further evidence on abnormal growth must be present, such as abnormal uterine and/or umbilical Doppler, arrest of rate of growth measured longitudinally. The CNGOF goes further, proposing IUGR should be used when fetus, placenta and amniotic fluid are all growth restricted and FGR should be used when the diagnosis is only based on BW.[45]

A risk assessment process is advised, based on parental history, by all the above guidelines. In addition, all, apart from the German guideline (DGGG),[44] propose the use of serial SFH measurements as a screening tool for restricted growth. The variation in the recommended week for the initial measurement is minimal (week 24 ± 2). The limitations of SFH determination are noted, but it is acknowledged as the only available physical examination screening test. US examination in the first trimester is the definitive method for growth restriction diagnosis, which is assisted further by additional examinations, such as amniotic fluid assessment and uterine and UAD studies.[42–45]

The HSIB 2021 National Learning Report, '*Intrapartum stillbirth: learning from maternity safety investigations that occurred during the COVID-19 pandemic, 1 April to 30 June 2020*',[1] identified 11 cases of stillborn SGA infants, of whom 8 were not detected until birth, a finding which could be viewed as further indication of the poor sensitivity of the methods currently used to detect SGA. The HSIB Report made eight safety recommendations, including Safety recommendation R/2021/148:

> *HSIB recommends that the Department of Health and Social Care commission a review to improve the reliability of existing assessment tools for fetal growth and fetal heart rate to minimise the risk for babies.*[1]

This systematic review has been commissioned and undertaken to summarise the available evidence to inform the fetal growth monitoring component of this recommendation. This systematic review focuses on methods used to diagnose a SGA fetus and FGR, as described in the RCOG guideline, and does not include methods used to assess parental risk for a SGA fetus/neonate, such as UtA Doppler or biomarkers (first trimester PAPP-A).[6]

Chapter 3 Assessment of clinical effectiveness

The systematic review followed the principles outlined in the Centre for Reviews and Dissemination (CRD) guidance for undertaking reviews in health care[46] and the Cochrane Handbook for Diagnostic Test Accuracy Reviews.[47]

Systematic review methods

Inclusion and exclusion criteria

Separate inclusion criteria were developed for each of the three research questions, and these are summarised in *Table 3*.

Search strategy

Search strategies were developed to identify studies of interventions used to identify SGA fetuses and FGR, as recommended in the CRD guidance for undertaking reviews in health care[46] and the Cochrane Handbook for Diagnostic Test Accuracy Reviews.[47]

Candidate search terms were identified from target references, browsing database thesauri [e.g. MEDLINE medical subject heading (MeSH) and EMBASE EMTREE], existing reviews and initial scoping searches. Strategy development involved an iterative approach testing candidate text and indexing terms across a sample of bibliographic databases, aiming to reach a satisfactory balance of sensitivity and specificity. Search strategies were developed specifically for each database, and the keywords and thesaurus terms were adapted according to the configuration of each database. No restrictions on language or publication status were applied. In order to maintain relevance to current clinical practice, searches were limited from 2000 to present year.

The following Resources were searched from 2000 to current year:

- MEDLINE (Ovid): 1946–20 March 2023
- MEDLINE In-Process Citations (Ovid): 1946–20 March 2023
- MEDLINE Daily Update (Ovid): 1946–20 March 2023
- MEDLINE Epub Ahead of Print (Ovid): 1946–20 March 2023
- EMBASE (Ovid): 1974–20 March 2023
- PubMed (NLM) (Internet): up to 22 March 2023
- Cochrane Database of Systematic Reviews (CDSR) (Wiley): up to March 2023/Iss3
- Cochrane Central Register of Controlled Trials (CENTRAL) (Wiley): up to February 2023/Iss2
- Database of Abstracts of Reviews of Effects (DARE) (Internet) (www.crd.york.ac.uk/CRDWeb/): up to 31 March 2015
- Kleijnen Systematic Reviews Ltd (KSR) Evidence (https://ksrevidence.com/): up to 21 March 2023
- Epistemonikos (Internet) (www.epistemonikos.org/): up to 22 March 2023
- PROSPERO (International Prospective Register of Systematic Reviews) (Internet) (http://www.crd.york.ac.uk/prospero/): up to 21 March 2023
- International Platform of Registered Systematic Review and Meta-Analysis Protocols (Internet) (https://inplasy.com/): up to 22 March 2023
- Latin American and Caribbean Health Sciences Literature (LILACS) (Internet) (http://regional.bvsalud.org/php/index.php?lang=en): up to 22 March 2023

Completed and ongoing trials were identified by searches of the following resources (no date limits were applied):

- National Institutes of Health (NIH) ClinicalTrials.gov (Internet) (www.clinicaltrials.gov/): up to 20 March 2023
- EU Clinical Trials Register (Internet) (www.clinicaltrialsregister.eu/ctr-search/search): up to 20 March 2023
- WHO International Clinical Trials Registry Platform (ICTRP) (Internet) (www.who.int/ictrp/en/): up to 20 March 2023
- ScanMedicine (Internet) (https://scanmedicine.com/): up to 20 March 2023

TABLE 3 Inclusion criteria

Item	Research question 1	Research question 2	Research question 3	Research question 4
Population	Included: • pregnant people, with and without risk factors for an FGR/SGA fetus/neonate Excluded: • pregnant people where there is a known fetal abnormality • pregnant people with monochorionic multiple pregnancies (in studies for diagnosis of selective FGR)			
Interventions	Iatrogenic delivery or other clinical intervention, following the identification of FGR/SGA by fetal growth monitoring	Fetal growth monitoring as published in practice guidelines	Included: • fetal growth monitoring methods (where the method of monitoring is defined as the complete process, including technology, timing, personnel and reference charts used) for the prediction of SGA/FGR Excluded: • tests used to assess the risk of SGA/FGR (uterine or UAD and parental blood biomarkers)	
Comparators	No iatrogenic delivery (or other intervention) for suspected FGR/SGA	SC before guideline implementation	Determination of SGA at delivery as the reference standard	
Outcomes	Rates of stillbirth; rates of neonatal death; clinical outcomes (e.g. neonatal morbidity, brain injuries, unplanned NICU admission and parental morbidity) in relation to both the intervention (e.g. iatrogenic delivery) and the result of fetal growth monitoring (e.g. TP, FP, TN and FN for FGR/SGA); rates of pre-term iatrogenic delivery and GA at iatrogenic delivery	Rates of stillbirth, rates of neonatal death, clinical outcomes (e.g. neonatal morbidity, NICU admission and parental morbidity), rates of pre-term iatrogenic delivery and GA at iatrogenic delivery before and after guideline implementation	Sufficient data to allow determination of the numbers of TP, FP, TN and FN determinations of FGR/SGA	Sufficient data to allow determination of the numbers of TP, FP, TN and FN determinations of FGR/SGA a. proportion of assessments, where a measurement cannot be obtained or where the reported result is not considered to be adequate for clinical decision-making; reported reasons why a measurement cannot be obtained or where the reported result is not considered to be adequate for clinical decision-making
Study designs	Any comparative study design (RCT, CCT or comparative observational study)		Diagnostic cohort and diagnostic case-control studies	Any study design

CCT, controlled clinical trial; FN, false negative; FP, false positive; RCT, randomised controlled trial, SC, standard care; TN, true negative; TP, true positive.

To identify conference proceedings, searches in EMBASE were not restricted to exclude conference abstracts. In addition, a search was undertaken of the following specific conference proceedings resource:

- Northern Light Life Sciences Conference Abstracts (Ovid): 2010–23/week10

The main EMBASE strategy for each search was independently peer reviewed by a second Information Specialist based on the Canadian Agency for Drugs and Technologies in Health Peer Review checklist.[48]

Additional searches

Searches of the following resources were conducted from 2013 to present year in order to identify the latest guidelines for the monitoring of fetal growth:

- TRIP database (Internet) (www.tripdatabase.com/): up to 16 March 2023
- Guidelines International Network (GIN) (Internet) (https://g-i-n.net/international-guidelines-library/): up to 16 March 2023

- National Institute for Health and Care Excellence (NICE) (Internet) (www.nice.org.uk/guidance/): up to March 2023/Iss3
- NIHR Health Technology Assessment (HTA) (Internet) (www.nihr.ac.uk/): up to 16 March 2023
- ECRI Guidelines Trust (Internet) (https://guidelines.ecri.org/): up to 31 March 2015
- International Network of Agencies for Health Technology Assessment (INAHTA) HTA Database (Internet) (https://database.inahta.org/): up to 20 March 2023
- Health Technology Assessment (HTA) Database (CRD): up to 22 March 2023

Update searches

In order to ensure no new relevant papers had been published since the original core strategies were run in March 2023, the main EMBASE and MEDLINE searches were rerun in their entirety in September 2023 before the submission of the final report. Results were deduplicated against the original search results.

- MEDLINE (Ovid): 1946–5 September 2023
- MEDLINE In-Process Citations (Ovid): 1946–5 September 2023
- MEDLINE Daily Update (Ovid): 1946–5 September 2023
- MEDLINE Epub Ahead of Print (Ovid): 1946–5 September 2023
- EMBASE (Ovid): 1974–5 September 2023

Search strategies for all the resources listed above are presented in *Appendix 1*.

Inclusion screening and data extraction

Two reviewers independently screened the titles and abstracts of all reports identified by the searches, and any discrepancies were discussed and resolved by consensus. Full copies of all studies deemed potentially relevant were obtained, and the same two reviewers independently assessed these for inclusion; any disagreements were resolved by consensus. Details of studies excluded at the full paper screening stage are presented in *Appendix 3*, *Table 21*, along with reasons for exclusion.

Where available, data were extracted on the following.

All included studies – study design/details, participant characteristics (e.g. parental characteristics, singleton or multiple pregnancy and risk factors for a SGA fetus/neonate), details of the method used to monitor fetal growth [e.g. US or SFH, frequency/timing of monitoring, training/experience of healthcare professions performing monitoring and definition of SGA/FGR (including threshold and type of reference chart used)].

Research question 1 – test performance outcome measures (numbers of TP, FP, FN and TN test results for fetal growth monitoring), details of any interventions (e.g. induction of labour and caesarean section), clinical outcomes (e.g. rates of brain injury, neonatal morbidity, NICU admission and parental morbidity) and rates of stillbirth and neonatal death.

Research question 2 – details of the intervention (implemented guideline), rates of labour interventions (e.g. induction of labour and caesarean section), clinical outcomes (e.g. rates of brain injury, neonatal morbidity, NICU admission and parental morbidity) and rates of stillbirth and neonatal death.

Research questions 3 and 4 – test performance outcome measures (data to calculate sensitivity and specificity).

Data were extracted by one reviewer, using the data extraction forms. A second reviewer checked data extraction, and any disagreements were resolved by consensus or discussion with a third reviewer. Full data extraction tables, including study details, baseline participant characteristics, details of monitoring methods and accuracy results for all monitoring strategies assessed, are provided in *Appendix 2*, *Tables 17*–*20*.

Quality assessment

The methodological quality of one included RCT (research question 2) was assessed using the revised Cochrane Risk of Bias Tool (RoB 2) for cluster randomised trials.[346] Diagnostic accuracy studies (research questions 1, 3 and 4) and RCTs, which were included for test accuracy data only (research questions 3 and 4), were assessed using Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2).[347] Separate QUADAS-2 assessments were completed for each publication from which data were extracted. Although some included studies reported comparative accuracy data, a pragmatic decision was taken not to undertake Quality Assessment of Diagnostic Accuracy Studies-Comparative (QUADAS-C) assessments. With the exception of two studies that assessed the effects of BMI[254] and diabetes[255] on test performance, included comparative accuracy studies were comparisons of the effects of applying multiple possible combinations threshold, reference chart(s) and parameters in the same cohort of study participants; as such, the application of QUADAS-C would have resulted in very large numbers of additional assessments with little informative value. The methodological quality of other study designs (research question 2) was assessed using topic-specific criteria. Quality assessment was undertaken by one reviewer and was checked by a second reviewer; any disagreements were resolved by consensus or discussion with a third reviewer. The results of the quality assessments are summarised and presented in tables (see *Study quality*).

Methods of analysis/synthesis

All meta-analyses of test accuracy studies (research question 3) involved fewer than four studies and estimated separate pooled estimates of sensitivity and specificity, using random-effects logistic regression.[348] Analyses were performed in MetaDisc.[349] Monitoring strategies were defined based on the seven key characteristics described in Monitoring methods. For comparative studies of guideline implementation (research question 2) that evaluated the same clinical outcome(s), intervention/monitoring method and comparator, summary effect estimates with 95% confidence intervals (CIs) were calculated via random-effects models using the Mantel–Haenszel (MH) method with a Paule-Mandel estimator.[350–352] R (version 4.1.3, The R Foundation for Statistical Computing, Vienna, Austria)[353] was used for meta-analysis via RStudio (RStudio version 2023.9.0.463),[354] using the additional packages meta,[355] dmetar[356] and ggplot2.[357] Where between-study heterogeneity limited the applicability of meta-analysis, a narrative summary utilising text, tables and figures has been provided (all research questions). Where data were available, the synthesis explored the impact, on growth monitoring performance, of the key variables described in Monitoring methods, as well as parental clinical characteristics/risk for a SGA fetus, singleton versus multiple pregnancies and key operational variables [e.g. timing of monitoring, type of BW growth reference chart used (customised or population-based), experience and training of clinical practitioner performing monitoring] (research question 4). The *Results of the assessment of clinical effectiveness* section of this report is structured by research question.

Results of the assessment of clinical effectiveness

The literature searches of bibliographic databases conducted for this assessment identified a total of 20,630 unique references, after deduplication. Following initial screening of titles and abstracts, 259 were considered to be potentially relevant and were ordered for full paper screening; of these, 78 were included in the review,[257,256,223–255,258–296,358–361] and 2 could not be obtained;[362,363] 4 duplicate references were identified at the full paper screening stage. *Figure 2* shows the flow of studies through the review process, and *Table 4* provides an overview of the included studies, and *Appendix 3*, *Table 21* provides details, with reasons for exclusion, of all publications excluded at the full paper screening stage.

Overview of included studies

Based on the searches and inclusion criteria described above, a total of 78 publications,[257,256,223–255,258–296,358–361] relating to 58 studies,[257,256,223–225,227–231,233,235,236,238–251,252,254,255,258,260–269,270,272,277,278,279,281,283,284–286,288,290–292,294,295,296] were included in the review; studies were cited using the primary publication (as indicated in *Table 4*) and, where this was different, the publication in which the referenced data were reported was used. Two studies could not be obtained.[362,363]

Five of the included studies provided information about the clinical effects of guideline implementation (Q2).[238,249,260,262,294] All of these studies assessed the effects of implementing the GAP care pathway,[40] and all reported rates of stillbirth and/or perinatal death before and after implementation. Three studies reported a variety of additional parental and neonatal outcomes (details provided in *Table 4*).[249,262,294] Two studies were conducted in the UK,[260,294] and the remaining studies were conducted in Australia,[262] New Zealand[249] and India.[238]

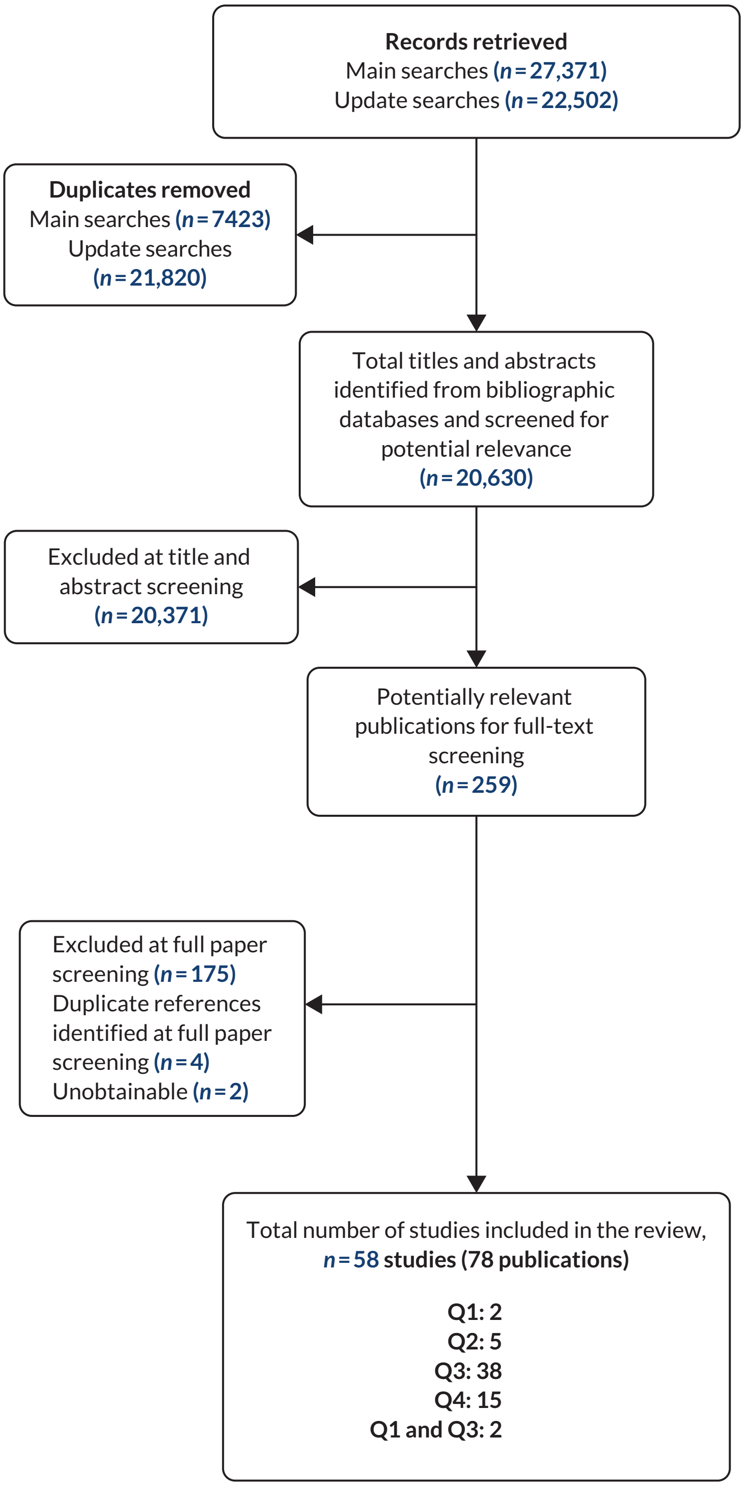


FIGURE 2 Flow of studies through the review process.

Thirty-eight studies provided information about the performance of one or more fetal growth monitoring methods for the prediction of SGA (Q3).[257,256,227–230,235,236,239–248,250,251,252,263–265,267,269,277,279,281,283,284–286,288,291,292,295,296] Two of these studies also reported some clinical outcomes conditional on test performance metrics (TP, FP, FN and TN) (Q1).[269,277] Fifteen further studies reported data to inform the effects, on the test performance of different methods of monitoring fetal growth, of key operational variables (details provided in *Table 4*).[223–225,231,232,254,255,258,261,266,268,270,272,278,290] Of the total of 53 studies that provided data on test performance, 3 were conducted in the UK,[268,278,288] 13 in other Western European or Scandinavian countries,[224,227,235,250,251,258,265–267,270,277,284,286] 19 in the USA or

TABLE 4 Overview of included studies

Study details	Country	*N*	Study design	Intervention(s) assessed	Outcome(s) reported
Q1 What are the effects, on clinical outcomes and rates of stillbirth and neonatal death, of interventions which are made based on the findings of fetal growth monitoring to detect SGA/FGR?					
Michaeli (2022)[269]	Israel	100,189	Diagnostic accuracy	EFW US	Spontaneous onset of labour; induction of labour; elective caesarean delivery; spontaneous delivery; instrumental delivery; caesarean delivery; placental complications; haemoglobin drop > 3 g/dl; use of blood products; puerperal fever; parental re-admission; BW; 1-minute Apgar score < 7; 5-minute Apgar score < 7; NICU admission; NICU for 72 hours; neonatal hospitalisation > 5 days
Nymark Hansen (2019)[277]	Denmark	2928	Diagnostic accuracy	EFW US	Caesarean delivery; elective caesarean (among caesareans); intended vaginal delivery; induction (among intended vaginal deliveries); vacuum (among intended vaginal deliveries); umbilical artery pH< 7.1; 5-minute Apgar score < 7; stillbirth; neonatal death; composite adverse neonatal outcome
Q2 What are the effects, on neonatal and parental outcomes, of implementing published guidelines for fetal growth monitoring?					
Chandra (2022)[238]	India	Before: 26,295 After: 31,264	Comparative observational	Implementation of GAP	Stillbirth at term; stillbirth from 28 weeks
Cowan (2021)[249]	New Zealand	Before: 1105 After: 1082	Comparative observational	Implementation of GAP	Stillbirth; neonatal death; induction of labour; caesarean birth; pre-term birth; post-term birth
Hugh (2021)[260]	UK	Before: 204,811 After: 199,575	Comparative observational	Implementation of GAP	Stillbirth
Jayawardena (2019)[262]	Australia	Before: 936 After: 882	Comparative observational	Implementation of GAP	Induction of labour; emergency caesarean; elective caesarean; 5-minute Apgar score < 7; admission to NISC; perinatal death; GA at birth; BW
Vieira (2022)[294]	UK	GAP: 11,096 SC: 13,810	Cluster RCT	Implementation of GAP	Induction of labour; spontaneous vaginal delivery; operative vaginal delivery; elective caesarean; emergency caesarean; postpartum haemorrhage > 1500 ml; third/fourth degree tears; epidural; episiotomy; GA at birth; pre-term birth; BW; 5-minute Apgar score < 7; umbilical artery pH < 7.1; respiratory support at birth; neonatal unit admission (HDU and SCBU); any major neonatal morbidity; any neonatal brain injury; supplemental oxygen > 28 days; necrotising enterocolitis; sepsis; retinopathy of prematurity; any minor neonatal morbidity; hypothermia; hypoglycaemia; nasogastric feeding; stillbirth; neonatal death; perinatal mortality
Q3 What is the accuracy of different methods of monitoring fetal growth for predicting SGA/FGR at delivery?					
Bais (2004)[227]	The Netherlands	NR	Diagnostic accuracy	Abdominal palpation	Test accuracy (TP, FP, FN, TN)
Barreto (2004)[228]	Brazil	250	Diagnostic accuracy	TCD/AC US	Test accuracy (TP, FP, FN, TN)
Bastek (2009)[229]	USA	93	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)

continued

DOI: 10.3310/AJLK7403

TABLE 4 Overview of included studies (*continued*)

Study details	Country	*N*	Study design	Intervention(s) assessed	Outcome(s) reported
Ben-Haroush (2007)[230]	Israel	259	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)
Bonnevier (2022)[235]	Sweden	59,452	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)
Carbone (2012)[236] Carbone (2011)[237]	USA	3329	Diagnostic accuracy	CRL US	Test accuracy (TP, FP, FN, TN)
Chang (2011)[239]	Taiwan	263	Diagnostic accuracy	st volume of upper arm US	Test accuracy (TP, FP, FN, TN)
Chang (2008)[240]	Taiwan	249	Diagnostic accuracy	RV US	Test accuracy (TP, FP, FN, TN)
Chang (2007)[241]	Taiwan	346	Diagnostic accuracy	FV US	Test accuracy (TP, FP, FN, TN)
Chang (2006a)[242]	Taiwan	417	Diagnostic accuracy	Liver volume US	Test accuracy (TP, FP, FN, TN)
Chang (2006b)[243]	Taiwan	300	Diagnostic accuracy	HV US	Test accuracy (TP, FP, FN, TN)
Chang (2005a)[244]	Taiwan	312	Diagnostic accuracy	TV US	Test accuracy (TP, FP, FN, TN)
Chang (2005b)[245]	Taiwan	482	Diagnostic accuracy	UAV US	Test accuracy (TP, FP, FN, TN)
Chauhan (2006)[246]	USA	1954	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)
Chauhan (2004)[247]	USA and Australia	178 (89 pregnancies)	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)
Chauhan (2003)[248]	USA	142	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)
De Jong (2000)[250]	The Netherlands	215	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)
De Reu (2008)[251]	The Netherlands	4224	Model development and validation	AC US; HC US	Test accuracy (TP, FP, FN, TN)
De Silva (2021)[252] De Silva (2017)[253]	Sri Lanka	508	Diagnostic accuracy	AC US; AC and HC US	Test accuracy (TP, FP, FN, TN)
Haragan (2015)[256]	USA	251	Diagnostic accuracy	EFW US; FH US; AC US; AC hh US	Test accuracy (TP, FP, FN, TN)
Harper (2013)[257]	USA	540 (270 twins)	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)
Leung (2008)[263]	China	2760	Diagnostic accuracy	CRL US	Test accuracy (TP, FP, FN, TN)
Li (2021)[264]	China	147	Diagnostic accuracy	FBV/FLV US	Test accuracy (TP, FP, FN, TN)
Lindström (2023)[265]	Sweden	31,521	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)
Martin-Palumbo (2022)[267]	Spain	488	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)
Michaeli (2022)[269]	Israel	100,189	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)

DOI: 10.3310/AJLK7403

TABLE 4 Overview of included studies (*continued*)

Study details	Country	*N*	Study design	Intervention(s) assessed	Outcome(s) reported
Nymark Hansen (2019)[277]	Denmark	2928	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)
Price (2022)[279] Price (2020)[280]	USA	612	Diagnostic accuracy	Change in EFW US; change in AC US	Test accuracy (TP, FP, FN, TN)
Rad (2018)[281] Rad (2015)[282]	USA	1594	Diagnostic accuracy	EFW US; AC US; EFW US and AC US	Test accuracy (TP, FP, FN, TN)
Ridha (2022)[283]	New Zealand	2061	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)
Salomon (2005)[284]	France	356	Diagnostic accuracy	FGP (model)	Test accuracy (TP, FP, FN, TN)
Sklar (2017)[285]	Canada	78	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)
Skrastad (2013)[286] Skrastad (2014)[287]	Norway	Study: 3190 Control: 3236	RCT	Study: Serial BPD US and mean abdominal diameter (second and third trimesters) Control: single BPD US and mean abdominal diameter (third trimester)	Test accuracy (TP, FP, FN, TN)
Sovio (2015)[288] Sovio (2014)[289]	UK	3977	Diagnostic accuracy	EFW US	Test accuracy (TP, FP, FN, TN)
Turitz (2014)[291]	USA	10,642	Diagnostic accuracy	EFW US and AC US	Test accuracy (TP, FP, FN, TN)
Tuuli (2011a)[292] Tuuli (2011b)[293]	USA	8433	Diagnostic accuracy	CRL US; EFW US; CRL US and HC US	Test accuracy (TP, FP, FN, TN)
Wan (2022)[295]	Australia	13,334	Diagnostic accuracy	SFH; EFW US and AC US; EFW US, AC US and SFH	Test accuracy (TP, FP, FN, TN)
Whitham (2023)[296]	USA	1587	Diagnostic accuracy	EFW US; EFW US and AC US	Test accuracy (TP, FP, FN, TN)
Q4 What are the effects, on the performance (accuracy) of different methods of monitoring fetal growth, of key operational variables and parental characteristics?					
Agrawal (2016)[223]	India	100	Diagnostic accuracy	TCD/AC US Variable(s) reported: Test timing	Test accuracy (TP, FP, FN, TN)
Badr (2023)[224]	Belgium	2378	Diagnostic accuracy	EFW US; EFW MRI Variable(s) reported: Reference chart for test Reference chart for BW	Test accuracy (TP, FP, FN, TN)

continued

TABLE 4 Overview of included studies (*continued*)

Study details	Country	*N*	Study design	Intervention(s) assessed	Outcome(s) reported
Baird (2016a)[225] Baird (2016b)[226]	Australia	107	Diagnostic accuracy	EFW US Variable(s) reported: Reference chart for test	Test accuracy (TP, FP, FN, TN)
Bergman (2022)[231]	Israel	6005	Diagnostic accuracy	EFW US Variable(s) reported: Reference chart for test	Test accuracy (TP, FP, FN, TN)
Blue (2019)[232] [a]**Blue (2018a)**[233] **Blue (2018b)**[234]	USA	831 1704 1226	Diagnostic accuracy	EFW US; EFW US and AC US Variable(s) reported: Reference chart for test	Test accuracy (TP, FP, FN, TN)
Dude (2021)[254]	USA	690	Diagnostic accuracy	EFW US Variable(s) reported: BMI	Test accuracy (TP, FP, FN, TN)
Dude (2018)[255]	USA	521	Diagnostic accuracy	EFW US Variable(s) reported: Diabetes	Test accuracy (TP, FP, FN, TN)
Henrichs (2019)[258] van Roekel (2023)[259]	The Netherlands	13,046	Diagnostic accuracy	AC US; change in AC US Variable(s) reported: Test timing	Test accuracy (TP, FP, FN, TN)
Humphries (2002)[261]	USA	238	Diagnostic accuracy	EFW US Variable(s) reported: Clinical practitioner	Test accuracy (TP, FP, FN, TN)
Marchand (2022)[266]	Germany	9292	Diagnostic accuracy	HC US; BPD US; AC US; FL US; TCD US; FL/AC US; TCD/AC US; HC/AC US Variable(s) reported: Test timing	Test accuracy (TP, FP, FN, TN)
Mathewlynn (2022)[268]	UK	17,678	Diagnostic accuracy	EFW US Variable(s) reported: Reference chart for test Reference chart for BW	Test accuracy (TP, FP, FN, TN)
[a]**Monier (2022a)**[270] **Monier (2022b)**[271]	France	9940	Diagnostic accuracy	EFW US Variable(s) reported: Reference chart for test	Test accuracy (TP, FP, FN, TN)

TABLE 4 Overview of included studies (*continued*)

Study details	Country	*N*	Study design	Intervention(s) assessed	Outcome(s) reported
[a]**Nwabuobi (2020)**[272] **Pressman (2022)**[273] **Odibo (2018a)**[274] Roeckner (2021)[275] Odibo (2018b)[156] Odibo (2018c)[155] Nwabuobi (2018)[276]	USA	1054	Diagnostic accuracy	EFW US; AC US; EFW US and AC US Variable(s) reported: Reference chart for test	Test accuracy (TP, FP, FN, TN)
Poljak (2017)[278]	UK	105	Diagnostic accuracy	EFW US; AC US Variable(s) reported: Reference chart for test Reference chart for BW	Test accuracy (TP, FP, FN, TN)
Temming (2017)[290]	USA	12,783	Diagnostic accuracy	EFW US Variable(s) reported: Reference chart for BW	Test accuracy (TP, FP, FN, TN)

CRL, crown–rump length; FBV, fetal brain volume; FGP, fetal growth potential; FLV, fetal liver volume; FV, femur volume; HDU, high-dependency unit; hh, hand-held; HV, humerus volume; MRI, magnetic resonance imaging; NISC, Neonatal Intensive and Special Care; RV, renal volume; SCBU, Special Care Baby Unit; st, soft tissue; TCD, transverse cerebellar diameter; TV, thigh volume; UAV, upper arm volume.
a Primary publication.

Note
Publications in bold have provided data for inclusion in this assessment.

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

Canada,[257,256,229,232,236,246–248,255,254,261,272,279,281,285,290–292,296] 3 in Australia or New Zealand,[225,283,295] 9 in China or Taiwan,[239–245,263,264] 3 in Israel[230,231,269] and 1 each in India,[223] Brazil[228] and Sri Lanka.[252]

Full details of the characteristics of study participants, study inclusion and exclusion criteria, monitoring method(s) evaluated and reference standard [for diagnostic test accuracy (DTA) studies] or comparator (for RCTs and before and after studies) are reported in the data extraction tables presented in *Appendix 2*, *Tables 17–20*.

Study quality

The methodological quality of the 53 studies that reported DTA data was assessed using QUADAS-2;[347] separate QUADAS-2 assessments were completed for each publication from which data were extracted, resulting in a total of 58 assessments. The main potential sources of bias related to patient spectrum and patient flow, while there were further concerns regarding the applicability of both the patient population and the implemented index test (growth monitoring strategy). The results of QUADAS-2 assessments are summarised in *Table 5*, and a summary of the risks of bias and applicability concerns within each QUADAS-2 domain is provided below.

Overall, the methodological quality of diagnostic accuracy studies included in this review was high; in 44 of the 58 QUADAS-2 assessments, studies were rated as having low risk of bias (RoB) and low concerns regarding applicability on all domains.

Patient spectrum

Two studies[225,229] were rated as high RoB for patient selection. In addition, two studies[250,255] did not provide sufficient information to make a judgement and were therefore rated as unclear RoB. One study[225] included only high-risk participants, defined as people with suspected FGR, based on clinical examination or an incidental finding of FGR on US, and also excluded people with PET (a high-risk characteristic). The second study[229] used a diagnostic case-control design that can lead to an exaggeration of the diagnostic accuracy.[364] One further study[251] was found on low RoB, but there were concerns around its applicability as it appeared that it did not include any nulliparous people.

Index test

Three studies[228,251,252] were assessed as high risk regarding the index test. Two[228,251] of the studies did not use a prespecified threshold, while the third one[252] used a non-standard method to define the threshold, which also raised concerns regarding its applicability. The majority of the studies reported, in detail, the index text characteristics (with respect to the seven key variables described in Monitoring methods); however, four studies[232,248,254,265] provided limited information and were assessed as unclear RoB, while in one[232] of them, the lack of information also raised concerns regarding its applicability.

Reference standard

One study[261] was assessed as high RoB because they used a specific weight as the reference standard and not a population reference chart, which is the norm for this area. The use of this reference might have introduced bias in the results. Two studies[248,254] did not report sufficient information and were assessed as unclear RoB.

Patient flow

Two studies[223,239] were assessed as high RoB because not all of the participants were included in the analysis. In one of the studies[223] which was prospective, participants were lost to the follow-up, while in the other,[239] fetuses of specific characteristics were excluded from the analysis. The remainder of the studies were low RoB.

The methodological quality of the included cluster RCT[258,286,294] (research question 2) was assessed using the revised Cochrane Risk of Bias Tool (RoB 2) for cluster randomised trials.[346] The results of this assessment are summarised in *Table 6*. The high overall RoB rating assigned to this trial arose primarily from (1) variable implementation of GAP in the intervention clusters and (2) variable implementation of the risk assessment and surveillance of FGR component of the SBLCB (commenced during the trial period) in the SC clusters.

Observational comparative studies provide a lower level of evidence with respect to the effects of an intervention than RCTs. Where observational study designs are used to provide estimates of effect, it is important to control, as

DOI: 10.3310/AJLK7403

TABLE 5 Summary of QUADAS-2 results

Study	RoB				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Agrawal (2016)[223]	✓	✓	✓	✗	✓	✓	✓
Badr (2023)[224]	✓	✓	✓	✓	✓	✓	✓
Baird (2016)[225]	✗	✓	✓	✓	✓	✓	✓
Bais (2004)[227]	✓	✓	✓	✓	✓	✓	✓
Barreto (2004)[228]	✓	✗	✓	✓✓	✓	✓	✓
Bastek (2009)[229]	✗	✓	✓	✓	✓	✓	✓
Ben-Haroush (2007)[230]	✓	✓	✓	✓	✓	✓	✓
Bergman (2022)[231]	✓	✓	✓	✓	✓	✓	✓
Blue (2018)[234]	✓	✓	✓	✓	✓	✓	✓
Blue (2018)[233]	✓	✓	✓	✓	✓	✓	✓
Blue (2019)[232]	✓	?	✓	✓	✓	✗	✓
Bonnevier (2022)[235]	✓	✓	✓	✓	✓	✓	✓
Carbone (2012)[236]	✓	✓	✓	✓	✓	✓	✓
Chang (2005)[245]	✓	✓	✓	✓	✓	✓	✓
Chang (2005)[244]	✓	✓	✓	✓	✓	✓	✓
Chang (2006)[243]	✓	✓	✓	✓	✓	✓	✓
Chang (2006)[242]	✓	✓	✓	✓	✓	✓	✓
Chang (2007)[241]	✓	✓	✓	✓	✓	✓	✓
Chang (2008)[240]	✓	✓	✓	✓	✓	✓	✓
Chang (2011)[239]	✓	✓	✓	✗	✓	✓	✓
Chauhan (2003)[248]	✓	?	?	✓	✓	✓	?
Chauhan (2004)[247]	✓	✓	✓	✓	✓	✓	✓
Chauhan (2006)[246]	✓	✓	✓	✓	✓	✓	✓
De Jong (2000)[250]	?	✓	✓	✓	✓	✓	✓
De Reu (2008)[251]	✓	✗	✓	✓	✗	✓	✓

continued

TABLE 5 Summary of QUADAS-2 results (*continued*)

	RoB				Applicability concerns		
Study	**Patient selection**	**Index test**	**Reference standard**	**Flow and timing**	**Patient selection**	**Index test**	**Reference standard**
de Silva (2021)[252]	✓	✗	✓	✓	✓	✗	✓
Dude (2018)[255]	?	✓	✓	✓	✓	✓	✓
Dude (2021)[254]	✓	?	?	✓	✓	✓	?
Haragan (2015)[256]	✓	✓	✓	✓	✓	✓	✓
Harper (2013)[257]	✓	✓	✓	✓	✓	✓	✓
Henrichs (2019)[258]	✓	✓	✓	✓	✓	✓	✓
Humphries (2002)[261]	✓	✓	✗	✓	✓	✓	✓
Leung (2008)[263]	✓	✓	✓	✓	✓	✓	✓
Li (2021)[264]	✓	✓	✓	✓	✓	✓	✓
Lindström (2023)[265]	✓	?	✓	✓	✓	✓	✓
Marchand (2022)[266]	✓	✓	✓	✓	✓	✓	✓
Martin-Palumbo (2022)[267]	✓	✓	✓	✓	✓	✓	✓
Mathewlynn (2022)[268]	✓	✓	✓	✓	✓	✓	✓
Michaeli (2022)[269]	✓	✓	✓	✓	✓	✓	✓
Monier (2022)[271]	✓	✓	✓	✓	✓	✓	✓
Monier (2022)[270]	✓	✓	✓	✓	✓	✓	✓
Nwabuobi (2020)[272]	✓	✓	✓	✓	✓	✓	✓
Nymark Hansen (2019)[277]	✓	✓	✓	✓	✓	✓	✓
Odibo (2018)[274]	✓	✓	✓	✓	✓	✓	✓
Poljak (2017)[278]	✓	✓	✓	✓	✓	✓	✓
Pressman (2022)[273]	✓	✓	✓	✓	✓	✓	✓
Price (2022)[279]	✓	✓	✓	✓	✓	✓	✓
Rad (2018)[281]	✓	✓	✓	✓	✓	✓	✓
Ridha (2022)[283]	✓	✓	✓	✓	✓	✓	✓
Salomon (2005)[284]	✓	✓	✓	✓	✓	✓	✓
Sklar (2017)[285]	✓	✓	✓	✓	✓	✓	✓

TABLE 5 Summary of QUADAS-2 results (*continued*)

Study	RoB				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Skrastad (2013)[286]	✓	✓	✓	✓	✓	✓	✓
Sovio (2015)[288]	✓	✓	✓	✓	✓	✓	✓
Temming (2017)[290]	✓	✓	✓	✓	✓	✓	✓
Turitz (2014)[291]	✓	✓	✓	✓	✓	✓	✓
Tuuli (2011)[292]	✓	✓	✓	✓	✓	✓	✓
Wan (2022)[295]	✓	✓	✓	✓	✓	✓	✓
Whitham (2023)[296]	✓	✓	✓	✓	✓	✓	✓

✓, low risk; ✗, high risk; ?, unclear risk; RoB, risk of bias.

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 6 Quality assessment of the included cluster RCT, Vieira *et al.* (2022)[294]

RoB arising from the randomisation process	***High***
1a.1 Was the allocation sequence random?	Yes
1a.2 Was the allocation sequence concealed until clusters were enrolled and assigned to interventions?	No (‘*due to the nature of the intervention, concealment was not possible*’)
1a.3 Did baseline differences between intervention groups suggest a problem with the randomisation process?	No
RoB arising from the timing of identification or recruitment of participants	***Some concerns***
1b.1 Were all the individual participants identified and recruited (if appropriate) before randomisation of clusters?	No (not appropriate)
1b.2 If no/probably no/no information to 1b.1: Is it likely that selection of individual participants was affected by knowledge of the intervention assigned to the cluster?	No (all women who gave birth in participating clusters during the year prior to randomisation and during the trial period were included)
1b.3 Were there baseline imbalances that suggest differential identification or recruitment of individual participants between intervention groups?	Yes (there were fewer participants of White ethnicity and in the least deprived quintile of the Index of Multiple Depravation in the intervention arm than in the control arm, during the outcome period)
RoB due to deviations from the intended interventions	***High***
2.1a Were participants aware that they were in a trial?	Probably no
2.1b If yes/probably yes/no information to 2.1a: Were participants aware of their assigned intervention during the trial?	N/A
2.2. Were carers and people delivering the interventions aware of participants’ assigned intervention during the trial?	Yes
2.3. If yes/probably yes/no information to 2.1b or 2.2: Were there deviations from the intended intervention that arose because of the trial context?	Yes (implementation of the intervention, GAP, was variable and implementation of the SBLCB commenced during the trial, with some clusters in the SC arm implementing the risk assessment and surveillance of FGR components of this care bundle)
2.4 if yes/probably yes to 2.3: Were these deviations likely to have affected the outcome?	Probably yes
2.5. If yes/probably yes/no information to 2.4: Were these deviations from intended intervention balanced between groups?	Probably yes
2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?	No
2.7 If no/probably no/no information to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomised?	No information
RoB due to missing outcome data	***Low***
3.1a Were data available for all clusters that recruited participants?	Yes
3.1b Were data available for all, or nearly all, participants within clusters?	Yes
3.2 If no/probably no/no information to 3.1a or 3.1b: Is there evidence that the result was not biased by missing data?	N/A
3.3 If no/probably no to 3.2 Could missingness in the outcome depend on its true value?	N/A
3.4 If yes/probably yes/no information to 3.3: Is it likely that missingness in the outcome depended on its true value?	N/A

TABLE 6 Quality assessment of the included cluster RCT, Vieira *et al.* (2022)[294] (*continued*)

RoB in measurement of the outcome	***Low***
4.1 Was the method of measuring the outcome inappropriate?	No
4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	Probably no
4.3a If no/probably no/no information to 4.1 and 4.2: Were outcome assessors aware that a trial was taking place?	Yes
4.3b If yes/probably yes/no information to 4.3a: Were outcome assessors aware of the intervention received by study participants?	Yes
4.4 If yes/probably yes/no information to 4.3b: Could assessment of the outcome have been influenced by knowledge of intervention received?	No (for the outcomes included in this systematic review)
4.5 If yes/probably yes/no information to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	N/A
RoB in selection of the reported result	
5.1 Were data analysed in accordance with a pre-specified analysis plan that was finalised before unblinded outcome data were available for analysis?	Yes
5.2. Were results likely to have been selected from multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	No
5.3. Were results likely to have been selected from multiple eligible analyses of the data?	No
Overall RoB	***High***

N/A, not applicable.

far as possible, for potential confounding factors (factors other than the intervention that may affect the outcome or outcomes being assessed), for example, by matching participants in the intervention and comparator groups on key risk factors.

The methodological quality of the four[238,249,260,262] observational comparative 'before-and-after' studies was assessed using a checklist, devised by the authors, for this review. The results of this assessment are described below and summarised in *Table 7*. Concerns were raised for one of the studies[262] as it excluded multiple pregnancies or known fetal anomalies, while the criteria used to select people for fetal growth monitoring before and after the introduction of the monitoring intervention were not the same. Assessing the quality of another study[238] was difficult due to the lack of information as only the abstract was available.

Research question 1

What are the effects, on clinical outcomes and rates of stillbirth and neonatal death, of interventions which are made based on the findings of fetal growth monitoring to detect SGA/FGR?

Two studies provided data to inform research question 1.[269,277] Both of these studies were retrospective cohort studies, which, in addition to providing some accuracy data for growth monitoring strategies using EFW US, also investigated the relationship between antenatal classification (fetal growth monitoring test result, SGA or AGA), whether or not this classification was correct in relation to BW classification of SGA or AGA (test result category TP, FP, FN or TN), and parental and neonatal clinical outcomes. Importantly, although both studies provided information about the relationship between fetal growth monitoring test result category and types of interventional delivery, it was not clear whether reported interventions were for planned pre-term delivery (iatrogenic delivery). In addition, neither of these two studies provided information about the relationship between the results of fetal growth monitoring and GA at delivery.

TABLE 7 Summary of quality assessment results for comparative observational 'before-and-after' studies

Study details	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Chandra *et al.* (2022)[238]	No	Unclear	Yes	Unclear	Unclear	Yes	Unclear	Unclear
Cowan *et al.* (2021)[249]	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hugh *et al.* (2021)[260]	No	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes
Jayawardena *et al.* (2019)[262]	Yes	No	No	Yes	Yes	Yes	Yes	Yes

Q, question.

Notes

Q1. Did the study have a prospective design?
Q2. Did the study population include an appropriate spectrum of patients (e.g. general or high-risk pregnant population, depending upon the intended application of the fetal growth monitoring intervention)?
Q3. Were the criteria used to select people for fetal growth monitoring similar, before and after the introduction of the monitoring intervention?
Q4. Were the study populations, before and after the introduction of the fetal growth monitoring intervention, similar with respect to baseline demographic characteristics (e.g. age and parity) and risk factors (e.g. HT, diabetes, smoking status and BMI)?
Q5. Other than the availability of fetal growth monitoring intervention, was the care pathway similar before and after the introduction of the monitoring intervention?
Q6. Was the implementation of the fetal growth monitoring intervention clearly described? (e.g. how were measurements and/or calculations made, who conducted examinations and what was the timing of examinations with respect to gestation)?
Q7. Were the outcomes measured in a valid and reliable way?
Q8. Was an appropriate statistical analysis used?

Michaeli *et al.* (2022)[269] evaluated the use of a single US examination (timing unclear) in an unselected general population with singleton pregnancies (i.e. universal testing), *n* = 100,198. EFW was calculated using the Hadlock 2c equation and a threshold of < 10th percentile, on local Israeli reference charts,[305] which was applied for both EFW and BW.[269] Accuracy data were reported for this test strategy.[269] This study[269] also used multivariable regression modelling to assess the association between fetal growth monitoring test classification (TP, FP and FN, with TN used as the reference group) and rates of caesarean delivery and rates of instrumental delivery; models were adjusted for maternal age, nulliparity, prior caesarean delivery, prior abortion, maternal hypertensive disorder, use of assisted reproductive technology, any background maternal clinical condition and neonatal BW. In addition, Michaeli *et al.* (2022)[269] reported simple, unadjusted comparisons of parental outcomes (rates of caesarean delivery and rates of instrumental delivery) by antennal classification [SGA (TP + FP) vs. AGA (TN + FN)] and neonatal outcomes [mean (SD) BW, 1-minute Apgar score < 7, 5-minute Apgar score < 7, NICU admission, NICU admission for ≥ 72 hours and prolonged (> 5 days) neonatal hospitalisation] by classification at birth [SGA (TP + FN) vs. AGA (FP + TN)].

Nymark Hansen *et al.* (2019)[277] evaluated the Danish national SGA screening program, wherein third-trimester growth monitoring, using EFW US, is undertaken only where clinically indicated based on the first trimester risk stratification (previous obstetric or medical history) and complications in the current pregnancy (i.e. selective testing); the evaluation was undertaken using a 1-year sample from a single university hospital, *n* = 2928. EFW was calculated using the Hadlock 1f equation and a threshold of ≤ 15th percentile on Scandinavian reference charts.[307] In Denmark, SGA is defined as BW ≤ 22nd percentile;[277] Nymark Hansen *et al.* (2019)[277] assessed the accuracy of the national SGA screening program for predicting SGA at this threshold and compared it to an alternative threshold of ≤ 15th percentile. Screening program test positive was defined as EFW ≤ threshold at the last US examination before birth, and screening test negative included those pregnancies where no EFW US was conducted after 24 weeks.[277] Sufficient data were reported to allow additional calculation of the accuracy of third-trimester EFW US, for predicting SGA at birth, in a high-risk population (i.e. the selected subgroup all of whom received third-trimester EFW US).[277] Nymark Hansen *et al.* (2019)[277] assessed the relationship between the results of fetal growth monitoring and parental and neonatal clinical outcomes by considering the consequences of an incorrect antenatal classification for each BW classification (SGA or AGA); adjusted odds ratios (ORs) (adjusted for GA at birth, percentage BW deviation, maternal BMI and parity), for each clinical outcome, were calculated for FP versus TN in AGA neonates and for FN versus TP in SGA neonates.

Both studies[269,277] reported sufficient data to allow all maternal and neonatal outcomes to be extracted by index test category (TP, FP, FN and TN), and these data are provided, in full, in *Appendix 2*, *Table 18*.

Accuracy of antennal growth monitoring for predicting smallness for gestational age at birth

Michaeli *et al.* (2022)[269] provided accuracy data for fetal growth monitoring using a single assessment of EFW at an unspecified time point, as described above; the sensitivity estimate for this test strategy was low, 26.3% (95% CI 25.5% to 27.1%), and the specificity estimate was 93.6% (95% CI 93.4% to 93.8%). The sensitivity and specificity estimates for the Danish national SGA screening program, when SGA at birth was defined as ≤ 22nd percentile, were 62.2% (95% CI 51.9% to 71.8%) and 94.4% (95% CI 93.5% to 95.2%), respectively.[277] Unsurprisingly, when the threshold for defining SGA at birth was lowered to ≤ 15th percentile, the estimated sensitivity of the screening program was reduced to 41.6% (95% CI 36.0% to 47.4%) and the estimated specificity was increased to 96.5% (95% CI 95.7% to 97.1%).[277] When sensitivity and specificity estimates were calculated for the high-risk population (i.e. the selected subgroup all of whom received third-trimester EFW US), the sensitivity estimates were higher and the specificity estimates lower than those for the unselected population at both thresholds; when SGA at birth was defined as ≤ 22nd percentile, the sensitivity and specificity were 79.2% (95% CI 68.5% to 87.6%) and 91.1% (95% CI 89.7% to 92.4%), and when the threshold was lowered to ≤ 15th percentile, the sensitivity and specificity were 85.7% (%5 CI 75.9% to 92.6%) and 83.0% (95% CI 81.2% to 84.7%).[277]

Parental clinical outcomes

The results of the multivariable regression analyses, presented in Michaeli *et al.* (2022),[269] indicated that: a FP result (incorrect antenatal prediction of SGA) was independently associated with an increased risk of caesarean delivery [OR, 1.72 (95% CI 1.56 to 1.88)] and a decreased risk of instrumental delivery [OR, 0.77 (95% CI 0.67 to 0.88)]; TP (correct antenatal prediction of SGA) was associated with an increased risk of instrumental delivery [OR, 1.40 (95% CI 1.16 to 1.68)] but not with an increased risk of caesarean delivery [OR, 1.06 (95% CI 0.9 to 1.22)]; FN (incorrect antenatal prediction of AGA) was associated with a decreased risk of caesarean delivery [OR, 0.78 (95% CI 0.69 to 0.88)] and an increased risk of instrumental delivery [OR, 1.50 (95% CI 1.32 to 1.71)].

Nymark Hansen *et al.* (2019)[277] reported that, in SGA pregnancies, failure to identify SGA antenatally, based on EFW US (FN), was associated with lower rates of induction among intended vaginal deliveries [35% (10/31)] than was the case for SGA pregnancies that were identified antenatally (TP) [83% (34/44), adjusted OR for FN vs. TP, 0.13 (95% CI 0.04 to 0.41)]. Rates of elective caesarean delivery were also lower (0) where SGA pregnancies were not identified antenatally (FN) than where SGA was identified (TP) [27% (7/26)]. Antenatal identification of SGA (TP) had no significant effects on the overall rates of caesarean delivery, intended vaginal delivery or rates of vacuum among vaginal deliveries, compared to SGA pregnancies that were not identified antenatally (FN). For AGA pregnancies, when EFW US resulting in an incorrect prediction of SGA (FP) was compared to a correct prediction of AGA (TN). Incorrect prediction of SGA was associated with increased rates of induction among intended vaginal deliveries [45% (57/128) vs. 28% (631/2221), adjusted OR for FP vs. TN, 2.51 (95% CI 1.70 to 3.71)] and non-statistically significant increase in the rates of caesarean delivery [23% (37/158) vs. 20% (535/2672), adjusted OR for FP vs. TN, 1.44 (95% CI 0.96 to 2.18)] and decrease in the rate of intended vaginal delivery [81% (128/158) vs. 83% (2221/2672), adjusted OR for FP vs. TN, 0.69 (95% CI 0.46 to 1.05)]. For all parental outcomes data, the definition of EFW US SGA was ≤ 15th percentile and the definition of BW SGA was ≤ 22nd percentile.

Where comparable parental outcomes were reported by both studies,[269,277] unadjusted ORs have been calculated from the data reported in Michaeli *et al.* (2022)[269] and results for both studies are provided in *Tables 8* and *9*.

In summary, data from both studies[269,277] indicated that both antenatal identification of SGA pregnancies and antenatal misclassification of AGA pregnancies as SGA were associated with increased rates of intervention (induction of labour and caesarean deliveries).

Neonatal outcomes

Nymark Hansen *et al.* (2019)[277] reported that, for SGA neonates, antenatal identification (TP) or non-identification (FN), by EFW US, was not associated with any significant differences in adverse neonatal outcomes (5-minute Apgar score < 7 or composite adverse outcome). Similarly, no significant differences in outcomes were observed for AGA neonates who were incorrectly predicted to be SGA (FP) and those who were correctly predicted to be AGA (TN). There was

TABLE 8 Effects of antenatal misclassification on parental outcomes in SGA pregnancies

Study	Total	SGA not identified by antenatal testing (FN)	SGA identified by antenatal testing (TP)	OR (95% CI), FN vs. TP	Adjusted OR[a] (95% CI), FN vs. TP
Caesarean delivery					
Nymark Hansen (2019)[277]	35/98	9/37	26/61	0.43 (0.17 to 1.07)	0.71 (0.24 to 2.13)
Michaeli (2022)[269]	1662/11,548	1106/8508	556/3040	0.67 (0.60 to 0.75)[b]	NR
Elective caesarean delivery					
Nymark Hansen (2019)[277]	7/35[c]	0	7/26	NC	NC
Michaeli (2022)[269]	788/11,548	498/8508	290/3040	0.59 (0.51 to 0.69)[b]	NR
Induction of labour					
Nymark Hansen (2019)[277]	44/75[d]	10/31	34/44	0.14 (0.05 to 0.39)	0.13 (0.04 to 0.41)
Michaeli (2022)[269]	2024/11,548	1055/8508	969/3040	0.30 (0.27 to 0.33)[b]	NR
Instrumental delivery					
Nymark Hansen (2019)[277]	8/63[e]	2/28	6/35	0.37 (0.07 to 2.01)	0.41 (0.07 to 2.30)
Michaeli (2022)[269]	62/11,548	51/8508	11/3040	1.09 (0.93 to 1.27)[b]	NR

NC, not calculable; NR, not reported.
a Adjusted for GA at birth, percentage BW deviation, maternal BMI and parity.
b Calculated value.
c Among all caesarean sections.
d Among intended vaginal deliveries.
e Vacuum among vaginal deliveries.

TABLE 9 Effects of incorrect antenatal prediction of SGA on parental outcomes in AGA pregnancies

Study	Total	Predicted SGA by antenatal testing (FP)	Predicted AGA by antenatal testing (TN)	OR (95% CI), FP vs. TN	Adjusted OR[a] (95% CI)
Caesarean delivery					
Nymark Hansen (2019)[277]	572/2830	37/158	535/2672	1.22 (0.84 to 1.79)	1.44 (0.96 to 2.18)
Michaeli (2022)[269]	7503/88,650	915/5671	6588/82,979	2.23 (2.07 to 2.41)[b]	NR
Elective caesarean delivery					
Nymark Hansen (2019)[277]	215/572[c]	12/37	203/535	0.79 (0.39 to 1.60)	1.49 (0.68 to 3.26)
Michaeli (2022)[269]	4461/88,650	679/5671	3782/82,979	2.85 (2.61 to 3.11)[b]	NR
Induction of labour					
Nymark Hansen (2019)[277]	688/2349[d]	57/128	631/2221	2.02 (1.41 to 2.90)	2.51 (1.70 to 3.71)
Michaeli (2022)[269]	7613/88,650	649/5671	6964/82,979	1.41 (1.30 to 1.54)[b]	NR
Instrumental delivery					
Nymark Hansen (2019)[277]	187/2258[e]	8/121	179/2137	0.77 (0.37 to 1.61)	0.66 (0.31 to 1.44)
Michaeli (2022)[269]	4794/88,650	248/5671	4546/82,979	0.79 (0.69 to 0.90)[b]	NR

a Adjusted for GA at birth, percentage BW deviation, maternal BMI and parity.
b Calculated value.
c Among all caesarean sections.
d Among intended vaginal deliveries.
e Vacuum among vaginal deliveries.

no evidence that the antenatal detection of SGE was associated with any change in the rate of stillbirths or neonatal deaths; however, it is important to note that these outcomes were only reported by Nymark Hansen *et al.* (2019)[277] and that this study was not adequately powered to detect any differences in these rare events. For all neonatal outcomes data, the definition of EFW US SGA was ≤ 15th percentile and the definition of BW SGA was ≤ 22nd percentile.

Where comparable neonatal outcomes were reported by both studies,[269,277] unadjusted ORs have been calculated from the data reported in Michaeli *et al.* (2022),[269] and results for both studies are provided in *Tables 10* and *11*. Unadjusted ORs have been calculated for NICU admission from the data reported in Michaeli *et al.* (2022).[269]

In summary, despite the apparently higher rates of intervention in SGA pregnancies that were identified antenatally, there was no clear evidence that the antenatal identification of SGA was associated with improved neonatal outcomes; the only statistically significant effect identified was an apparently lower rate of NICU admission in SGA neonates who were not identified antenatally (FN) compared with those who were correctly identified (TP), unadjusted OR, calculated from the data reported in Michaeli *et al.* (2022),[269] 0.63 (0.52, 0.76). Conversely, there was no evidence that the antenatal misclassification of AGA neonates as SGA had any statistically significant detrimental effects on outcomes. With respect to stillbirths and neonatal deaths, only the smaller of the two studies (Nymark Hansen *et al.* (2019)[277] *n* = 2928) reported data for these outcomes, and it is important to note that this study was not adequately powered to detect any differences in these rare events.

Research question 2

What are the effects, on neonatal and parental outcomes, of implementing published guidelines for fetal growth monitoring?

Five studies, four comparative observational[238,249,260,262] and one RCT,[294] assessed the clinical effects of implementing guidance on fetal growth monitoring, and all of these studies were concerned with the effects of implementing GAP. Five key neonatal and parental outcomes, which were assessed by more than one study, were synthesised in meta-analyses: stillbirths, induction of labour, caesarean births, Apgar score and admission to NICU/HDUs and SCBUs. Individual study results for the remaining neonatal and parental outcomes assessed are provided in *Appendix 2*, *Table 20*.

TABLE 10 Effects of antenatal misclassification on outcomes in SGA neonates

Study	Total	SGA not identified by antenatal testing (FN)	SGA identified by antenatal testing (TP)	OR (95% CI), FN vs. TP	Adjusted OR[a] (95% CI)
Stillbirth					
Nymark Hansen (2019)[277]	2/98	1/37	1/61	1.67 (0.10 to 27.47)	0.66 (0.02 to 27.39)
Neonatal death					
Nymark Hansen (2019)[277]	1/98	0	1/61	NC	NC
5-minute Apgar score < 7					
Nymark Hansen (2019)[277]	4/95	1/35	3/60	0.56 (0.06 to 5.59)	0.50 (0.04 to 5.78)
Michaeli (2022)[269]	185/11,548	143/8508	42/3040	1.22 (0.86 to 1.73)[b]	NR
Composite adverse neonatal outcome[c]					
Nymark Hansen (2019)[277]	11/98	3/37	8/61	0.58 (0.14 to 2.36)	0.53 (0.12 to 2.37)
NICU admission					
Michaeli (2022)[269]	510/11,548	328/8508	182/3040	0.63 (0.52 to 0.76)[b]	NR

a Adjusted for GA at birth, percentage BW deviation, maternal BMI and parity.
b Calculated value.
c Umbilical artery pH < 7.1, Apgar score < 7 after 5 minutes, stillborn and neonatal death in one variable.

TABLE 11 Effects of incorrect antenatal prediction of SGA on outcomes in AGA neonates

Study	Total	Predicted SGA by antenatal testing (FP)	Predicted AGA by antenatal testing (TN)	OR (95% CI)	Adjusted OR[a] (95% CI), FP vs. TN
Stillbirth					
Nymark Hansen (2019)[277]	9/2830	0	9/2672	NC	NC
Neonatal death					
Nymark Hansen (2019)[277]	2/2830	0	2/2672	NC	NC
5-minute Apgar score < 7					
Nymark Hansen (2019)[277]	22/2812	1/158	21/2654	0.80 (0.11 to 5.98)	0.65 (0.08 to 5.22)
Michaeli (2022)[269]	405/88,650	22/5671	383/82,979	0.84 (0.55 to 1.29)[b]	NR
Composite adverse neonatal outcome[c]					
Nymark Hansen (2019)[277]	144/2830	6/158	138/2672	0.72 (0.31 to 1.67)	0.63 (0.27 to 1.50)
NICU admission					
Michaeli (2022)[269]	755/88,650	51/5671	704/82,979	1.06 (0.80 to 1.41)[b]	NR

a Adjusted for GA at birth, percentage BW deviation, maternal BMI and parity.
b Calculated value.
c Umbilical artery pH < 7.1, Apgar score < 7 after 5 minutes, stillborn and neonatal death in one variable.

Growth Assessment Protocol implementation effect on stillbirth outcomes

The rates of stillbirths before and after GAP implementation were reported by four studies.[238,249,260,294] Two studies were conducted in the UK (one in England[294] and one in England and Wales[260]), one in India[238] and one in New Zealand.[249] Sample size varied from 2187 to 564,260 births. The MH method for binary outcomes was used instead of the generic inverse variance, fitting random-effects models. Fixed-effects models were also fitted, but the difference in the results and heterogeneity were minimal. The results of the risk ratio (RR) meta-analysis are illustrated in *Figure 3*. The pooled effect size of RR 0.79 (95% CI 0.74 to 0.84); $p = 0.0011$, was found to be statistically significant. Heterogeneity between studies was minimal, $I^2 = 0.0\%$ (95% CI 0.0% to 84.7%). No statistical outliers were identified. The results of this analysis indicate that the implementation of the GAP program reduced the risk of stillbirths. It should be noted that this analysis pooled data from studies of different designs; three of the studies were comparative observational[238,249,260] and one was a cluster RCT.[294] As such, different RoB tools were used to assess study quality (see *Study quality*). The small number of studies did not allow for meaningful subgroup analysis. However, an influence analysis using the leave-one-out method was fitted for both an effect size and I^2 influence (see *Appendix 4*). The results of which indicated that, while removal of Hugh *et al.* (2021)[260] resulted in wider CIs which crossed the line of null effect, and the direction of effect remained consistent throughout. For Vieira *et al.* (2022),[294] the only RCT, the difference in stillbirth rate (GAP–SC, adjusted for age, ethnicity, parity and stratification factor) reported in the study was −0.07 (95% CI −0.14 to −0.01); $p = 0.03$.

Study	With GAP Events	With GAP Total	Without GAP Events	Without GAP Total	Weight	RR MH, random, 95% CI
Chandra and Aziz (2022)	147	31,624	162	26,295	10.6%	0.75 [0.60 to 0.94]
Cowan *et al.* (2021)	7	1082	6	1105	0.4%	1.19 [0.40 to 3.53]
Hugh *et al.* (2021)	1095	274,721	1465	289,539	86.2%	0.79 [0.73 to 0.85]
Vieira *et al.* (2022)	34	11,096	50	13,810	2.8%	0.85 [0.55 to 1.31]
Total (95% CI)		**31,8523**		**330,749**	**100.0%**	**0.79 [0.74 to 0.84]**

Heterogeneity: $tau^2 = 0$; $chi^2 = 0.80$, $df = 3$ ($p = 0.85$); $I^2 = 0\%$

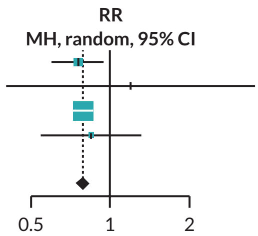


FIGURE 3 Forest plot, RR of stillbirths after GAP implementation. CI, confidence interval; GAP, Growth Assessment Protocol; MH, Mantel-Haenszel; RR, risk ratio.

Publication bias was examined in a funnel plot (*Figure 4*), where we see that the studies are following the expected shape and fall within the funnel, which does not indicate publication bias; however, the informative value of this plot is limited by the small number of studies. Egger's test was not fitted due to the small number of studies (< 10).

Growth Assessment Protocol implementation effect on induction of labour

Three studies reported the outcome of labour induction in relation of the implementation of GAP. One study was conducted in England,[294] one in Australia[262] and one in New Zealand.[249] The results of the three studies were pooled using a random-effects model.[350] The pooled OR was 1.16 (95% CI 1.01 to 1.34); $p = 0.0451$ (*Figure 5*). The result of this analysis indicated that the initiation of GAP was associated with a small increase in the number of labour inductions. This finding should be interpreted in conjunction with the rest of the perinatal outcomes. It should be noted that, as with the analysis of stillbirths, this analysis combined data from a RCT with data from observational studies. For Vieira *et al.* (2022),[294] the only RCT, the difference in the rate of induction of labour (GAP–SC, adjusted for age, ethnicity, parity and stratification factor) reported in the study was 1.7 (95% CI −0.4 to 3.8).

Heterogeneity between studies was minimal, $I^2 = 13.7\%$ (95% CI 0.0 to 91.0%). No statistical outliers or studies with excess influence on the effect size were identified (see *Appendix 4*). Publication bias was not identified in a funnel plot (*Figure 6*); however, the informative value of this plot is limited by the small number of studies.

Growth Assessment Protocol implementation effect on caesarean section operations

Caesarean section operations as an outcome of GAP implementation was reported by the same three studies.[249,262,294] Two studies[262,294] reported data separately for elective and emergency caesareans and these were summed for use in

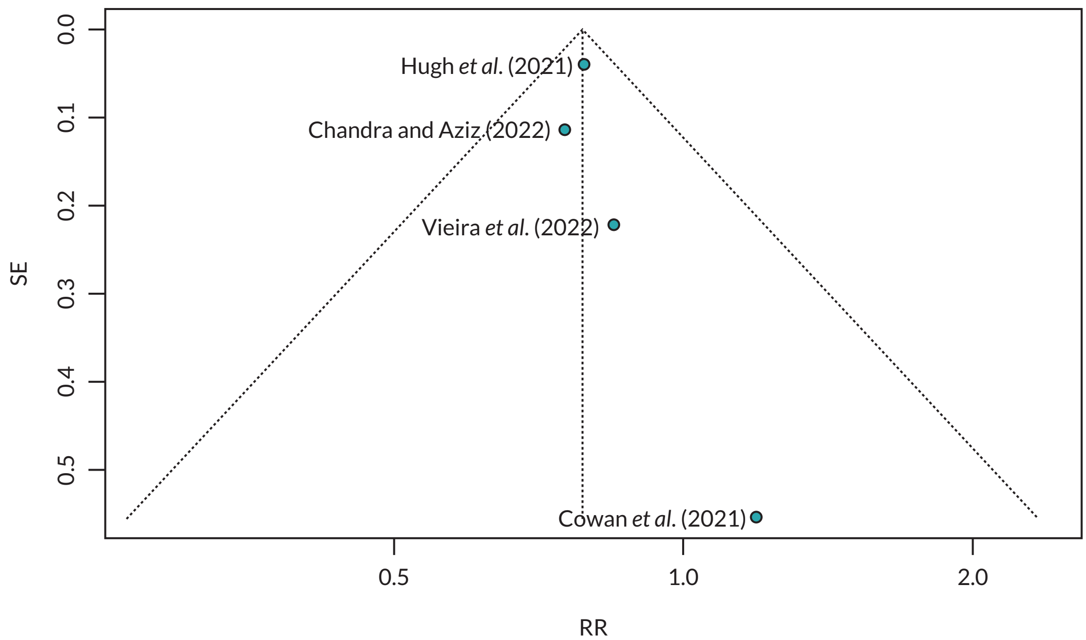


FIGURE 4 Funnel plot, RR of stillbirths after GAP implementation. GAP, Growth Assessment Protocol; RR, risk ratio; SE, standard error.

	With GAP		Without GAP			OR
Study	Events	Total	Events	Total	Weight	MH, random, 95% CI
Cowan *et al.* (2021)	368	1082	312	1105	12.0%	1.31 [1.09 to 1.57]
Jayawardena *et al.* (2019)	324	882	305	936	10.6%	1.20 [1.99 to 1.46]
Vieira *et al.* (2022)	3273	11,096	3715	13,810	77.4%	1.14 [1.08 to 1.20]
Total (95% CI)		**13,060**		**15,851**	**100.0%**	**1.16 [1.01 to 1.34]**

Heterogeneity: $tau^2 = 0.0006$; $chi^2 = 2.32$, $df = 2$ ($p = 0.31$); $I^2 = 14\%$

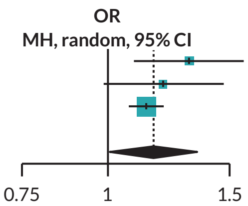


FIGURE 5 Forest plot, OR of induction of labour after GAP implementation. CI, confidence interval; GAP, Growth Assessment Protocol; MH, Mantel-Haenszl; OR, odds ratio.

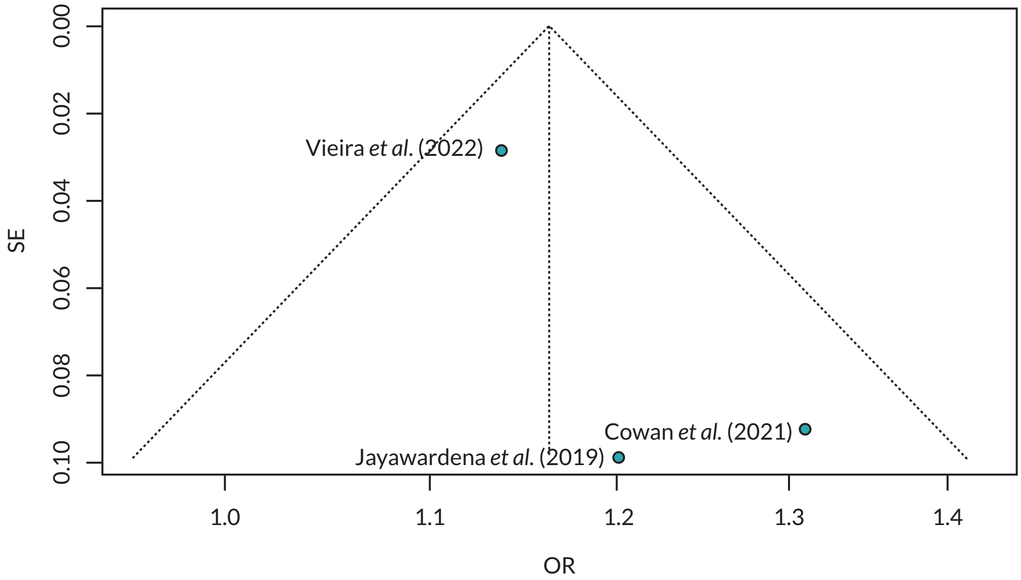


FIGURE 6 Funnel plot, OR of induction of labour after GAP implementation. GAP, Growth Assessment Protocol; MH, Mantel-Haenszel; RR, risk ratio.

this analysis. The results of the three studies were pooled using a random-effects model.[350] The pooled RR was 1.0460 (95% CI 0.86 to 1.27); $p = 0.4271$ (*Figure 7*). As shown in the forest plot, there was high between-study heterogeneity ($I^2 = 61\%$). Two studies, Jayawardena *et al.* (2019)[264] and Vieira *et al.* (2022),[294] found no difference in the rate of caesarean sections before and after the implementation of GAP, while the remaining study, Cowan *et al.* (2021),[249] found a small increase following implementation. The effect of this study in the pooled effect size is also illustrated in the influence analysis using the leave-one-out method (see *Appendix 4*). It is not clear whether this inconsistency could be partially attributed to clinical heterogeneity, related to clinical practises in New Zealand. These results should be examined in the context of the rates of caesarean section operations in these counties (England, Australia and New Zealand), during these periods, irrespective of GAP implementation. Publication bias was not identified in a funnel plot (*Figure 8*), however, the information value of this plot is limited by the small number of studies.

Growth Assessment Protocol implementation effect on Apgar score outcomes

The outcome of 5-minute Apgar score < 7 in relation to GAP implementation was reported by two of the studies[262,294] included in the previous analyses. Apgar score is used as a method of reporting the status of newborn infants and response to resuscitation.[365] An Apgar score of 4–6 is considered to be moderately abnormal. The meta-analysis of these two studies resulted in a pooled RR of 0.78 (95% CI 0.64 to 0.95) with a *p*-value = 0.0396 (*Figure 9*). Consistent with the result for the analysis of stillbirths, the result of this analysis indicates that the implementation of GAP was associated with a reduction in adverse neonatal outcomes as indicated by Apgar score. As was the case for previous analyses, it should be noted that this analysis combines results from a RCT[294] with those of an observational study;[262] while the direction of effect was consistent, only the (much larger) RCT[294] indicated a statistically significant effect. However, it should be noted that the difference in the rate of Apgar score < 7 (GAP – SC, adjusted for age, ethnicity, parity and stratification factor) reported in the RCT was not statistically significant, −0.2 (95% CI −0.4 to 0.1).[294]

	With GAP		Without GAP			RR
Study	Events	Total	Events	Total	Weight	MH, random, 95% CI
Cowan *et al.* (2021)	395	1082	348	1105	28.1%	1.16 [1.03 to 1.30]
Jayawardena *et al.* (2019)	250	882	265	936	22.1%	1.00 [0.86 to 1.16]
Vieira *et al.* (2022)	3473	11,096	4295	13,810	49.8%	1.01 [0.97 to 1.04]
Total (95% CI)		**13,060**		**15,851**	**100.0%**	**1.05 [0.86 to 1.27]**

Heterogeneity: $tau^2 = 0.0038$; $chi^2 = 5.13$, $df = 2$ ($p = 0.08$); $I^2 = 61\%$

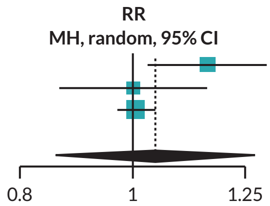


FIGURE 7 Forest plot, RR of caesarean section operations after GAP implementation. CI, confidence interval; GAP, Growth Assessment Protocol; MH, Mantel-Haenszel; RR, risk ratio.

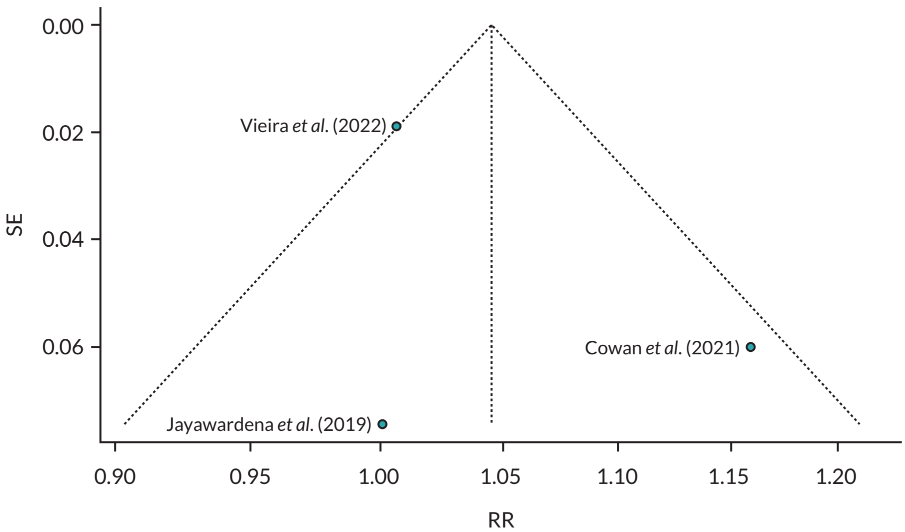


FIGURE 8 Funnel plot, RR of caesarean section operations after GAP implementation. GAP, Growth Assessment Protocol; RR, risk ratio.

Study	With GAP Events	With GAP Total	Without GAP Events	Without GAP Total	Weight	RR MH, random, 95% CI
Jayawardena *et al.* (2019)	14	882	18	936	6.3%	0.83 [0.41 to 1.65]
Vieira *et al.* (2022)	189	11,096	304	13,810	93.7%	0.77 [0.65 to 0.93]
Total (95% CI)		**11,978**		**14,746**	**100.0%**	**0.78 [0.64 to 0.95]**

Heterogeneity: tau^2 = 0; chi^2 = 0.03, *df* = 1 (*p* = 0.86); I^2 = 0%

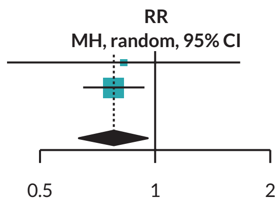


FIGURE 9 Forest plot, RR of 5-minute Apgar score < 7 after GAP implementation. CI, confidence interval; GAP, Growth Assessment Protocol; MH, Mantel-Haenszel; RR, risk ratio.

Growth Assessment Protocol implementation effect on neonatal admissions

Two studies reported admissions to NICU/HDU and SCBU.[262,294] The pooled effect estimate from these studies was an RR of 0.59 (95% CI 0.02 to 20.03); *p* = 0.3099 (*Figure 10*). The very wide CI reflects the high level of between-study heterogeneity. It should be noted that, as with all analyses reported in this section, this analysis combined data from a RCT[294] with data from an observational study.[262] The direction of effect was consistent between these two studies, however, only the RCT[294] reported a statistically significant effect. The individual data from these two studies, as used in this analysis, are broadly consistent with an association between implementation of GAP and a reduction in neonatal admissions. However, it should be noted that when the effect estimate was adjusted for age, ethnicity, parity and stratification factor (as reported in the RCT), there was no significant difference in the rate of neonatal unit admissions before and after the implementation of GAP 0.4 (95% CI −0.8 to 1.7); *p* = 0.48.[294]

Study	With GAP Events	With GAP Total	Without GAP Events	Without GAP Total	Weight	RR MH, random, 95% CI
Jayawardena *et al.* (2019)	75	882	100	936	46.8%	0.80 [0.60 to 1.06]
Vieira *et al.* (2022)	821	11,096	2237	13,810	53.2%	0.46 [0.42 to 0.49]
Total (95% CI)		**11,978**		**14,746**	**100.0%**	**0.59 [0.02 to 20.03]**

Heterogeneity: tau^2 = 0.1428; chi^2 = 13.62, *df* = 1 (*p* = < 0.01); I^2 = 93%

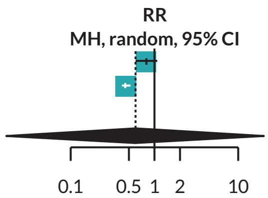


FIGURE 10 Forest plot, RR of infant admission to intensive care units after GAP implementation. CI, confidence interval; GAP, Growth Assessment Protocol; MH, Mantel-Haenszel; RR, risk ratio.

Research question 3

What is the accuracy of different methods of monitoring fetal growth for predicting smallness for gestational age at delivery?

In total, 53 studies reported data on the accuracy of fetal growth monitoring strategies for the prediction of SGA at delivery.[257,256,223–225,227–231,233,235,236,239–248,250,251,252,254,255,258,261,263–269,270,272,277,278,279,281,283,284–286,288,290–292,295,296] A subset of these studies also provided information about the effects of key operational variables on test performance (research question 4). Studies which provided information to address research question 4 are further described in Research question 4; however, some results from these studies (for individual growth monitoring strategies) are also included in this section.

Test parameter

Thirty-six different tests and combinations were used. The most frequently evaluated test was EFW US, as illustrated in *Figure 11*. Focusing only on single tests evaluated in an unselected, population, for singleton pregnancies, during the third trimester, while using a < 10th percentile threshold for defining SGA by BW, 16 different tests, and their combinations, were used: AC US, BPD US, EFW MRI, EFW US, FBV/FLV US, FV US, FL US, HC US, HV US, liver volume US, RV US, SFH, st volume of upper arm US, TCD US, TV US, UAV US. The index test as well as the type of reference growth charts used (for the test) are displayed in the tree network diagram in *Figure 12*.

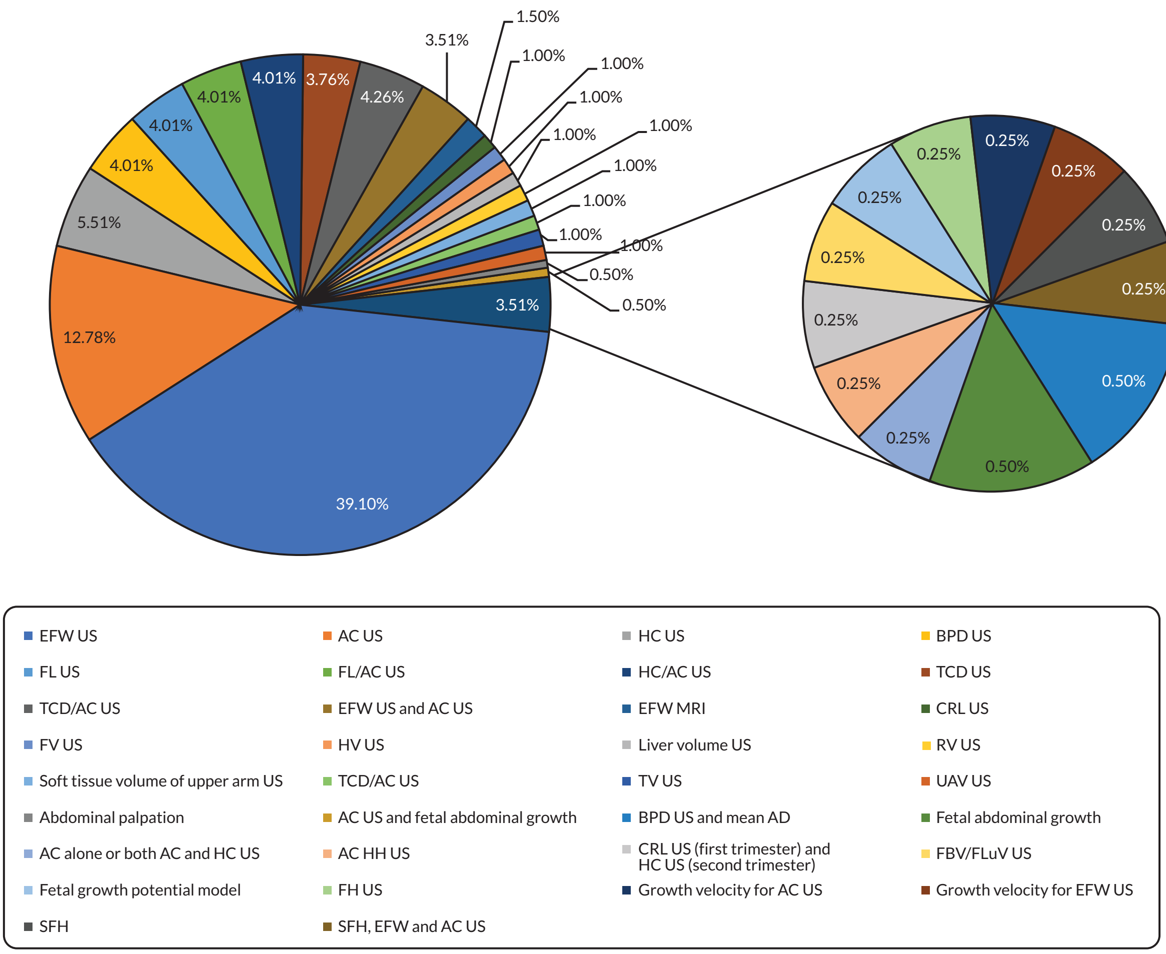


FIGURE 11 Percentage of tests applied in studies overall. AC, abdominal circumference; BPD, biparietal diameter; CRL, crown-rump length; EFW, estimated fetal weight; FBV, fetal brain volume; FLuV, fetal lung volume; FL, femur length; HC, head circumference; MRI, magnetic resonance imaging; SFH, symphysis fundal height; TCD, transverse cerebellar diameter; US, ultrasound.

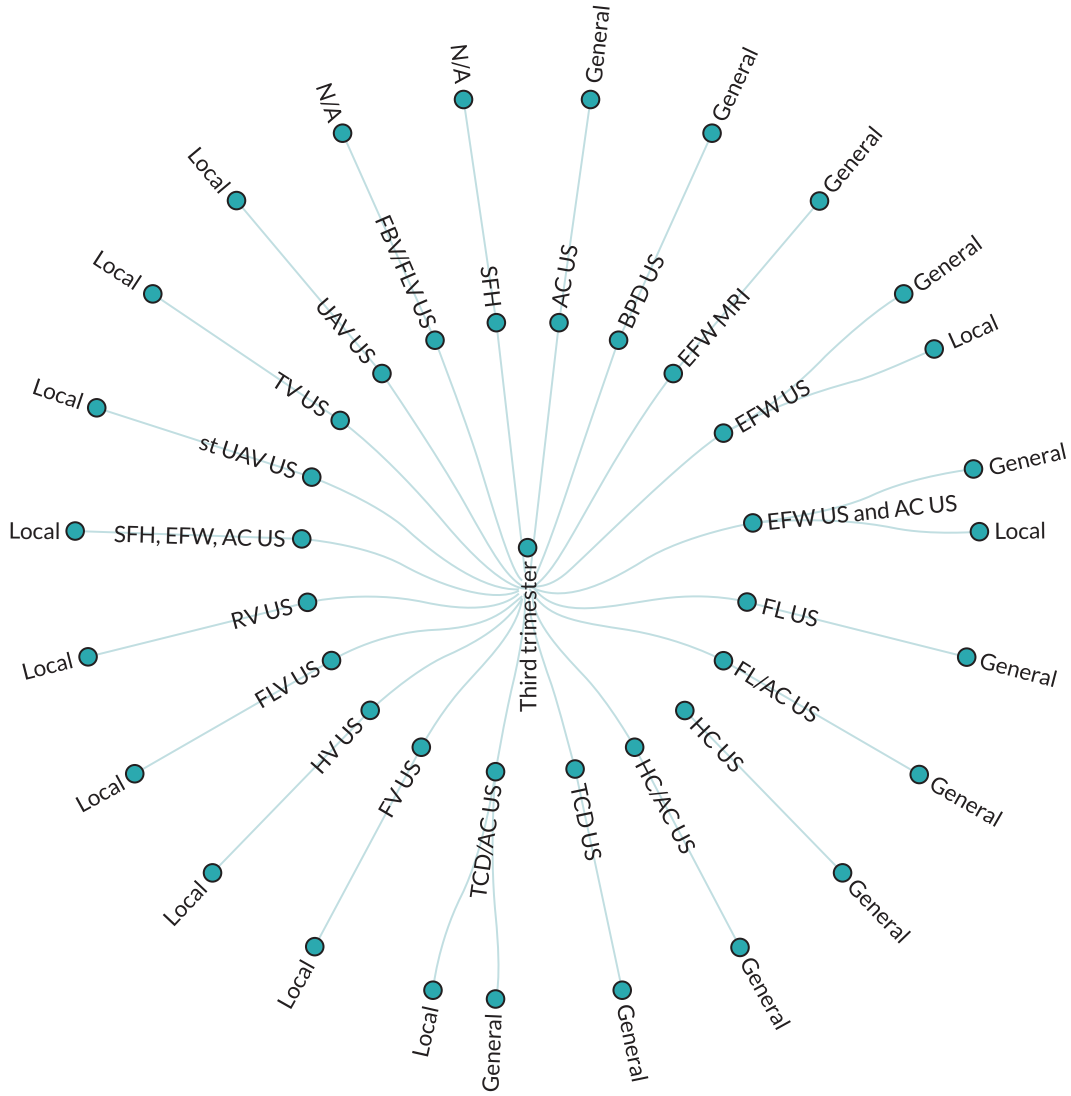


FIGURE 12 Tree network diagram, index tests. AC, abdominal circumference; BPD, biparietal diameter; EFW, estimated fetal weight; FV, femur volume; FL, femur length; FLV, fetal liver volume; HC, head circumference; hh, hand-held; HV, humerus volume; NA, not applicable, RV, renal volume; SFH, symphysis fundal height, UAV, upper arm volume; TCD, transverse cerebellar diameter; TV, thigh volume.

Single examination test strategies during third trimester

Table 12 summarises the available accuracy data for growth monitoring strategies (those based on third-trimester US assessment of EFW and/or individual biometric parameters used in formulae for EFW) for the prediction of BW below the 10th percentile (UK definition of SGA) for both unselected and high-risk populations. Summary estimates (indicated in bold) are provided where more than one study evaluated the same fetal growth monitoring strategy and used the same or comparable growth charts to determine SGA at delivery. The three most examined single US examination test strategies, both in unselected and high-risk populations, were EFW US, AC US and combinations of these parameters. A variation of the Hadlock equation was used by most of the studies for calculating EFW US, Hadlock 2c being the most popular, followed by Hadlock 2d.[29] The most frequently used threshold for the test was < 10th percentile followed by the < 5th percentile (excluding ratio tests). A comparison of the outcomes, irrespective of the reference charts used for both test and BW threshold is illustrated in *Figure 13* for unselected populations and is illustrated in *Figure 14* for high-risk populations. A further key variation in the test strategy was the choice of reference charts for both the test and neonatal BW. Studies used a wide variety of local and general reference charts, some of which were further customisable for fetal/neonatal sex and paternal characteristics (see *Research question 4*).

Unselected populations

Although the most frequently used tests were EFW US and AC US, the most accurate results reported by studies conducted in unselected populations were for tests that are not routinely used in clinical practice. The vast majority of

TABLE 12 Accuracy of single US (3rd trimester) test strategies for predicting SGA at birth in singleton pregnancies for a BW threshold of < 10th percentile

Test strategy							
Test parameter	**Estimation method (EFW)**	**Test threshold**	**Test chart category**	**BW chart category**	**Author (date)**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**
Unselected populations							
AC US	Hadlock 2c	< 10th percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Rad *et al.* (2018)[281]	64 (58 to 70)	97 (96 to 98)
AC US	Hadlock 2c	< 3rd percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Rad *et al.* (2018)[281]	37 (31 to 43)	99 (98 to 99)
AC US	Hadlock 2c	< 5th percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Rad *et al.* (2018)[281]	44 (38 to 50)	99 (98 to 99)
AC US	N/A	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	83 (78 to 88)	89 (87 to 91)
AC US	N/A	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	84 (81 to 87)	90 (89 to 91)
AC US	N/A	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	90 (84 to 94)	91 (88 to 93)
BPD US	N/A	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	65 (59 to 71)	87 (85 to 89)
BPD US	N/A	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	65 (61 to 69)	87 (85 to 88)
BPD US	N/A	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	67 (58 to 75)	86 (83 to 89)
EFW MRI	Baker (1994)[366]	< 10th percentile	General; FMF[298]	General; FMF[298]	Badr *et al.* (2023)[224]	59 (53 to 64)	98 (98 to 99)
EFW MRI	Baker (1994)[366]	< 10th percentile	General; INTERGROWTH-21st[299]	General; INTERGROWTH-21st, Villar (2014)[299]	Badr *et al.* (2023)[224]	24 (17 to 33)	100 (100 to 100)
EFW MRI	Baker (1994)[366]	< 10th percentile	General; WHO, Kiserud (2017)[4]	General; WHO, Kiserud (2017)[4]	Badr *et al.* (2023)[224]	40 (32 to 49)	100 (99 to 100)
EFW US	Hadlock 1f	< 10th percentile	Local; Duryea (2014)[345]	General; Fenton (2013)[344]	Whitham *et al.* (2023)[296]	23 (15 to 33)	100 (99 to 100)
EFW US	Hadlock 2c	< 10th percentile	General sex-specific; WHO, Kiserud (2017)[4]	Local; French BW reference charts, Mamelle (1996)[330]	Monier *et al.* (2022)[270]	36 (33 to 39)	94 (94 to 95)
EFW US	Hadlock 2c	< 10th percentile	General; FMF[298]	General; FMF[298]	Badr *et al.* (2023)[224]	70 (65 to 75)	89 (87 to 90)
EFW US	Hadlock 2c	< 10th percentile	General; FMF[298]	Local; French BW reference charts, Mamelle (1996)[330]	Monier *et al.* (2022)[271]	55 (52 to 59)	86 (86 to 87)
EFW US	Hadlock 2c	< 10th percentile	General; GROW, Gardosi (2020)[300]	General; GROW, Gardosi (2020)[300]	Mathewlynn *et al.* (2022)[268]	22 (20 to 24)	99 (99 to 99)

TABLE 12 Accuracy of single US (3rd trimester) test strategies for predicting SGA at birth in singleton pregnancies for a BW threshold of < 10th percentile (*continued*)

Test strategy							
Test parameter	**Estimation method (EFW)**	**Test threshold**	**Test chart category**	**BW chart category**	**Author (date)**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**
EFW US	Hadlock 2c	< 10th percentile	General; GROW, Gardosi (2020)[300]	General; INTERGROWTH-21st, Villar (2014)[299]	Mathewlynn *et al.* (2022)[268]	32 (28 to 35)	98 (98 to 98)
EFW US	Hadlock 2c	< 10th percentile	General; GROW, Gardosi (2020)[300]	Local; UK growth chart, Freeman (1995)[328]	Mathewlynn *et al.* (2022)[268]	22 (20 to 25)	98 (98 to 98)
EFW US	Hadlock 2c	< 10th percentile	General; GROW, Gardosi (2020)[300]	Local; UK local reference, Mathewlynn (2022)[268]	Mathewlynn *et al.* (2022)[268]	18 (16 to 20)	98 (98 to 98)
EFW US	Hadlock 2c	< 10th percentile	General; GROW, Gardosi (2020)[300]	Local; WHO, SACN, RCPCH, UK growth chart[329]	Mathewlynn *et al.* (2022)[268]	37 (33 to 40)	98 (98 to 98)
EFW US	Hadlock 2c	< 10th percentile	General; Hadlock (1991)[30]	General; GROW, Gardosi (2020)[300]	Mathewlynn *et al.* (2022)[268]	21 (19 to 23)	99 (99 to 99)
EFW US	Hadlock 2c	< 10th percentile	General; Hadlock (1991)[30]	General; INTERGROWTH-21st, Villar (2014)[299]	Mathewlynn *et al.* (2022)[268]	38 (34 to 41)	98 (98 to 98)
EFW US	Hadlock 2c	< 10th percentile	General; Hadlock (1991)[30]	Local; French BW reference charts, Mamelle (1996)[330]	Monier *et al.* (2022)[270]	38 (34 to 41)	94 (93 to 94)
EFW US	Hadlock 2c	< 10th percentile	General; Hadlock (1991)[30]	Local; UK growth chart, Freeman (1995)[328]	Mathewlynn *et al.* (2022)[268]	27 (25 to 30)	98 (98 to 99)
EFW US	Hadlock 2c	< 10th percentile (last US); universal US	General; Hadlock (1991)[30]	Local; UK growth chart, Freeman (1995)[328]	Sovio *et al.* (2015)[288]	57 (51 to 62)	90 (89 to 91)
EFW US	Hadlock 2c	< 10th percentile	General; Hadlock (1991)[30]	Local; UK growth chart, Freeman (1995)[328]	Mathewlynn *et al.* (2022)[268] and Sovio *et al.* (2015)[288]	**33.4 (31.1 to 35.8)**	**96.9 (96.6 to 97.1)**
EFW US	Hadlock 2c	< 10th percentile (last US); selective US	General; Hadlock (1991)[30]	Local; UK growth chart, Freeman (1995)[328]	Sovio *et al.* (2015)[288]	20 (16 to 24)	98 (98 to 99)
EFW US	Hadlock 2c	< 10th percentile	General; Hadlock (1991)[30]	Local; UK local reference, Mathewlynn (2022)[268]	Mathewlynn *et al.* (2022)[268]	21 (19 to 23)	99 (98 to 99)
EFW US	Hadlock 2c	< 10th percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Rad *et al.* (2018)[281]	51 (44 to 57)	98 (97 to 99)
EFW US	Hadlock 2c	< 10th percentile	General; Hadlock (1991)[30]	Local; WHO, SACN, RCPCH, UK growth chart[329]	Mathewlynn *et al.* (2022)[268]	44 (41 to 48)	98 (98 to 99)
EFW US	Hadlock 2c	< 10th percentile	General; INTERGROWTH-21st[23]	General; GROW, Gardosi (2020)[300]	Mathewlynn *et al.* (2022)[268]	11 (10 to 12)	100 (99 to 100)

continued

TABLE 12 Accuracy of single US (3rd trimester) test strategies for predicting SGA at birth in singleton pregnancies for a BW threshold of < 10th percentile (*continued*)

Test strategy							
Test parameter	Estimation method (EFW)	Test threshold	Test chart category	BW chart category	Author (date)	Sensitivity % (95% CI)	Specificity % (95% CI)
EFW US	Hadlock 2c	< 10th percentile	General; INTERGROWTH-21st[23]	General; INTERGROWTH-21st, Villar (2014)[299]	Mathewlynn *et al.* (2022)[268]	21 (18 to 24)	99 (99 to 99)
EFW US	Hadlock 2c	< 10th percentile	General; INTERGROWTH-21st[299]	General; INTERGROWTH-21st, Villar (2014)[299]	Badr *et al.* (2023)[224]	38 (29 to 47)	98 (98 to 99)
EFW US	Hadlock 2c	< 10th percentile	General; INTERGROWTH-21st[299]	General; INTERGROWTH-21st, Villar (2014)[299]	Badr *et al.* (2023)[224] and Mathewlynn *et al.* (2022)[268]	**23.4 (20.6** to **26.3)**	**99.2 (99** to **99.3)**
EFW US	Hadlock 2c	< 10th percentile	General; INTERGROWTH-21st[23]	Local; French BW reference charts, Mamelle (1996)[330]	Monier *et al.* (2022)[271]	20 (17 to 23)	98 (97 to 98)
EFW US	Hadlock 2c	< 10th percentile	General; INTERGROWTH-21st[23]	Local; UK growth chart, Freeman (1995)[328]	Mathewlynn *et al.* (2022)[268]	14 (12 to 16)	99 (99 to 100)
EFW US	Hadlock 2c	< 10th percentile	General; INTERGROWTH-21st[23]	Local; UK local reference, Mathewlynn (2022)[268]	Mathewlynn *et al.* (2022)[268]	11 (9 to 12)	100 (99 to 100)
EFW US	Hadlock 2c	< 10th percentile	General; INTERGROWTH-21st[23]	Local; WHO, SACN, RCPCH, UK growth chart[329]	Mathewlynn *et al.* (2022)[268]	25 (22 to 29)	100 (99 to 100)
EFW US	Hadlock 2c	< 10th percentile	General; WHO, Kiserud (2017)[4]	General; GROW, Gardosi (2020)[300]	Mathewlynn *et al.* (2022)[268]	21 (19 to 23)	99 (99 to 99)
EFW US	Hadlock 2c	< 10th percentile	General; WHO, Kiserud (2017)[4]	General; INTERGROWTH-21st, Villar (2014)[299]	Mathewlynn *et al.* (2022)[268]	38 (34 to 41)	98 (98 to 98)
EFW US	Hadlock 2c	< 10th percentile	General; WHO, Kiserud (2017)[4]	General; WHO, Kiserud (2017)[4]	Badr *et al.* (2023)[224]	43 (35 to 51)	94 (93 to 95)
EFW US	Hadlock 2c	< 10th percentile	General; WHO, Kiserud (2017)[4]	Local; French BW reference charts, Mamelle (1996)[330]	Monier *et al.* (2022)[271]	38 (35 to 41)	94 (93 to 94)
EFW US	Hadlock 2c	< 10th percentile	General; WHO, Kiserud (2017)[4]	Local; French BW reference charts, Mamelle (1996)[330]	Monier *et al.* (2022)[270]	35 (32 to 39)	94 (94 to 95)
EFW US	Hadlock 2c	< 10th percentile	General; WHO, Kiserud (2017)[4]	Local; UK growth chart, Freeman (1995)[328]	Mathewlynn *et al.* (2022)[268]	27 (24 to 29)	98 (98 to 99)
EFW US	Hadlock 2c	< 10th percentile	General; WHO, Kiserud (2017)[4]	Local; UK local reference, Mathewlynn (2022)[268]	Mathewlynn *et al.* (2022)[268]	21 (19 to 23)	99 (98 to 99)
EFW US	Hadlock 2c	< 10th percentile	General; WHO, Kiserud (2017)[4]	Local; WHO, SACN, RCPCH, UK growth chart[329]	Mathewlynn *et al.* (2022)[268]	44 (40 to 48)	98 (98 to 99)
EFW US	Hadlock 2c	< 10th percentile	Local ethnicity-specific; NICHD-White, Buck Louis (2015)[331]	Local; French BW reference charts, Mamelle (1996)[330]	Monier *et al.* (2022)[271]	51 (47 to 54)	89 (88 to 89)

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 12 Accuracy of single US (3rd trimester) test strategies for predicting SGA at birth in singleton pregnancies for a BW threshold of < 10th percentile (*continued*)

Test strategy							
Test parameter	**Estimation method (EFW)**	**Test threshold**	**Test chart category**	**BW chart category**	**Author (date)**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**
EFW US	Hadlock 2c	< 10th percentile	Local sex-specific; Epopé, Ego (2016)[333]	Local; French BW reference charts, Mamelle (1996)[330]	Monier *et al.* (2022)[270]	20 (17 to 22)	98 (98 to 98)
EFW US	Hadlock 2c	< 10th percentile	Local; CFEF, Massoud (2016)[332]	Local; French BW reference charts, Mamelle (1996)[330]	Monier *et al.* (2022)[271]	35 (32 to 38)	95 (94 to 95)
EFW US	Hadlock 2c	< 10th percentile	Local; Epopé, Ego (2016)[333]	Local; French BW reference charts, Mamelle (1996)[330]	Monier *et al.* (2022)[271]	20 (18 to 23)	98 (97 to 98)
EFW US	Hadlock 2c	< 10th percentile	Local; Epopé, Ego (2016)[333]	Local; French BW reference charts, Mamelle (1996)[330]	Monier *et al.* (2022)[270]	20 (17 to 23)	98 (97 to 98)
EFW US	Hadlock 2c	< 10th percentile	Local; Mathewlynn (2022)[268]	General; GROW, Gardosi (2020)[300]	Mathewlynn *et al.* (2022)[268]	41 (39 to 43)	94 (94 to 95)
EFW US	Hadlock 2c	< 10th percentile	Local; Mathewlynn (2022)[268]	General; INTERGROWTH-21st, Villar (2014)[299]	Mathewlynn *et al.* (2022)[268]	64 (60 to 67)	93 (92 to 93)
EFW US	Hadlock 2c	< 10th percentile	Local; Mathewlynn (2022)[268]	Local; UK growth chart, Freeman (1995)[328]	Mathewlynn *et al.* (2022)[268]	50 (47 to 53)	94 (93 to 94)
EFW US	Hadlock 2c	< 10th percentile	Local; Mathewlynn (2022)[268]	Local; UK local reference, Mathewlynn (2022)[268]	Mathewlynn *et al.* (2022)[268]	44 (41 to 46)	94 (94 to 95)
EFW US	Hadlock 2c	< 10th percentile	Local; Mathewlynn (2022)[268]	Local; WHO, SACN, RCPCH, UK growth chart[329]	Mathewlynn *et al.* (2022)[268]	67 (63 to 70)	93 (93 to 93)
EFW US	Hadlock 2c	< 10th percentile	Local; Salomon (2007)[20]	Local; French BW reference charts, Mamelle (1996)[330]	Monier *et al.* (2022)[271]	48 (44 to 51)	91 (90 to 91)
EFW US	Hadlock 2c	< 3rd percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Rad *et al.* (2018)[281]	16 (12 to 21)	100 (99 to 100)
EFW US	Hadlock 2c	< 5th percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Rad *et al.* (2018)[281]	24 (19 to 30)	100 (99 to 100)
EFW US	Hadlock 2d	< 10th percentile on neonatal BW curve alone	Local; Dollberg (2005)[305]	Local; Israeli growth curve, Dollberg (2005)[305]	Bergman *et al.* (2022)[231]	33 (27 to 39)	95 (88 to 98)
EFW US	Hadlock 2d	< 10th percentile on sonographic EFW curve or neonatal BW curve	Local; Sapir (2017), Dollberg (2005)[305,306]	Local; Israeli growth curve, Dollberg (2005)[305]	Bergman *et al.* (2022)[231]	46 (41 to 52)	90 (86 to 93)

continued

TABLE 12 Accuracy of single US (3rd trimester) test strategies for predicting SGA at birth in singleton pregnancies for a BW threshold of < 10th percentile (*continued*)

Test strategy							
Test parameter	Estimation method (EFW)	Test threshold	Test chart category	BW chart category	Author (date)	Sensitivity % (95% CI)	Specificity % (95% CI)
EFW US	Hadlock 2d	< 10th percentile	General; Hadlock (1991)[30]	Local; French BW reference charts, Mamelle (1996)[330]	Monier *et al.* (2022)[271]	38 (34 to 41)	94 (93 to 94)
EFW US and AC US	Hadlock 1f	AC < 10th percentile OR EFW < 10th percentile	General; Hadlock (1991)[30]	General; Fenton (2013)[344]	Whitham *et al.* (2023)[296]	30 (21 to 41)	97 (97 to 98)
EFW US and AC US	Hadlock 2b	AC < 5th percentile AND EFW < 10th percentile	Local; Alexander (1996)[303]	Local; US growth curve, Alexander (1996)[303]	Turitz *et al.* (2014)[291]	27 (25 to 29)	98 (98 to 98)
EFW US and AC US	Hadlock 2b	AC < 5th percentile OR EFW < 10th percentile	Local; Alexander (1996)[303]	Local; US growth curve, Alexander (1996)[303]	Turitz *et al.* (2014)[291]	43 (40 to 45)	94 (93 to 94)
EFW US and AC US	Hadlock 2c	AC < 10th percentile AND EFW < 10th percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Rad *et al.* (2018)[281]	47 (41 to 53)	98 (98 to 99)
EFW US and AC US	Hadlock 2c	AC < 10th percentile OR EFW < 10th percentile	General; Hadlock (1991)[30]	Local; Australian reference chart, Dobbins (2012)[301]	Wan *et al.* (2022)[295]	35 (33 to 38)	96 (96 to 96)
EFW US and AC US	Hadlock 2c	AC < 10th percentile OR EFW < 10th percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Rad *et al.* (2018)[281]	68 (62 to 73)	97 (96 to 98)
FBV/FLV US	Ratio	4.1	N/A	Local; Chinese growth chart, Dai (2014)[323]	Li *et al.* (2021)[264]	93 (81 to 99)	95 (89 to 98)
FV US	3D-view software	< 10th percentile	Local; Chang (2007)[241]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2007)[241]	71 (55 to 84)	94 (91 to 96)
FV US	3D-view software	< 25th percentile	Local; Chang (2007)[241]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2007)[241]	95 (84 to 99)	78 (73 to 82)
FV US	3D-view software	< 50th percentile	Local; Chang (2007)[241]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2007)[241]	100 (93 to 100)	59 (53 to 65)
FV US	3D-view software	< 5th percentile	Local; Chang (2007)[241]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2007)[241]	36 (22 to 52)	95 (92 to 97)
FL US	N/A	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	67 (58 to 75)	89 (86 to 92)

TABLE 12 Accuracy of single US (3rd trimester) test strategies for predicting SGA at birth in singleton pregnancies for a BW threshold of < 10th percentile (*continued*)

Test strategy							
Test parameter	Estimation method (EFW)	Test threshold	Test chart category	BW chart category	Author (date)	Sensitivity % (95% CI)	Specificity % (95% CI)
FL US	N/A	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	70 (64 to 76)	90 (88 to 92)
FL US	N/A	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	70 (66 to 74)	89 (88 to 90)
FL/AC US	Ratio	> 90th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	57 (52 to 61)	63 (61 to 65)
FL/AC US	Ratio	> 90th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	60 (53 to 66)	58 (55 to 61)
FL/AC US	Ratio	> 90th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	88 (81 to 93)	40 (36 to 44)
HC US	Sniders and Nicolaides (1994)[367]	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	64 (55 to 72)	90 (87 to 93)
HC US	Sniders and Nicolaides (1994)[367]	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	65 (59 to 71)	93 (91 to 95)
HC US	Sniders and Nicolaides (1994)[367]	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	65 (61 to 69)	91 (90 to 92)
HC/AC US	Ratio	> 90th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	27 (20 to 35)	94 (92 to 96)
HC/AC US	Ratio	> 90th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	33 (29 to 37)	92 (90 to 93)
HC/AC US	Ratio	> 90th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	34 (28 to 40)	91 (89 to 93)
HV US	Chang *et al.* (2003)[311]	< 10th percentile	Local; Chang (2003)[311]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2006)[243]	98 (87 to 100)	87 (83 to 91)
HV US	Chang *et al.* (2003)[311]	< 25th percentile	Local; Chang (2003)[311]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2006)[243]	100 (93 to 100)	73 (67 to 79)
HV US	Chang *et al.* (2003)[311]	< 50th percentile	Local; Chang (2003)[311]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2006)[243]	100 (93 to 100)	53 (47 to 59)
HV US	Chang *et al.* (2003)[311]	< 5th percentile	Local; Chang (2003)[311]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2006)[243]	36 (22 to 52)	93 (90 to 96)

continued

TABLE 12 Accuracy of single US (3rd trimester) test strategies for predicting SGA at birth in singleton pregnancies for a BW threshold of < 10th percentile (*continued*)

Test strategy							
Test parameter	Estimation method (EFW)	Test threshold	Test chart category	BW chart category	Author (date)	Sensitivity % (95% CI)	Specificity % (95% CI)
FLV US	**Chang (1997)[368]**	< 10th percentile	Local; Chang (2003)[312]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2006)[242]	98 (87 to 100)	94 (91 to 96)
FLV US	**Chang (1997)[368]**	< 25th percentile	Local; Chang (2003)[312]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2006)[242]	100 (93 to 100)	77 (73 to 81)
FLV US	**Chang (1997)[368]**	< 50th percentile	Local; Chang (2003)[312]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2006)[242]	100 (93 to 100)	38 (33 to 43)
FLV US	**Chang (1997)[368]**	< 5th percentile	Local; Chang (2003)[312]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2006)[242]	86 (71 to 95)	99 (97 to 100)
RV US	Chang (1997)[369] and Yu *et al.* (2000)[313]	< 10th percentile	Local; Yu (2000)[313]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2008)[240]	96 (82 to 100)	96 (92 to 98)
RV US	Chang (1997)[369] and Yu *et al.* (2000)[313]	< 25th percentile	Local; Yu (2000)[313]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2008)[240]	100 (90 to 100)	81 (76 to 86)
RV US	Chang (1997)[369] and Yu *et al.* (2000)[313]	< 50th percentile	Local; Yu (2000)[313]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2008)[240]	100 (90 to 100)	41 (35 to 48)
RV US	Chang (1997)[369] and Yu *et al.* (2000)[313]	< 5th percentile	Local; Yu (2000)[313]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2008)[240]	7 (1 to 24)	99 (96 to 100)
SFH	N/A	SFH 3 cm ≤ GA	N/A	Local; Australian reference chart, Dobbins (2012)[301]	Wan *et al.* (2022)[295]	22 (20 to 24)	92 (91 to 92)
SFH, EFW and AC US	Hadlock 2c	SFH 3 cm ≤ GA OR EFW < 10th percentile OR AC < 10th percentile	Local; EFW: Hadlock (1991) customised for the Australian population, AC: Chitty (1994)[30,338]	Local; Australian reference chart, Dobbins (2012)[301]	Wan *et al.* (2022)[295]	40 (37 to 42)	90 (89 to 90)
stUAV US	UAV-HV, Chang *et al.* (2002),[308] Chang *et al.* (2003)[311]	< 10th percentile	Local; Chang (2011)[239]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2011)[239]	84 (70 to 93)	89 (85 to 93)
stUAV US	UAV-HV, Chang *et al.* (2002),[308] Chang *et al.* (2003)[311]	< 25th percentile	Local; Chang (2011)[239]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2011)[239]	89 (75 to 96)	75 (69 to 81)

TABLE 12 Accuracy of single US (3rd trimester) test strategies for predicting SGA at birth in singleton pregnancies for a BW threshold of < 10th percentile (*continued*)

Test strategy							
Test parameter	**Estimation method (EFW)**	**Test threshold**	**Test chart category**	**BW chart category**	**Author (date)**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**
stUAV US	UAV-HV, Chang *et al.* (2002),[308] Chang *et al.* (2003)[311]	< 50th percentile	Local; Chang (2011)[239]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2011)[239]	95 (85 to 99)	49 (43 to 56)
stUAV US	UAV-HV, Chang *et al.* (2002),[308] Chang *et al.* (2003)[311]	< 5th percentile	Local; Chang (2011)[239]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2011)[239]	84 (70 to 93)	94 (90 to 96)
TCD US	N/A	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	20 (15 to 26)	92 (90 to 94)
TCD US	N/A	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	25 (18 to 33)	91 (88 to 93)
TCD US	N/A	< 10th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	25 (21 to 29)	91 (90 to 92)
TCD/AC US	Ratio	> 90th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	13 (8 to 20)	96 (94 to 98)
TCD/AC US	Ratio	> 90th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	24 (19 to 30)	97 (96 to 98)
TCD/AC US	Ratio	> 90th percentile	General; Hadlock (1991)[30]	Local; German growth chart, Voigt (2006)[326]	Marchand *et al.* (2022)[266]	24 (21 to 28)	96 (95 to 97)
TCD/AC US	Ratio	> 90th percentile	Local; Agrawal (2016)[223]	General; Lubchenco (1963)[297]	Agrawal *et al.* (2016)[223]	80 (52 to 96)	91 (82 to 96)
TV US	3D-view software	< 10th percentile	Local; Chang (2003)[310]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2005)[244]	87 (69 to 96)	91 (87 to 94)
TV US	3D-view software	< 25th percentile	Local; Chang (2003)[310]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2005)[244]	100 (90 to 100)	80 (75 to 85)
TV US	3D-view software	< 50th percentile	Local; Chang (2003)[310]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2005)[244]	100 (90 to 100)	54 (48 to 60)
TV US	3D-view software	< 5th percentile	Local; Chang (2003)[310]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2005)[244]	13 (4 to 31)	94 (91 to 97)
UAV US	Chang *et al.* (2002)[308]	< 10th percentile	Local; Chang (2002)[308]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2005)[245]	98 (87 to100)	93 (90 to 95)
UAV US	Chang *et al.* (2002)[308]	< 25th percentile	Local; Chang (2002)[308]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2005)[245]	100 (93 to 100)	87 (83 to 90)

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 12 Accuracy of single US (3rd trimester) test strategies for predicting SGA at birth in singleton pregnancies for a BW threshold of < 10th percentile (*continued*)

Test strategy							
Test parameter	**Estimation method (EFW)**	**Test threshold**	**Test chart category**	**BW chart category**	**Author (date)**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**
UAV US	Chang *et al.* (2002)[308]	< 50th percentile	Local; Chang (2002)[308]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2005)[245]	100 (93 to 100)	64 (59 to 69)
UAV US	Chang *et al.* (2002)[308]	< 5th percentile	Local; Chang (2002)[308]	Local; Taiwanese growth curve, Hsieh (1991)[309]	Chang *et al.* (2005)[245]	20 (9 to 36)	96 (94 to 98)
High-risk populations							
AC hh-US	N/A	< 5th percentile	General; Hadlock (1991)[30]	Local; US growth curves, Olsen (2010)[320]	Haragan *et al.* (2015)[256]	74 (54 to 89)	90 (85 to 93)
AC US	Hadlock 2c	< 5th percentile	General; Hadlock (1991)[30]	Local; US growth curves, Olsen (2010)[320]	Haragan *et al.* (2015)[256]	56 (35 to 75)	96 (93 to 98)
AC US	Hadlock 2d	< 10th percentile	General; Hadlock (1991)[30]	Local sex-specific; US growth curve, Alexander (1996)[303]	Blue *et al.* (2018)[233]	59 (52 to 65)	92 (91 to 93)
AC US	Hadlock 2d	< 5th percentile	General; Hadlock (1991)[30]	Local sex-specific; US growth curve, Alexander (1996)[303]	Blue *et al.* (2018)[233]	47 (40 to 53)	95 (94 to 96)
AC US	N/A	< 10th percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Pressman *et al.* (2022)[273]	50 (42 to 59)	94 (93 to 96)
EFW US	Combs *et al.* (1993)[14]	< 10th percentile	Local; Alexander (1996)[303]	Local; US growth curve, Alexander (1996)[303]	Chauhan *et al.* (2003)[248]	73 (48 to 91)	88 (77 to 95)
EFW US	Combs *et al.* (1993)[14]	< 10th percentile	Local; Alexander (1996)[303]	Local; US growth curve, Alexander (1996)[303]	Chauhan *et al.* (2003)[248]	80 (48 to 97)	65 (52 to 76)
EFW US	Customised standard	< 10th percentile	Individualised; Gardosi (2009)[334]				
EFW US	Hadlock 2b	< 10th percentile	Local; Alexander (1996)[303]	Local; US growth curve, Alexander (1996)[303]	Chauhan *et al.* (2006)[246]	64 (60 to 68)	94 (93 to 95)
EFW US	Hadlock 2c	< 10th percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Nwabuobi *et al.* (2020)[272]	42 (33 to 50)	96 (95 to 98)
EFW US	Hadlock 2c	< 10th percentile	General; Hadlock (1991)[30]	Local; US growth curves, Olsen (2010)[320]	Haragan *et al.* (2015)[256]	22 (9 to 42)	100 (98 to 100)
EFW US	Hadlock 2c	< 10th percentile	General; INTERGROWTH-21st[299]	Local; US growth curve, Alexander (1996)[303]	Nwabuobi *et al.* (2020)[272]	24 (18 to 32)	97 (95 to 98)
EFW US	Hadlock 2d	< 10th percentile	General; Hadlock (1991)[30]	Local sex-specific; US growth curve, Alexander (1996)[303]	Blue *et al.* (2018)[233]	59 (52 to 65)	94 (92 to 95)
EFW US	Hadlock 2d	< 10th percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Blue *et al.* (2019)[232]	71 (63 to 78)	92 (90 to 94)

TABLE 12 Accuracy of single US (3rd trimester) test strategies for predicting SGA at birth in singleton pregnancies for a BW threshold of < 10th percentile (*continued*)

Test strategy							
Test parameter	**Estimation method (EFW)**	**Test threshold**	**Test chart category**	**BW chart category**	**Author (date)**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**
EFW US	Hadlock 2d	< 15th percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Blue *et al.* (2019)[232]	84 (77 to 90)	86 (83 to 88)
EFW US	Hadlock 2d	< 5th percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Blue *et al.* (2019)[232]	55 (46 to 64)	95 (93 to 97)
EFW US	INTERGROWTH-21st[23]	< 10th percentile	Local; INTERGROWTH-21st[23]	Local; US growth curve, Alexander (1996)[303]	Blue *et al.* (2019)[232]	58 (49 to 66)	94 (92 to 96)
EFW US	INTERGROWTH-21st[23]	< 22nd percentile	Local; INTERGROWTH-21st[23]	Local; US growth curve, Alexander (1996)[303]	Blue *et al.* (2019)[232]	83 (75 to 89)	87 (84 to 89)
EFW US	INTERGROWTH-21st[23]	< 5th percentile	Local; INTERGROWTH-21st[23]	Local; US growth curve, Alexander (1996)[303]	Blue *et al.* (2019)[232]	46 (38 to 55)	96 (94 to 97)
EFW US	[a]Mikolajcyzk *et al.* (2011)[336]	< 10th percentile	General ethnicity-specific; GROW (with ethnicity), Gardosi (2020)[300]	Local; Australian reference chart, Dobbins (2012)[301]	Baird *et al.* (2016)[225]	87 (78 to 94)	28 (13 to 47)
EFW US	[a]Mikolajcyzk *et al.* (2011)[336]	< 10th percentile	General; GROW (no ethnicity), Gardosi (2020)[300]	Local; Australian reference chart, Dobbins (2012)[301]	Baird *et al.* (2016)[225]	88 (79 to 95)	28 (13 to 47)
EFW US	[a]Mikolajcyzk *et al.* (2011)[336]	< 10th percentile	General; Hadlock (1991)[30]	Local; Australian reference chart, Dobbins (2012)[301]	Baird *et al.* (2016)[225]	77 (66 to 86)	45 (26 to 64)
EFW US	[a]Mikolajcyzk *et al.* (2011)[336]	< 10th percentile	Local; Dobbins (2012)[301]	Local; Australian reference chart, Dobbins (2012)[301]	Baird *et al.* (2016)[225]	68 (56 to 78)	69 (49 to 85)
EFW US	Salomon *et al.* (2007)[20]	< 10th percentile	Local; Salomon (2007)[20]	Local; US growth curve, Alexander (1996)[303]	Blue *et al.* (2019)[232]	43 (34 to 52)	96 (95 to 98)
EFW US	Salomon *et al.* (2007)[20]	< 53rd percentile	Local; Salomon (2007)[20]	Local; US growth curve, Alexander (1996)[303]	Blue *et al.* (2019)[232]	74 (66 to 81)	75 (72 to 78)
EFW US	Salomon *et al.* (2007)[20]	< 5th percentile	Local; Salomon (2007)[20]	Local; US growth curve, Alexander (1996)[303]	Blue *et al.* (2019)[232]	34 (26 to 43)	98 (96 to 99)
EFW US and AC US	Hadlock 2c	AC < 10th percentile OR EFW < 10th percentile	General; Hadlock (1991)[30]	Local; US growth curve, Alexander (1996)[303]	Pressman *et al.* (2022)[273]	55 (46 to 63)	93 (92 to 95)
EFW US and AC US	Hadlock 2d	AC < 10th percentile AND EFW < 10th percentile	General; Hadlock (1991)[30]	Local sex-specific; US growth curve, Alexander (1996)[303]	Blue *et al.* (2018)[233]	52 (45 to 58)	95 (94 to 96)

continued

DOI: 10.3310/AJLK7403

TABLE 12 Accuracy of single US (3rd trimester) test strategies for predicting SGA at birth in singleton pregnancies for a BW threshold of < 10th percentile (*continued*)

Test strategy							
Test parameter	**Estimation method (EFW)**	**Test threshold**	**Test chart category**	**BW chart category**	**Author (date)**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**
EFW US and AC US	Hadlock 2d	AC < 10th percentile AND normal EFW	General; Hadlock (1991)[30]	Local sex-specific; US growth curve, Alexander (1996)[303]	Blue *et al.* (2018)[233]	7 (4 to 11)	97 (96 to 98)
EFW US and AC US	Hadlock 2d	AC < 10th percentile OR EFW < 10th percentile	General; Hadlock (1991)[30]	Local sex-specific; US growth curve, Alexander (1996)[303]	Blue *et al.* (2018)[233]	66 (60 to 72)	91 (89 to 92)
EFW US and AC US	Hadlock 2d	AC < 5th percentile AND normal EFW	General; Hadlock (1991)[30]	Local sex-specific; US growth curve, Alexander (1996)[303]	Blue *et al.* (2018)[233]	3 (1 to 7)	99 (98 to 100)
EFW US and AC US	Hadlock 2d	AC < 5th percentile OR EFW < 10th percentile	General; Hadlock (1991)[30]	Local sex-specific; US growth curve, Alexander (1996)[303]	Blue *et al.* (2018)[233]	62 (56 to 68)	93 (91 to 94)
FH US	N/A	FH (size less than dates)	General; Hadlock (1991)[30]	Local; US growth curves, Olsen (2010)[320]	Haragan *et al.* (2015)[256]	37 (19 to 58)	95 (91 to 98)
Low-risk populations							
AC US	N/A	< 10th percentile	Local; Verburg (2008)[321]	Local; Dutch growth curve, Verburg (2008)[321]	Henrichs *et al.* (2019)[258]	11 (9 to 14)	99 (99 to 99)
AC US	N/A	< 10th percentile	Local; Verburg (2008)[321]	Local; Dutch growth curve, Verburg (2008)[321]	Henrichs *et al.* (2019)[258]	13 (11 to 16)	98 (98 to 98)
AC US	N/A	< 25th percentile	Local; Derivation cohort	Local; Dutch growth chart, Visser (2009)[315]	De Reu *et al.* (2008)[251]	59 (46 to 71)	87 (84 to 90)
EFW US	Hadlock 2b	< 10th percentile	Local; Leiberman (1993) cited in paper	Local; Israeli growth chart, Leiberman (1993) cited in paper	Ben-Haroush *et al.* (2007)[230]	21 (6 to 46)	97 (94 to 99)
EFW US	Hadlock 2b	< 35th percentile	Local; Leiberman (1993) cited in paper	Local; Israeli growth chart, Leiberman (1993) cited in paper	Ben-Haroush *et al.* (2007)[230]	42 (20 to 67)	90 (85 to 93)
EFW US	Hadlock 2d	< 10th percentile	General; Hadlock (1991)[30]	Local; Yudkin (1987) cited in paper	Martin-Palumbo *et al.* (2022)[267]	31 (21 to 42)	99 (97 to 100)

CFEF, Collège Français d'Échographie Foetale; FLV, fetal liver volume; NICHD, National Institute of Child Health and Human Development; RCPCH, Royal College of Paediatrics and Child Health; SACN, Scientific Advisory Committee on Nutrition; TCD, transverse cerebellar diameter.
a Adapted for the Australian population by anchorage of the EFW formula to the mean BW at 40.0 weeks for Australian births, based on Dobbins *et al.* (2012).[301]

Note
Values in bold are pooled estimates.

DOI: 10.3310/AJLK7403

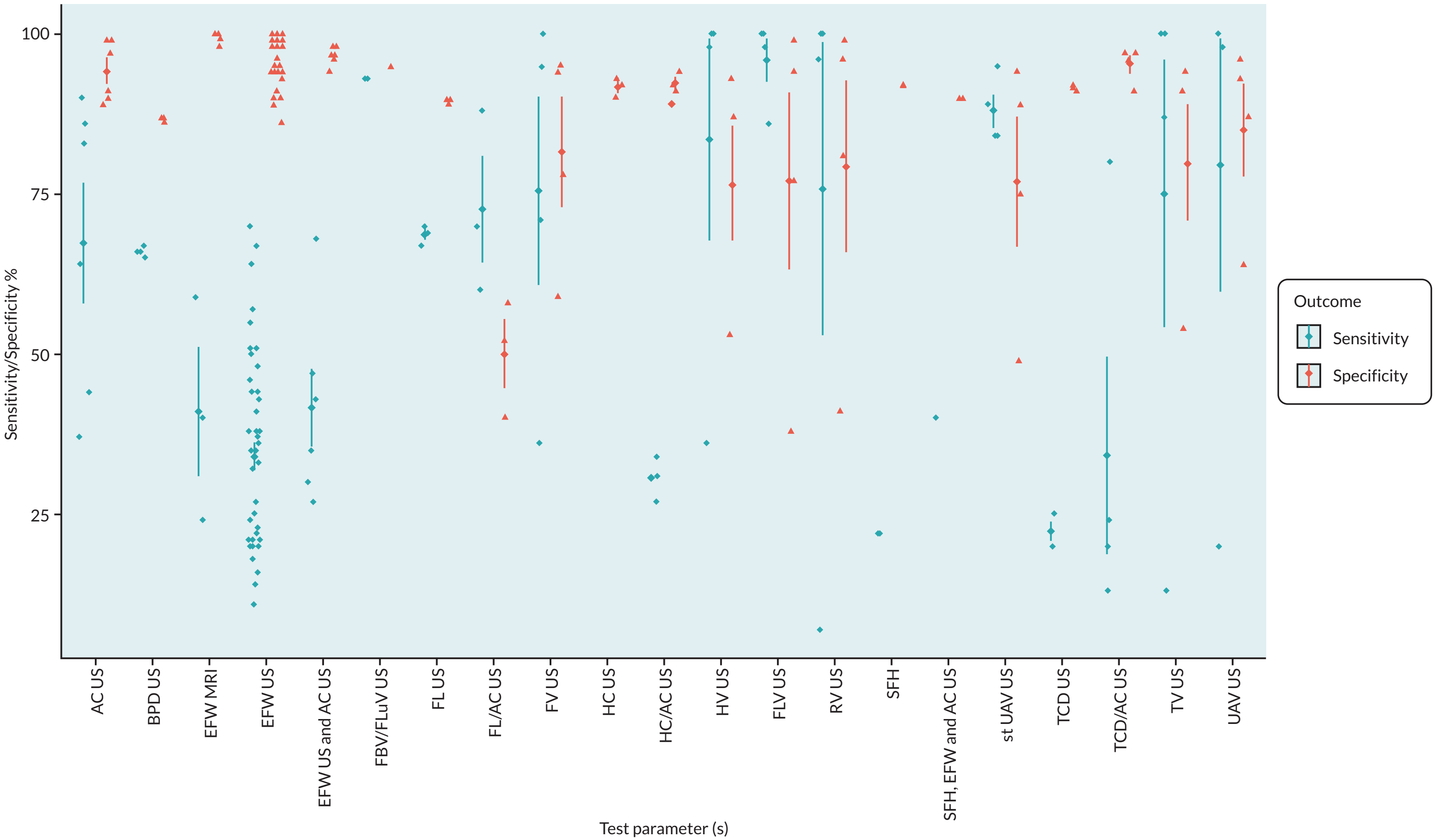


FIGURE 13 Accuracy outcomes of single test strategies, applied in unselected populations, during the 3rd trimester targeting a BW < 10th percentile. Note: the lines illustrate the mean and SE for each test. AC, abdominal circumference; BPD, biparietal diameter; EFW, estimated fetal weight; FBV, fetal brain volume; FLuV, fetal lung volume; FH, fundal height; FL, femur length; FV, femur volume; HC, head circumference; HV, humerus volume; LV, liver volume; MRI, magnetic resonance imaging; RV, renal volume; SE, standard error; SFH, symphysis fundal height; st, soft tissue; TCD, transverse cerebral diameter; TV, thigh volume; UAV, upper arm volume; US, ultrasound.

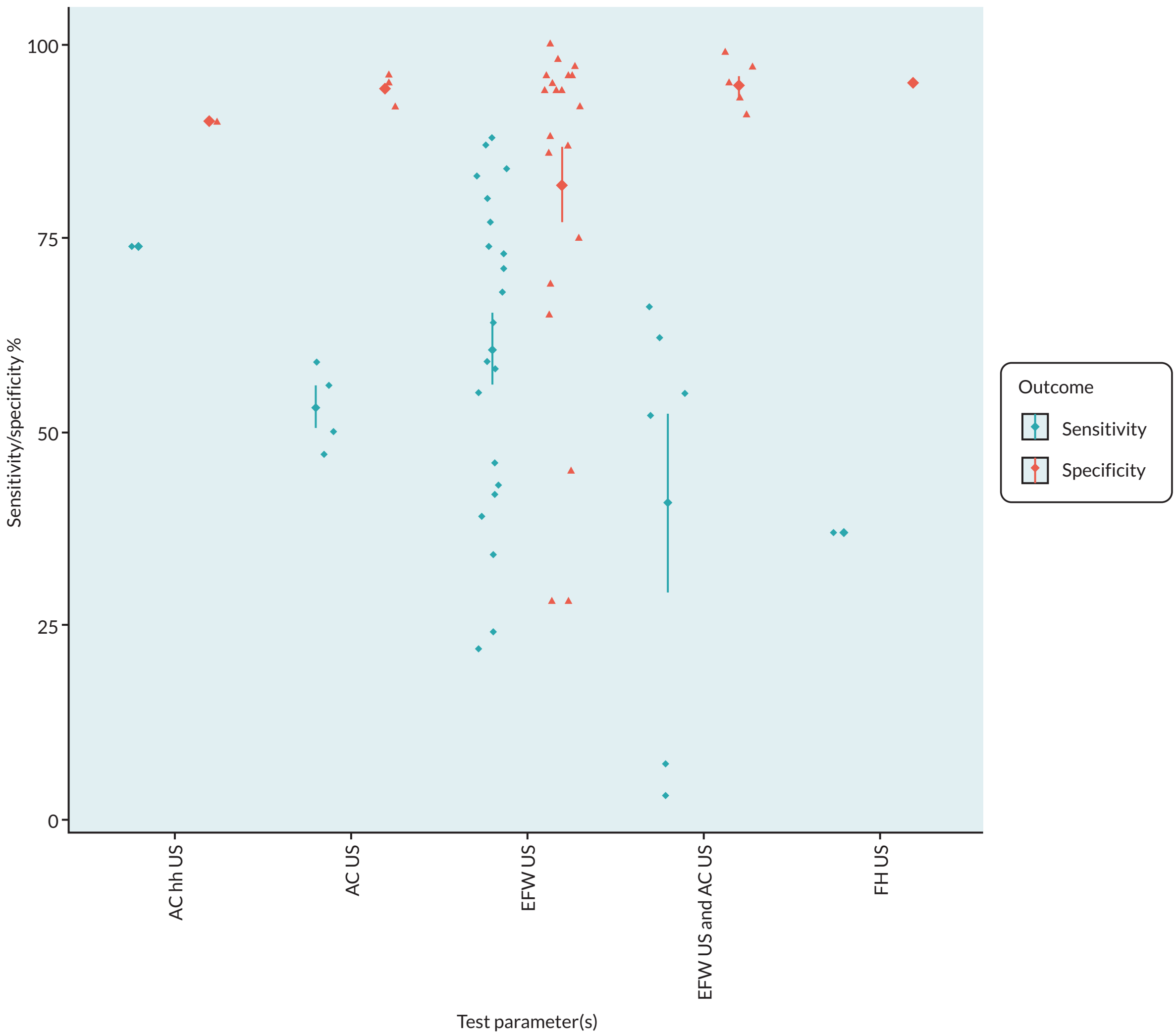


FIGURE 14 Accuracy outcomes of single test strategies, applied in high-risk populations, during the third trimester targeting a BW < 10th percentile. Note: the lines illustrate the mean and SE for each test. AC, abdominal circumference; BW, birth weight; EFW, estimated fetal weight; FH, fundal height; hh, hand-held; SE, standard error; US, ultrasound.

these tests were examined in the study series by Chang *et al.* (2005–2011).[239–245] UAV US had reported a sensitivity of 100% and 87% specificity, while RV US had 100% sensitivity and 81% specificity, when both were applied with a test threshold < 25th percentile. It should be noted that these results came from two studies[245,240] which used their own methods for calculating the outcome, as well as using local charts,[313,308] for the tests and a local BW chart[309] which was tailored to the Taiwanese population.

Among studies evaluating EFW US, the best overall performance reported was 70% sensitivity and 89% specificity.[224] This study used the Hadlock 2c equation to calculate EFW and the test threshold was < 10th percentile, using a general reference chart[298] for both test and BW. Among studies evaluating AC US, the best overall performance reported was 90% sensitivity and 91% specificity.[266] The test threshold was again < 10th percentile, using a general test reference chart[30] and a local BW reference chart[326] tailored to the German population. A third different study[281] reported 68% sensitivity and 97% specificity, for a combination of these two parameters, where the threshold was EFW OR AC < 10th percentile, using a general reference chart[30] for the test and a local BW reference chart for the USA.[303] Based on this between-study comparison of the best results reported for test, the single-parameter AC US appeared to perform better than EFW US or a combination of the two measures for predicting SGA at birth. However, between-study comparisons of test accuracy should always be interpreted with caution, because of the potential for differences in test accuracy to be driven by differences between the studies (other than the test being evaluated).

Of note, one study, Rad *et al.* (2018)[281] included in this section reported the results of a within-study comparison of EFW US calculated using Hadlock 2c, AC US and combinations of the two parameters, using test and BW thresholds of < 10th percentile, a general reference chart for tests[30] and a local BW reference chart[303] for the USA. This study found that a combination of EFW and AC, where the test threshold was defined as both parameters < 10th percentile, gave the lowest sensitivity, 47%, with AC alone having a sensitivity of 64%, EFW having a sensitivity of 51% and the highest sensitivity of 68% being obtained using a combination where the threshold was defined as either parameter < 10th percentile; there was little variation in specificity, with the corresponding specificity values being 98%, 97%, 98% and 97%. Within each of these three most frequently evaluated tests (EFW, AC and EFW plus AC), the use of local reference charts appears to achieve higher accuracies than the general charts both for test and BW. The results of within-study comparisons of reference charts are presented in Research question 4.

All of the studies described in this section evaluated the accuracy of growth monitoring strategies when applied universally in unselected general populations. One UK study, Sovio *et al.* (2015),[288] compared the accuracy of thirds trimester EFW US, when applied universally in a generally population, to its accuracy when applied only where clinically indicated. The clinically indicated/selective application of US growth monitoring is consistent with current UK practice. For clinically indicated/selective US growth monitoring, participants who did not receive the third-trimester EFW US were classified as test negative.[288] This study used a general reference chart[30] and a threshold of < 10th percentile for EFW, calculated using Hadlock 2c, and a local UK BW chart.[328] The performance of growth monitoring to predict both SGA (BW < 10th percentile) and severe SGA (BW < 3rd percentile) was assessed.[288] For both BW thresholds, the sensitivity of third-trimester EFW US was substantially lower when applied selectively (as clinically indicated) than when the testing was applied universally to the whole unselected population. For BW < 10th percentile, the sensitivity was 57% for universal testing and was 20% for clinically indicated testing, with corresponding specificities of 90% and 98%.[288] For BW third percentile, the sensitivity was 77% for universal testing and was 32% for clinically indicated testing, with corresponding specificities of 87% and 97%.[288]

High-risk populations

In high-risk populations, only four parameters were examined (AC US, AC US using a portable hh device, EFW US, AC and EFW US and FH US) (*Table 12*) in nine publications.[256,225,232,233,248,246,272,274,273] As illustrated in *Figure 14*, the highest sensitivities were obtained using EFW US. Studies evaluating EFW used a variety of different equations and test thresholds ranging from < 5th percentile to < 53rd percentile. Unsurprisingly, raising the test threshold generally resulted in increases in sensitivity and corresponding decreases in specificity (see *Table 12*). When considering the test threshold of < 10th percentile for EFW, the best overall performance was reported by a study using the Hadlock 2d and a general reference chart for EFW[30] with a local BW reference chart for the US population;[303] this study[232] reported a sensitivity of 71% and a specificity of 92%. Studies evaluating AC US in high-risk populations, using a test threshold of < 10th percentile, reported sensitivities which were lower than those reported for the general population, 50%[273] and 59%[233] with corresponding specificities of 94% and 92%.

Two studies, Blue *et al.* (2018) and Nwabuobi *et al.* (2020), provided within-study comparisons of EFW US, AC US and combinations of these two parameters, which were reported in three publications.[233,272,273] Both studies used a general test reference chart[30] and a local BW chart for the US population,[303] with one study[233] using a sex-specific version of this chart. Nwabuobi *et al.* (2020)[272] used Hadlock 2c to calculate EFW and Blue *et al.* (2018)[233] used Hadlock 2d. Both studies reported the highest sensitivity, 66%[233] and 55%,[273] respectively, when using a combination of EFW US and AC US with the test threshold defined as EFW OR AC < 10th percentile. As was the case for the general population, these within-study comparisons indicated that the lowest sensitivity estimates were associated with a combination of parameters using the test threshold EFW AND AC < 10th percentile where this was assessed, sensitivity 52%,[233] or with EFW alone, sensitivity 42%,[272] where the AND combination was not assessed. Again, similarly to the general population, these within-study comparisons indicated little variation in specificity across test strategies, 91–95% for Blue *et al.* (2018)[233] and 93–96% for Nwabuobi *et al.* (2020).[272,273] The results of within-study comparisons of reference charts are presented in Research question 4.

Two studies[255,254] evaluated the accuracy of growth monitoring in populations with specific risk factors, diabetes mellitus (DM) (gestational and pre-gestational), high BMI (≥ 35.0 kg/m^2) and PET. The results show low efficiency across all populations (sensitivity range 0–57%, specificity 94–100%) (*Figure 15*).

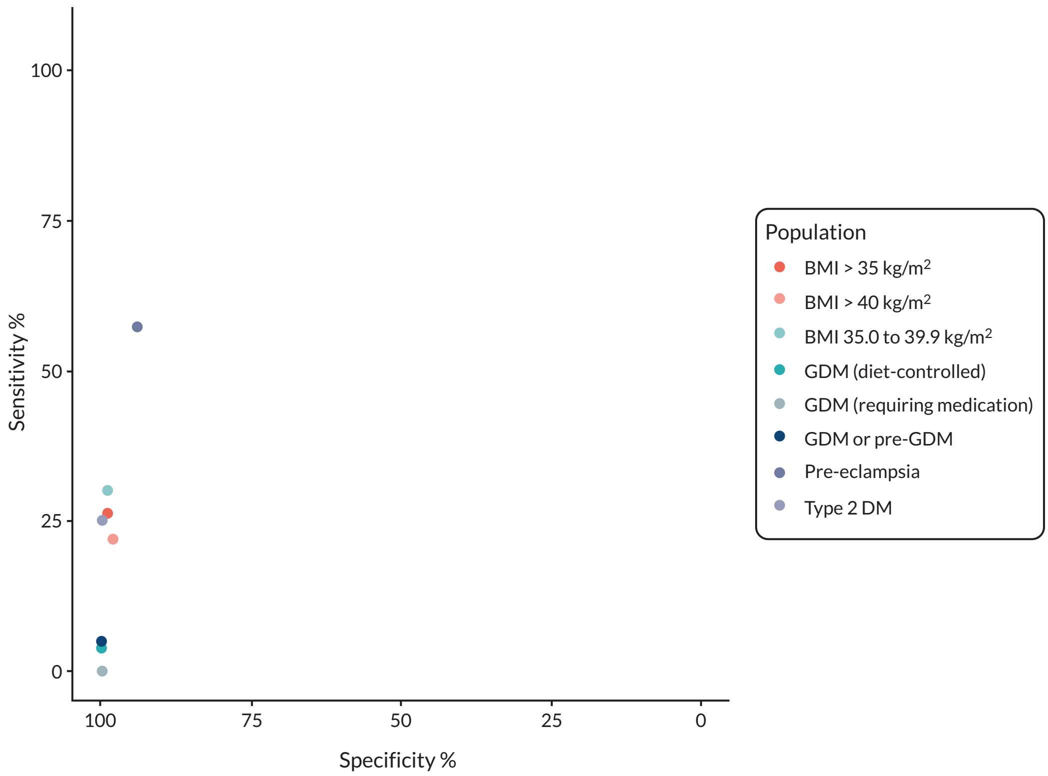


FIGURE 15 The ROC space plot comparing test results for pregnant people with different risk factors. BMI, body mass index (in kg/m2); DM, diabetes mellitus; GDM, gestational diabetes mellitus; ROC, receiver operating characteristic.

Low-risk populations

Three parameters were evaluated by four studies in low-risk populations (study populations from which participants with obstetric or medical risk factors had been excluded).[230,251,258,267] Three different thresholds for the tests were examined (< 10th, < 25th and < 35th percentiles). The best overall performance was sensitivity 59% and specificity 97%, reported for a threshold of < 25th percentile of the AC US test, using a chart developed for a low-risk population in the Netherlands and a local BW chart for the Dutch population.[315] The remaining tests evaluated had very low reported sensitivities 11–31%, with specificities between 90% and 99%.

Henrichs *et al.* (2019)[258] also provided a comparison routine US in the third trimester (28–30 and 34–36 weeks), in addition to usual care, versus usual care (serial FH measurements with clinically indicated third-trimester US). The parameters evaluated were AC US, threshold < 10th percentile and fetal abdominal growth (change in AC US), threshold decrease of at least 20 percentiles. The sensitivity estimates reported in this study were very low for all test metrics assessed and were consistently higher for routine/universal third-trimester US (sensitivity range 14–32%) compared to clinically indicated US/usual care (sensitivity range 8–19%); specificity ranged from 90% to 98% for universal testing and from 97% to 99% for clinically indicated testing.

Testing strategies in multiple gestations

Three studies evaluated growth monitoring strategies in the context of multiple gestation pregnancies.[257,247,285] All of these studies evaluated third-trimester US EFW for the prediction of SGA, defined as BW < 10th percentile. *Table 13* summarises the available data about the accuracy of fetal growth monitoring for the prediction of SGA in multiple gestation pregnancies. The best overall performance for twin pregnancies, sensitivity 70.7% and specificity 85.4%,[257] was reported when using the equation suggested by Shepard *et al.* (1982),[22] with a < 10th percentile threshold for the local chart for the US[303] for both test and BW. Between-study comparisons appear to indicate that the use of twin-specific BW charts may result in lower sensitivity and higher specificity (see *Table 13*).

TABLE 13 Accuracy of single US (third trimester) test strategies for predicting SGA at birth in multiple pregnancies

Test strategy	Test timing (weeks)	Pregnancy	Target condition	Number of studies	Sensitivity % (95% CI)	Specificity % (95% CI)
EFW estimated using Hadlock 1e[a]; threshold < 10th percentile; twin-specific reference charts[314]	30–35	Twin	BW < 10th percentile; twin-specific growth charts[314]	1[247]	40.0 (12.2, 73.8)[b]	93.7 (85.8, 97.9)[b]
					46.2 (19.2, 74.9)[c]	94.7 (87.1, 98.5)[c]
EFW estimated using Hadlock 2c[a]; threshold < 10th percentile; local reference charts (USA)[303]	34–36 + 6		BW < 10th percentile; local growth charts (USA)[303]	1[257]	70.1 (62.2, 77.1)	86.4 (82.6, 89.7)
EFW estimated using Shepard (1982)[d]; threshold < 10th percentile; local reference charts (USA)[303]					70.7 (62.9, 77.7)	85.4 (81.4, 88.8)
EFW estimated using Ong (1999)[a]; threshold < 10th percentile; local reference charts (USA)[303]					66.9 (58.9, 74.2)	89.0 (85.5, 92.0)
EFW estimated using Hadlock 2c[a]; threshold < 10th percentile (at least 1 fetus); local reference charts (Canada)[341]	28–40 +	Triplet	BW < 10th percentile (at least 1 fetus); local growth charts (Canada)[341]	1[285]	55.6 (35.3, 74.5)	100 (94.3, 100)

a *Table 2*.
b Twin A.
c Twin B.
d *Table 1*.

Single testing strategies during first or second trimester

Seven studies reported data about the accuracy of early growth monitoring strategies (applied during the first or second trimester) for the prediction of BW < 10th percentile.[223,236,263,266,283,290,292] These results are summarised in *Table 14*. US measurement of CRL was the only first trimester test to be evaluated; reported sensitivities for this test were very low, ranging from 10% to 24%, with specificities ranging from 83% to 93%. When considering second trimester testing, one study[266] compared the accuracy of five individual parameters (AC US, BPD US, FL US, HC US and TCD US), at a test threshold of < 10th percentile, and three ratios (FL/AC, HC/AC and TCD/AC), at a test threshold of > 90th percentile, using a local BW chart for the German population;[326] the best overall performance was achieved using AC US, which gave a sensitivity of 62% and a specificity of 90%. Two further studies[290,292] assessed the accuracy of EFW US in the second trimester, calculated using Hadlock 2c, for predicting BW < 10th percentile. Sensitivity estimates for EFW were very low, 10% using a threshold of < 10th percentile[290] and 37% using a threshold of GA-adjusted z-score < −1,[292] with corresponding specificities of 98% and 86%.

Multiple-serial testing strategies

Six studies evaluated the performance of growth monitoring strategies based on multiple US examinations for the prediction of BW < 10th percentile,[234,250,258,278,279,292] and two of these studies evaluated methods of estimating growth velocity.[278,279] These results are summarised in *Table 15*. Serial testing strategies were evaluated in unselected, high-risk as well as low-risk populations. Accuracy in unselected and in low-risk populations was low (6–32% sensitivity; 89–99% specificity). In high-risk populations, sensitivity reached 68% (89% specificity), when EFW US was applied multiple times in the third trimester, with a < 10th percentile threshold, using an individually customised reference chart for both EFW and BW.[250] The remaining tests evaluated in the high-risk population had a very low accuracy; sensitivity range: 15–24%, specificity range: 89–95%. Across all populations, all but one[250] of the multiple-serial testing strategies evaluated included US examinations in the first and/or second trimester.

Individual study results for the accuracy of all reported growth monitoring strategies are provided in *Appendix 2*, *Table 20*.

TABLE 14 Accuracy of single US (other timings) test strategies for predicting SGA at birth in singleton pregnancies

Test strategy	Test timing (weeks)	Population	Target condition	Number of studies	Sensitivity % (95% CI)	Specificity % (95% CI)
1st trimester testing						
CRL US; threshold GA-adjusted *z*-score < -1	1–12	General	BW < 10th percentile; local growth charts (USA)[303]	2[236,292]	**22.0 (19.1 to 25.2)**	**84.0 (83.3 to 84.7)**
CRL US; <expected by 3–6 days	11 + 0–13 + 6		BW < 10th percentile; local growth charts (China)[322]	1[263]	10.4 (6.8 to 14.9)	92.7 (91.7 to 93.7)
2nd trimester testing						
AC US; threshold < 10th percentile; Hadlock reference charts[30]	14 + 0-23 + 6	General	BW < 10th percentile; local growth charts (Germany)[326]	1[266]	62.1 (57.6 to 66.5)	89.8 (88.9 to 90.6)
BPD US; threshold < 10th percentile; Hadlock reference charts[30]					34.7 (30.4 to 39.1)	88.0 (87.1 to 88.8)
FL US; threshold < 10th percentile; Hadlock reference charts[30]					58.5 (53.9 to 63.0)	89.8 (89.0 to 90.6)
HC US; threshold < 10th percentile; Hadlock reference charts[30]					42.5 (38.0 to 47.1)	90.8 (90.0 to 91.5)
TCD US; threshold < 10th percentile; Hadlock reference charts[30]					23.2 (19.4 to 27.2)	90.4 (89.6 to 91.2)
FL/AC US; threshold > 90th percentile; Hadlock reference charts[30]					18.3 (14.9 to 22.1)	84.4 (83.5 to 85.4)
HC/AC US; threshold > 90th percentile; Hadlock reference charts[30]					28.2 (24.2 to 32.4)	90.8 (90.0 to 91.5)
TCD/AC US; threshold > 90th percentile; Hadlock reference charts[30]					21.5 (17.9 to 25.5)	93.6 (92.9 to 94.2)
EFW estimated using Hadlock 2c[a]; threshold GA-adjusted *z*-score < -1	18–22		BW < 10th percentile; local growth charts (USA)[303]	1[292]	37.2 (33.2 to 41.4)	85.5 (84.7 to 86.3)
EFW estimated using Hadlock 2c[a]; threshold < 10th percentile; Warsof reference charts[25] before 20 + 7 WG Hadlock reference charts[30] after 20 + 7 WG	13–27		BW < 10th percentile; local growth charts (USA)[303]	1[290]	10 (9 to 12)	98 (98 to 98)
EFW estimated using Hadlock 2c[a]; threshold < 10th percentile; Warsof reference charts[25] before 20 + 7 WG Hadlock reference charts[30] after 20 + 7 WG			BW < 10th percentile; general growth chart[344]	1[290]	10 (8 to 12)	98 (98 to 98)
EFW estimated using Hadlock 2c[a]; threshold < 10th percentile; Warsof reference charts[25] before 20 + 7 WG Hadlock reference charts[30] after 20 + 7 WG			BW < 10th percentile; local growth charts (USA)[319]	1[290]	10 (8 to 11)	98 (98 to 98)

TABLE 14 Accuracy of single US (other timings) test strategies for predicting SGA at birth in singleton pregnancies (*continued*)

Test strategy	Test timing (weeks)	Population	Target condition	Number of studies	Sensitivity % (95% CI)	Specificity % (95% CI)
2nd or 3rd trimester testing						
EFW estimated using Hadlock 2d[a]; threshold < 10th percentile; customised individual growth chart	20–43	General	BW < 10th percentile; customised individual growth chart	1[283]	50.8 (46.4 to 55.1)	97.3 (96.4 to 98.1)
TCD/AC US; threshold > 90th percentile derived from study population	20–28		BW < 10th percentile; general growth chart[297]	1[223]	60 (32.2 to 83.6)	94.1 (86.8 to 98)

a *Table 2*.

Note

Values in bold are pooled estimates.

TABLE 15 Accuracy of multiple US test strategies for predicting SGA at birth in singleton pregnancies

Test strategy	Test timing (weeks)	Population	Target condition	Number of studies	Sensitivity % (95% CI)	Specificity % (95% CI)
EFW estimated using Hadlock (equation NR); threshold < 10th percentile on first and all subsequent US examinations; Hadlock growth chart[30]	18–22 and unclear	General	BW < 10th percentile; local growth charts (USA)[303]	1[234]	5.9 (2.6 to 11.3)	99.0 (98.2 to 99.5)
1st trimester CRL and 2nd trimester HC; threshold 1st-2nd trimester growth lag (*z*-score < −1)	10–14 and 18–22			1[292]	25.6 (22.0 to 29.4)	89.5 (88.8 to 90.2)
EFW estimated using Hadlock 1d[a] or Cambell and Wilkin (1979) when FL unavailable; threshold < 10th percentile on one or more scans; customised individual growth chart	3rd trimester	High risk	BW < 10th percentile; customised individual growth chart	1[250]	67.7 (54.9 to 78.8)	88.7 (82.5 to 93.3)
EFW (2nd–3rd trimester) growth velocity; threshold < 10th percentile	18–22 and 26–36		BW < 10th percentile; local growth charts (USA)[303]	1[279]	17.6 (9.3 to 28.9)	91.0 (88.2 to 93.3)
AC (2nd–3rd trimester) growth velocity; threshold < 10th percentile					14.7 (7.1 to 25.6)	90.9 (88.2 to 93.2)
AC (2nd–3rd trimester) growth velocity threshold < 10th percentile, which triggered iatrogenic delivery (induction of labour or elective caesarean section)	18–22 and 34–38		BW < 10th percentile; general growth chart[335]	1[278]	23 (13 to 35)	95 (84 to 99)
AC (2nd–3rd trimester) growth velocity threshold < 10th percentile, which triggered iatrogenic delivery (induction of labour or elective caesarean section)	18–22 and 34–38		BW < 10th percentile; general growth chart[299]	1[278]	24 (13 to 37)	94 (83 to 99)
AC (2nd–3rd trimester) growth velocity threshold < 10th percentile, which triggered iatrogenic delivery (induction of labour or elective caesarean section)	18–22 and 34–38		BW < 10th percentile; local growth chart UK[337]	1[278]	24 (13 to 37)	93 (82 to 98)

continued

TABLE 15 Accuracy of multiple US test strategies for predicting SGA at birth in singleton pregnancies (*continued*)

Test strategy	Test timing (weeks)	Population	Target condition	Number of studies	Sensitivity % (95% CI)	Specificity % (95% CI)
AC (2nd–3rd trimester) growth velocity threshold < 10th percentile, which triggered iatrogenic delivery (induction of labour or elective caesarean section)	18–22 and 34–38		BW < 10th percentile; general growth chart[336]	1[278]	21 (11 to 34)	92 (80 to 98)
AC (3rd trimester) growth velocity; decrease in AC of at least 20 centiles over a minimum 2-week period; local growth charts (the Netherlands)[321]	18–30 and 34–36	Low risk	BW < 10th percentile; local growth charts (the Netherlands)[321]	1[258]	13.7 (10.9 to 16.8)	91.2 (90.4 to 91.9)
AC (3rd trimester); threshold < 10th percentile; local growth charts (the Netherlands)[321]					21.9 (18.6 to 25.6)	97.5 (97.1 to 97.9)
AC (3rd trimester); threshold < 10th percentile AND decrease in AC of at least 20 centiles over a minimum 2-week period; local growth charts (the Netherlands)[321]					32.2 (28.3 to 36.3)	89.8 (89.0to 90.5)

a *Table 2*.

Research question 4

What are the effects, on the performance (accuracy) of different methods of monitoring fetal growth, of key operational variables and parental characteristics?

Fifteen studies reported data to inform the effects, on the test performance of different methods of monitoring fetal growth, of key operational variables (e.g. test timing, reference chart used, healthcare professional performing the monitoring tests and patient characteristics).[223–225,231,232,254,255,258,261,266,268,270,272,278,290] We did not identify any studies which reported factors affecting failure to obtain a satisfactory measurement or clinical confidence in the reported result. With the exception of two studies that assessed the effects of BMI[254] and diabetes[255] on test performance, all comparisons reported in this section were within cohort (i.e. for each study, all strategies being compared were assessed in the same participants).

Reference charts

Nine studies[224,225,231,232,268,271,272,278,290] compared the use of different reference charts either for the index test, the reference standard (BW) or both in singleton pregnancies. Five studies[224,231,268,271,290] evaluated single testing in an unselected population, four[224,231,268,271] during the third trimester and one[290] during the second trimester. The remaining four studies[225,232,272,278] evaluated single and serial testing in high-risk populations at various time points.

Overall, the differences between the perspectives of the reference charts lay in whether they addressed a general (international) population or a local one as well as whether they were adjustable for fetal/neonatal characteristics and parental characteristics. The variation of charts is illustrated in the tree network diagram in *Figure 16*. Three of the reference charts for the index texts were adjusted for the sex of the fetus[4,333,335] and only one of them (GROW[335]) was reported to have been further adjusted for parental characteristics in one study.[278] Regarding the reference charts for BW, eight were adjusted for the sex of the neonate.[4,268,303,299,300,328–330] All the studies used the < 10th percentile threshold for defining SGA neonates, with the exception of Badr *et al.* (2023)[224] where the < 5th percentile was used.

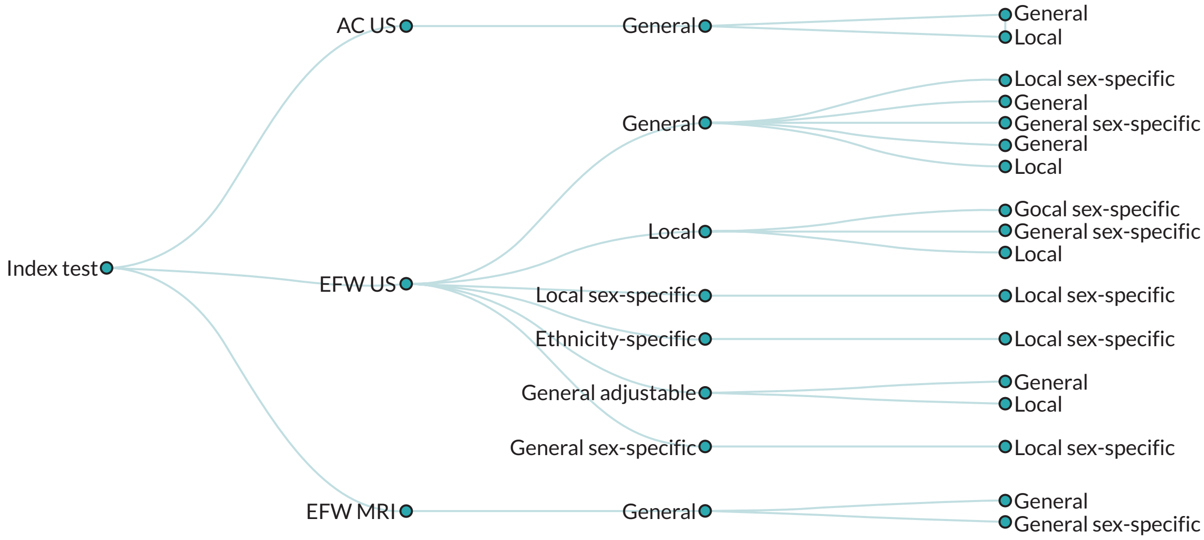


FIGURE 16 Tree network diagram of reference charts for index tests and neonatal weight. AC, abdominal circumference; EFW, estimated fetal weight; MRI, magnetic resonance imaging; US, ultrasound.

Unselected populations

One study[268] evaluated and compared both local and general standards for US-derived EFW (EFW US) and SGA by BW. The Hadlock 2c equation was used to calculate EFW for all combinations, while the threshold for both EFW and BW SGA was < 10th percentile. Four general reference charts were used for EFW centiles Hadlock,[30] WHO,[4] GROW[300] and INTERGROWTH-21st[23] and one local chart, which was created by the authors[268] following the methods by Altman and Chitty.[370] None of the EFW reference charts were adjusted for parental or neonatal characteristics. Regarding BW, two general reference charts were used: GROW[300] and INTERGROWTH-21st[299] as well as three local ones, the UK growth chart[328] the locally derived BW chart[268] and the WHO SACN chart.[329] All the BW reference charts were adjusted for the sex of the neonate. A comparison of the different combinations of reference charts is illustrated in *Figure 17*. The most effective (maximum sensitivity and specificity) combination was the local EFW chart and BW charts (Mathewlynn[268] + WHO,[329] sensitivity 67% and specificity 93%), followed by a combination of a local EFW chart and a general chart (Mathewlynn[268] + INTERGROWTH-21st,[299] sensitivity 64% and specificity 93%). Specificity for all combinations was relatively high, 93–100%. However, sensitivity varied much more widely and was generally low, ranging from 11% to 67%. As can be seen from *Figure 17*, the use of local reference charts for EFW (either with local or general reference charts for BW) appeared to provide the highest sensitivity (41–67%); as would be expected, these combinations were also associated with the lowest specificity values (93–94%).

Badr *et al.* (2023)[224] compared three reference charts all for general populations: WHO,[4] INTERGROWTH-21st[299] and Fetal Medicine Foundation (FMF) by Nicolaides *et al.* (2018).[298] None of these reference charts were adjusted for parental or neonatal characteristics when applied to the index test (EFW), but two charts (WHO[4] and INTERGROWTH-21st[299]) were adjusted for neonatal sex when applied to the BW. The study assessed EFW US and EFW using MRI, implementing two thresholds for both reference charts and both imaging modalities, < 5th and < 10th percentile. The Hadlock 2c equation was used to calculate EFW US and the equation by Baker *et al.* (1994)[366] was used for EFW MRI. The results are illustrated in *Figure 18*. For both thresholds, the highest sensitivity was achieved by the EFW US using the Nicolaides[298] reference charts, reaching a sensitivity of 71% and 70% for < 5th and < 10th percentiles, respectively (specificity 92% and 89%). The use of the FMF reference charts also produced the highest sensitivity and specificity estimates for EFW using MRI.

Monier *et al.* (2022)[271,270] compared eight different charts (*Table 16*) for determining a EFW US threshold of < 10th percentile, which were either general, local or ethnicity-specific. For two of the charts (Epopé,[333] WHO[4]), both the unisex and the sex-specific versions were evaluated (see *Appendix 2*, *Table 17*). The EFW US reference charts were combined with one sex-specific BW chart created for the French population.[330] The Hadlock 2c equation was used for

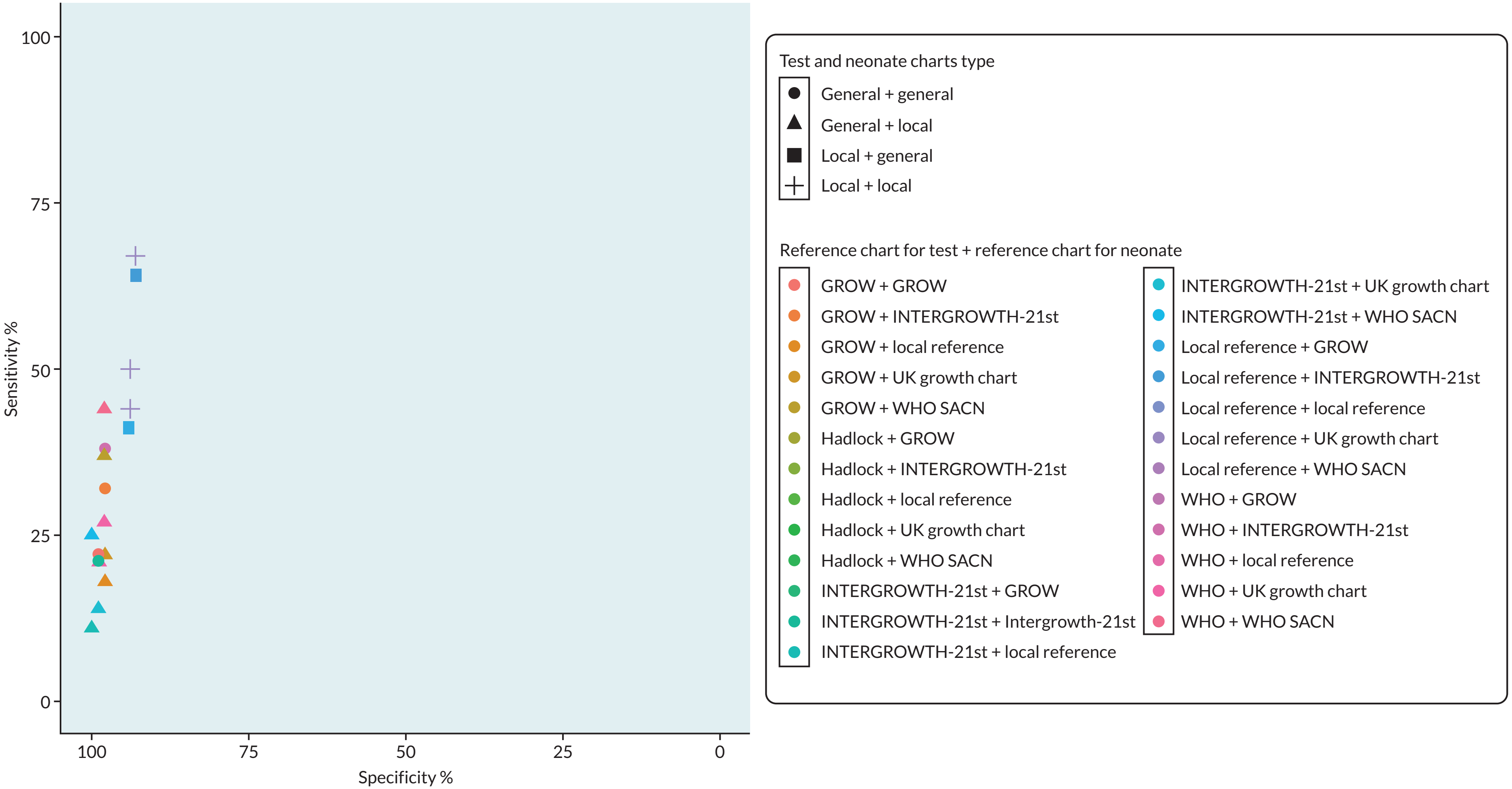


FIGURE 17 The ROC space plot comparing reference charts of index tests and BWs by Mathewlynn *et al.* (2022).[268] Reference charts for test: GROW;[300] Hadlock;[30] INTERGROWTH-21st;[23] local reference;[268] WHO.[4] Reference charts for BW: GROW;[300] INTERGROWTH-21st;[299] local reference;[268] UK growth chart;[328] WHO SACN.[329] GROW, Gestation Related Optimal Weight; ROC, receiver operating characteristics; WHO SACN, World Health Organization Scientific Advisory Committee on Nutrition.

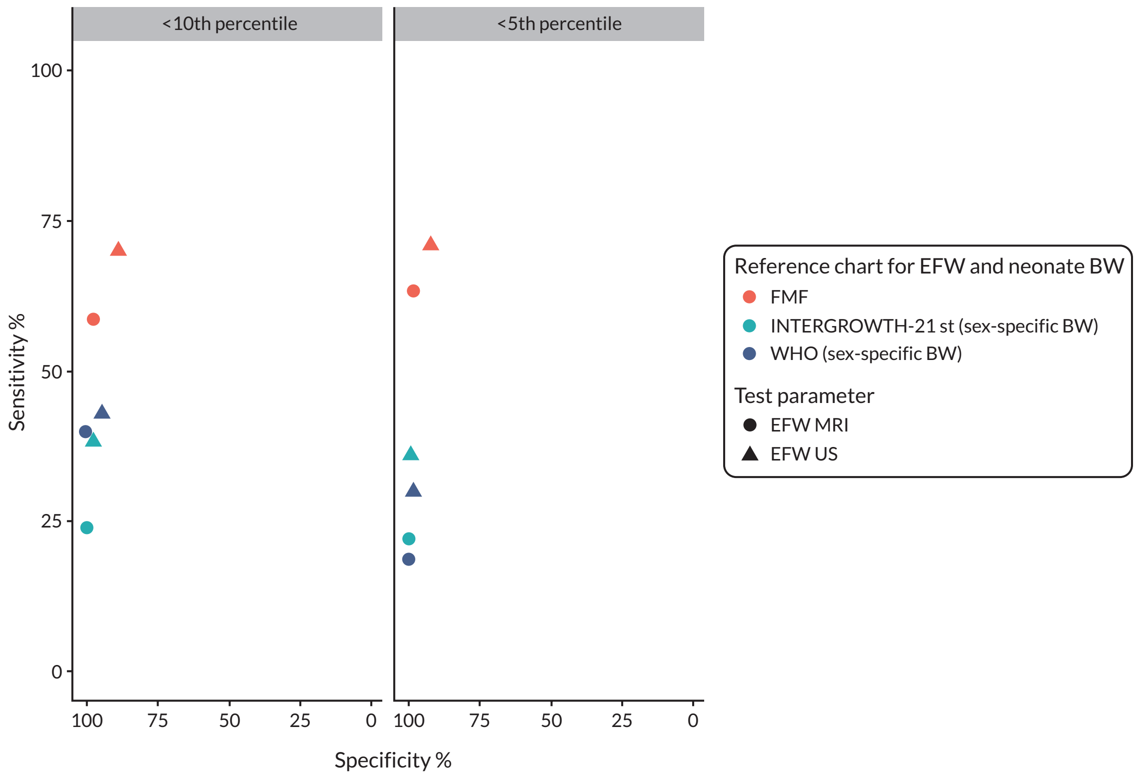


FIGURE 18 The ROC space plot comparing reference charts of index tests and BWs by Badr *et al.* (2023).[224] Reference charts: FMF;[298] INTERGROWTH-21st;[299] WHO.[4] BW, birth weight; EFW, estimated fetal weight; FMF, Fetal Medicine Foundation; MRI, magnetic resonance imaging; ROC, receiver operating characteristics; US, ultrasound; WHO, World Health Organization.

TABLE 16 Index test reference charts used by Monier *et al.* (2022)[271] (both publications)

Reference chart	Type	BW chart adjusted
FMF[298]	General	No
INTERGROWTH-21st[23]	General	No
Hadlock *et al.* (1991)[30]	General	No
French chart, Salomon[20]	Local	No
French College of Fetal Ultrasound (CFEF)[332]	Local	No
Epopé[333]	Local	No
		Sex-specific
WHO, Kiserud[4]	General	No
		Sex-specific
NICHD-White[331]	Ethnicity-specific	No

calculating all EFW US test results, apart from one where Hadlock 2d was used. The results of this study are illustrated in *Figure 19*. The performance of different BW reference charts was considered for all neonates as well as separately for females/males. Overall, the general EFW chart by FMF[298] exhibited the highest sensitivity of 55% and specificity of 86%. Sensitivity was low for all charts, ranging from 14% to 55%, while specificity was high, ranging from 86% to 98%. For female neonates, the general non sex-specific chart by Hadlock *et al.* (1991)[30] had provided the highest sensitivity of 40% (specificity 92%), while for male neonates, the general sex-specific chart by WHO exhibited the highest sensitivity of 35% (specificity 93%).

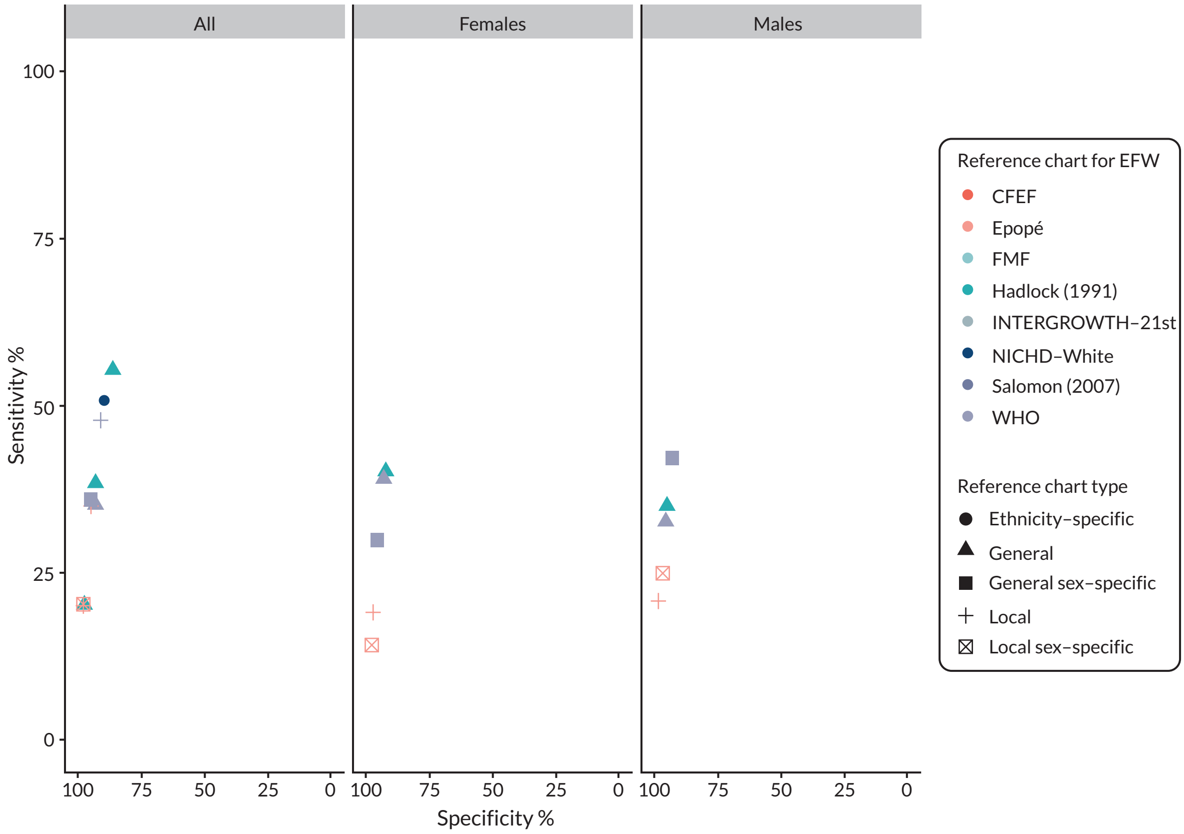


FIGURE 19 The ROC space plot comparing reference charts of index tests by Monier *et al.* (2022).[107] Reference charts: CFEF;[332] Epopé;[333] FMF;[298] Hadlock;[30] INTERGROWTH-21st;[23] NICHD;[331] Salomon;[20] WHO.[4] EFW, estimated fetal weight; CFEF, French College of Fetal Ultrasound; FMF, Fetal Medicine Foundation; NICHD, National Institute of Child Health and Human Development; ROC, receiver operating characteristics; WHO, World Health Organization.

Bergman *et al.* (2022)[231] compared the use of one local reference chart[305] for determining a EFW US threshold of < 10th percentile to the use of two local reference charts[305,306] together. Both were combined with a local unisex BW chart created for the Israeli population.[305] The Hadlock 2d equation was used for calculating EFW US. The sensitivity of the two EFW reference charts together was higher (46%) than one of them alone (33%); specificity was 90% and 95%, respectively.

Temming *et al.* (2017)[290] was the only study that used the same reference charts for EFW US (one for before 20 0/7 weeks of gestation[25] and one for 20 0/7 weeks of gestation onwards[30]) to compare three different BW reference charts, two local[303,319] and one general.[344] This single test was evaluated in an unselected population, during the second trimester, using the Hadlock 2c equation to calculate the EFW US and < 10th percentile threshold for all reference charts. The differences between the charts were minimal, all having a very low sensitivity of 10% and specificity of 98%.

High-risk populations

Poljak *et al.* (2017)[278] compared different combinations of reference charts, using mostly general charts and one local chart,[337] applied in a high-risk population. At the same time, this study also compared two different test parameters, EFW US and US-derived AC (AC US). The Hadlock 1f equation was used for calculating EFW US. The threshold for both tests and BW was set at < 10th percentile. The reference charts used for EFW US were GROW,[335] Mikolajczyk *et al.* (2011),[336] Hadlock,[30] and the reference charts for AC US were INTERGROWTH-21st,[299] Hadlock,[30] Chitty *et al.* (1994)[338] and Sovio *et al.* (2016).[288] One of the charts (GROW[335]) was adjustable and was individualised for parental characteristics. Unfortunately, further details on how the GROW chart was customised in this study were not reported.[278] The BW reference charts were the general GROW,[335] INTERGROWTH-21st,[299] Mikolajczyk *et al.* (2011)[336] and the local for the UK chart by WHO.[337] None of the BW reference charts were adjusted for the neonates' sex. The results are illustrated in *Figure 20*. The combination of AC US, using the Hadlock reference chart,[30] and the INTERGROWTH-21st BW reference chart[299] appeared to provide the overall best overall performance, with a sensitivity of 78% and a specificity of 98%. By comparison, the combination of AC US, using the Hadlock reference

DOI: 10.3310/AJLK7403

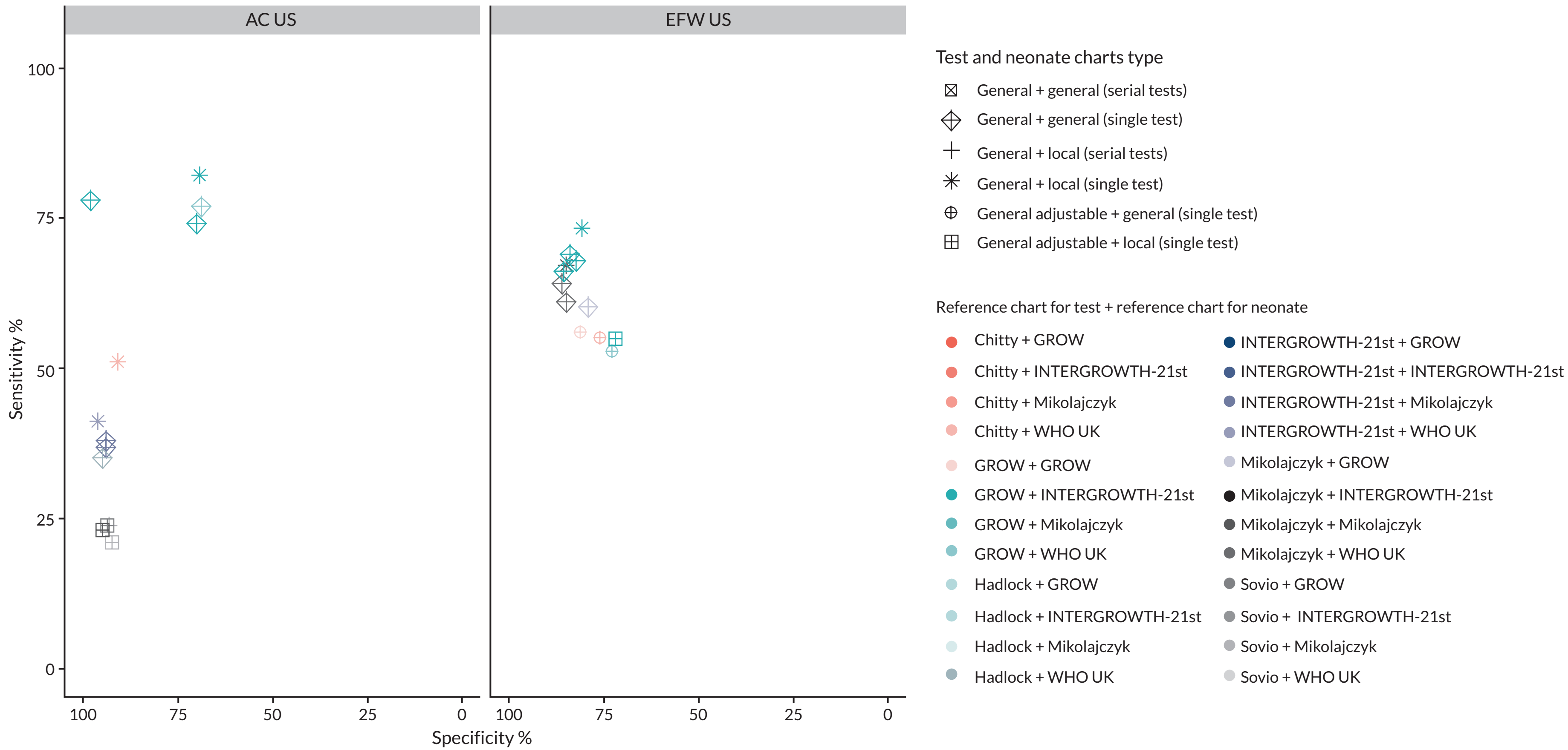


FIGURE 20 The ROC space plot comparing reference charts of index tests and BWs by Poljak *et al.* (2017).[278] Reference charts for test: Chitty;[338] GROW;[335] Hadlock;[30] INTERGROWTH-21st;[371] Mikolajczyk;[336] Sovio.[288] Reference charts for BW: GROW;[335] INTERGROWTH-21st;[299] Mikolajczyk;[336] WHO UK.[337] AC, abdominal circumference; EFW, estimated fetal weight; GROW, Gestation Related Optimal Weight; ROC, receiver operating characteristics; US, ultrasound; WHO, World Health Organization.

chart[30] and local WHO BW reference chart,[337] provided an increased sensitivity of 82% but with a marked reduction in specificity (specificity 69%). When EFW US was the test parameter evaluated, the best sensitivity of 73% was achieved using the Hadlock reference chart[30] and the local WHO BW reference chart[337] (specificity 81%); overall, this strategy performed less well than the best AC US strategy, having lower estimates for both sensitivity and specificity. Single AC US tests appeared to be more effective than serial AC US testing (*Figure 20*); the timing of serial testing was not reported.

Baird *et al.* (2016)[225] compared four different charts for determining an EFW US threshold of < 10th percentile, which were either general, local or ethnicity-specific and which were combined with one sex-specific BW chart created for the Australian population.[301] These single test strategies were evaluated in a high-risk population during the third trimester. The EFW US calculations were based on the method by Mikolajczyk *et al.* (2011)[336] adapted for the Australian population by anchorage of the EFW formula to the mean BW at 40.0 weeks for Australian births.[301] The results of this study are presented in *Figure 21*. The combination of GROW (2020) + Dobbins (2012) exhibited the highest sensitivity of 88%, followed closely by the combination of the ethnicity-specific version of GROW (2020) + Dobbins (2012) with a sensitivity of 87%; both strategies had a low specificity of 28%.

Blue *et al.* (2019)[232] evaluated and compared both local and general standards for EFW US combined with one local, sex-specific BW chart for the USA.[303] The three reference charts compared for EFW US were the general by Hadlock[30] and the local INTERGROWTH-21st[23] and by Salomon,[20] and the equation proposed by each was used for the calculation of the EFW US. This single test was applied in a high-risk population during the third trimester. The threshold for BW SGA was set at < 10th percentile, but five different EFW US (index test) thresholds were assessed (< 5th, < 10th, < 15th, < 22nd and < 53rd percentiles). The highest accuracy test strategy for the < 10th percentile EFW threshold was provided by the use of the general reference chart by Hadlock,[30] with a sensitivity of 71% and a specificity of 92%.

Nwabuobi *et al.* (2020)[272] compared two general EFW US reference charts, Hadlock[30] and INTERGROWTH-21st (2014),[299] combined with a local BW chart developed for the USA.[303] This study evaluated EFW US as a single test, applied in a high-risk population, during the third trimester, and using the Hadlock 2c equation to calculate EFW.

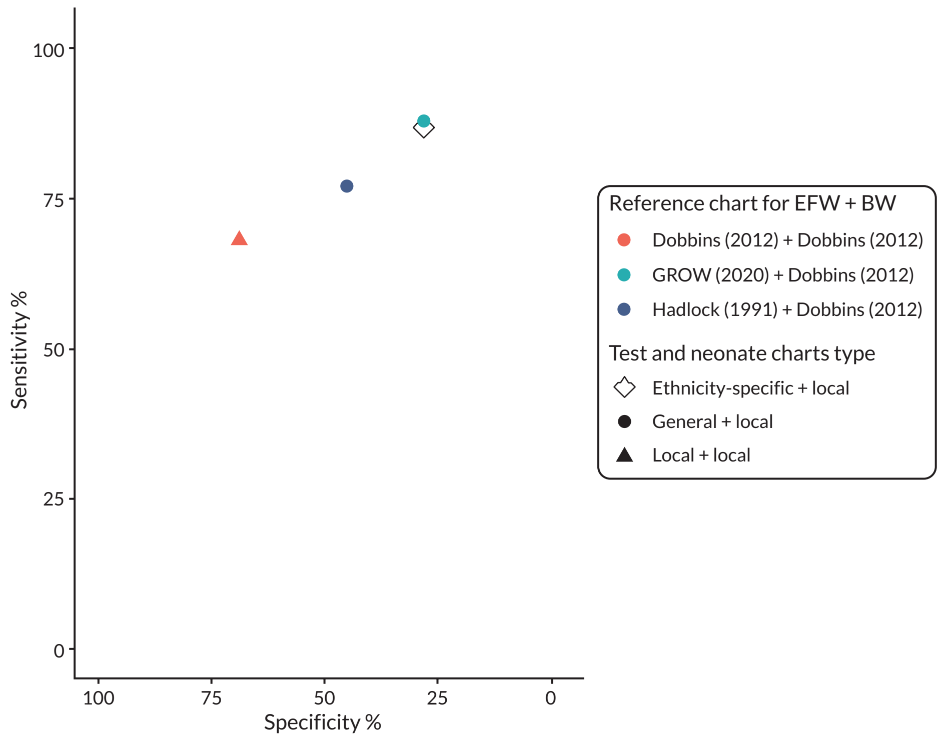


FIGURE 21 The ROC space plot comparing reference charts of index tests by Baird *et al.* (2016).[225] Reference charts: Dobbins;[301] GROW;[300] Hadlock.[30] BW, birth weight; EFW, estimated fetal weight; GROW, Gestation Related Optimal Weight; ROC, receiver operating characteristics.

The sensitivity of the test when using Hadlock reference charts[30] was higher at 42%, specificity 96%, compared to INTERGROWTH-21st (2014),[299] which gave a sensitivity of 24% and a specificity of 97%.

Reference charts summary of results

Overall, the nine studies described above[224,225,231,232,268,271,272,278,290] used 18 different reference charts for the index test and 15 different BW reference charts. This level of variation makes meaningful comparisons challenging.

Regarding the unselected populations, a comparison between four studies[224,231,268,271,270] using a single EFW US test and a threshold of < 10th percentile for both the index test and the BW for SGA, during the third trimester, as representing the most frequently evaluated test strategy, is provided in *Figure 22*, where nine different combinations of EFW and BW charts are illustrated. All the studies used a variation of the Hadlock formula, either equation 2c or 2d.[29] The highest sensitivity across all studies was reported using the general FMF reference chart[298] for both EFW and BW; sensitivity 70% with a specificity of 89%. The second highest was reached using the combination of the local EFW chart by Mathewlynn[268] and the local, sex-specific WHO SACN BW chart,[329] both developed for the UK;[268] although the sensitivity was lower, for this combination, at 67%, the specificity was higher at 93%. Overall, no clear trend in test

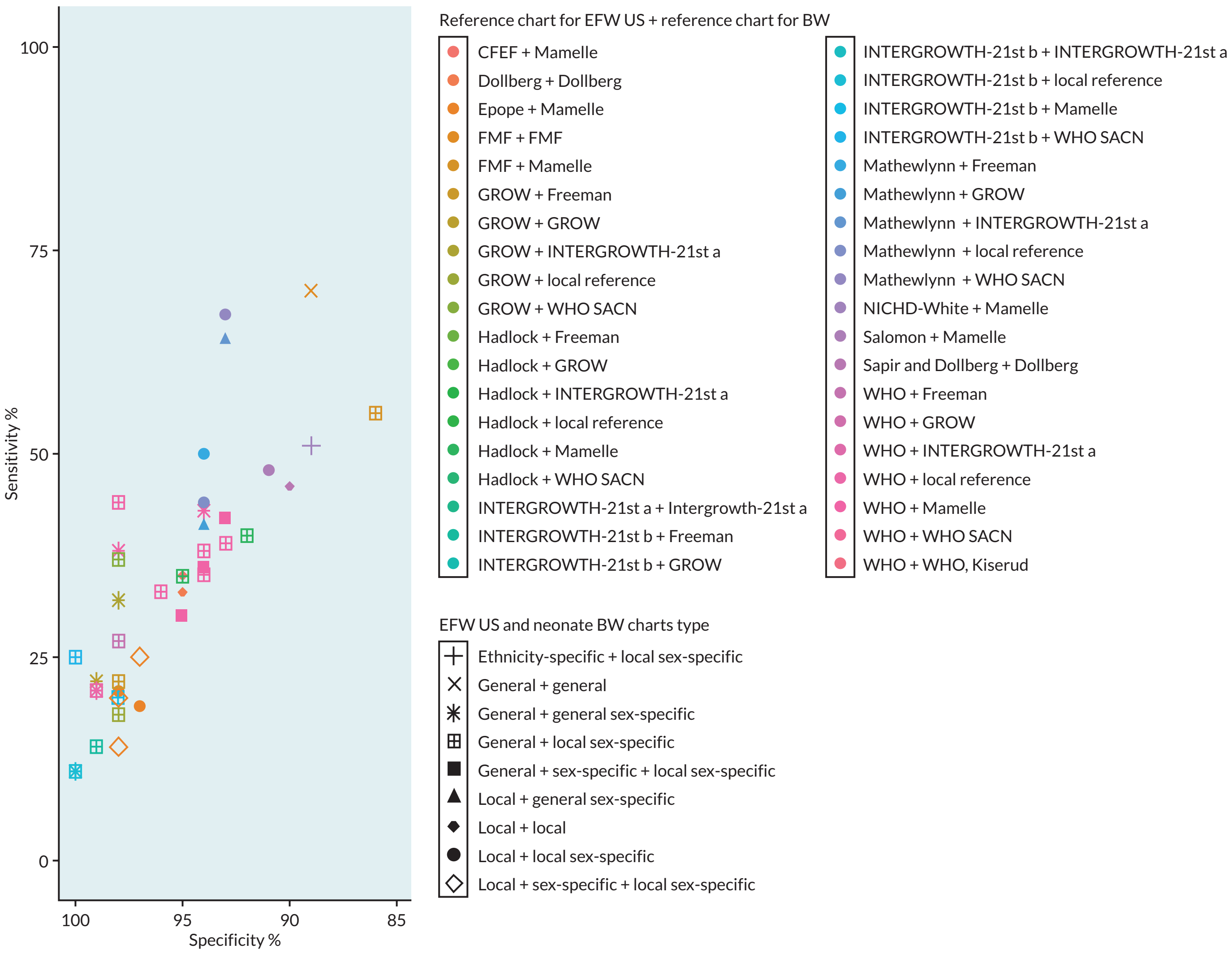


FIGURE 22 The ROC space plot comparing efficiency of reference charts using EFW US for detecting SGA neonates in unselected populations. Reference charts: Epopé,[333] FMF,[298] French BW reference charts by Mamelle,[330] French EFW chart by Salomon,[20] French College of Fetal Ultrasound (CFEF), by Massoud,[332] GROW (2019),[300] GROW (1995),[335] Hadlock,[30] INTERGROWTH-21st (2017),[23] INTERGROWTH-21st (2014),[299] local reference by Mathewlynn,[268] Mikolajczyk,[336] modelled chart adapted for the French population by Mornier,[271] NICHD,[331] UK growth chart,[328] WHO sex-specific,[4] WHO by Cole,[337] WHO by Kiserud,[4] WHO, the SACN and the RCPCH.[329] BW, birth weight; EFW, estimated fetal weight; GROW, Gestation Related Optimal Weight; SGA, small for gestational age; ROC, receiver operating characteristics; US, ultrasound.

strategy performance was observable across studies, either for the use of general versus local reference charts (either EFW or BW) or for the use of non- specific versus sex-specific BW reference charts. However, cross-study comparisons are particularly difficult because each different BW reference chart essentially defines a different target condition, that is a different definition of SGA. This is illustrated by the large UK study by Mathewlynn *et al.* (2022),[268] which used five different BW reference charts giving prevalences of SGA (< 10th percentile) at birth ranging from 4.2%[329] to 10.9%.[300] Of note, within this study, the use of the locally derived reference chart for EFW always resulted in the highest sensitivity and lowest specificity estimates, for any given BW reference chart.

Among the studies with the high-risk populations, three[225,232,272] applied single EFW US tests during the third trimester with a EFW US and BW threshold < 10th percentile. Meaningful comparison between the charts used in these studies was not feasible because five different equations were used for the calculation of EFW US.

Timing

Three studies[226,246,268] reported data to inform the effects of test timing on the accuracy of fetal growth monitoring. Agrawal (2016)[223] compared the accuracy of a single TCD/AC US and a test threshold of > 90th percentile for prediction of SGA at birth (BW < 10th percentile) in an unselected population, where the test was undertaken at 20–28 weeks versus 30–36 weeks; sensitivity increased with GA (60% vs. 80%), while the corresponding specificity estimates were similar (94% and 91%). Henrichs (2019)[258] compared the accuracy of a single AC US and a test threshold of < 10th percentile for prediction of SGA at birth (BW < 10th percentile) in a low-risk population, where the test was undertaken at 28–30 weeks versus 34–36 weeks; the sensitivity was low (13% and 11%) and specificity high (98% and 99%) at both time points. Marchand (2022)[266] assessed the variation in accuracy of eight different US parameters (HC, BPD, AC, FL, TCD, FL/AC, TCD/AC and HC/AC) with the GA at which the assessment was undertaken (14 + 0–18 + 6, 19 + 0–23 + 6, 24 + 0–28 + 6, 29 + 0–33 + 6 and 34 + 0–38 + 6 weeks); for most parameters assessed, there was a trend towards increasing sensitivity with increasing GA, while specificity tended to remain constant (> 85% in most cases). Full data are provided in *Appendix 2*, *Table 20*.

Operators

One study[261] compared the accuracy of testing according to the US operators, specifically, registered diagnostic medical sonographers versus board-certified maternal–fetal medicine specialists. Both groups implemented a single EFW US during the third trimester in a general population. The test threshold was < 2500 g for both index text and BW SGA. Sensitivity was higher in the group of board-certified maternal–fetal medicine specialists (78% vs. 72%), but specificity was higher in the group of registered diagnostic medical sonographers (99% vs. 90%).

Patient characteristics

Two studies[255,254] provided some information about the accuracy of fetal growth monitoring, a single EFW US during the third trimester and a test threshold of < 10th percentile, for prediction of SGA at birth (BW < 10th percentile) in patients with specific risk factors; Dude (2018)[255] compared accuracy in patients with different types of diabetes (gestational diabetes managed with diet, gestational diabetes requiring medication and type 2 diabetes), and Dude (2021)[254] compared accuracy between different categories of BMI in patients with obesity. Dude (2018)[255] found that sensitivity was very low (0–25%) and specificity was 100% in all subgroups of patients with diabetes. Similarly, Dude (2021)[254] found that sensitivity was low (22–30%) with specificity values of 98% or 99% in all subgroups of patients with obesity. Neither of these studies included a comparison with accuracy in patients without the risk factor of interest. Sensitivity and specificity values are illustrated in a ROC space plot (*Figure 15*), and full data are provided in *Appendix 2*, *Table 20*.

Two studies[232,265] compared the equation used for calculating EFW US in singleton pregnancies using a < 10th percentile threshold for BW reference charts. The first study[232] compared the equations Hadlock 2d,[29] INTERGROWTH-21st[23] and Salomon *et al.* (2007),[20] using the corresponding EFW reference charts and a < 10th percentile threshold. The BW reference chart was a local reference chart, Alexander *et al.* (1996).[303] The Hadlock 2d equation[29] was the most accurate of the three, with a sensitivity of 71% and a specificity of 92%. The second study[265] compared Hadlock 2c,[29] Persson and Weldner (1986)[325] and Shepard *et al.* (1982).[22] They used a < third percentile threshold based on the EFW reference chart by Lindstrom *et al.* (2021).[324] In this study, the most accurate prediction was achieved by the equation by Persson and Weldner (1986),[325] with a sensitivity of 79% and a specificity of 91%. A direct comparison between the two studies is not feasible since different EFW thresholds and BW reference charts were used.

Chapter 4 Discussion

Statement of principal findings

Overall, 58 studies reported in 78 publications were included in this systematic review. In some cases, the studies' evidence addressed more than one of the four review questions and as such were included in the appropriate sections. Meta-analyses were only conducted for review questions 2 and 3. Our study differs from previous systematic reviews in that we have focused on identifying and assessing evidence to inform the question of whether and how fetal growth monitoring and the results of monitoring may be associated with rates of still birth and neonatal clinical outcomes (research question 1). We have also assessed the available evidence about the real-world effects of implementing clinical guidelines/standardise care pathways (research question 2). When considering the diagnostic performance of fetal growth monitoring tests to predict SGA at birth, previous systematic reviews and meta-analyses have combined data from primary studies which used different reference charts[12,31] and/or different anthropometric formulae.[12] With respect to accuracy data, our systematic review focuses on providing a thorough exploration of potential sources of variation in test performance rather than generation of overall summary estimates (research questions 3 and 4).

Research question 1

What are the effects, on clinical outcomes and rates of stillbirth and neonatal death, of interventions which are made based on the findings of fetal growth monitoring to detect SGA/FGR?

Two studies[269,277] provided evidence regarding the effect of interventions, that were made based on SGA/FGR monitoring, on neonatal and parental clinical outcomes, where clinical outcomes were reported by test result (2 × 2 data). This type of study is particularly useful, because it can provide information about the clinical consequences of antenatal identification versus non-identification of SGA babies as well as any clinical consequences of antenatal misclassification of AGA babies. The findings of both studies indicated that, as might be expected, antenatal SGA classification, whether correct (TP) or incorrect (FP), was associated with increased rates of parental interventions (induction of labour and caesarean section operations) compared to missed SGA (FN) or correct classification of AGA babies (TN), respectively. However, it is important to note that, although both studies provided information about the relationship between fetal growth monitoring test result category and types of interventional delivery, it was not clear whether reported interventions were for planned pre-term delivery (iatrogenic delivery). In addition, neither of these two studies provided information about the relationship between the results of fetal growth monitoring and GA at delivery. Regarding, fetal/neonatal clinical outcomes, both correct and incorrect antenatal SGA classifications did not have a statistically significant effect on almost all of the clinical outcomes under examination. Thus, the available evidence did not indicate that fetal growth monitoring strategies and increased parental interventions translated to improved fetal/neonatal clinical outcomes for SGA babies.[269,277] Conversely, the increased parental intervention rates observed in AGA babies who were misclassified as SGA antenatally did not appear to be associated with significant adverse effects on fetal/neonatal clinical outcomes.[269,277] There was no evidence that antenatal detection of SGE was associated with any change in the rate of stillbirths or neonatal deaths; however, it is important to note that these outcomes were only reported by Nymark Hansen *et al.* (2019)[277] and that this study was not adequately powered to detect any differences in these rare events. It should be further noted, that neither of these two studies are directly applicable to UK clinical practice. One of the studies[269] examined growth monitoring in unselected general populations (universal third-trimester testing), which is not current practice in the UK. The second study[277] evaluated third-trimester growth monitoring based on clinical indication, which is more closely aligned to UK practice; however, the thresholds used to define SGA were ≤ 15th and ≤ 22nd percentile thresholds, higher than the ≤ 10th percentile usually used in the UK.

Research question 2

What are the effects, on neonatal and parental outcomes, of implementing published guidelines for fetal growth monitoring?

A recent systematic review[372] focusing on CPGs on FGR found differences in key aspects of the CPGs in defining, classifying and managing FGR pregnancies. Differences were identified in the frequency of fetal monitoring, the criteria for hospital admission as well as the timing and mode of delivery.

We identified five studies[238,249,260,262,294] which examined the effects of guideline implementation. All of these studies were concerned with the clinical effects of GAP implementation in four countries: UK, Australia, New Zealand and India. We did not identify any RCTs, CCTs or observational comparative studies that assessed the effects on clinical outcomes of implementation of UK clinical guidelines (as described in Clinical guidelines on fetal growth monitoring), or any other clinical guidelines implemented in other countries.

Meta-analysis was executed for stillbirth rates, two neonatal outcomes and two parental outcomes. Stillbirth rates were found to be reduced after the implementation of GAP with a RR of 0.79 (95% CI 0.74 to 0.84). Four studies were included in this meta-analysis (two conducted in the in the UK,[260,294] one in India[238] and one in New Zealand[249]). Similarly, the 5-minute Apgar score < 7, was also reduced after GAP implementation with an RR of 0.78 (95% CI 0.63 to 0.95). This meta-analysis included two studies (one conducted in the UK[294] and one in Australia[262]). Statistical heterogeneity was minimal in both models. Apgar score does not predict individual neonatal mortality or neurologic outcome, but it is a standardised tool for the assessment of neonates' status.[365] According to the ACOG Committee on Obstetric Practice, a 5-minute Apgar score > 7 confers a highly reduced risk for neonatal encephalopathy.[365] The second neonatal outcome that was pooled was admissions to NICU/HDU and SCBU reported by two studies (one executed in England[294] and one in Australia[262]). Although the direction of the effect was the same for both studies, indicating a reduction of neonatal admissions after the implementation of GAP, the effect estimate was not statistically significant with a pooled RR of 0.59 (95% CI 0.02 to 20.03); the very wide CIs reflect the high level of between-study heterogeneity. While the individual results from these two studies appear consistent with an association between implementation of GAP and a reduction in neonatal admissions, it should be noted that when the effect estimate was adjusted for age, ethnicity, parity and stratification factor (as reported in the RCT), there was no significant difference in the rate of neonatal unit admissions before and after the implementation of GAP 0.4 (95% CI −0.8 to 1.7); $p = 0.48$.[294] It should be noted that all the above pooling included both observational studies and a RCT. The results of the meta-analyses broadly indicate that the implementation of GAP was associated with reductions in adverse neonatal outcomes.

Regarding paternal interventions, two outcomes were synthesised, induction of labour and caesarean section operations (elective or emergency). The effects of GAP implementation on induction of labour was pooled from three studies (one in England,[294] one in Australia[262] and one in New Zealand[249]). With a pooled OR of 1.16 (95% CI 1.0082 to 1.3419), the implementation of GAP is associated with a small increase in the rate of labour induction, with the direction of effect being consistent across all three studies. On the pooled estimate for rates of caesarean section, operations from the same studies indicated no difference between GAP and usual care, RR of 1.0460 (95% CI 0.8601 to 1.2719). As was the case for research question 1, the timing of the parental interventions reported in these studies was unclear. The relationship, if any, between the findings of growth monitoring conducted in the context of GAP and parental interventions was also unclear.

Taken together, the results of the meta-analyses on both parental and fetal/neonatal outcomes indicate that the implementation of GAP in these countries was associated with a decrease on the rate of stillbirths and of adverse fetal/neonatal outcome (5-minute Apgar score < 7) without substantially increasing parental interventions (induction of labour and caesarean section operations). However, the results of these meta-analyses should be interpreted with caution, as all involved combining data from observational studies with data from a RCT. In addition, there appears to be some inconsistency between the results of these meta-analyses and the findings of the studies included for research question 1, which did not support a link between antenatal detection of SGA infants' and reduction in stillbirths or improved neonatal outcomes; however, it is important to note that (for research question 1) stillbirth rates were only reported by Nymark Hansen *et al.* (2019)[277] and that this study was not adequately powered to detect any differences in these rare events. Viera *et al.* (2022)[294] was the only included RCT to evaluate GAP implementation and was also conducted in the UK. This study found that GAP was not superior to usual care in predicting SGA at birth [antenatal detection of SGA was 25.9% in the intervention and 27.7% in the SC arm (adjusted difference 2.2%, 95% CI −6.4% to 10.7%)] and concluded that the reduced adverse neonatal/perinatal outcomes found in the intervention group were not achieved by improved SGA detection.[294] The study authors noted that the benefit of earlier-than-term iatrogenic birth offered to women with suspected SGA is under dispute, as, although it might decrease stillbirth and perinatal mortality, it could adversely increase the short-term neonatal morbidity or later adverse developmental outcomes.[294] The only other study focusing on GAP implementation in the UK was by Hugh *et al.* (2021);[260] this study found that implementation was associated with a reduction in the rate of stillbirths. Both studies acknowledged that there was a significant temporal overlap between the study period and the NHS England SBLCB, which included three more elements/interventions (smoking cessation, reduced fetal movements and intrapartum monitoring). These parallel interventions are important potential confounding factors that have the potential to drive changes in clinical outcomes. It should also be noted that the GAP care pathway is a complex, multifactorial intervention and is not equivalent to the simple third-trimester screening strategies evaluated in the studies included for research question 1 or any of the test strategies evaluated by studies addressing research questions 3 and 4. Given the apparent lack of evidence to support a direct link between increased antenatal SGA detection and clinical outcomes,[269,277,294] it is plausible that the observed improvements in clinical outcomes following GAP implementation[260,294] may be attributable to confounding parallel interventions/changes in routine care and/or to other properties of the GAP care pathway, such as increased contact and engagement with patients or improved standardisation/consistency of care arising from the audit and quality control elements of the protocol.

Research question 3

What is the accuracy of different methods of monitoring fetal growth for predicting smallness for gestational age at delivery?

Data from the vast majority of the studies (53 out of 58 studies) included in the systematic review was used to inform research question 3 on the accuracy of different methods of monitoring fetal growth for predicting SGA at delivery. A high degree of variation of the monitoring strategies evaluated was apparent with respect to all the seven main components defined in Monitoring methods. Studies also varied with respect to the population in which monitoring strategies were evaluated (e.g. unselected, high-risk, low-risk, singleton or multiple pregnancy). The Results section of this report predominantly focused on strategies which adopted the < 10th percentile threshold for postnatal SGA categorisation as this is the definition used in UK clinical practice.

Within single examination strategies applied during the third trimester on singleton pregnancies, the three most examined US examination test strategies, both in unselected and high-risk populations, were EFW US, AC US and combinations of these parameters. The majority of the EFW US test outcomes was derived using a variation of the Hadlock equations.[29] The most frequently used threshold for all the tests was < 10th percentile. In general, between-study comparisons of test accuracy are of limited value in that DTA studies are observational studies and hence comparing the performance of tests based on evaluations in different studies does not take into account of the differences in study population or other factors that may affect test performance. For this systematic review, between-study comparisons of the performance of different growth parameters were particularly problematic (even for the selected data set described above) due to the high degree of variation in other components of the test strategy (EFW equations, test reference charts and BW reference charts). Three studies, one conducted in a general unselected population[281] and two reported in three publications[233,272,273] conducted in high-risk populations, reported the results of within-study comparisons of EFW US calculated using Hadlock 2c, AC US and OR/AND combinations of the two parameters, using test and BW thresholds of < 10th percentile. The results of these studies consistently indicated that, as might be expected, the highest sensitivity was achieved when using a combination of EFW and AC where the test threshold was defined as either parameter < 10th percentile (the OR combination) and that a combination where the test threshold was defined as both parameters < 10th percentile (the AND combination) gave the lowest sensitivity; EFW US alone consistently had either lower or equal sensitivity to AC US alone. The results of all of three within-study comparisons reported generally high specificity values (> 90%) with little variation in specificity across test strategies (EFW US calculated using Hadlock 2c, AC US and OR/AND combinations of the two parameters).

Four studies evaluated the performance of growth monitoring in low-risk populations (study populations from which participants with obstetric or medical risk factors had been explicitly excluded).[230,251,258,267] Reported sensitivities of growth monitoring were generally very low (11–31%) across all test strategies evaluated, with the exception of one study, which used a test threshold of < 25th percentile on a local reference chart developed for a low-risk population in the Netherlands and reported a sensitivity of 59% and a specificity of 97%.

Comparison of different testing timings between first, second and third trimesters indicated that the third-trimester testing strategies performed better with respect to maximising the antenatal detection of SGA babies. In terms of single versus multiple/serial testing, sensitivity estimates were very low for all strategies other than where multiple testing occurred during the third trimester. Both these outcomes suggest that there is more to be gained from testing during the third trimester alone; the limited available data do not support a benefit from earlier testing.

Results on multiple non-monochorionic multiple pregnancies were scarce. All identified studies evaluated EFW US executed during the third trimester. Between-study comparison indicated that highest performance was achieved when using twin-specific BW charts.

All studies, included in this systematic review, that evaluated the accuracy of growth monitoring strategies in general, unselected pregnant populations evaluated universal testing (screening) applied to the whole population, rather than testing based on clinical indication/risk factors as recommended by the RCOG Green-top Guideline No. 31[6] endorsed by NICE guideline NG201.[36] Studies evaluating universal testing could be considered to be of limited applicability to the UK setting based on current UK clinical guidance and practice. One UK study, Sovio *et al.* (2015),[288] compared the accuracy of third-trimester EFW US when applied universally in a general population to its accuracy when applied only where clinically indicated. The clinically indicated/selective application of US growth monitoring is consistent with the current UK practice. This study used a general reference chart[30] and a threshold of < 10th percentile for EFW, calculated using Hadlock 2c, and a local UK BW chart.[328] The sensitivity of third-trimester EFW US to predict SGA (BW < 10th percentile) was substantially lower when applied selectively (as clinically indicated) than when testing was applied universally to the whole.[288] One of the studies conducted in a low-risk population, Henrichs *et al.* (2019),[258] also provided a comparison routine US in the third trimester (28–30 and 34–36 weeks), in addition to usual care, versus usual care (serial FH measurements with clinically indicated third-trimester US). US evaluations were based on measures of AC and abdominal growth. The sensitivity estimates reported in this study were very low for all test metrics assessed and were consistently higher for routine/universal third-trimester US (sensitivity range 14–32%) compared to clinically indicated US/usual care (sensitivity range 8–19%); specificity ranged from 90% to 98% for universal testing and from 97% to 99% for clinically indicated testing.[258] Neither of these two studies[258,288] provided any information linking clinical outcomes to test result (2 × 2 data) for research question 1.

Research question 4

What are the effects, on the performance (accuracy) of different methods of monitoring fetal growth, of key operational variables and parental characteristics?

There were insufficient data to adequately inform assessment of how the accuracy of individual test strategies for fetal growth monitoring may vary with maternal characteristics (e.g. BMI/obesity and ethnicity) or the type, training and experience of clinical practitioner performing the monitoring.

Most of the available data for research question 4 were concerned with variations in the accuracy of the metrics used to monitor fetal growth (EFW US and AC US) with the reference chart used for the index test or the reference chart used for BW. Nine included studies[224,225,231,232,268,271,272,278,290] used 18 different reference charts for the index test and 15 different BW reference charts, rendering meaningful between-study comparisons challenging. Overall, no clear trend in test strategy performance was observable across studies, either for the use of general versus local reference charts (either EFW or BW) or for the use of non- specific versus sex-specific BW reference charts. Within this group of studies, the large UK study by Mathewlynn *et al.* (2022)[268] used five different BW reference charts, resulting in prevalences of SGA (< 10th percentile) at birth ranging from 4.2%[329] to 10.9%.[300] In Mathewlynn *et al.* (2022),[268] the use of the locally derived reference chart for EFW always resulted in the highest sensitivity and lowest specificity estimates for any given BW reference chart (definition of SGA). Considering the data for the UK BW reference chart,[328] which gave the median SGA prevalence of 7.1%, and applying the reported test strategy performance characteristics to a hypothetical population cohort of 100,000 pregnancies, switching from the least sensitive EFW reference chart (INTERGROWTH-21st[23]) to the local chart which was created by the authors[268] would result in the antenatal detection of an additional 2506 SGA babies; however, this switch would also result in the incorrect classification of 4645 AGA babies as SGA.

We did not identify any studies which reported factors affecting failure to obtain a satisfactory measurement or clinical confidence in the reported result.

Strengths and limitations of assessment

Extensive literature searches were conducted in order to maximise the retrieval of relevant studies. These included electronic searches of a variety of bibliographic databases as well as screening of clinical trials registers and conference abstracts to identify unpublished studies. Because of the known difficulties in identifying test accuracy studies using study design-related search terms,[373] search strategies were developed to maximise sensitivity at the expense of reduced specificity. Thus, large numbers of citations were identified and screened, relatively few of which met the inclusion criteria of the review. In addition, our systematic review included up-to-date searches conducted in September 2023.

The possibility of publication bias remains a potential problem for all systematic reviews. Considerations may differ for systematic reviews of test accuracy studies. It is relatively simple to define a positive result for studies of treatment, for example, a significant difference between the treatment and control groups which favours treatment. This is not the case for test accuracy studies, which measure agreement between index test and reference standard. It would seem likely that studies finding greater agreement (high estimates of sensitivity and specificity) will be published more often. In addition, test accuracy data are often collected as part of routine clinical practice, or by retrospective review of records; test accuracy studies are not subject to the formal registration procedures applied to RCTs and are therefore more easily discarded when results appear unfavourable. The extent to which publication bias occurs in studies of test accuracy remains unclear; however, simulation studies have indicated that the effect of publication bias on meta-analytic estimates of test accuracy is minimal.[374] Formal assessment of publication bias in systematic reviews of test accuracy studies remains problematic and reliability is limited.[47] We did not undertake a statistical assessment of publication bias in this review. However, our search strategy included a variety of routes to identify unpublished studies and resulted in the inclusion of a number of conference abstracts.

Clear inclusion criteria were specified in the protocol for this review, the review has been registered on PROSPERO (CRD42023408030). The eligibility of studies for inclusion is therefore transparent. In addition, we have provided specific reasons for exclusion for all of the studies which were considered potentially relevant at initial citation screening and were subsequently excluded on assessment of the full publication (see *Appendix 3*, *Table 21*). The review process followed recommended methods to minimise the potential for error and/or bias;[46] studies were independently screened for inclusion by two reviewers, and data extraction and quality assessment were done by one reviewer and checked by a second reviewer. Any disagreements were resolved by consensus.

To our knowledge, this is the first systematic review and meta-analysis to address the clinical effects related to the implementation of GAP.

The data available to inform this assessment were reduced by poor reporting of primary studies, in particular failure to report sufficient detail to define the growth monitoring text strategy or strategies being evaluated and/or failure to provide a definition of the target condition (SGA at birth).

A further key limitation was the paucity of data that could be considered directly applicable to the UK setting; we did not identify any studies that assessed the effects implementing UK clinical guidelines, and only one of the included studies that assessed the accuracy of growth monitoring strategies in a general, unselected pregnant population[288] considered the use of differential monitoring strategies determined by clinical indication as recommended by the RCOG Green-Top Guideline No. 31.[6]

Uncertainties

Despite the large volume of published literature around fetal growth monitoring, there remains a lack of evidence linking fetal growth monitoring and the results of tests used to monitor fetal growth to changes in rates of stillbirth, perinatal death or adverse neonatal clinical outcomes. Very few studies have compared clinical outcomes by the result of fetal growth monitoring test (see *Implications for service provision*, research question 1), and where this has been done, the results have indicated that, while antenatal prediction of SGA [whether correct (TP) or incorrect (FP)] appeared to be associated with increased intervention (e.g. induction of labour and caesarean birth), there was no evidence of an associated change in rates of stillbirth and little evidence of any change in neonatal outcomes. The studies of the effects of guideline implementation included in our systematic review appeared to indicate that implementation of the GAP care pathway in UK settings may be associated with some reductions in adverse neonatal outcomes (see *Implications for service provision*, research question 2); however, it was unclear to what extent the observed changes were attributable to GAP or other contemporaneous changes in care arising from the NHS England SBLCB, and evidence about reductions in stillbirths was inconsistent. The extent to which any effects of GAP, which is a complex, multifactorial intervention, may be attributable to antenatal detection of SGA is also unknown. One of the studies which contributed accuracy data to this report, Henrichs *et al.* (2019),[258] was primarily a report of results from the International Randomized Study of Interferon and STI571 (IRIS) study (a nationwide, stepped wedge cluster randomised trial conducted in 60 primary care midwifery practices in the Netherlands and evaluating the addition of routine third-trimester US to usual care in low-risk pregnancies), which reported overall comparisons of outcomes between the intervention and control group.[258] Studies of this type, which assess the overall effects of different monitoring strategies, without providing a link to a correct or incorrect test result (TP, FP, FN and TN), were outside the scope of this systematic review. However, it may be of interest to note that Henrichs *et al.* (2019)[258] reported an adjusted OR (adjusted for clustering, maternal age, BMI, smoking, alcohol or recreational drug use, midwifery practice size, parity; educational level, employment status, marital status and sex of infant) of 0.88 (95% CI 0.70 to 1.20) for a composite of severe adverse perinatal outcomes (one or more of perinatal death, Apgar score < 4 at 5 minutes, impaired consciousness, asphyxia, seizures, > 24-hour assisted ventilation, haemorrhage, meningitis, bronchopulmonary dysplasia, intraventricular haemorrhage, cystic periventricular leucomalacia or necrotising enterocolitis). This result is consistent with the findings of a systematic review and meta-analysis,[375] which included the IRIS study and six further RCTs comparing routine third-trimester US to usual care of serial SFH measurement in low-risk pregnant populations. This study reported the results of meta-analyses indicating that the intervention was associated with an increased identification of EFW < 10th percentile [RR 2.11 (95% CI 1.86 to 2.39), five studies, n = 19,409], but reported no statistically significant differences in rates of perinatal death or any of neonatal outcomes assessed (resuscitation, NICU admission, respiratory distress syndrome, grade III or IV intraventricular haemorrhage and sepsis) and did not provide any information about the proportions of pregnancies misclassified in either the intervention or control groups.[375] Overall, the increased antenatal prediction of SGA with the routine third-trimester US observed in these studies did not translate into reductions in stillbirth or adverse neonatal outcomes.

A further key area of uncertainty concerns the effectiveness of guideline implementation in the UK; we did not identify any RCTs, CCTs or comparative observational studies that assessed the effects of implementing any UK clinical guideline on the rates of stillbirth, rates of neonatal death, clinical outcomes (e.g. neonatal morbidity, NICU admission and parental morbidity), rates of pre-term iatrogenic delivery or GA at iatrogenic delivery.

Finally, there was a lack of data to inform research question 4 in respect of how the accuracy of individual test strategies for fetal growth monitoring may vary with maternal characteristics (e.g. BMI/obesity and ethnicity) or the type, training and experience of clinical practitioner performing the monitoring.

Equality, diversity and inclusion

This study was secondary research and followed a scope defined by the Department of Health and Social Care. The research questions specified (see *Chapter 1*) included consideration of the potential for differential effects of interventions between different ethnic groups in people with BMI classification of overweight/obese and in people with obstetric risk factors.

Chapter 5 Conclusions

Implications for service provision

There is a lack of evidence linking fetal growth monitoring and the results of tests used to monitor fetal growth to changes in rates of stillbirth, perinatal death or adverse neonatal clinical outcomes.

There is some evidence to suggest that implementation of the GAP care pathway in UK settings may be associated with a reduction in adverse neonatal outcomes, but evidence about the effects of implementation on stillbirth was inconsistent. It is also unclear to what extent observed effects were attributable to GAP or other contemporaneous changes to routine care. The extent to which any effects of GAP, which is a complex, multifactorial intervention, may be attributable to antenatal detection of SGA is also unknown.

There is lack of evidence to assess the effectiveness of implementing UK clinical guidelines with respect to reducing rates of stillbirth and adverse neonatal outcomes.

Where fetal growth monitoring is to be implemented, there is insufficient evidence to support strongly favouring any one test strategy [combination of test timing, parameter measured (including any formula used, as for estimating EFW) and threshold and reference chart] for use in fetal growth monitoring. The available evidence suggests that testing during the third trimester is likely to result in maximal antenatal detection of SGA babies with no clear benefit from earlier or serial testing and that (for the general pregnant population) universal third-trimester US is likely to offer improved sensitivity compared to testing based on clinical indication. Evidence from within-study comparisons of different test parameters suggest that, for a single US examination in the third trimester, a combination of EFW OR AC < 10th percentile could offer increased sensitivity relative to either parameter used alone. There is also some evidence to suggest that, when using EFW US to monitor fetal growth, the use of a locally derived reference chart for EFW may result in the highest sensitivity for a given BW reference chart (definition of SGA).

Suggested research priorities

Large diagnostic cohort studies, which assess the accuracy of fetal growth monitoring strategies as implemented in the UK and which record clinical outcomes by growth monitoring test result (TP, FP, FN and TN) are needed to adequately resolve the question of whether and how fetal growth monitoring and the results of tests used to monitor fetal growth affect clinical outcomes. These studies should be adequately powered to detect differences in low prevalence outcomes such as stillbirth.

Comparative studies are needed to assess whether the implementation of UK clinical guidance, such as RCOG Green-top Guideline No. 31[6] endorsed by NICE guideline NG201,[36] or the NHS England SBLCB version 3,[38] 2023, has been effective in reducing rates of stillbirth or adverse neonatal outcomes.

Studies of the effects of GAP implementation in the UK setting, specifically taking into consideration any overlap with recent recommendations for UK practice, such as the NHS England SBLCB, may be useful to inform choices about the implementation of specific protocols.

Studies of guideline implementation with more than one time point before and after implementation (i.e. interrupted time series designs) might also be useful to look at trends over time.

Additional diagnostic cohort studies are needed to assess the accuracy of fetal growth monitoring strategies for predicting SGA at birth in multiple pregnancies and to adequately assess the effects on accuracy of maternal characteristics (e.g. BMI/obesity and ethnicity) and the type, training and experience of clinical practitioner performing the monitoring.

Additional information

CRediT contribution statement

Evangelos Danopoulos (https://orcid.org/0000-0002-9186-8811): Conceptualisation (supporting), Data curation (equal), Formal analysis (lead), Investigation (equal), Project administration (supporting), Validation (equal), Visualisation (lead), Writing – original draft (equal), Writing – reviewing and editing (equal).

Nigel Armstrong (https://orcid.org/0000-0002-7443-4798): Conceptualisation (supporting), Data curation (equal), Project administration (supporting), Validation (equal), Writing – reviewing and editing (equal).

Kevin McDermott (https://orcid.org/0000-0002-9415-8647): Investigation (equal), Validation (equal), Writing – reviewing and editing (equal).

Jiongyu Chen (https://orcid.org/0009-0004-4518-1463): Investigation (equal), Validation (equal), Writing – reviewing and editing (equal).

Xiaoyu Tian (https://orcid.org/0009-0000-0993-2187): Investigation (equal), Validation (equal), Writing – reviewing and editing (equal).

Caro Noake (https://orcid.org/0000-0003-0329-4772): Investigation (equal), Validation (equal), Writing – original draft (supporting), Writing – reviewing and editing (equal), Specialist information provision (design and conduct of literature searches and management of bibliographic database (lead).

Marie Westwood (https://orcid.org/0000-0002-6257-0653): Conceptualisation (lead), Data curation (equal), Formal analysis (supporting), Investigation (equal), Project administration (lead), Supervision (lead), Validation (equal), Visualisation (supporting), Writing – original draft (equal), Writing – reviewing and editing (equal).

Patient data statement

This study was secondary research (a systematic review), with no primary data collection component. Opportunity and potential for patient and public involvement in this type of project are limited and generally focus on the question setting/protocol development phase. The scope/research questions for this project were defined by the requirements of the policy customer (Department of Health and Social Care) who commissioned the work following a recommendation in the HSIB 2021 National Learning Report, *'Intrapartum stillbirth: learning from maternity safety investigations that occurred during the COVID-19 pandemic, 1 April to 30 June 2020'*.[1] The HSIB report drew on evidence from 37 maternity investigations, each of which incorporated engagement with patients and families. The experiences of pregnant people and their families have, therefore, contributed indirectly to the process of specifying and commissioning this project.

Data-sharing statement

Requests for access to data should be addressed to the corresponding author.

Ethics statement

This report concerns secondary research, for which ethics approval is not required.

Information governance statement

This study was secondary research and did not handle any personal information.

Disclosure of interests

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at https://doi.org/10.3310/AJLK7403.

Primary conflicts of interest: None declared.

Disclaimer

Every effort has been made to obtain the necessary permissions for reproduction, to credit original sources appropriately and to respect copyright requirements. However, despite our diligence, we acknowledge the possibility of unintentional omissions or errors and we welcome notifications of any concerns regarding copyright or permissions.

References

1. Healthcare Safety Investigation Branch. *Intrapartum Stillbirth: Learning from Maternity Safety Investigations That Occurred during the COVID-19 Pandemic, 1 April to 30 June 2020 (National Learning Report) [Internet]*. Reading: HSIB; 2021. URL: https://hsib-kqcco125-media.s3.amazonaws.com/assets/documents/HSIB_Intrapartum_Stillbirth_Report_web.pdf (accessed 16 January 2023).
2. Flenady V, Koopmans L, Middleton P, Frøen JF, Smith GC, Gibbons K, *et al*. Major risk factors for stillbirth in high-income countries: a systematic review and meta-analysis. *Lancet* 2011;**377**:1331–40.
3. Moraitis AA, Wood AM, Fleming M, Smith GCS. Birth weight percentile and the risk of term perinatal death. *Obstet Gynecol* 2014;**124**:274–83.
4. 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.
5. Office for National Statistics (ONS). *ONS Website, Statistical Bulletin: Child and Infant Mortality in England and Wales: 2021. Stillbirths, Infant and Childhood Deaths Occurring Annually in England and Wales, and Associated Risk Factors: 4. Neonatal and Infant Mortality Risk Factors [Internet]*. 2023. URL: www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/childhoodinfantandperinatalmortalityinenglandandwales/2021#neonatal-and-infant-mortality-risk-factors (accessed 7 September 2023).
6. Royal College of Obstetricians & Gynaecologists. *The Investigation and Management of the Small-for-Gestational-Age Fetus: Green-Top Guideline No. 31 (2nd ed.) [Internet]*. London: RCOG; 2014. URL: www.rcog.org.uk/guidance/browse-all-guidance/green-top-guidelines/small-for-gestational-age-fetus-investigation-and-management-green-top-guideline-no-31/ (accessed 17 January 2023).
7. Imdad A, Yakoob MY, Siddiqui S, Bhutta ZA. Screening and triage of intrauterine growth restriction (IUGR) in general population and high risk pregnancies: a systematic review with a focus on reduction of IUGR related stillbirths. *BMC Public Health* 2011;**11**:S1.
8. Goto E. Comparing the accuracy of maternal, clinical, and ultrasound estimations to predict birthweight: a meta-analysis. *Acta Obstet Gynecol Scand* 2017;**96**:1289–99.
9. Kose S. Diagnostic and prognostic power of the first biometric measurements and doppler examination in fetal growth restriction. *J Clin Obstet Gynecol* 2019;**29**:100–9.
10. Peter JR, Ho JJ, Valliapan J, Sivasangari S. Symphysial fundal height (SFH) measurement in pregnancy for detecting abnormal fetal growth. *Cochrane Database Syst Rev* 2015;**2015**:CD008136. https://doi.org/10.1002/14651858.CD008136.pub3
11. Goto E. Prediction of low birthweight and small for gestational age from symphysis-fundal height mainly in developing countries: a meta-analysis. *J Epidemiol Community Health* 2013;**67**:999–1005.
12. Goto E. Usefulness of ultrasound fetal anthropometry in primary and secondary screening to identify small for gestational age: a meta-analysis. *J Clin Ultrasound* 2019;**47**:212–8.
13. Campbell S, Wilkin D. Ultrasonic measurement of fetal abdomen circumference in the estimation of fetal weight. *Br J Obstet Gynaecol* 1975;**82**:689–97.
14. Combs CA, Jaekle RK, Rosenn B, Pope M, Miodovnik M, Siddiqi TA. Sonographic estimation of fetal weight based on a model of fetal volume. *Obstet Gynecol* 1993;**82**:365–70.
15. Higginbottom J, Slater J, Porter G, Whitfield CR. Estimation of fetal weight from ultrasonic measurement of trunk circumference. *Br J Obstet Gynaecol* 1975;**82**:698–701.

16. Hsieh FJ, Chang FM, Huang HC, Lu CC, Ko TM, Chen HY. Computer-assisted analysis for prediction of fetal weight by ultrasound-comparison of biparietal diameter (BPD), abdominal circumference (AC) and femur length (FL). *Taiwan Yi Xue Hui Za Zhi* 1987;**86**:957–64.
17. Jordaan HV. Estimation of fetal weight by ultrasound. *J Clin Ultrasound* 1983;**11**:59–66.
18. Ott WJ, Doyle S, Flamm S. Accurate ultrasonic estimation of fetal weight. *Am J Perinatol* 1985;**2**:178–82.
19. Sabbagha RE, Minogue J, Tamura RK, Hungerford SA. Estimation of birth weight by use of ultrasonographic formulas targeted to large-, appropriate-, and small-for-gestational-age fetuses. *Am J Obstet Gynecol* 1989;**160**:854–60; discussion 860.
20. Salomon LJ, Bernard JP, Ville Y. Estimation of fetal weight: reference range at 20–36 weeks' gestation and comparison with actual birth-weight reference range. *Ultrasound Obstet Gynecol* 2007;**29**:550–5.
21. Shinozuka N, Okai T, Kohzuma S, Mukubo M, Shih CT, Maeda T, *et al*. Formulas for fetal weight estimation by ultrasound measurements based on neonatal specific gravities and volumes. *Am J Obstet Gynecol* 1987;**157**:1140–5.
22. 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.
23. Stirnemann J, Villar J, Salomon LJ, Ohuma E, Ruyan P, Altman DG, *et al*.; International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). International estimated fetal weight standards of the INTERGROWTH-21st Project. *Ultrasound Obstet Gynecol* 2017;**49**:478–86.
24. Vintzileos AM, Campbell WA, Rodis JF, Bors-Koefoed R, Nochimson DJ. Fetal weight estimation formulas with head, abdominal, femur, and thigh circumference measurements. *Am J Obstet Gynecol* 1987;**157**:410–4.
25. Warsof SL, Gohari P, Berkowitz RL, Hobbins JC. The estimation of fetal weight by computer-assisted analysis. *Am J Obstet Gynecol* 1977;**128**:881–92.
26. Warsof SL, Wolf P, Coulehan J, Queenan JT. Comparison of fetal weight estimation formulas with and without head measurements. *Obstet Gynecol* 1986;**67**:569–73.
27. Woo JS, Wan CW, Cho KM. Computer-assisted evaluation of ultrasonic fetal weight prediction using multiple regression equations with and without the fetal femur length. *J Ultrasound Med* 1985;**4**:65–7.
28. Hadlock FP, Harrist RB, Carpenter RJ, Deter RL, Park SK. Sonographic estimation of fetal weight. The value of femur length in addition to head and abdomen measurements. *Radiology* 1984;**150**:535–40.
29. 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.
30. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology* 1991;**181**:129–33.
31. Caradeux J, Martinez-Portilla RJ, Peguero A, Sotiriadis A, Figueras F. Diagnostic performance of third-trimester ultrasound for the prediction of late-onset fetal growth restriction: a systematic review and meta-analysis. *Am J Obstet Gynecol* 2019;**220**:449–59.e19.
32. Smith GC, 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**:1–190.
33. Chiossi G, Pedroza C, Costantine MM, Truong VTT, Gargano G, Saade GR. Customized vs population-based growth charts to identify neonates at risk of adverse outcome: systematic review and Bayesian meta-analysis of observational studies. *Ultrasound Obstet Gynecol* 2017;**50**:156–66.
34. Gardosi J. Customized fetal growth standards: rationale and clinical application. *Semin Perinatol* 2004;**28**:33–40.

35. Gestation Network. *Growth Charts [Internet]*. Birmingham: Perinatal Institute; n.d. URL: www.gestation.net/growthcharts.htm (accessed 25 September 2023).

36. National Institute for Health and Care Excellence. *Antenatal Care. NICE Guideline 201 [Internet]*. London: NICE; 2021. URL: www.nice.org.uk/guidance/ng201 (accessed 7 September 2023).

37. NHS England. *Saving Babies' Lives Version Two: A Care Bundle for Reducing Perinatal Mortality (SBLCBv2) [Internet]*. NHS England; 2020. URL: www.england.nhs.uk/publication/saving-babies-lives-version-two-a-care-bundle-for-reducing-perinatal-mortality/ (accessed 7 September 2023).

38. NHS England. *Saving Babies' Lives Version Three: A Care Bundle for Reducing Perinatal Mortality (SBLCBv3) [Internet]*. NHS England; 2023. URL: www.england.nhs.uk/wp-content/uploads/2023/05/PRN00130-Saving-babies-lives-version-three-a-care-bundle-for-reducing-perinatal-mortality-v3.1-july-23.pdf (accessed 18 September 2023).

39. Perinatal Institute. *Growth Assessment Protocol (GAP) [Internet]*. Birmingham: Perinatal Institute; n.d. URL: www.perinatal.org.uk/GAP/programme (accessed 19 September 2023).

40. Perinatal Institute. *Growth Assessment Protocol: GAP Guidance v3 [Internet]*. Birmingham: Perinatal Institute; 2023. URL: https://perinatal.org.uk/GAPguidance.pdf (accessed 19 September 2023).

41. Lees CC, Stampalija T, Baschat A, da Silva Costa F, Ferrazzi E, Figueras F, *et al*. ISUOG Practice Guidelines: diagnosis and management of small-for-gestational-age fetus and fetal growth restriction. *Ultrasound Obstet Gynecol* 2020;**56**:298–312.

42. Lausman A, Kingdom J; Maternal–Fetal Medicine Committee. Intrauterine growth restriction: screening, diagnosis, and management. *J Obstet Gynaecol Can* 2013;**35**:741–8.

43. American College of Obstetricians and Gynecologists. Fetal growth restriction: ACOG Practice Bulletin, Number 227 [Interim Update]. *Obstet Gynecol* 2021;**137**:e16–28.

44. Kehl S, Dötsch J, Hecher K, Schlembach D, Schmitz D, Stepan H, Gembruch U. Intrauterine growth restriction. Guideline of the German Society of Gynecology and Obstetrics (S2k-Level, AWMF Registry No. 015/080, October 2016). *Geburtshilfe Frauenheilkd* 2017;**77**:1157–73.

45. Vayssiere C, Sentilhes L, Ego A, Bernard C, Cambourieu D, Flamant C, *et al*. Fetal growth restriction and intra-uterine growth restriction: guidelines for clinical practice from the French College of Gynaecologists and Obstetricians. *Eur J Obstet Gynecol Reprod Biol* 2015;**193**:10–8.

46. Centre for Reviews and Dissemination. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care [Internet]*. York: University of York; 2009. URL: www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm (accessed 23 March 2011).

47. Cochrane Diagnostic Test Accuracy Working Group. *Handbook for DTA Reviews [Internet]*. The Cochrane Collaboration; 2009. URL: http://srdta.cochrane.org/handbook-dta-reviews (accessed 23 March 2011).

48. Canadian Agency for Drugs and Technologies in Health. *PRESS – Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration (PRESS E&E) [Internet]*. Ottawa: CADTH; 2016. URL: www.cadth.ca/resources/finding-evidence/press (accessed 16 November 2020).

49. Abulhaj Martínez M, Martínez Chica S, Rodríguez García I, Redondo Aguilar RM, Teva García MJ. Analysis of the detection rate by ultrasound of intrauterine growth restricted and small for gestational age fetuses. *Rev Chil Obstet Ginecol* 2012;**77**:259–62.

50. Australian New Zealand Clinical Trials Registry (ACTRN). *Investigating the Utility of the Customised Fetal Growth Chart: A Randomised Controlled Trial. ACTRN12612001215819. In WHO International Clinical Trials Registry Platform (ICTRP) [Internet]*. Geneva: World Health Organization (WHO); 2012. URL: https://trialsearch.who.int/Trial2.aspx?TrialID=ACTRN12612001215819 (accessed 16 August 2023).

51. Adam S, Lombaard HAD, van Zyl DG. Are we missing at-risk babies? Comparison of customised growth charts v. standard population charts in a diabetic population. *S Afr J Obstet Gynaecol* 2014;**20**:88–90.

52. Adams JH, Racine JL, Rhoades J, Meyer M, Poehlmann J, Hoppe KK, *et al*. 991: Ahead of the curve? Hadlock versus Brenner for predicting small for gestational age infants. Presented at SMFM 40th Annual Meeting: The Pregnancy Meeting; 3–8 Feb 2020; Grapevine United States. *Am J Obstet Gynecol* 2020;**222**:S616.
53. Aderoba AK, Ioannou C, Kurinczuk JJ, Quigley MA, Cavallaro A, Impey L; Oxford Growth Restriction Identification Programme (OxGRIP) Group. The impact of a universal late third-trimester scan for fetal growth restriction on perinatal outcomes in term singleton births: a prospective cohort study. *BJOG* 2023;**130**:791–802.
54. Aiken C, Rowley A, Dyer E, Scott J. 869: Can we improve ultrasonographic estimation of fetal weight by reducing operator bias? Presented at SMFM 39th Annual Meeting: The Pregnancy Meeting; 11–16 Feb 2019; Las Vegas, USA. *Am J Obstet Gynecol* 2019;**220**:S565.
55. Akolekar R, Panaitescu AM, Ciobanu A, Syngelaki A, Nicolaides KH. Two-stage approach for prediction of small-for-gestational-age neonate and adverse perinatal outcome by routine ultrasound examination at 35–37 weeks' gestation. *Ultrasound Obstet Gynecol* 2019;**54**:484–91.
56. Ali M, Miller J, Chan C, Fields J, Houston L, Bernhard K, *et al*. Prenatal detection of fetal growth restriction in twins: the TWIG study. *Prenat Diagn* 2017;**37**:101–2.
57. Andreasen LA, Tabor A, Nørgaard LN, Ringsted C, Sandager P, Rosthøj S, Tolsgaard MG. Multicenter randomized trial exploring effects of simulation-based ultrasound training on obstetricians' diagnostic accuracy: value for experienced operators. *Ultrasound Obstet Gynecol* 2020;**55**:523–9.
58. Andrews C, Burns A, Middleton P, Ellwood D, Gordon A, Davies-Tuck M, *et al*. Benchmarking to drive service quality improvement to reduce stillbirths and monitor impact of the safer baby bundle. Presented at Perinatal Society of Australia and New Zealand Annual Congress (PSANZ 2022); 16–18 May 2023; Adelaide, SA Australia. *J Paediatr Child Health* 2022;**58**:48–9.
59. Angarita AM, Mehta V, Roman A, Jayakumaran JS, Ramirez A, Martins JG, Al-Kouatly HB. Prediction of SGA at anatomy ultrasound: EFW percentiles and estimated due date difference. Presented at SMFM 42nd Annual Meeting: The Pregnancy Meeting; 31 Jan–5 Feb 2022; Orlando, US. *Am J Obstet Gynecol* 2022;**226**:S463–4.
60. Ashfaq N, Ghani N, Zafar R, Aalam W, Touqir S. HC/AC ratio as a predictor of intrauterine growth retardation during pregnancy: diagnostic tool. *Pak J Med Health Sci* 2022;**16**:138–9.
61. Ashfaq N, Akram U, Zafar R, Khalid R. Diagnostic accuracy of TCD/AC ratio in predicting intrauterine growth retardation. *Pak J Med Health Sci* 2022;**16**:124–5.
62. Bailey E, Bosse B, Wetterhahn M, Iruretagoyena JI, Rhoades J, Adams JH. Is an isolated short femur an indication for growth ultrasounds? Presented at SMFM 42nd Annual Meeting: The Pregnancy Meeting; 31 Jan–5 Feb 2022; Orlando, US. *Am J Obstet Gynecol* 2022;**226**:S489.
63. Barel O, Maymon R, Elovits M, Smorgick N, Tovbin J, Vaknin Z. Evaluation of fetal weight estimation formulas in assessing small-for-gestational-age fetuses. *Ultraschall Med* 2016;**37**:283–9.
64. Basuki TR, Caradeux J, Eixarch E, Gratacos E, Figueras F. Longitudinal assessment of abdominal circumference versus estimated fetal weight in the detection of late fetal growth restriction. *Fetal Diagn Ther* 2019;**45**:230–7.
65. Beddow ME, Blue NR, Savabi M, Katukuri VR, Fritts C, Chao CR. Ultrasound prediction of small-for-gestational age at birth: the more, the merrier? Presented at 66th Annual Clinical and Scientific Meeting of the American College of Obstetricians and Gynecologists; 27–30 April 2018; Austin, TX. *Obstet Gynecol* 2018;**131**:194S–194S.
66. Bellussi F, Cataneo I, Visentin S, Simonazzi G, Lenzi J, Fantini MP, *et al*. Clinical validation of the INTERGROWTH-21st standards of fetal abdominal circumference for the prediction of small-for-gestational-age neonates in Italy. *Fetal Diagn Ther* 2017;**42**:198–203.
67. Bergman E, Kieler H, Petzold MG, Sonesson C, Axelsson O. Symphysis-fundus measurements for detection of small for gestational age pregnancies. *Acta Obstet Gynecol Scand* 2006;**85**:407–12.

68. Bhimarao, Nagaraju RM, Bhat V, Gowda PV. Efficacy of transcerebellar diameter/abdominal circumference versus head circumference/abdominal circumference in predicting asymmetric intrauterine growth retardation. *J Clin Diagn Res* 2015;**9**:TC01–5.

69. Bikmetova E, Trishkin A, Artymuk N. Fetal growth restriction: informativity of modern diagnosis methods. Presented at 11th World Congress of Perinatal Medicine 2013; 19–23 June 2013; Moscow, Russian Federation. *J Perinat Med* 2013;**41**.

70. Blue NR, Savabi M, Beddow ME, Katukuri VK, Fritts C, Izquierdo LA, Chao CR. Fetal growth surveillance in at-risk pregnancies: how soon is too soon? Presented at 38th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting 29 Jan–3 Feb 2018; Dallas, TX. *Am J Obstet Gynecol* 2018;**218**:S270.

71. Blue NR, Beddow ME, Savabi M, Katukuri VR, Fritts CM, Izquierdo LA, Chao CR. RCOG versus ACOG: whose method is better for diagnosing fetal growth restriction? Presented at 38th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 29 Jan–3 Feb 2018; Dallas, TX. *Am J Obstet Gynecol* 2018;**218**:S269.

72. Bocchi C, Azzolini E, Vannuccini S, Orlandini C, Boni C, Severi FM, *et al*. How to improve the accuracy of diagnostic ultrasound to identify SGA fetuses. Presented at 58th Annual Scientific Meeting of the Society for Gynecologic Investigation (SGI 2011); 16–19 March 2010; Miami Beach, FL. *Reprod Sci* 2011;**18**:90A.

73. Burd I, Srinivas S, Pare E, Dharan V, Wang E. Is sonographic assessment of fetal weight influenced by formula selection? *J Ultrasound Med* 2009;**28**:1019–24.

74. Callec R, Lamy C, Perdriolle-Galet E, Patte C, Heude B, Morel O; EDEN Mother–Child Cohort Study Group. Impact on obstetric outcome of third-trimester screening for small-for-gestational-age fetuses. *Ultrasound Obstet Gynecol* 2015;**46**:216–20.

75. Caradeux J, Eixarch E, Mazarico E, Basuki TR, Gratacos E, Figueras F. Second- to third-trimester longitudinal growth assessment for prediction of small-for-gestational age and late fetal growth restriction. *Ultrasound Obstet Gynecol* 2018;**51**:219–24.

76. Carlin A, Kadji C, De Angelis R, Cannie MM, Jani JC. Prenatal prediction of small-for-gestational age neonates using MR imaging: comparison with conventional 2D ultrasound. *J Matern Fetal Neonatal Med* 2019;**32**:1673–81.

77. Cassells K. Utility of ultrasound to predict aberrant fetal growth in women attending a dedicated antenatal clinic for morbid obesity. Presented at Annual Scientific Meeting of the Royal Australian and New Zealand College of Obstetricians and Gynaecologists (RANZCOG 2021); 15–18 Feb 2021; Virtual. *Aust N Z J Obstet Gynaecol* 2021;**61**:143.

78. Chauhan SP, Cole J, Sanderson M, Magann EF, Scardo JA. Suspicion of intrauterine growth restriction: use of abdominal circumference alone or estimated fetal weight below 10%. *J Matern Fetal Neonatal Med* 2006;**19**:557–62.

79. Chauhan S, Tajik P, Boers K, Van Wyk L, Mol BW, Scherjon S. Differentiating newborns with birth weight <vs> 3 percentiles for gestational age: secondary analysis of a randomized clinical trial (DIGITAT). Presented at 32nd Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 6–11 Feb 2012; Dallas, TX. *Am J Obstet Gynecol* 2012;**206**:S177–8.

80. Chauhan S, Abuhamad A, Chang E, Lutton T, Morrison J, Magsann E. Identification of growth restricted newborns at risk for neonatal death. Presented at 31st Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting 2011; 7–12 Feb 2011; San Francisco, CA. *Am J Obstet Gynecol* 2011;**204**:S176.

81. Chen C, Kon M, Rathi S, Vincent SJ, Ram KT. Effect of race in the diagnosis of fetal growth restriction in a multicultural urban population. Presented at 2022 Annual Clinical and Scientific Meeting of the American College of Obstetricians and Gynecologists; 6–8 May 2022; San Diego, CA. *Obstet Gynecol* 2022;**139**:58S–9S.

82. Choi SKY, Gordon A, Hilder L, Henry A, Hyett JA, Brew BK, *et al*. Performance of six birth-weight and estimated-fetal-weight standards for predicting adverse perinatal outcome: a 10-year nationwide population-based study. *Ultrasound Obstet Gynecol* 2021;**58**:264–77.

83. Ciobanu A, Khan N, Syngelaki A, Akolekar R, Nicolaides KH. Routine ultrasound at 32 vs 36 weeks' gestation: prediction of small-for-gestational-age neonates. *Ultrasound Obstet Gynecol* 2019;**53**:761–8.

84. Ciobanu A, Formuso C, Syngelaki A, Akolekar R, Nicolaides KH. Prediction of small-for-gestational-age neonates at 35–37 weeks' gestation: contribution of maternal factors and growth velocity between 20 and 36 weeks. *Ultrasound Obstet Gynecol* 2019;**53**:488–95.

85. Cordiez S, Deruelle P, Drumez E, Bodart S, Subtil D, Houfflin-Debarge V, Garabedian C. Impact of customized growth curves on screening for small for gestational age twins. *Eur J Obstet Gynecol Reprod Biol* 2017;**215**:28–32.

86. Craig AM, Kaur K, Heerboth SA, Chen H, Lauderdale CJ, Shannon C, Zuckerwise LC. Clinical implications of second and third trimester surveillance ultrasounds of growth-restricted fetuses. *AJP Rep* 2021;**11**:e132–6.

87. De Silva S, Adams F, Gullam J. Accuracy of estimated fetal weight measurement in the extremely obese. Presented at Royal Australian and New Zealand College of Obstretricans and Gynaecologists 2017 Annual Scientific Meeting (RANZCOG 2017); 29 Oct–1 Nov 2017; Auckland, New Zealand. *Aust N Z J Obstet Gynaecol* 2017;**57**:25.

88. de Sousa Basso NA, Morceli G, Costa R, Dias A, Rudge MVC, Calderon IMP. Validation of a symphysis-fundal height chart developed for pregnancy complicated by diabetes and hyperglycemia: an observational study. *Reprod Health* 2016;**13**:89.

89. Doulaveris G, Gallagher P, Romney E, Richley M, Gebb J, Rosner M, Dar P'er. Fetal abdominal circumference in the second trimester and prediction of small for gestational age at birth. *J Matern Fetal Neonatal Med* 2020;**33**:2415–21.

90. Dowe M, Tang AW, Arya R. Are we performing growth scans for the right reasons? Presented at British Maternal & Fetal Medicine Society (BMFMS) Annual Conference 2016; 21–22 April 2016; Birmingham, UK. *BJOG* 2016;**123**:42.

91. Dromey B, Nelson J, Everett T, Cohen K. Customised growth charts-effects on outcomes in a large teaching hospital. Presented at Annual Meeting of the Blair Bell Research Society 2019; 17–18 January 2019; Manchester, UK. *BJOG* 2019;**126**:e133.

92. Dude A, Davis B, Delaney K, Yee LM. Identifying fetal growth disorders using ultrasound in women with obesity. Presented at 65th Annual Scientific Meeting of the Society for Gynecologic Investigation (SGI 2018); 6–10 March 2018; San Diego, CA. *Reprod Sci* 2018;**25**:173A.

93. Dude AM, Yee LM. Identifying fetal growth disorders using ultrasound in women with diabetes. Presented at 64th Annual Scientific Meeting of the Society for Gynecologic Investigation (SGI 2017); 15–18 March 2017; Orlando, FL. *Reprod Sci* 2017;**24**:250A–1A.

94. Febres-Cordero DA, Ajayi A, Reforma LG, Trochtenberg AL, Modest AM, Spiel MH, Collier AY. Is fetal abdominal circumference predictive of small for gestational age birthweight in twin pregnancies? Presented at SMFM 42nd Annual Meeting: The Pregnancy Meeting; 31 Jan–5 Feb 2022; Orlando; US. *Am J Obstet Gynecol* 2022;**226**:S626.

95. Figueras FR. Comparison of the performance of estimated fetal weight charts for the detection of small- and large-for-gestational-age newborns with adverse outcomes: a French population-based study. *BJOG* 2022;**129**:949.

96. Finneran MM, Ware CA, Russo J, Webster S, Mathew S, Buhimschi IA, Buhimschi CS. Use of birth weight- vs. ultrasound-derived fetal weight classification methods: implications for detection of abnormal umbilical artery Doppler. *J Perinat Med* 2020;**48**:615–24.

97. Fitzgerald Z, Davies J, Goodfellow L. Detection of small for gestational age pregnancies is improved with customised growth charts without impacting upon the false positive rate. Presented at RCOG World Congress 2013; 24–26 June 2013; Liverpool, UK. *BJOG* 2013;**120**:83–4.

98. Francis A, Gardosi J. Effectiveness of ultrasound biometry at 34–36 weeks in the detection of small-for-gestational age at birth. Presented at 18th Annual Conference of the British Maternal and Fetal Medicine Society, BMFMS 2016; Birmingham, UK. *BJOG* 2016;**123**:86–7.

99. Francis A, Gardosi J. Effectiveness of ultrasound biometry at 34-36 weeks in the detection of SGA at birth. Presented at RCOG World Congress 2016; 20-22 June 2016; Birmingham, UK. BJOG 2016;123:22

100. Fuchs KM, D'Alton M. Can sonographer education and image review standardize image acquisition and caliper placement in 2D ultrasounds? Experience from the NICHD Fetal Growth Study. Presented at Society for Maternal–Fetal Medicine (SMFM) Annual Meeting 2012; 6–11 Feb 2012; Dallas, TX. *Am J Obstet Gynecol* 2012;**206**:S15–6.

101. Gabbay-Benziv R, Aviram A, Hadar E, Chen R, Bardin R, Wiznitzer A, Yogev Y. Pregnancy outcome after false diagnosis of fetal growth restriction. *J Matern Fetal Neonatal Med* 2017;**30**:1916–9.

102. Gabbay-Benziv R, Aviram A, Bardin R, Ashwal E, Melamed N, Hiersch L, *et al*. Prediction of small for gestational age: accuracy of different sonographic fetal weight estimation formulas. *Fetal Diagn Ther* 2016;**40**:205–13.

103. Gabbay-Benziv R, Aviram A, Ashwal E, Hiersch L, Melamed N, Hadar E, Yogev Y. Sonographic prediction of small for gestational age-which formula is more accurate? Presented at 36th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 1–6 Feb 2016; Atlanta, GA. *Am J Obstet Gynecol* 2016;**214**:S356.

104. Gardosi J, Giddings S, Clifford S, Wood L, Francis A. Association between reduced stillbirth rates in England and regional uptake of accreditation training in customised fetal growth assessment. *BMJ Open* 2013;**3**:e003942.

105. Gardosi J, Hugh O. Stillbirth risk and smallness for gestational age according to Hadlock, INTERGROWTH-21st, WHO, and GROW fetal weight standards: analysis by maternal ethnicity and body mass index. *Am J Obstet Gynecol* 2023;**79**:250–2.

106. Gawie-Rotman M, Menashe S, Haggiag N, Shrim A, Hallak M, Gabbay-Benziv R. The accuracy of sonographically estimated fetal weight and prediction of small for gestational age in twin pregnancy – comparison of the first and second twins. *J Clin Med* 2023;**12**:06.

107. Gelman M, Wilkof-Segev R, Gawie-Rotman M, Nadir E, Shrim A, Hallak M, Gabbay-Benziv R. Abdominal circumference discordance for prediction of small for gestational age at birth in twin pregnancies. *J Matern Fetal Neonatal Med* 2022;**35**:3573–8.

108. Gilmore S, Dougan C. Getting a better picture: a prospective audit on the accuracy in ultrasound-detection in patients managed as small for gestational age. *BMC Pregnancy Childbirth* 2017;**17**:P32.

109. Goetzinger KR, Tuuli MG, Odibo AO, Roehl KA, Macones GA, Cahill AG. Screening for fetal growth disorders by clinical exam in the era of obesity. *J Perinatol* 2013;**33**:352–7.

110. Goldshore M, Henderson J, Ehsanipoor R, Baghlaf H, Werner E. Sensitivity and specificity of ultrasound in obese gravida. Presented at 33rd Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 11–16 Feb 2013; San Francisco, CA. *Am J Obstet Gynecol* 2013;**208**:S166.

111. Grantz KL, Ortega-Villa AM, Pugh SJ, Bever A, Grobman W, Newman RB, *et al*. Combination of fundal height and ultrasound to predict small for gestational age at birth. *Am J Perinatol* 2023;**40**:155–62.

112. Green B, Brownfoot F. Detecting the small for gestational age fetus at term: third-trimester ultrasound versus clinical evaluation. Presented at Royal Australian and New Zealand College of Obstretricans and Gynaecologists 2017 Annual Scientific Meeting (RANZCOG 2017); 29 Oct–1 Nov 2017; Auckland, New Zealand. *Aust N Z J Obstet Gynaecol* 2017;**57**:45.

113. Grzybek A, Orr K, Nicoll A. Accurately identifying the small-for-gestational-age fetus: determining the ideal method for our population in NHS Tayside. *BJOG* 2019;**126**:67–8.

114. Gupta LM, Gaston L, Chauhan SP. Detection of fetal growth restriction with preterm severe preeclampsia: experience at two tertiary centers. *Am J Perinatol* 2008;**25**:247–9.

115. Hebbar S, Rai L, Ali Z. Critical evaluation of various sonologic parameters of early foetal growth discrepancies in predicting adverse pregnancy outcomes. *J Clin Diagn Res* 2018;**12**:QC06–10.

116. Hairston JC, Lu L, Bejerano S, Fuchs K, Miller RS, Simpson L, *et al*. Expanded ultrasound criteria for fetal growth restriction and detection of small for gestational age neonates. Presented at SMFM 42nd Annual Meeting: The Pregnancy Meeting; 31 Jan–5 Feb 2022; Orlando, US. *Am J Obstet Gynecol* 2022;**226**:S461–2.

117. Haragan A, Hulsey T, Newman R, Chang E. Screening for fetal growth abnormalities in the obese woman: a comparison of fundal height to bedside fetal abdominal circumference. Presented at 35th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 2–7 Feb 2015; San Diego, CA. *Am J Obstet Gynecol* 2015;**212**:S151.

118. Honart AW, Reichert E, Price JT, Kumwenda A, Appiagyei A, Lubeya MK, *et al*. 976 The utility of third trimester fetal biometry in predicting SGA infants in a Zambian population. Presented at SMFM 41st Annual Meeting: The Pregnancy Meeting; 25–30 Jan 2021; Virtual, Online. *Am J Obstet Gynecol* 2021;**224**:S605–6.

119. Horwitz G, Trostle ME, Finning S, Stein G, Roman AS. 1123 Midtrimester fetal growth restriction as predictor of small for gestational age and adverse obstetrical outcomes. Presented at SMFM 41st Annual Meeting: The Pregnancy Meeting; 25–30 Jan 2021; Virtual, Online. *Am J Obstet Gynecol* 2021;**224**:S692.

120. Iliodromiti S, Smith GCS, Lawlor DA, Pell JP, Nelson SM. UK stillbirth trends in over 11 million births provide no evidence to support effectiveness of Growth Assessment Protocol program. *Ultrasound Obstet Gynecol* 2020;**55**:599–604.

121. Ivars J, Houfflin-Debarge V, Vaast P, Deruelle P. Accuracy of ultrasonography fetal weight estimation in twin pregnancies. *Gynecol Obstet Fertil* 2010;**38**:740–6.

122. Kabiri D, Romero R, Gudicha DW, Hernandez-Andrade E, Pacora P, Benshalom-Tirosh N, *et al*. Prediction of adverse perinatal outcome by fetal biometry: comparison of customized and population-based standards. *Ultrasound Obstet Gynecol* 2020;**55**:177–88.

123. Khan DB, Bari V, Chishty IA. Ultrasound in the diagnosis and management of intrauterine growth retardation. *J Coll Physicians Surg Pak* 2004;**14**:601–4.

124. Kim MA, Han GH, Kim YH. Prediction of small-for-gestational age by fetal growth rate according to gestational age. *PLOS ONE* 2019;**14**:e0215737.

125. Konwar R, Basumatary B, Dutta M, Mahanta P. Accuracy of fetal weight estimation by ultrasonographic evaluation in a northeastern region of India. *Int J Biomater* 2021;**2021**:9090338.

126. Kopparam R, Matthews K, Khan R, Grossman TB, Gelber S. Predictive value of fetal abdominal circumference in diagnosing small for gestational age infants. Presented at 67th Annual Clinical and Scientific Meeting of the American College of Obstetricians and Gynecologists; 3–6 May 2019; Nashville, TN. *Obstet Gynecol* 2019;**133**:17S.

127. Landres IV, Clark A, Chasen ST. Improving antenatal prediction of small-for-gestational-age neonates by using customized versus population-based reference standards. *J Ultrasound Med* 2013;**32**:1581–6.

128. Landres I, Clark A, Chasen ST. Improving antenatal detection of intra uterine growth restriction by using customized versus population based growth curves. Presented at 31st Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 7–12 Feb 2011; San Francisco, CA. *Am J Obstet Gynecol* 2011;**204**:S126–7.

129. Larkin JC, Chauhan SP, Simhan HN. Small for gestational age: the differential mortality when detected versus undetected antenatally. *Am J Perinatol* 2017;**34**:409–14.

130. Larkin J, Chauhan S, Simhan H. Small for gestational age: the differential mortality when detected versus undetected antenatally. Presented at 35th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 2–7 Feb 2015; San Diego, CA. *Am J Obstet Gynecol* 2015;**212**:S44.

131. Law LW, Leung TY, Sahota DS, Chan LW, Fung TY, Lau TK. Which ultrasound or biochemical markers are independent predictors of small-for-gestational age? *Ultrasound Obstet Gynecol* 2009;**34**:283–7.

132. Leon-Martinez D, Lundsberg LS, Culhane J, Zhang J, Son M, Reddy UM. Fetal growth restriction and small for gestational age as predictors of neonatal morbidity: which growth nomogram to use? *Am J Obstet Gynecol* 2023;**229**:678.e1–16.

133. Lesmes C, Gallo DM, Panaiotova J, Poon LC, Nicolaides KH. Prediction of small-for-gestational-age neonates: screening by fetal biometry at 19–24 weeks. *Ultrasound Obstet Gynecol* 2015;**46**:198–207.

134. Luo H, Zhou S, Yang T, Liu S, Xu H. Ultrasonic multiple measurement estimation of fetal weight in parturient. *Hua Xi Yi Ke Da Xue Xue Bao* 2001;**32**:618–20.

135. Ma J, Cheng D, Zhang Z, Cai B, Xu X. Evaluating the accuracy of sonographic fetal weight estimations using the Hadlock IV formula in a Chinese population. *Quant Imaging Med Surg* 2023;**13**:3726–34.

136. Mailath-Pokorny M, Polterauer S, Worda K, Springer S, Bettelheim D. Isolated short fetal femur length in the second trimester and the association with adverse perinatal outcome: experiences from a tertiary referral center. *PLOS ONE* 2015;**10**:e0128820.

137. Maines JL, Mirault J, Ibekwe T, Hameroff A, Ural S, Pauli JM, Curtin W. 358: declining NICU admissions in uncomplicated fetal growth restriction following adoption of ACOG/SMFM practice guidelines. Presented at SMFM 40th Annual Meeting: The Pregnancy Meeting; 3–8 Feb 2020; Grapevine, USA. *Am J Obstet Gynecol* 2020;**222**:S239–40.

138. Malin G, Anblagan D, Bugg G, Gowland P, Jones NW, Robinson R. Diagnostic accuracy of antenatal magnetic resonance imaging (MRI) to predict birth weight > 90th centile or < 10th centile in the third trimester. Presented at Perinatal Medicine 2014; 9–11 June 2014; Harrogate, UK. *Arch Dis Child Fetal Neonatal Ed* 2014;**99**:A94.2–5.

139. Matthews KC, Kopparam R, Khan R, Sholle E, Gelber SE. 1045: which is better: fetal AC or EFW in predicting SGA infants? Presented at SMFM 39th Annual Meeting: The Pregnancy Meeting; 11–16 Feb 2019; Las Vegas, US. *Am J Obstet Gynecol* 2019;**220**:S670–1.

140. McKenna D, Tharmaratnam S, Mahsud S, Bailie C, Harper A, Dornan J. A randomized trial using ultrasound to identify the high-risk fetus in a low-risk population. *Obstet Gynecol* 2003;**101**:626–32.

141. McKenna D, Tharmaratnam S, Mahsud S, Bailie C, Harper A, Dornan J, *et al*. Ultrasound scans reduced the risk of small for gestational age infants and increased antenatal interventions in low-risk pregnancies. *Evid Based Obstet Gynecol* 2004;**6**:6–7.

142. McKeown CD, NiLaighin C, Ismail KI, Hannigan A, O'Donoghue K, Cotter A, *et al*. Antenatal detection of small for gestational age by universal single-scan screening in early third trimester. *BJOG* 2018;**125**:27.

143. Mendoza-Carrera CE, Acevedo-Gallegos S, Lumbreras-Marquez M, Gallardo-Gaona JM, Copado-Mendoza DY, Rodriguez-Sibaja MJ. Comparison of four fetal growth charts in the prediction of adverse perinatal outcomes in a tertiary hospital in Mexico. *Ginecol Obstet Mex* 2021;**89**:704–14.

144. Merriam A, Jain V, Shlossman P, Hoffman M. Does abdominal circumference less than the 5 percentile improve detection of fetal growth restriction? Presented at 34th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 3–8 Feb 2014; New Orleans, LA. *Am J Obstet Gynecol* 2014;**210**:S102.

145. Meyer M, Adams JH, Iruretagoyena JI, Antony KM, Schierloh H, Poehlmann J, *et al*. 734: Which fetal growth curve best predicts a small for gestational age infant in obese women? Presented at SMFM 40th Annual Meeting: The Pregnancy Meeting; 3–8 Feb 2020; Grapevine, Texas. *Am J Obstet Gynecol* 2020;**222**:S464–5.

146. Michaeli J, Rozitzky A, Grisaru-Granovsky S, Moshe NS. 308: Fetal weight estimation, mode of delivery and neonatal outcome: effect of cognitive bias in obstetrics. Presented at SMFM 40th Annual Meeting: The Pregnancy Meeting; 3–8 Feb 2020; Grapevine, US. *Am J Obstet Gynecol* 2020;**222**:S207.

147. Miranda J, Rodriguez-Lopez M, Triunfo S, Sairanen M, Kouru H, Parra-Saavedra M, *et al*. Prediction of fetal growth restriction using estimated fetal weight vs a combined screening model in the third trimester. *Ultrasound Obstet Gynecol* 2017;**50**:603–11.

148. Miranda J, Maestre N, Paternina-Caicedo A, Parra-Saavedra M, Caradeux J, Sepulveda-Martinez A, *et al*. Performance of the INTERGROWTH-21st and World Health Organization fetal growth charts for the detection of small-for-gestational age neonates in Latin America. *Int J Gynaecol Obstet* 2023;**161**:1083–91.

149. Monier I, Ancel PY, Ego A, Jarreau PH, Lebeaux C, Kaminski M, *et al*.; EPIPAGE 2 Study Group. Fetal and neonatal outcomes of preterm infants born before 32 weeks of gestation according to antenatal vs postnatal assessments of restricted growth. *Am J Obstet Gynecol* 2017;**216**:516.e1–516.e10.

150. Mori S, Kanamori S, Shinoda S, Inamura T, Shibata T, Ueno T, *et al*. Ultrasonographic assessment of FGR for prediction of SGA. Presented at 72nd Annual Congress of the Japan Society of Obstetrics and Gynecology; 23–28 April 2020; Tokyo, Japan. *J Obstet Gynaecol Res* 2020;**46**:1515.

151. Muller B, Crowe R, Haney A, Chang E, Newman R, Finneran MM. 929 Predictive performance of race/ethnicity based growth velocities in pregnancies complicated by diabetes. Presented at SMFM 41st Annual Meeting: The Pregnancy Meeting; 25–30 Jan 2021; Virtual, Online. *Am J Obstet Gynecol* 2021;**224**:S576–7.

152. Nadham AN, Naas BAM, Dayoub N. Factors affecting the accuracy of ultrasound fetal weight assessment among diabetic patients. *Bahrain Med Bull* 2019;**41**:21–4.

153. Najafzadeh A, Graves AR. Screening for fetal growth restriction with universal third trimester ultrasonography in nulliparous women in the Pregnancy Outcome Prediction (POP) study: a prospective cohort study. *Sonography* 2016;**3**:70–1.

154. Neel A, Cunningham CE, Teale GR. A routine third trimester growth ultrasound in the obese pregnant woman does not reliably identify fetal growth abnormalities: a retrospective cohort study. *Aust N Z J Obstet Gynaecol* 2021;**61**:116–22.

155. Odibo AO, Nwabuobi C, Sinkey R, Odibo L, Tuuli MG. INTERGROWTH-21st (IG-21) compared to the NICHD race-specific growth standard for identifying pregnancies at risk for neonatal small for gestational age (SGA) and adverse outcomes? Presented at 38th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 29 Jan–3 Feb 2018; Dallas, TX. *Am J Obstet Gynecol* 2018;**218**:S306.

156. Odibo A, Obican S, Odibo L, Sinkey R, Tuuli MG. Customized versus NICHD race-specific growth charts for identifying pregnancies at risk for small for gestational age and neonatal adverse outcomes. Presented at 38th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 29 Jan–3 Feb 2018; Dallas, TX. *Am J Obstet Gynecol* 2018;**218**:S305–6.

157. Ott WJ. Diagnosis of intrauterine growth restriction: comparison of ultrasound parameters. *Am J Perinatol* 2002;**19**:133–7.

158. Owen P, Ogah J, Bachmann LM, Khan KS. Prediction of intrauterine growth restriction with customised estimated fetal weight centiles. *BJOG* 2003;**110**:411–5.

159. Owen P, Burton K, Ogston S, Khan KS, Howie PW. Using unconditional and conditional standard deviation scores of fetal abdominal area measurements in the prediction of intrauterine growth restriction. *Ultrasound Obstet Gynecol* 2000;**16**:439–44.

160. Padilla-Amigo C, Lopez-Felix J, Martfnez-Villafana E, Sierra-Lozada N. Prediction of fetal growth restriction with the 11–14 sieve algorithm. *Ginecol Obstet Mex* 2022;**90**:1–7.

161. Palaniappan S, Heathcote J, Gandham S, Reddy A. Obstetric consequence of antenatal detection of small for gestational infants. Presented at RCOG World Congress 2013; 24–26 June 2013; Liverpool, UK. *BJOG* 2013;**120**:148–9.

162. Papastefanou I, Thanopoulou V, Dimopoulou S, Syngelaki A, Akolekar R, Nicolaides KH. Competing-risks model for prediction of small-for-gestational-age neonate at 36 weeks' gestation. *Ultrasound Obstet Gynecol* 2022;**60**:612–9.

163. Papastefanou I, Nowacka U, Buerger O, Akolekar R, Wright D, Nicolaides KH. Evaluation of the RCOG guideline for the prediction of neonates that are small for gestational age and comparison with the competing risks model. *BJOG* 2021;**128**:2110–5.

164. Papastefanou I, Nowacka U, Syngelaki A, Mansukhani T, Karamanis G, Wright D, Nicolaides KH. Competing risks model for prediction of small-for-gestational-age neonates from biophysical markers at 19 to 24 weeks' gestation. *Am J Obstet Gynecol* 2021;**225**:530.e1–19.

165. Papastefanou I, Nowacka U, Syngelaki A, Dragoi V, Karamanis G, Wright D, Nicolaides KH. Competing-risks model for prediction of small-for-gestational-age neonate from estimated fetal weight at 19–24 weeks' gestation. *Ultrasound Obstet Gynecol* 2021;**57**:917–24.

166. Papastefanou I, Souka AP, Eleftheriades M, Pilalis A, Chrelias C, Kassanos D. Predicting fetal growth deviation in parous women: combining the birth weight of the previous pregnancy and third trimester ultrasound scan. *J Perinat Med* 2015;**43**:485–92.

167. Patel V, Resnick K, Liang C, Smith M, Haghpeykar HS, Mastrobattista JM, Gandhi M. Midtrimester ultrasound predictors of small-for-gestational-age neonates. *J Ultrasound Med* 2020;**39**:2027–31.

168. Paulsen CB, Nielsen BB, Msemo OA, Moller SL, Ekmann JR, Theander TG, *et al*. Anthropometric measurements can identify small for gestational age newborns: a cohort study in rural Tanzania. *BMC Pediatr* 2019;**19**:120.

169. Pay ASD, Froen JF, Staff AC, Jacobsson B, Gjessing HK. Symphysis-fundus measurement–the predictive value of a new reference curve. *Tidsskr Nor Laegeforen* 2017;**137**:717–20.

170. Pay A, Froen JF, Staff AC, Jacobsson B, Gjessing HK. Prediction of small-for-gestational-age status by symphysis-fundus height: a registry-based population cohort study. *BJOG* 2016;**123**:1167–73.

171. Pils S, Springer S, Seemann R, Wehrmann V, Worda C, Ott J. Reliability of sonographic fetal weight estimation in triplet pregnancies: a retrospective cohort study. *Arch Gynecol Obstet* 2018;**297**:1441–7.

172. Policiano C, Mendes JM, Fonseca A, Barros J, Vargas S, Cal M, *et al*. Routine ultrasound at 30th–33rd weeks versus 30th–33rd and 35th–37th weeks in low-risk pregnancies: a randomized trial. *Fetal Diagn Ther* 2022;**49**:425–33.

173. Prechapanich J, Thitadilok W. Comparison of the accuracy of fetal weight estimation using clinical and sonographic methods. *J Med Assoc Thai* 2004;**87**:S1–7.

174. Premru-Srsen T, Brzan Simenc G, Lucovnik M, Trojner Bregar A, Verdenik I. Antenatal detection of a small-for-gestational-age fetus and consequences for obstetric management and neonatal outcomes. Presented at 14th World Congress of Perinatal Medicine; 11–14 Sept 2019; Istanbul, Turkey. *J Perinat Med* 2019;**47**:eA110.

175. Priya Appanamuthu Mari B, Rathore AM, Gupta MM, Agarwal K, Bhasin S, Singh P. Screening for small-for-gestational age fetus using rcog risk factors and symphysis-fundal height. Presented at Royal College of Obstetricians and Gynaecologists World Congress (RCOG 2021); 9–12 June 2021; Virtual. *BJOG* 2021;**128**:102.

176. Procas-Ramon B, Hierro-Espinosa C, Salim I, Impey L, Ioannou C. The impact of individual sonographer variation on the detection of small for gestational age fetuses using a third trimester growth scan. *J Clin Ultrasound* 2021;**49**:442–50.

177. Quinton A, Cook C, Peek M. The prediction of the small for gestational age fetus with the head circumference to abdominal circumference (HC/AC) ratio: a new look at an old measurement. *Sonography* 2015;**2**:27–31.

178. Radhika B, Suseela V, Nirmalan PK. Identification of fetal growth patterns with customized growth charts: a prospective study in South India. *Int J Infertil Fetal Med* 2015;**6**:30–4.

179. Ravangard S, Leftwich H, Schmidt B, Hibbard J. Risk of small-for-gestational-age among pregnancies with an estimated fetal weight between the 10th to 20th percentile. Presented at 34th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 3–8 Feb 2014; New Orleans, LA. *Am J Obstet Gynecol* 2014;**210**:S102–3.

180. Reboul Q, Delabaere A, Luo ZC, Nuyt AM, Wu Y, Chauleur C, *et al*. Prediction of small-for-gestational-age neonate by third-trimester fetal biometry and impact of ultrasound-delivery interval. *Ultrasound Obstet Gynecol* 2017;**49**:372–8.

181. Roma E, Arnau A, Berdala R, Bergos C, Montesinos J, Figueras F. Ultrasound screening for fetal growth restriction at 36 vs 32 weeks' gestation: a randomized trial (ROUTE). *Ultrasound Obstet Gynecol* 2015;**46**:391–7.

182. Safonova I, Grebinichenko G. Antenatal ultrasound monitoring and prediction of perinatal outcomes in the early and late onset of fetal growth retardation. Presented at 25th European Congress of Perinatal Medicine; 15–18 June 2016; Maastricht, Netherlands. *J Matern Fetal Neonatal Med* 2016;**29**:137.

183. Sanchez Fernandez M, Garcia Cotes AE, Munoz Gallardo MDC, Corral Segade ME, Mozas Moreno J, Jimenez-Moleon JJ. Screening for intrauterine growth restriction: which is its diagnostic validity through two-dimensional ultrasound? Presented at 12th World Congress of Perinatal Medicine 2015; 3–6 Nov 2015; Madrid, Spain. *J Perinat Med* 2015;**43**.

184. Sandlin AT, Magann EF, Ounpraseuth ST, Dahlke JD, Igwe E, Chang E, *et al*. Significant variability in antenatal detection of small-for-gestational age neonates among tertiary centers. Presented at 59th Annual Scientific Meeting of the Society for Gynecologic Investigation (SGI 2012); 21–24 March 2012; San Diego, CA. *Reprod Sci* 2012;**19**:371A.

185. Saviron-Cornudella R, Esteban LM, Aznar-Gimeno R, Dieste-Perez P, Perez-Lopez FR, Campillos JM, *et al*. Prediction of late-onset small for gestational age and fetal growth restriction by fetal biometry at 35 weeks and impact of ultrasound-delivery interval: comparison of six fetal growth standards. *J Clin Med* 2021;**10**:2984.

186. Schreiber V, Hurst C, da Silva Costa F, Stoke R, Turner J, Kumar S. Definitions matter: detection rates and perinatal outcome for infants classified prenatally as having late fetal growth restriction using SMFM biometric vs ISUOG/Delphi consensus criteria. *Ultrasound Obstet Gynecol* 2023;**61**:377–85.

187. Schreiber V, Hurst C, Stoke R, Turner J, Kumar S. Definitions matter: detection rates and perinatal outcomes for infants prenatally identified with late FGR using biometric versus consensus delphi criteria. Presented at Perinatal Society of Australia and New Zealand Annual Congress (PSANZ 2022); 15–18 May 2022; Adelaide, SA. *J Paediatr Child Health* 2022;**58**:56.

188. Scott KH, Hopper L, Huda S. Late preterm inductions of labour: a 12-month retrospective audit. *BJOG* 2019;**126**:70.

189. Sekar R, Khatun M, Barrett HL, Duncombe G. A prospective pilot study in assessing the accuracy of ultrasound estimated fetal weight prior to delivery. *Aust N Z J Obstet Gynaecol* 2016;**56**:49–53.

190. Severi FM, Azzolini E, Vannuccini S, Vellucci F, Imperatore A, Bocchi C, *et al*. Spontaneous preterm birth and fetal weight. Presented at 58th Annual Scientific Meeting of the Society for Gynecologic Investigation (SGI 2011); 16–19 March 2011; Miami Beach, FL. *Reprod Sci* 2011;**18**:310A–1A.

191. Shen Y, Zhao W, Lin J, Liu F. Accuracy of sonographic fetal weight estimation prior to delivery in a Chinese Han population. *J Clin Ultrasound* 2017;**45**:465–71.

192. Shmueli A, Chen R, Aviram A, Wiznitzer A, Meizner I, Bardin R, Gabbay-Benziv R. Predictors for false sonographic estimation of fetal weight extremes – why are we wrong? Presented at 37th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 26–28 Jan 2017; Las Vegas, NV. *Am J Obstet Gynecol* 2017;**216**:S151.

193. Sim D, Beattie RB, Dornan JC. Evaluation of biophysical fetal assessment in high-risk pregnancy to assess ultrasound parameters suitable for screening in the low-risk population. *Ultrasound Obstet Gynecol* 1993;**3**:11–7.

194. Sinkey RG, Odibo L, Tuuli MG, Odibo AO. NICHD racial/ethnic standards for fetal growth-do they predict small for gestational age fetuses and short-term neonatal outcomes better than Hadlock's chart? Presented at 37th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 23–28 Jan 2017; Las Vegas, NV. *Am J Obstet Gynecol* 2017;**216**:S157–8.

195. Sokol Karadjole V, Agarwal U, Berberovic E, Poljak B, Alfirevic Z. Does serial 3rd trimester ultrasound improve detection of small for gestational age babies: comparison of screening policies in 2 European maternity units. *Eur J Obstet Gynecol Reprod Biol* 2017;**215**:45–9.

196. Sotiriadis A, Figueras F, Eleftheriades M, Papaioannou GK, Chorozoglou G, Dinas K, Papantoniou N. First-trimester and combined first- and second-trimester prediction of small-for-gestational age and late fetal growth restriction. *Ultrasound Obstet Gynecol* 2019;**53**:55–61.

197. Souka AP, Papastefanou I, Pilalis A, Michalitsi V, Kassanos D. Performance of third-trimester ultrasound for prediction of small-for-gestational-age neonates and evaluation of contingency screening policies. *Ultrasound Obstet Gynecol* 2012;**39**:535–42.

198. Sovio U, Smith GCS. Comparison of estimated fetal weight percentiles near term for predicting extremes of birthweight percentile. *Am J Obstet Gynecol* 2021;**224**:292.e1–19.

199. Sparks TN, Cheng YW, McLaughlin B, Esakoff TF, Caughey AB. Fundal height: a useful screening tool for fetal growth? *J Matern Fetal Neonatal Med* 2011;**24**:708–12.

200. Price K, Cotton J, Bapir M. *Audit of the Detection of Small for Gestational Age (SGA) and Fetal Growth Restriction (FGR) in Babies Born in North Tees and Hartlepool NHS Foundation Trust. Presented at Royal College of Obstetricians and Gynaecologists (RCOG) International Congress 2016; 20–22 Jun 2016; Birmingham, UK, 20-Jun-2016.* 2016. URL: www.epostersonline.com/rcog2016/node/1731?view=true (accessed 8 July 2025).

201. Stanger E, Verma A, Pinto S. Identification of SGA (small for gestational age) – a comparative study of fetal weight estimation at birth using population based chart and customised centiles. Presented at Royal College of Obstetricians and Gynaecologists (RCOG) International Congress 2013; 24–26 June 2013; Liverpool, UK. *BJOG* 2013;**120**:132.

202. Stirnemann JJ, Benoist G, Salomon LJ, Bernard JP, Ville Y. Optimal risk assessment of small-for-gestational-age fetuses using 31–34-week biometry in a low-risk population. *Ultrasound Obstet Gynecol* 2014;**43**:311–6.

203. Stout MJ, Tuuli MG, Young O, Macones GA, Dicke JM, Cahill AG. Accuracy of growth restriction diagnosis-implications for testing and intervention. Presented at 36th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 1–6 Feb 2016; Atlanta, GA. *Am J Obstet Gynecol* 2016;**214**:S120.

204. Strassberg ER, Schuster M, Rajaram AM, Paglia MJ, Neubert AG, Ross JW, *et al*. Comparing diagnosis of fetal growth restriction and the potential impact on management and outcomes using different growth curves. *J Ultrasound Med* 2019;**38**:3273–81.

205. Swanzy L, Larkin J. Neonatal morbidity is increased with the inaccurate diagnosis of fetal growth restriction. Presented at 37th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 23–28 Jan 2017; Las Vegas, NV. *Am J Obstet Gynecol* 2017;**216**:S139–40.

206. Temming LA, Fears F, Rampersad R, Tuuli MG, Macones GA, Cahill AG. Midtrimester fetal growth restriction: who is at risk of remaining small? Presented at 64th Annual Scientific Meeting of the Society for Gynecologic Investigation (SGI 2017); 15–18 March 2017; Orlando, FL. *Reprod Sci* 2017;**24**:135A.

207. Thiebaugeorges O, Fresson J, Audibert F, Guihard-Costa AM, Frydman R, Droulle P. Diagnosis of small-for-gestational-age fetuses between 24 and 32 weeks, based on standard sonographic measurements. *Ultrasound Obstet Gynecol* 2000;**16**:49–55.

208. Trudell A, Tuuli M, Cahill A, MacOnes G, Odibo A. Stillbirth prediction and small for gestational age: comparison of a gender specific versus a non-gender specific growth curve. Presented at 33rd Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 11–16 Feb 2013; San Francisco, CA. *Am J Obstet Gynecol* 2013;**208**:S135.

209. Turitz A, Quant H, Schwartz N, Elovitz M, Bastek J. Isolated abdominal circumference <5% is an independent risk factor for small for gestational age infants. Presented at 33rd Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 11–16 Feb 2013; San Francisco, CA. *Am J Obstet Gynecol* 2013;**208**:S158.

210. Turitz A, Quant H, Schwartz N, Elovitz M, Bastek J. Estimated fetal weight 10–19% is an independent risk factor for small for gestational age infants. Presented at 33rd Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 11–16 Feb 2013; San Francisco, CA. *Am J Obstet Gynecol* 2013;**208**:S157–8.

211. Turner S, Williams M, Hugh O, Gardosi J. Antenatal detection of SGA and false-positive rates in the GAP programme. *BJOG* 2019;**126**:64.

212. Turner S, Williams M, Wood L, Gardosi J. Small-for-gestational-age referral rates according to maternal body mass index before and after growth assessment protocol training. *BJOG* 2016;**123**:39.

213. Turner S, Williams M, Hugh O, Gardosi J. Improvement in antenatal detection of small-for-gestational-age babies following implementation of the growth assessment protocol (GAP) in the UK. Presented at 2019 World Congress of the Royal College of Obstretriscians and Gynaecologists (RCOG 2019); 17–19 June 2019; London, UK. *BJOG* 2019;**126**:51.

214. Turner S, Williams M, Wood L, Gardosi J. SGA referral rates before and after implementation of the growth assessment protocol. Presented at RCOG World Congress 2016; 20–22 June 2016; Birmingham, UK. *BJOG* 2016;**123**:13.

215. Van De Kamp K, Pajkrt E, Zwinderman AH, Van Der Post JA, Snijders RJM. Validation of reference charts for mid-trimester fetal biometry. *Fetal Diagn Ther* 2019;**45**:42–9.

216. Verger C, Moraitis AA, Barnfield L, Sovio U, Bamfo J. Performance of different fetal growth charts in prediction of large-for-gestational age and associated neonatal morbidity in multiethnic obese population. *Ultrasound Obstet Gynecol* 2020;**56**:73–7.

217. Vieira LA, Skupski DW, Lin SN, Eglinton GS. Does abdominal circumference <20th percentile predict fetal growth restriction? Presented at 61st Annual Scientific Meeting of the Society for Gynecologic Investigation (SGI 2014); 26–29 March 2014; Florence, Italy. *Reprod Sci* 2014;**21**:356A–7A.

218. Walid MS, Pomortsev AV. Early screening for IUGR: comparison of two related echographic markers. *Arch Gynecol Obstet* 2009;**279**:551–6.

219. Wanyonyi SZ, Orwa J, Ozelle H, Martinez J, Atsali E, Vinayak S, *et al*. Routine third-trimester ultrasound for the detection of small-for-gestational age in low-risk pregnancies (ROTTUS study): randomized controlled trial. *Ultrasound Obstet Gynecol* 2021;**57**:910–6.

220. Wright D, Wright A, Smith E, Nicolaides KH. Impact of biometric measurement error on identification of small- and large-for-gestational-age fetuses. *Ultrasound Obstet Gynecol* 2020;**55**:170–6.

221. Yordan JMP, Blue NR, Beddow ME, Savabi M, Fritts C, Izquierdo LA, *et al*. Does timing of ultrasound improve the predictive value of a small for gestational age infant? Presented at 38th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 29 Jan–3 Feb 2018; Dallas, TX. *Am J Obstet Gynecol* 2018;**218**:S310.

222. Yoshida S, Unno N, Kagawa H, Shinozuka N, Kozuma S, Taketani Y. Prenatal detection of a high-risk group for intrauterine growth restriction based on sonographic fetal biometry. *Int J Gynaecol Obstet* 2000;**68**:225–32.

223. Agrawal C, Agrawal KK, Gandhi S. Assessment of fetal growth using the ratio of the transverse cerebellar diameter to abdominal circumference. *Int J Gynaecol Obstet* 2016;**135**:33–7.

224. Badr DA, Cannie MM, Kadji C, Kang X, Carlin A, Jani JC. The impact of different growth charts on birthweight prediction: obstetric ultrasound versus magnetic resonance imaging. *Am J Obstet Gynecol* 2023;**5**:101123.

225. Baird SM, Davies-Tuck M, Coombs P, Knight M, Wallace EM. Detection of the growth-restricted fetus: which centile charts? *Sonography* 2016;**3**:81–6.

226. Baird SM, Davies-Tuck M, Coombs P, Knight M, Wallace EM. Detection of the growth restricted fetus: which centile charts. Presented at 20th Annual Meeting of the Perinatal Society of Australia and New Zealand (PSANZ 2016): 22–25 May 2016; Townsville, QLD. *J Paediatr Child Health* 2016;**52**:108.

227. Bais JM, Eskes M, Pel M, Bonsel GJ, Bleker OP. Effectiveness of detection of intrauterine growth retardation by abdominal palpation as screening test in a low risk population: an observational study. *Eur J Obstet Gynecol Reprod Biol* 2004;**116**:164–9.

228. Barreto JA, Alencar Júnior CA. Diagnosis of fetal growth restriction by transverse cerebellar diameter/abdominal circumference ratio. *Rev Bras Ginecol Obstet* 2004;**26**:535–41.

229. Bastek JA, Pare E, Wang E, Elovitz MA, Srinivas SK. Limitations of ultrasound in diagnosing intrauterine growth restriction in severe preeclampsia. *J Matern Fetal Neonatal Med* 2009;**22**:1039–44.

230. 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.

231. Bergman M, Reichman O, Farkash R, Bin-Nun A, Samueloff A, Sapir AZ, Sela HY. Sonographic growth curves versus neonatal birthweight growth curves for the identification of fetal growth restriction. *J Matern Fetal Neonatal Med* 2022;**35**:4558–65.

232. Blue NR, Savabi M, Beddow ME, Katukuri VR, Fritts CM, Izquierdo LA, Chao CR. The Hadlock method is superior to newer methods for the prediction of the birth weight percentile. *J Ultrasound Med* 2019;**38**:587–96.

233. Blue NR, Beddow ME, Savabi M, Katukuri VR, Mozurkewich EL, Chao CR. A comparison of methods for the diagnosis of fetal growth restriction between the Royal College of Obstetricians and Gynaecologists and the American College of Obstetricians and Gynecologists. *Obstet Gynecol* 2018; **131**:835–41.

234. Blue NR, Beddow ME, Savabi M, Katukuri VR, Fritts CM, Izquierdo LA, Chao CR. Should we care about fetal growth percentiles at 18–22 weeks? Presented at 38th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 29 Jan–3 Feb 2018; Dallas, TX. *Am J Obstet Gynecol* 2018;**218**:S307–8.

235. Bonnevier A, Marsal K, Kallen K. Detection and clinical outcome of small-for-gestational-age fetuses in the third trimester – a comparison between routine ultrasound examination and examination on indication. *Acta Obstet Gynecol Scand* 2022;**101**:102–10.

236. Carbone JF, Tuuli MG, Bradshaw R, Liebsch J, Odibo AO. Efficiency of first-trimester growth restriction and low pregnancy-associated plasma protein-A in predicting small for gestational age at delivery. *Prenat Diagn* 2012;**32**:724–9.

237. Carbone J, Tuuli M, Bradshaw R, Liebsch J, Odibo A. Are first trimester growth restriction (FGR) and low pregnancy-associated plasma protein A (PAPP-A) efficient for predicting small for gestational (SGA) at delivery? Presented at 31st Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; San Francisco, CA United States. *Am J Obstet Gynecol* 2011;**204**:S141–2.

238. Chandra P, Aziz N. Impact of growth assessment protocol on SGA detection rates and perinatal outcomes. Presented at Royal College of Obstretriscians and Gynaecologists World Congress (RCOG 2022); 13–15 June 2022; London, UK. *BJOG* 2022;**129**:80–1.

239. Chang CH, Tsai PY, Yu CH, Ko HC, Chang FM. Soft tissue volume of upper arm in predicting small-for-gestational-age fetuses using three-dimensional ultrasound. *J Clin Ultrasound* 2011;**39**:21–6.

240. Chang CH, Tsai PY, Yu CH, Ko HC, Chang FM. Predicting fetal growth restriction with renal volume using 3-D ultrasound: efficacy evaluation. *Ultrasound Med Biol* 2008;**34**:533–7.

241. Chang CH, Tsai PY, Yu CH, Ko HC, Chang FM. Prenatal detection of fetal growth restriction by fetal femur volume: efficacy assessment using three-dimensional ultrasound. *Ultrasound Med Biol* 2007;**33**:335–41.

242. Chang CH, Yu CH, Ko HC, Chen CL, Chang FM. Predicting fetal growth restriction with liver volume by three-dimensional ultrasound: efficacy evaluation. *Ultrasound Med Biol* 2006;**32**:13–7.

243. Chang CH, Yu CH, Ko HC, Chen CL, Chang FM. Predicting fetal growth restriction by humerus volume: a three-dimensional ultrasound study. *Ultrasound Med Biol* 2006;**32**:791–5.

244. Chang CH, Yu CH, Ko HC, Chen CL, Chang FM. The efficacy assessment of thigh volume in predicting intrauterine fetal growth restriction by three-dimensional ultrasound. *Ultrasound Med Biol* 2005;**31**:883–7.

245. Chang CH, Yu CH, Ko HC, Chen CL, Chang FM. Fetal upper arm volume in predicting intrauterine growth restriction: a three-dimensional ultrasound study. *Ultrasound Med Biol* 2005;**31**:1435–9.

246. 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.

247. Chauhan SP, Shields D, Parker D, Sanderson M, Scardo JA, Magann EF. Detecting fetal growth restriction or discordant growth in twin gestations stratified by placental chorionicity. *J Reprod Med* 2004;**49**:279–84.

248. Chauhan SP, Magann EF, Velthius S, Nunn SL, Reynolds D, Scardo JA, *et al*. Detection of fetal growth restriction in patients with chronic hypertension: is it feasible? *J Matern Fetal Neonatal Med* 2003;**14**:324–8.

249. Cowan FJ, McKinlay CJD, Taylor RS, Wilson J, McAra-Couper J, Garrett N, *et al*. Detection of small for gestational age babies and perinatal outcomes following implementation of the Growth Assessment Protocol at a New Zealand tertiary facility: an observational intervention study. *Aust N Z J Obstet Gynaecol* 2021;**61**:339–46.

250. De Jong CL, Francis A, Van Geijn HP, Gardosi J. Customized fetal weight limits for antenatal detection of fetal growth restriction. *Ultrasound Obstet Gynecol* 2000;**15**:36–40.

251. De Reu PA, 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.

252. de Silva RS, Perera H. Serial measurements of fetal head circumference and abdominal circumference to predict fetal growth restriction in a Sri Lankan study population. *J SAFOG* 2021;**13**:202–6.

253. De Silva R, Perera H. Serial ultrasound measurements of fetal head circumference and abdominal circumference to predict fetal growth restriction in a Sri Lankan study population. Presented at Royal College of Obstetricians and Gynaecologists World Congress (RCOG 2017); 20–22 March 2017; Cape Town, SA. *BJOG* 2017;**124**:12.

254. Dude AM, Davis B, Delaney K, Yee LM. Identifying fetal growth disorders using ultrasound in obese nulliparous women. *J Matern Fetal Neonatal Med* 2021;**34**:1768–73.

255. Dude AM, Yee LM. Identifying fetal growth disorders using ultrasonography in women with diabetes. *J Ultrasound Med* 2018;**37**:1103–8.

256. Haragan AF, Hulsey TC, Hawk AF, Newman RB, Chang EY. Diagnostic accuracy of fundal height and handheld ultrasound-measured abdominal circumference to screen for fetal growth abnormalities. *Am J Obstet Gynecol* 2015;**212**:820.e1–8.

257. Harper LM, Roehl KA, Tuuli MG, Odibo AO, Cahill AG. Sonographic accuracy of estimated fetal weight in twins. *J Ultrasound Med* 2013;**32**:625–30.

258. Henrichs J, Verfaille V, Jellema P, Viester L, Pajkrt E, Wilschut J, *et al*.; IRIS Sudy Group. Effectiveness of routine third trimester ultrasonography to reduce adverse perinatal outcomes in low risk pregnancy (the IRIS study): nationwide, pragmatic, multicentre, stepped wedge cluster randomised trial. *Br Med J* 2019;**367**:l5517.

259. van Roekel M, Henrichs J, Franx A, Verhoeven CJ, de Jonge A. Implication of third-trimester screening accuracy for small-for-gestational age and additive value of third-trimester growth-trajectory indicators in

predicting severe adverse perinatal outcome in low-risk population: pragmatic secondary analysis of IRIS study. *Ultrasound Obstet Gynecol* 2023;**62**:209–18.

260. Hugh O, Williams M, Turner S, Gardosi J. Reduction of stillbirths in England from 2008 to 2017 according to uptake of the Growth Assessment Protocol: 10-year population-based cohort study. *Ultrasound Obstet Gynecol* 2021;**57**:401–8.

261. Humphries J, Reynolds D, Bell-Scarbrough L, Lynn N, Scardo JA, Chauhan SP. Sonographic estimate of birth weight: relative accuracy of sonographers versus maternal–fetal medicine specialists. *J Matern Fetal Neonatal Med* 2002;**11**:108–12.

262. Jayawardena L, Sheehan P. Introduction of a customised growth chart protocol increased detection of small for gestational age pregnancies in a tertiary Melbourne hospital. *Aust N Z J Obstet Gynaecol* 2019;**59**:493–500.

263. Leung TY, Sahota DS, Chan LW, Law LW, Fung TY, Leung TN, Lau TK. Prediction of birth weight by fetal crown-rump length and maternal serum levels of pregnancy-associated plasma protein-A in the first trimester. *Ultrasound Obstet Gynecol* 2008;**31**:10–4.

264. Li K, Yan G, Zheng W, Sun J, Zou Y. Measurement of the brain volume/liver volume ratio by three-dimensional MRI in appropriate-for-gestational age fetuses and those with fetal growth restriction. *J Magn Reson Imaging* 2021;**54**:1796–801.

265. Lindstrom L, Cnattingius S, Axelsson O, Granfors M. Accuracy and precision of sonographic fetal weight estimation in Sweden. *Acta Obstet Gynecol Scand* 2023;**102**:699–707.

266. Marchand C, Koppe J, Koster HA, Oelmeier K, Schmitz R, Steinhard J, *et al*. Fetal growth restriction: comparison of biometric parameters. *J Pers Med* 2022;**12**:1125.

267. Martin-Palumbo G, Atanasova VB, Rego Tejeda MT, Antolin Alvarado E, Bartha JL. Third trimester ultrasound estimated fetal weight for increasing prenatal prediction of small-for-gestational age newborns in low-risk pregnant women. *J Matern Fetal Neonatal Med* 2022;**35**:6721–6.

268. Mathewlynn S, Impey L, Ioannou C. Detection of small- and large-for-gestational age using different combinations of prenatal and postnatal charts. *Ultrasound Obstet Gynecol* 2022;**60**:373–80.

269. Michaeli J, Michaeli O, Rozitzky A, Grisaru-Granovsky S, Feldman N, Srebnik N. Application of prospect theory in obstetrics by evaluating mode of delivery and outcomes in neonates born small or appropriate for gestational age. *JAMA Network Open* 2022;**5**:e222177.

270. Monier I, Ego A, Benachi A, Hocquette A, Blondel B, Goffinet F, Zeitlin J. Unisex vs sex-specific estimated fetal weight charts for fetal growth monitoring: a population-based study. *Am J Obstet Gynecol* 2022;**4**:100527.

271. Monier I, Ego A, Benachi A, Hocquette A, Blondel B, Goffinet F, Zeitlin J. Comparison of the performance of estimated fetal weight charts for the detection of small- and large-for-gestational age newborns with adverse outcomes: a French population-based study. *BJOG* 2022;**129**:938–48.

272. Nwabuobi C, Odibo L, Camisasca-Lopina H, Leavitt K, Tuuli M, Odibo AO. Comparing INTERGROWTH-21st Century and Hadlock growth standards to predict small for gestational age and short-term neonatal outcomes. *J Matern Fetal Neonatal Med* 2020;**33**:1906–12.

273. Pressman K, Odibo L, Duncan JR, Odibo AO. Impact of using abdominal circumference independently in the diagnosis of fetal growth restriction. *J Ultrasound Med* 2022;**41**:157–62.

274. Odibo AO, Nwabuobi C, Odibo L, Leavitt K, Obican S, Tuuli MG. Customized fetal growth standard compared with the INTERGROWTH-21st century standard at predicting small-for-gestational-age neonates. *Acta Obstet Gynecol Scand* 2018;**97**:1381–7.

275. Roeckner JT, Pressman K, Odibo L, Duncan JR, Odibo AO. Outcome-based comparison of SMFM and ISUOG definitions of fetal growth restriction. *Ultrasound Obstet Gynecol* 2021;**57**:925–30.

276. Nwabuobi C, Camisasca-Lopina H, Leavitt K, Odibo L, Tuuli M, Odibo A. INTERGROWTH-21st and Hadlock growth standards to predict neonatal small for gestational age and short-term neonatal outcomes. Presented

at Conference: 38th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting. Dallas, TX United States. *Am J Obstet Gynecol* 2018;**218**:S310.

277. Nymark Hansen D, Sand Odgaard H, Uldbjerg N, Sinding M, Sorensen A. Screening for small-for-gestational-age fetuses. *Acta Obstet Gynecol Scand* 2019;**99**:503–9.
278. Poljak B, Agarwal U, Jackson R, Alfirevic Z, Sharp A. Diagnostic accuracy of individual antenatal tools for prediction of small-for-gestational age at birth. *Ultrasound Obstet Gynecol* 2017;**49**:493–9.
279. Price CR, Roeckner J, Odibo L, Odibo A. Comparing fetal biometric growth velocity versus estimated fetal weight for prediction of neonatal small for gestational age. *J Matern Fetal Neonatal Med* 2022;**35**:3931–6.
280. Price CR, Odibo L, Odibo A. 265: Comparing fetal biometric growth trajectory versus estimated fetal weight for prediction of neonatal SGA. Presented at Conference: SMFM 40th Annual Meeting: The Pregnancy Meeting; 3–8 Feb 2020; Grapevine, US. *Am J Obstet Gynecol* 2020;**222**:S181–2.
281. Rad S, Beauchamp S, Morales C, Mirocha J, Esakoff TF. Defining fetal growth restriction: abdominal circumference as an alternative criterion. *J Matern Fetal Neonatal Med* 2018;**31**:3089–94.
282. Rad S, Beauchamp S, Morales C, Mirocha J, Esakoff T. Defining fetal growth restriction: Should fetal abdominal circumference be an alternate criterion? Presented at 35th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 2–7 Feb 2015; San Diego, CA. *Am J Obstet Gynecol* 2015;**212**:S358.
283. Ridha N, Bergin CJ, Kelly J, Tarr GP, Anderson N, Sadler L. Accuracy of ultrasound in the estimation of customised birth weight in a public hospital service. *J Med Imaging Radiat Oncol* 2022;**66**:1044–51.
284. Salomon LJ, Bernard JP, Duyme M, Ville Y. Predicting late-onset growth abnormalities using growth velocity between trimesters. *J Matern Fetal Neonatal Med* 2005;**17**:193–7.
285. Sklar C, Yaskina M, Ross S, Naud K. Accuracy of prenatal ultrasound in detecting growth abnormalities in triplets: a retrospective cohort study. *Twin Res Hum Genet* 2017;**20**:84–9.
286. Skrastad RB, Eik-Nes SH, Sviggum O, Johansen OJ, Salvesen KA, Romundstad PR, Blaas HK. A randomized controlled trial of third-trimester routine ultrasound in a non-selected population. *Acta Obstet Gynecol Scand* 2013;**92**:1353–60.
287. Skrastad RB, Eik-Nes SH, Sviggum O, Johansen OJ, Salvesen KA, Romundstad PR, Blaas HK. A randomized controlled trial of third-trimester routine ultrasound in a nonselected population. *Obstet Gynecol Surv* 2014;**69**:185–7.
288. 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.
289. Sovio U, Smith G, Dacey A. Level 1 evidence for the diagnostic effectiveness of routine sonography as a screening test for small for gestational age (SGA) infants. Presented at 34th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting; 3–8 Feb 2014; New Orleans, LA. *Am J Obstet Gynecol* 2014;**210**:S408.
290. Temming LA, Dicke JM, Stout MJ, Rampersad RM, Macones GA, Tuuli MG, Cahill AG. Early second-trimester fetal growth restriction and adverse perinatal outcomes. *Obstet Gynecol* 2017;**130**:865–9.
291. Turitz AL, Quant H, Schwartz N, Elovitz M, Bastek JA. Isolated abdominal circumference < 5% or estimated fetal weight 10 to 19% as predictors of small for gestational age infants. *Am J Perinatol* 2014;**31**:469–76.
292. Tuuli MG, Cahill A, Stamilio D, Macones G, Odibo AO. Comparative efficiency of measures of early fetal growth restriction for predicting adverse perinatal outcomes. *Obstet Gynecol* 2011;**117**:1331–40.
293. Tuuli M, Cahill A, Stamilio D, Macones G, Odibo A. Early fetal growth restriction (FGR) for predicting adverse perinatal outcomes. Presented at 31st Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting 2011; 7–12 Feb 2011; San Francisco, CA. *Am J Obstet Gynecol* 2011;**204**:S149.

294. Vieira MC, Relph S, Muruet-Gutierrez W, Elstad M, Coker B, Moitt N, *et al*.; DESiGN Collaborative Group. Evaluation of the Growth Assessment Protocol (GAP) for antenatal detection of small for gestational age: the DESiGN cluster randomised trial. *PLOS Med* 2022;**19**:e1004004.

295. Wan A, Zampogna C, Reddy M, Robinson A, Hodges R, Rolnik DL. Performance of a risk factor-based approach in the detection of small for gestational age neonates: a cohort study. *Aust N Z J Obstet Gynaecol* 2022;**63**:321–7.

296. Whitham MD, Reynolds DM, Urban AR, Ennen CS, Dudley DJ. Comparative diagnostic performance of estimated fetal weight and isolated abdominal circumference for the detection of fetal growth restriction. *J Ultrasound Med* 2023;**42**:477–85.

297. Lubchenco LO, Hansman C, Dressler M, Boyd E. Intrauterine growth as estimated from liveborn birth-weight data at 24 to 42 weeks of gestation. *Pediatrics* 1963;**32**:793–800.

298. Nicolaides KH, Wright D, Syngelaki A, Wright A, Akolekar R. Fetal Medicine Foundation fetal and neonatal population weight charts. *Ultrasound Obstet Gynecol* 2018;**52**:44–51.

299. Villar J, Cheikh Ismail L, Victora CG, Ohuma EO, Bertino E, Altman DG, *et al*.; International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet* 2014;**384**:857–68.

300. Gardosi J. *GROW Documentation [Internet]*. Birmingham: Gestation Network, Perinatal Institute; 2020. URL: www.gestation.net/GROW_documentation.pdf (accessed 9 October 2023).

301. Dobbins TA, Sullivan EA, Roberts CL, Simpson JM. Australian national birthweight percentiles by sex and gestational age, 1998-2007. *Med J Aust* 2012;**197**:291–4.

302. Bleker OP, Buimer M, van der Post JAM, van der Veen F. Ted (G.J.) Kloosterman: On intrauterine growth. The significance of prenatal care. Studies on birth weight, placental weight and placental index. *Placenta* 2006;**27**:1052–4.

303. Alexander GR, Himes JH, Kaufman RB, Mor J, Kogan M. A United States national reference for fetal growth. *Obstet Gynecol* 1996;**87**:163–8.

304. Leiberman JR, Fraser D, Weitzman S, Glezerman M. Birthweight curves in southern Israel populations. *Isr J Med Sci* 1993;**29**:198–203.

305. Dollberg S, Haklai Z, Mimouni FB, Gorfein I, Gordon ES. Birth weight standards in the live-born population in Israel. *Isr Med Assoc J* 2005;**7**:311–4.

306. Sapir AZ, Khayyat I, Rabinowitz R, Samueloff A, Drukker L, Sela HY. New Israeli sonographic estimated fetal weight growth curves as compared to current birth weight growth curves: on what should diagnosis of intrauterine growth disorders be based? *Isr Med Assoc J* 2017;**19**:604–9.

307. Marsál K, Persson PH, Larsen T, Lilja H, Selbing A, Sultan B. Intrauterine growth curves based on ultrasonically estimated foetal weights. *Acta Paediatr* 1996;**85**:843–8.

308. Chang CH, Yu CH, Chang FM, Ko HC, Chen HY. Assessment of normal fetal upper arm volume by three-dimensional ultrasound. *Ultrasound Med Biol* 2002;**28**:859–63.

309. Hsieh TT, Hsu JJ, Chen CJ, Chiu TH, Liou JD, Hsieh CC, *et al*. Analysis of birth weight and gestational age in Taiwan. *J Formos Med Assoc* 1991;**90**:382–7.

310. Chang CH, Yu CH, Chang FM, Ko HC, Chen HY. Three-dimensional ultrasound in the assessment of normal fetal thigh volume. *Ultrasound Med Biol* 2003;**29**:361–6.

311. Chang CH, Yu CH, Ko HC, Chang FM, Chen HY. Prenatal assessment of normal fetal humerus volume by three-dimensional ultrasound. *Ultrasound Med Biol* 2003;**29**:1675–80.

312. Chang CH, Yu CH, Chang FM, Ko HC, Chen HY. The assessment of normal fetal liver volume by three-dimensional ultrasound. *Ultrasound Med Biol* 2003;**29**:1123–9.

313. Yu C, Chang C, Chang F, Ko H, Chen H. Fetal renal volume in normal gestation: a three-dimensional ultrasound study. *Ultrasound Med Biol* 2000;**26**:1253–6.

314. Ananth CV, Vintzileos AM, Shen-Schwarz S, Smulian JC, Lai YL. Standards of birth weight in twin gestations stratified by placental chorionicity. *Obstet Gynecol* 1998;**91**:917–24.

315. Visser GH, Eilers PH, Elferink-Stinkens PM, Merkus HM, Wit JM. New Dutch reference curves for birthweight by gestational age. *Early Hum Dev* 2009;**85**:737–44.

316. Loughna P, Chitty L, Evans T, Chudleigh T. Fetal size and dating: charts recommended for clinical obstetric practice. *Ultrasound* 2009;**17**:160–6.

317. Landmann E, Reiss I, Misselwitz B, Gortner L. Ponderal index for discrimination between symmetric and asymmetric growth restriction: percentiles for neonates from 30 weeks to 43 weeks of gestation. *J Matern Fetal Neonatal Med* 2006;**19**:157–60.

318. Brenner WE, Edelman DA, Hendricks CH. A standard of fetal growth for the United States of America. *Am J Obstet Gynecol* 1976;**126**:555–64.

319. Oken E, Kleinman KP, Rich-Edwards J, Gillman MW. A nearly continuous measure of birth weight for gestational age using a United States national reference. *BMC Pediatr* 2003;**3**:6.

320. Olsen IE, Groveman SA, Lawson ML, Clark RH, Zemel BS. New intrauterine growth curves based on United States data. *Pediatrics* 2010;**125**:e214–24.

321. Verburg BO, Steegers EA, De Ridder M, Snijders RJ, Smith E, Hofman A, *et al*. New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population-based cohort study. *Ultrasound Obstet Gynecol* 2008;**31**:388–96.

322. Fok TF, So HK, Wong E, Ng PC, Chang A, Lau J, *et al*.; Hong Kong Neonatal Measurements Working Group. Updated gestational age specific birth weight, crown-heel length, and head circumference of Chinese newborns. *Arch Dis Child Fetal Neonatal Ed* 2003;**88**:F229–36.

323. Dai L, Deng C, Li Y, Zhu J, Mu Y, Deng Y, *et al*. Birth weight reference percentiles for Chinese. *PLOS ONE* 2014;**9**:e104779.

324. Lindstrom L, Ageheim M, Axelsson O, Hussain-Alkhateeb L, Skalkidou A, Wikstrom AK, Bergman E. Swedish intrauterine growth reference ranges for estimated fetal weight. *Sci Rep* 2021;**11**:12464.

325. Persson PH, Weldner BM. Intra-uterine weight curves obtained by ultrasound. *Acta Obstet Gynecol Scand* 1986;**65**:169–73.

326. Voigt M, Fusch C, Olbertz D, Hartmann K, Rochow N, Renken C, *et al*. Analysis of the newborn collective of the Federal Republic of Germany. *Geburtshilfe Frauenheilkd* 2006;**66**:956–70.

327. Yudkin PL, Aboualfa M, Eyre JA, Redman CW, Wilkinson AR. New birthweight and head circumference centiles for gestational ages 24 to 42 weeks. *Early Hum Dev* 1987;**15**:45–52.

328. Freeman JV, Cole TJ, Chinn S, Jones PR, White EM, Preece MA. Cross sectional stature and weight reference curves for the UK, 1990. *Arch Dis Child* 1995;**73**:17–24.

329. WHO, The Scientific Advisory Committee on Nutrition (SACN), Royal College of Paediatrics and Child Health (RCPCH). *SACN Application of WHO Growth Standards in the UK*. London: The Stationery Office; 2007.

330. Mamelle N, Munoz F, Grandjean H. Fetal growth from the AUDIPOG study. I. Establishment of reference curves. *J Gynecol Obstet Biol Reprod (Paris)* 1996;**25**:61–70.

331. Buck Louis GM, Grewal J, Albert PS, Sciscione A, Wing DA, Grobman WA, *et al*. Racial/ethnic standards for fetal growth: the NICHD Fetal Growth Studies. *Am J Obstet Gynecol* 2015;**213**:449.e1–41.

332. Massoud M, Duyme M, Fontanges M, Combourieu D; French College of Fetal Sonography (CFEF). Chart for estimation of fetal weight 2014 by the French College of Fetal Sonography (CFEF). *J Gynecol Obstet Biol Reprod (Paris)* 2016;**45**:80–5.

333. Ego A, Prunet C, Lebreton E, Blondel B, Kaminski M, Goffinet F, Zeitlin J. Customized and non-customized French intrauterine growth curves. I – methodology. *J Gynecol Obstet Biol Reprod (Paris)* 2016;**45**:155–64.

334. Gardosi J, Francis A. A customized standard to assess fetal growth in a US population. *Am J Obstet Gynecol* 2009;**201**:25.e1–7.

335. Gardosi J, Mongelli M, Wilcox M, Chang A. An adjustable fetal weight standard. *Ultrasound Obstet Gynecol* 1995;**6**:168–74.

336. Mikolajczyk RT, Zhang J, Betran AP, Souza JP, Mori R, Gulmezoglu AM, Merialdi M. A global reference for fetal-weight and birthweight percentiles. *Lancet* 2011;**377**:1855–61.

337. Cole TJ, Freeman JV, Preece MAB. 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med* 1998;**17**:407–29.

338. Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 3. Abdominal measurements. *Br J Obstet Gynaecol* 1994;**101**:125–31.

339. Papageorghiou AT, Ohuma EO, Altman DG, Todros T, Cheikh Ismail L, Lambert A, *et al.*; International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. *Lancet* 2014;**384**:869–79.

340. Leroy B, Lefort F. The weight and size of newborn infants at birth. *Rev Fr Gynecol Obstet* 1971;**66**:391–6.

341. Kramer MS, Platt RW, Wen SW, Joseph KS, Allen A, Abrahamowicz M, *et al.*; Fetal/Infant Health Study Group of the Canadian Perinatal Surveillance System. A new and improved population-based Canadian reference for birth weight for gestational age. *Pediatrics* 2001;**108**:E35.

342. Eik-Nes SH, Gröttum P, Persson PH, Marsál K. Prediction of fetal growth deviation by ultrasonic biometry. I. Methodology. *Acta Obstet Gynecol Scand* 1982;**61**:53–8.

343. Eik-Nes SH, Persson PH, Gröttum P, Marsál K. Prediction of fetal growth deviation by ultrasonic biometry. II. Clinical application. *Acta Obstet Gynecol Scand* 1983;**62**:117–23.

344. Fenton TR, Kim JH. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC Pediatr* 2013;**13**:59.

345. Duryea EL, Hawkins JS, McIntire DD, Casey BM, Leveno KJ. A revised birth weight reference for the United States. *Obstet Gynecol* 2014;**124**:16–22.

346. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, *et al.* RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;**366**:l4898.

347. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al.*; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529–36.

348. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Med Res Methodol* 2007;**7**:3.

349. Zamora J, Abraira V, Nuriel A, Khan KS, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006;**6**:31. https://doi.org/10.1186/1471-2288-6-31

350. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. *Doing Meta-Analysis with R: A Hands-On Guide [Internet]*. London: Chapman & Hall/CRC Press; 2021. URL: https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/ (accessed 12 October 2023).

351. Paule RC, Mandel J. Consensus values and weighting factors. *J Res Natl Bur Stand* 1977;**87**:377–85.

352. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;**22**:719–48.

353. R Core Team. *The R Project for Statistical Computing [Internet]*. Vienna, Austria: R Foundation for Statistical Computing; 2022. URL: www.r-project.org/ (accessed 3 November 2023).

354. Posit Team. *RStudio: Integrated Development Environment for R [Internet]*. Boston, MA: Posit Team; 2023. URL: https://posit.co/ (accessed 3 November 2023).

355. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health* 2019;**22**:153–60.

356. Harrer M, Cuijpers P, Furukawa T, Ebert DD. *dmetar: Companion R Package for the Guide 'Doing Meta-Analysis in R'. R Package Version 0.1.0 [Internet]*. Protectlab; 2019. URL: http://dmetar.protectlab.org/ (accessed 3 November 2023).

357. Wickham H, Chang W, Lionel H, Pederson T, Takahashi K, Wilke C, *et al*. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer; 2016. URL: https://ggplot2.tidyverse.org/ (accessed 8 July 2025).

358. Odibo A, Nwabuobi C, Odibo L, Tuuli MG. Is the intergrowth-21st growth standard superior to the customized approach for identifying pregnancies at risk for neonatal small for gestational age (SGA) and adverse outcomes? Presented at 38th Annual Meeting of the Society for Maternal–Fetal Medicine: The Pregnancy Meeting 29 Jan–3 Feb 2018; Dallas, TX. *Am J Obstet Gynecol* 2018;**218**:S272–3.

359. Pressman K, Odibo L, Duncan JR, Odibo A. 725 Impact of using abdominal circumference independently in the diagnosis of fetal growth restriction. Presented at SMFM 41st Annual Meeting: The Pregnancy Meeting; 25–30 Jan 2021; Virtual, Online. *Am J Obstet Gynecol* 2021;**224**:S454–5.

360. Roeckner JT, Pressman K, Duncan JR, Odibo L, Odibo A. 179 An outcome-based comparison of ISUOG vs SMFM guidelines for fetal growth restriction. Presented at SMFM 41st Annual Meeting: The Pregnancy Meeting; 25–30 Jan 2021; Virtual, Online. *Am J Obstet Gynecol* 2021;**224**:S121.

361. Wan A, Zampogna C, Reddy M, Robinson A, Hodges R, Rolnik DL. Performance of a risk factor-based approach in the detection of small for gestational age neonates. Presented at 24th Annual Congress of the Perinatal Society of Australia and New Zealand (PSANZ); 5–8 April 2020; Sydney, NSW. *J Paediatr Child Health* 2020;**56**:50.

362. Atallah A, Butin M, Moret S, Claris O, Gaucherand P, Doret-Dion M. Fetal growth restriction: underdiagnosed condition with non-optimal screening. *J Matern Fetal Neonatal Med* 2022;**35**:8237–44.

363. Lesser HT, D'Adamo C, O'Reilly G, Arrabal P, Ehsanipoor RM. Diagnosis of mid-second trimester fetal growth restriction and associated outcomes. *J Matern Fetal Neonatal Med* 2022;**35**:10168–72.

364. Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. *Clin Chem* 2005;**51**:1335–41.

365. Committee on Obstetric Practice. *The Apgar Score: Committee Opinion No.664 [Internet]*. Washington, DC: American College of Obstetricians and Gynecologists; 2015. URL: www.acog.org/clinical/clinical-guidance/committee-opinion/articles/2015/10/the-apgar-score (accessed 12 October 2023).

366. Baker PN, Johnson IR, Gowland PA, Hykin J, Harvey PR, Freeman A, *et al*. Fetal weight estimation by echo-planar magnetic resonance imaging. *Lancet* 1994;**343**:644–5.

367. Snijders RJ, Nicolaides KH. Fetal biometry at 14–40 weeks' gestation. *Ultrasound Obstet Gynecol* 1994;**4**:34–48.

368. Chang FM, Hsu KF, Ko HC, Yao BL, Chang CH, Yu CH, Chen HY. Three-dimensional ultrasound assessment of fetal liver volume in normal pregnancy: a comparison of reproducibility with two-dimensional ultrasound and a search for a volume constant. *Ultrasound Med Biol* 1997;**23**:381–9.

369. Chang FM, Liang RI, Ko HC, Yao BL, Chang CH, Yu CH. Three-dimensional ultrasound-assessed fetal thigh volumetry in predicting birth weight. *Obstet Gynecol* 1997;**90**:331–9.

370. Altman DG, Chitty LS. Charts of fetal size: 1. Methodology. *Br J Obstet Gynaecol* 1994;**101**:29–34.

371. Papageorgiou A. Intrauterine growth restriction: short and long term implications for the newborn. Presented at 11th World Congress of Perinatal Medicine 2013; Moscow, Russian Federation. *J Perinat Med* 2013;**41**.

372. Alameddine S, Capannolo G, Rizzo G, Khalil A, Di Girolamo R, Iacovella C, *et al*. A systematic review and critical evaluation of quality of clinical practice guidelines on fetal growth restriction. *J Perinat Med* 2023;**51**:970–80.

373. Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *J Clin Epidemiol* 2011;**64**:602–7.

374. 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.

375. Al-Hafez L, Chauhan SP, Riegel M, Balogun OA, Hammad IA, Berghella V. Routine third-trimester ultrasound in low-risk pregnancies and perinatal death: a systematic review and meta-analysis. *Am J Obstet Gynecol* 2020;**2**:100242.

Appendix 1 Literature search strategies

Appendix 1: Search strategies

Original searches

Database	Dates covered	Hits
EMBASE	1974–20 March 2023	13,344
MEDLINE + PreMedline	1946–20 March 2023	8591
PubMed	Up to 22 March 2023	529
CDSR	Up to March 2023/Iss3	25
CENTRAL	Up to February 2023/Iss2	800
DARE (CRD)	Up to 31 March 2015	23
KSR Evidence	Up to 21 March 2023	139
Epistemonikos	Up to 22 March 2023	138
PROSPERO	Up to 21 March 2023	11
INPLASY	Up to 22 March 2023	0
LILACS	Up to 22 March 2023	123
ClinTrials.gov	Up to 20 March 2023	467
EUCTR	Up to 20 March 2023	101
WHO ICTRP	Up to 20 March 2023	614
ScanMedicine	Up to 20 March 2023	1358
Northern Light	2012–3/Wk10	443
TRIP	Up to 16 March 2023	564
GIN	Up to 16 March 2023	0
NICE	Up to March 2023/Iss3	38
NIHR HTA	Up to 16 March 2023	12
ECRI	Up to 31 March 2015	11
INAHTA	Up to 20 March 2023	23
HTA	Up to 22 March 2023	17
Total		**27,371**

EMBASE (Ovid): 1974–20 March 2023

Searched 21 March 2023

1 exp intrauterine growth retardation/(54,264)
2 ((fetal or foetal or fetus or foetus or embryo or "in utero" or intra?uterine or intrauterine or prenatal) adj3 (growth or grow or grown or growing or size)).ti,ab,ot,hw. (76,170)
3 (small adj3 (age or date)).ti,ab,ot,hw. (28,730)
4 (SGA or FGR or IUGR).ti,ab. (30,428)

5 fetus growth/(17,200)
6 or/1-5 (108,827)
7 (measur$ adj3 tape$).ti,ab,ot. (2286)
8 uterus fundus height/(182)
9 ((fundus or fundal) adj3 height).ti,ab,ot. (572)
10 (SFH or "McDonald's rule").ti,ab. (489)
11 ultrasound/ or fetus echography/ or (phonophoresis or sonication or sonification or sonograph$ or ultra?sound or ultrashell or ultrasonic$ or ultrasonograph$ or echograph$).ti,ab,ot. (762,513)
12 abdominal circumference/ or ((abdomen or abdominal) adj3 circumference).ti,ab,ot. (7941)
13 fetus weight/ or ((fetal or fetus or foetal or foetus or embryo) adj3 weight).ti,ab,ot. (16,060)
14 7 or 8 or 9 or 10 or 11 or 12 or 13 (781,724)
15 6 and 14 (19,282)
16 limit 15 to yr="2000 -Current" (15,654)
17 animal/(1,601,216)
18 animal experiment/(3,030,188)
19 (rat or rats or mouse or mice or murine or rodent or rodents or hamster or hamsters or pig or pigs or porcine or rabbit or rabbits or animal or animals or dogs or dog or cats or cow or bovine or sheep or ovine or monkey or monkeys).ti,ab,ot,hw. (7,568,080)
20 or/17-19 (7,568,080)
21 exp human/(25,229,095)
22 human experiment/(641,172)
23 or/21-22 (25,231,429)
24 20 not (20 and 23) (5,686,821)
25 16 not 24 (13,840)
26 (letter or editorial or note).pt. (2,994,297)
27 **25 not 26 (13,344)**

MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations and Daily (Ovid): 1946–20 March 2023

Searched 21 March 2023

1 Fetal Growth Retardation/(18,559)
2 ((fetal or foetal or fetus or foetus or embryo or "in utero" or intra?uterine or intrauterine or prenatal) adj3 (growth or grow or grown or growing or size)).ti,ab,ot,hw. (46,904)
3 (small adj3 (age or date)).ti,ab,ot,hw. (18,568)
4 Infant, Small for Gestational Age/(8534)
5 (SGA or FGR or IUGR).ti,ab. (18,464)
6 exp *Fetal Development/(14,821)
7 or/1-6 (75,852)
8 (measur$ adj3 tape$).ti,ab,ot. (1487)
9 ((fundus or fundal) adj3 height).ti,ab,ot,hw. (385)
10 (SFH or "McDonald's rule").ti,ab. (359)
11 exp Ultrasonography, Prenatal/ (37,021)
12 (phonophoresis or sonication or sonification or sonograph$ or ultra?sound or ultrashell or ultrasonic$ or ultrasonograph$ or echograph$).ti,ab,ot. (487,849)
13 ((abdomen or abdominal) adj3 circumference).ti,ab,ot. (3734)
14 ((fetal or fetus or foetal or foetus or embryo) adj3 weight).ti,ab,ot. (8414)
15 or/8-14 (511,337)
16 7 and 15 (13,938)
17 limit 16 to yr="2000 –Current" (9812)
18 (letter or editorial or note).pt. (1,852,157)

19 exp animals/ not (exp animals/ and humans/) (5,104,267)
20 **17 not (18 or 19) (8591)**

PubMed (NLM) (Internet): Up to 22 March 2023

Searched 22 March 23

(Top up search)

#10Search: #8 and #9529
#9Search: (pubstatusaheadofprint OR publisher[sb] OR pubmednotmedline[sb])5,146,247
#8Search: #4 and #79,176
#7Search: #5 or #61,431,135
#6Search: measure[Title/Abstract] OR measuring[Title/Abstract] OR "fundal height"[Title/Abstract] OR "fundus height"[Title/Abstract] OR SFH[Title/Abstract] OR "McDonald's rule"[Title/Abstract] OR phonophoresis[Title/Abstract] OR sonication[Title/Abstract] OR sonification[Title/Abstract] OR sonograph*[Title/Abstract] OR ultrasound[Title/Abstract] OR ultrashell[Title/Abstract] OR ultrasonic*[Title/Abstract] OR ultrasonograph*[Title/Abstract] OR echograph*[Title/Abstract] OR "abdominal circumference"[Title/Abstract]1,418,216
#5Search: "Ultrasonography, Prenatal"[Mesh] Sort by: Most Recent37,013
#4Search: #1 or #348,482
#3Search: ("Fetal Growth"[Title/Abstract] OR "foetal growth"[Title/Abstract] OR "small for Gestational Age"[Title/Abstract] OR SGA[Title/Abstract] OR FGR[Title/Abstract] OR IUGR[Title/Abstract])38,918
#1Search: ("Fetal Growth Retardation"[Mesh]) OR "Infant, Small for Gestational Age"[Mesh] Sort by: Most Recent24,979

Cochrane Database of Systematic Reviews (CDSR) (Wiley): up to March 2023/Iss3

Cochrane Central Register of Controlled Trials (CENTRAL) (Wiley): up to February 2023/Iss2

Search 21 March 2023

#1MeSH descriptor: [Fetal Growth Retardation] this term only527
#2((fetal or foetal or fetus or foetus or embryo or "in utero" or intra?uterine or intrauterine or prenatal) near/3 (growth or grow or grown or growing or size)):ti,ab,kw2423
#3MeSH descriptor: [Infant, Small for Gestational Age] explode all trees356
#4(small adj3 (age or date)):ti,ab,kw0
#5(SGA or FGR or IUGR):ti,ab1831
#6MeSH descriptor: [Fetal Development] explode all trees3386
#7#1 or #2 or #3 or #4 or #5 or #66761
#8(measur* near/3 tape*):ti,ab,kw1373
#9((fundus or fundal) near/3 height):ti,ab,kw92
#10(SFH or "McDonald's rule"):ti,ab39
#11MeSH descriptor: [Ultrasonography, Prenatal] explode all trees667
#12(phonophoresis or sonication or sonification or sonograph* or ultra?sound or ultrashell or ultrasonic* or ultrasonograph* or echograph*):ti,ab50,478
#13((abdomen or abdominal) near/3 circumference):ti,ab910
#14((fetal or fetus or foetal or foetus or embryo) near/3 weight):ti,ab609
#15#8 or #9 or #10 or #11 or #12 or #13 or #1453,019
#16#7 and #15 with Cochrane Library publication date Between Jan 2000 and Apr 2023903
#17nct*:au241117
#18#16 not #17 in Cochrane Reviews, Cochrane Protocols, Trials825

CDSR retrieved 25 records

CENTRAL retrieved 800 records

Database of Abstracts of Reviews of Effects (DARE)(CRD): up to 31 March 2015

Searched 20 March 2023

1 MeSH DESCRIPTOR Fetal Growth Retardation EXPLODE ALL TREES 25 Delete
2 (((fetal or foetal or fetus or foetus or embryo or "in utero" or intra?uterine or intrauterine or prenatal) adj3 (growth or grow or grown or growing or size)))88Delete
3 ((small adj3 (age or date)))90Delete
4 MeSH DESCRIPTOR Infant, Small for Gestational Age EXPLODE ALL TREES21Delete
5 ((SGA or FGR or IUGR))41Delete
6 MeSH DESCRIPTOR Fetal Development EXPLODE ALL TREES158Delete
7 #1 OR #2 OR #3 OR #4 OR #5 OR #6330Delete
8 **((SGA or FGR or IUGR)) IN DARE FROM 2000 TO 202323Delete**

KSR Evidence (https://ksrevidence.com/): up to 21 March 2023

Searched 21 March 2023

1 ((fetal or foetal or fetus or foetus or embryo or "in utero" or intra?uterine or intrauterine or prenatal) adj3 (growth or grow or grown or growing or size)) in All text497 results
2 (small adj3 (age or date)) in All text723 results
3 (SGA or FGR or IUGR) in All text461 results
4 #1 or #2 or #3 in All text1238 results
5 (measur* adj3 tape*) in All text21 results
6 ((fundus or fundal) adj3 height) in All text7 results
7 (SFH or "McDonald's rule") in All text7 results
8 (phonophoresis or sonication or sonification or sonograph* or ultra?sound or ultrashell or ultrasonic* or ultrasonograph* or echograph*) in All text4826 results
9 ((abdomen or abdominal) adj3 circumference) in All text54 results
10 ((fetal or fetus or foetal or foetus or embryo) adj3 weight) in All text116 results
11 #5 or #6 or #7 or #8 or #9 or #10 in All text4972 results
12 **#4 and #11 in All text139 results**

Epistemonikos (Internet) (www.epistemonikos.org/): up to 22 March 2023

Searched 22 March 2023

(title:(("Fetal Growth" OR "foetal growth" OR "small for Gestational Age" OR SGA OR FGR OR IUGR) AND (measure OR measuring OR "fundal height" OR "fundus height" OR SFH OR "McDonald's rule" OR phonophoresis OR sonication OR sonification OR sonograph* OR ultrasound OR ultrashell OR ultrasonic* OR ultrasonograph* OR echograph* OR "abdominal circumference")) OR abstract:(("Fetal Growth" OR "foetal growth" OR "small for Gestational Age" OR SGA OR FGR OR IUGR) AND (measure OR measuring OR "fundal height" OR "fundus height" OR SFH OR "McDonald's rule" OR phonophoresis OR sonication OR sonification OR sonograph* OR ultrasound OR ultrashell OR ultrasonic* OR ultrasonograph* OR echograph* OR "abdominal circumference"))) [Filters: protocol=no, classification=systematic-review, cochrane=missing, min_year=2000, max_year=2023]

Results found = 138

PROSPERO (International Prospective Register of Systematic Reviews) (CRD): up to 21 March 2023

Searched: 23 March 2023

#1MeSH DESCRIPTOR Fetal Growth Retardation EXPLODE ALL TREES48
#2(((fetal or foetal) near3 growth)):TI67
#3("Small for Gestational Age"):TI42
#4((SGA or FGR or IUGR)):TI23
#5MeSH DESCRIPTOR Infant, Small for Gestational Age EXPLODE ALL TREES19
#6#1 OR #2 OR #3 OR #4 OR #5146
#7MeSH DESCRIPTOR Ultrasonography, Prenatal EXPLODE ALL TREES52
#8(((fundus or fundal) NEAR3 height)):TI1
#9((measur* NEAR3 tape*)):TI0
#10(SFH or "McDonalds rule"):TI0
#11((phonophoresis or sonication or sonification or sonograph* or ultra?sound or ultrashell or ultrasonic* or ultrasonograph* or echograph*)):TI1297
#12(((abdomen or abdominal) NEAR3 circumference)):TI2
#13(((fetal or fetus or foetal or foetus or embryo) NEAR3 weight)):TI7
#14#7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #131330
#15**#6 AND #1411**

INPLASY (Internet) (https://inplasy.com/): up to 22 March 2023

Searched: 22 March 2023

Searched using following keywords and the results browsed for relevant papers

Search terms	Hits
fetal growth	0
Foetal growth	0
Small for gestational age	0
SGA	0
FGR	0
IUGR	0
Results	**0**

LILACS (Internet) (https://regional.bvsalud.org/php/index.php?lang=en): up to 22 March 2023

Searched: 22 March 2023

(("Fetal Growth" OR "foetal growth" OR "small for Gestational Age" OR sga OR fgr OR iugr)) AND ((measure OR measuring OR "fundal height" OR "fundus height" OR sfh OR "McDonald's rule" OR phonophoresis OR sonication OR sonification OR sonograph* OR ultrasound OR ultrashell OR ultrasonic* OR ultrasonograph* OR echograph* OR "abdominal circumference")) AND (db:("LILACS")) AND (year_cluster:[2000 TO 2012])

Results = 123

ClinicalTrials.gov (Internet) (https://clinicaltrials.gov/): up to 20 March 2023

Date searched: 20 March 2023

Search terms	Hits
"fetal growth" OR "foetal growth"	276
"Small for gestational age"	159
SGA	64
FGR	11
IUGR	270
Results before deduplication	**780**
Results after deduplication	**467**

EU Clinical Trials Register (Internet) (www.clinicaltrialsregister.eu/ctr-search/search): up to 20 March 2023

Date searched: 20 March 2023

Search terms	Hits
"fetal growth" OR "foetal growth"	25
"Small for gestational age"	31
SGA	57
FGR	10
IUGR	18
Results before deduplication	**141**
Results after deduplication	**101**

World Health Organization International Clinical Trials Registry Platform (WHO ICTRP) (Internet) (https://trialsearch.who.int/Default.aspx): up to 20 March 2023

Searched: 20 March 2023

Search terms	Hits
"fetal growth" OR "foetal growth"	279
"Small for gestational age"	226
SGA	141
FGR	25
IUGR	88
Results before deduplication	**759**
Results after deduplication	**614**

ScanMedicine (Internet) (https://scanmedicine.com/): up to 20 March 2023

Searched: 20 March 2023

This exact phrase (in document)	Hits trial	Devices
fetal growth	455	1
Foetal growth	41	0
Small for gestational age	379	0
SGA	381	0
FGR	64	0
IUGR	365	0
Results	**1685**	**1**
Results after deduplication	**1357**	**1**

Northern Light Life Sciences Conference Abstracts (Ovid): 2010–23/Wk10

Searched 21 March 2023

1 Fetal Growth Retardation/(814)
2 ((fetal or foetal or fetus or foetus or embryo or "in utero" or intra?uterine or intrauterine or prenatal) adj3 (growth or grow or grown or growing or size)).ti,ab. (2543)
3 (small adj3 (age or date)).ti,ab. (1126)
4 Infant, Small for Gestational Age/(0)
5 (SGA or FGR or IUGR).ti,ab. (3177)
6 exp Fetal Development/(0)
7 or/1-6 (5861)
8 (measur$ adj3 tape$).ti,ab. (38)
9 ((fundus or fundal) adj3 height).ti,ab. (28)
10 (SFH or "McDonald's rule").ti,ab. (50)
11 (phonophoresis or sonication or sonification or sonograph$ or ultra?sound or ultrashell or ultrasonic$ or ultrasonograph$ or echograph$).ti,ab. (44,582)
12 ((abdomen or abdominal) adj3 circumference).ti,ab. (192)
13 ((fetal or fetus or foetal or foetus or embryo) adj3 weight).ti,ab. (444)
14 or/8-13 (45,153)
15 **7 and 14 (443)**

TRIP Database (Internet) (www.tripdatabase.com): up to 16 March 2023

Searched: 16 March 2023

Advanced Search (TRIP PRO)

Limits: Publication year – 2013–21

This exact phrase (in document)	Hits guidelines	Hits regulatory G
fetal growth	293	20
Foetal growth	293	20
Small for gestational age	238	3
SGA	157	4
FGR	54	3

This exact phrase (in document)	Hits guidelines	Hits regulatory G
IUGR	122	10
Results	**1157**	**60**
Results after deduplication	**529**	**35**

Records found: 564

Guidelines International Network (GIN) (Internet) (https://g-i-n.net/international-guidelines-library/): up to 16 March 2023

Searched: 16 March 2023

Limits:

Publication year – 2013–current

Guideline publication status – Published

Search terms	Hits
fetal growth	0
Foetal growth	0
Small for gestational age	0
SGA	0
FGR	0
IUGR	0
Results	**0**

National Institute for Health and Care Excellence (NICE) (Internet) (www.nice.org.uk/guidance/): up to 16 March 2023

Searched: 16 March 2023

Search terms	Hits
fetal growth	26
Foetal growth	0/22 (dupes)
Small for gestational age	11/25
SGA	0/8
FGR	1/10
IUGR	0/2
Results	**93**
Results without duplicates	**38**

NIHR Health Technology Assessment (HTA) (Internet) (www.nihr.ac.uk/): up to 16 March 2023

Searched 16 March 2023

Home/Researchers/Data and publications

Search term	Results (journals and research)
fetal growth	5
Foetal growth	0
Small for gestational age	2
SGA	4/6
FGR	1/3
IUGR	0
Total	**16**
Total after deduplication	**12**

ECRI Guidelines Trust (Internet) (https://guidelines.ecri.org/): up to 20 March 2023

https://guidelines.ecri.org/

Searched: 20 March 2023

Limits

Date: 1 January 2013–20 March 2023

Resource type: Guidance

Search term	Results (screened for relevancy)
fetal growth	8
Foetal growth	0
Small for gestational age	3/6
SGA	0/9
FGR	0/2
IUGR	0
Total after deduplication	**11**

International HTA Database (INAHTA) (Internet) (https://database.inahta.org/): up to 20 March 2023

Date searched: 20 March 2023

Limits: Publication year – 2013–21

((“Infant, Small for Gestational Age”[mhe] or “Fetal Growth Retardation”[mhe])) OR (fetal growth) OR (Foetal growth) OR (“Small for gestational age”) OR (SGA or FGR or IUGR) FROM 2013 TO 2023

Records found = 23

HTA Database (CRD): up to 31 March 2018*

Searched: 20 March 2023

1 (MeSH DESCRIPTOR Infant, Small for Gestational Age EXPLODE ALL TREES)21Delete
2 (MeSH DESCRIPTOR Fetal Growth Retardation EXPLODE ALL TREES)25Delete
3 ((((fetal or foetal or fetus or foetus or "in utero" or intra?uterine or intrauterine or prenatal) adj3 (growth or grow or grows or growing or grown or size))))88Delete
4 (((small adj3 (age or date))))90Delete
5 (SGA or FGR or IUGR):TI2Delete
6 #1 OR #2 OR #3 OR #4 OR #5163Delete
7 **(#6) IN HTA17Delete**

*Archival resource only.

Update Searches

Database	Dates covered	Hits
EMBASE	1974–5 September 2023	13,641
MEDLINE + PreMedline	1946–5 September 2023	8861
Total		**22,502**

EMBASE (Ovid): 1974–5 September 2023

Searched 6.9.23

1 exp intrauterine growth retardation/(39,201)
2 ((fetal or foetal or fetus or foetus or "in utero" or intra?uterine or intrauterine or prenatal) adj3 (growth or grow or grown or growing or size)).ti,ab,ot,hw. (73,258)
3 (small adj3 (age or date)).ti,ab,ot,hw. (29,270)
4 (SGA or FGR or IUGR).ti,ab. (30,878)
5 fetus growth/ (17,458)
6 or/1-5 (106,528)
7 (measur$ adj3 tape$).ti,ab,ot. (2333)
8 uterus fundus height/(185)
9 ((fundus or fundal) adj3 height).ti,ab,ot. (576)
10 (SFH or "McDonald's rule").ti,ab. (487)
11 ultrasound/ or fetus echography/ or (phonophoresis or sonication or sonification or sonograph$ or ultra?sound or ultrashell or ultrasonic$ or ultrasonograph$ or echograph$).ti,ab,ot. (775,955)
12 abdominal circumference/ or ((abdomen or abdominal) adj3 circumference).ti,ab,ot. (8217)
13 fetus weight/ or ((fetal or fetus or foetal or foetus or embryo) adj3 weight).ti,ab,ot. (16,377)
14 7 or 8 or 9 or 10 or 11 or 12 or 13 (795,618)
15 6 and 14 (19,482)
16 limit 15 to yr="2000–Current" (15,901)
17 animal/(1,632,873)
18 animal experiment/(3,038,754)

19 (rat or rats or mouse or mice or murine or rodent or rodents or hamster or hamsters or pig or pigs or porcine or rabbit or rabbits or animal or animals or dogs or dog or cats or cow or bovine or sheep or ovine or monkey or monkeys).ti,ab,ot,hw. (7,632,969)
20 or/17-19 (7,632,969)
21 exp human/(25,504,725)
22 human experiment/(640,512)
23 or/21-22 (25,507,076)
24 20 not (20 and 23) (5,728,250)
25 16 not 24 (14,145)
26 (letter or editorial or note).pt. (3,022,559)
27 **25 not 26 (13,641)**

MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations and Daily (Ovid): 1946–5 September 2023

Searched 6.9.23

1 Fetal Growth Retardation/(18,883)
2 ((fetal or foetal or fetus or foetus or embryo or "in utero" or intra?uterine or intrauterine or prenatal) adj3 (growth or grow or grown or growing or size)).ti,ab,ot,hw. (47,854)
3 (small adj3 (age or date)).ti,ab,ot,hw. (19,040)
4 Infant, Small for Gestational Age/ (8646)
5 (SGA or FGR or IUGR).ti,ab. (19,015)
6 exp *Fetal Development/(14,876)
7 or/1-6 (77,307)
8 (measur$ adj3 tape$).ti,ab,ot. (1530)
9 ((fundus or fundal) adj3 height).ti,ab,ot,hw. (392)
10 (SFH or "McDonald's rule").ti,ab. (367)
11 exp Ultrasonography, Prenatal/ (37,460)
12 (phonophoresis or sonication or sonification or sonograph$ or ultra?sound or ultrashell or ultrasonic$ or ultrasonograph$ or echograph$).ti,ab,ot. (500,648)
13 ((abdomen or abdominal) adj3 circumference).ti,ab,ot. (3864)
14 ((fetal or fetus or foetal or foetus or embryo) adj3 weight).ti,ab,ot. (8595)
15 or/8-14 (524,515)
16 7 and 15 (14,223)
17 limit 16 to yr="2000 –Current" (10,094)
18 (letter or editorial or note).pt. (1,890,307)
19 exp animals/ not (exp animals/and humans/) (5,152,274)
20 **17 not (18 or 19) (8861)**

Appendix 2 Data extraction tables

TABLE 17 Baseline study details

Study details	Selection criteria	Participant details
Agrawal (2016)[223]	**Inclusion criteria:** Singleton; women with a regular menstrual history and known date of LMP; gestation 20–36 weeks	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Major structural anomaly	**Singleton/multiple:** Singleton
Country: India	**Research question:** 4	**Parental age, mean (SD) years:** 24.82 (3.31)
Setting: Prenatal clinic (tertiary care teaching hospital)		**Ethnicity:** NR
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 100		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Badr (2023)[224]	**Inclusion criteria:** Women aged ≥ 18 years with single live fetuses at 36 + 0/7–36 + 6/7 WG; dated by first trimester US; planning to deliver in the study centre	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Patients with learning difficulties or mental handicap; major fetal abnormality; known contraindication to MRI; failure to complete the MRI examination; intrauterine fetal demise; neonatal weight measured > 6 hours after the birth	**Singleton/multiple:** Singleton
Country: Belgium	**Research question:** 4	**Parental age:** NR
Setting: Department of Obstetrics and Gynecology (university hospital)		**Ethnicity:** NR
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 2378		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Baird (2016)[225]	**Inclusion criteria:** Women who gave birth at ≥ 24 + 0 weeks; US within 2 weeks of birth (the scan performed closest to birth was used); at least one USS for suspected FGR or an USS showing FGR as an incidental finding	**Participant category:** High risk

continued

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Baird (2016b)[226]	**Exclusion criteria:** Multiple pregnancy; pre-eclampsia; fetal abnormality (not specified); uncertain gestation	**Singleton/multiple:** Singleton
Publication type: Full paper	**Research question:** 4	**Parental age, mean (SD) years:** 31.8 (5.4)
Country: Australia		**Ethnicity (country of birth):** Australia 46; South Asia 28; Southeast- East Asia 21; Africa 5; Other 7
Setting: University teaching hospital		**Parity:** NR
Funding: NHMRC and the Victorian Government's Operational Infrastructure Support Program		**BMI, median (IQR):** 22.8 (21–26)
Study design: Diagnostic accuracy		**Diabetes:** NR
Recruitment: Retrospective		**Smoking:** NR
Number of participants: 107		**PET/gestational HT:** 0
		Female neonate: 64
Bais (2004)[227]	**Inclusion criteria:** Singleton pregnancies; GA confirmed by US; prenatal care started before 20 weeks; low risk at 20 weeks	**Participant category:** Low risk
Publication type: Full paper	**Exclusion criteria:** Delivery between 16 and 20 WG; BW < 500 g	**Singleton/multiple:** Singleton
Country: The Netherlands	**Research question:** 3	**Parental age:** NR
Setting: NR		**Ethnicity:** NR
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Retrospective		**Diabetes:** NR
Number of participants: Unclear		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Barreto[228]	**Inclusion criteria:** High-risk pregnancies admitted to a clinical obstetric ward; US dating before week 22 concord-ant with last known menses; singleton pregnancy	**Participant category:** High risk
Publication type: Full paper	**Exclusion criteria:** Fetal anomalies of the central nervous system, abdomen or both; fetal death; < 14 days from last US to birth	**Singleton/multiple:** Singleton
Country: Brazil	**Research question:** 3	**Parental age:** NR

Study details	Selection criteria	Participant details
Setting: University hospital		**Ethnicity:** NR
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 250		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Bastek (2009)[229]	**Inclusion criteria:** Women diagnosed with PET	**Participant category:** High risk (PET)
Publication type: Full paper	**Exclusion criteria:** Multifetal gestation; delivery outside hospital; lack of a growth US at the study institution within 3 weeks of delivery	**Singleton/multiple:** Singleton
Country: USA	**Research question:** 3	**Parental age:** NR
Setting: Tertiary care hospital		**Ethnicity:** African American 76
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI (≥ 30):** 58
Recruitment: Prospective		**Diabetes (type 2):** 20
Number of participants: 93		**Smoking:** 16
		PET/gestational HT: 93
		Female neonate: NR
Ben-Haroush (2007)[230]	**Inclusion criteria:** Healthy, singleton pregnancy; US documentation of the fetal BPD, HC, AC and FL performed at 28 and 34 WG	**Participant category:** Low risk
Publication type: Full paper	**Exclusion criteria:** HT; diabetes	**Singleton/multiple:** Singleton
Country: Israel	**Research question:** 3	**Parental age, median (min, max):** 29 (17–42)
Setting: Department of Obstetrics and Gynaecology, Rabin Medical Centre, Israel		**Ethnicity:** NR
Funding: NR		**Parity (nulliparous):** 91
Study design: Diagnostic accuracy		**BMI, median (min, max):** 23.8 (16.8 to 35.2)
Recruitment: Prospective		**Diabetes:** NR

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Number of participants: 259		**Smoking:** 0
		PET/gestational HT: 0
		Female neonate: NR
Bergman (2022)[231]	**Inclusion criteria:** Singleton pregnancy; S-EFW in the Fetal Medicine Unit between 24 and 36.6 weeks of gestation; delivery in research centre within 18 weeks from the first sonographic weight estimation	**Participant category:** General populations
Publication type: Full paper	**Exclusion criteria:** Examinations without GA documentation; examinations with missing or apparently false values	**Singleton/multiple:** Singleton
Country: Israel	**Research question:** 4	**Parental age, mean (SD) years:** 29.1 (NR)
Setting: Fetal Medicine Unit of a Medical Centre		**Ethnicity:** NR
Funding: NR		**Parity (nulliparous):** 1724
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Retrospective		**Diabetes:** NR
Number of participants: 6005		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Blue (2019)[232]	**Inclusion criteria:** Singleton live births; delivery at the study institution; delivery within 2 weeks of a US-derived EFW	**Participant category:** High risk; high risk; general population
[a]**Blue (2018a)**[233]	**Exclusion criteria:** Fetal hydrops; intrauterine fetal demise; inconsistent GA documentation; missing US or BW data; congenital anomalies not allowing for accurate assessment of the BPD, HC, AC or FL (e.g. hydrocephalus or severe ventriculomegaly, holoprosencephaly, bony cranial abnormality, abdominal wall defect, limb–body wall anomaly, skeletal dysplasia and caudal regression syndrome)	**Singleton/multiple:** Singleton
Blue (2018b)[234]	**Research question:** 4	**Parental age, mean (SD) years:** 28.9 (6.5); 28.8 (6.5); NR
Publication type: Full paper; full paper; conference abstract		**Ethnicity:**
Country: USA		White 189; Hispanic 397; Native American 109; African American 18; Asian 18; Other/missing 87
Setting: Department of Obstetrics and Gynecology (university hospital)		White 406; Hispanic 844; Native American 184; Black 32; Asian 41; other or missing 195
Funding: NR		NR

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Study design: Diagnostic accuracy		**Parity (nulliparous):** 229; 461; NR
Recruitment: Retrospective		**BMI:** NR
Number of participants: 831; 1704; 1226		**Diabetes (pre-gestational):** 56; 81; NR
		Diabetes (gestational): 149; 279; NR
		Smoking: 113; 239; NR
		PET/gestational HT: 152; 225; NR
		Female neonate: 418; 824; NR
Bonnevier (2022)[235]	**Inclusion criteria:** Pregnant women in the catchment area of Malmö-Lund who received routine US examination with EFW at 32 + 0 34 + 6 weeks (completed gestational weeks + days), as part of the free antenatal care program	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Multiple births; infants with unknown BW or unknown sex; GA was not estimated by US in the second trimester; US examination for an indication other than screening	**Singleton/multiple:** Singleton
Country: Sweden	**Research question:** 3	**Parental age:** NR
Setting: Obstetric and neonatal units		**Ethnicity:** NR
Funding: Region Skane, Department of Research and Development		**Parity (nulliparous):** 27,771
Study design: Diagnostic accuracy		**BMI, < 18.5; 18.5–24.9; 25–29.9; ≥ 30; missing:** 1524; 33,429; 12,548; 5281; 6670
Recruitment: Retrospective		**Diabetes:** NR
Number of participants: 59,452		**Smoking:** 6202
		PET/gestational HT: NR
		Female neonate: NR
Carbone (2012)[236]	**Inclusion criteria:** Women with reliably dated pregnancies based on LMPs and consistent with first trimester US	**Participant category:** General population
Carbone (2011)[237]	**Exclusion criteria:** Multiple gestations; pregnancies with aneuploidy and major structural anomalies	**Singleton/multiple:** Singleton
Publication type: Full paper	**Research question:** 3	**Parental age, mean (SD) years:** 34.9 (4.3)
Country: USA		**Ethnicity:** African American 292; Caucasian 2780; Asian 89; Hispanic 52; other 103

continued

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Setting: Division of Maternal–Fetal Medicine and Ultrasound, Department of Obstetrics and Gynecology		**Parity (nulliparous):** 664
Funding: None		**BMI, < 18.5; 18.5–24.9; 25–29.9; ≥ 30; missing**: 83; 1613; 750; 618
Study design: Diagnostic accuracy		**Diabetes, pre-gestational type II, gestational:** 78; 277
Recruitment: Prospective		**Smoking:** 184
Number of participants: 3329		**PET/gestational HT:** 270
		Female neonate: NR
Chandra (2022)[238]	**Inclusion criteria:** NR	**Participant category:** NR
Publication type: Conference abstract	**Exclusion criteria:** NR	**Singleton/multiple:** NR
Country: India	**Research question:** 2	**Parental age:** NR
Setting: NR		**Ethnicity:** NR
Funding: NR		**Parity:** NR
Study design: Comparative observational		**BMI:** NR
Recruitment: Retrospective, database study		**Diabetes:** NR
Number of participants (before GAP implementation; after GAP implementation): 26,295; 31,264		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Chang (2011)[239]	**Inclusion criteria:** Women who attended the prenatal clinic and were referred for a US examination for fetal biometry; defined LMP; dating confirmed by early pregnancy US using either by CRL or BPD; GA 21–40 weeks	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** NR	**Singleton/multiple:** Singleton
Country: Taiwan	**Research question:** 3	**Parental age:** NR
Setting: University hospital		**Ethnicity:** NR
Funding: National Science Council		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 263		**Smoking:** NR

DOI: 10.3310/AJLK7403

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
		PET/gestational HT: NR
		Female neonate: NR
Chang (2008)[240]	**Inclusion criteria:** Women who attended the prenatal clinic and were referred for a US examination for fetal biometry; defined LMP; dating confirmed by early pregnancy US using either by CRL or BPD; GA 21–40 weeks	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Fetal renal anomalies	**Singleton/multiple:** Singleton
Country: Taiwan	**Research question:** 3	**Parental age:** NR
Setting: University Hospital		**Ethnicity:** NR
Funding: National Science Council		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 249		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Chang (2007)[241]	**Inclusion criteria:** Women who attended the prenatal clinic and were referred for a US examination for fetal biometry; defined LMP; dating confirmed by early pregnancy US using either by CRL or BPD; GA 21–40 weeks	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** NR	**Singleton/multiple:** Singleton
Country: Taiwan	**Research question:** 3	**Parental age, range years:** 16–42
Setting: University hospital		**Ethnicity:** NR
Funding: National Science Council		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 346		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR

continued

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Chang (2006a)[242]	**Inclusion criteria:** Women who attended the prenatal clinic and were referred for a US examination for fetal biometry; defined LMP; dating confirmed by early pregnancy US using either by CRL or BPD; GA 21–40 weeks	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** NR	**Singleton/multiple:** Singleton
Country: Taiwan	**Research question:** 3	**Parental age:** NR
Setting: University Hospital		**Ethnicity:** NR
Funding: National Science Council		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 417		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Chang (2006b)[243]	**Inclusion criteria:** Women who attended the prenatal clinic and were referred for a US examination for fetal biometry; defined LMP; dating confirmed by early pregnancy US using either by CRL or BPD; GA 21–40 weeks	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** NR	**Singleton/multiple:** Singleton
Country: Taiwan	**Research question:** 3	**Parental age:** NR
Setting: University hospital		**Ethnicity:** NR
Funding: National Science Council		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 300		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Chang (2005a)[244]	**Inclusion criteria:** Women who attended the prenatal clinic and were referred for a US examination for fetal biometry; defined LMP; dating confirmed by early pregnancy US using either by CRL or BPD; GA 21–40 weeks	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** NR	**Singleton/multiple:** Singleton
Country: Taiwan	**Research question:** 3	**Parental age, mean (SD) years:** 30.8 (4.5)

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Setting: University hospital		**Ethnicity:** NR
Funding: National Science Council		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 312		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Chang (2005b)[245]	**Inclusion criteria:** Women who attended the prenatal clinic and were referred for a US examination for fetal biometry; defined LMP; dating confirmed by early pregnancy US using either by CRL or BPD; GA 21–40 weeks	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** NR	**Singleton/multiple:** Singleton
Country: Taiwan	**Research question:** 3	**Parental age:** NR
Setting: University hospital		**Ethnicity:** NR
Funding: National Science Council		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 482		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Chauhan (2006)[246]	**Inclusion criteria:** Singleton pregnancy; non-anomalous fetus; reliable GA based on clinical history, physical examination and sonographic examination before 22 weeks; medical or obstetric complication requiring antepartum testing; delivery at the study hospital within 4 weeks of biometric measurements	**Participant category:** High risk
Publication type: Full paper	**Exclusion criteria:** NR	**Singleton/multiple:** Singleton
Country: USA	**Research question:** 3	**Parental age:** NR
Setting: Regional medical centre		**Ethnicity:** NR
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Recruitment: Retrospective		**Diabetes (pre-gestational and gestational):** 247
Number of participants: 1954		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Chauhan (2004)[247]	**Inclusion criteria:** Dichorionic twin pregnancy; reliable determination of GA; no known fetal abnormalities; sonographic or histopathological determination of chorionicity; EFW performed within 21 days of delivery; live delivery of both twin fetuses (after 26 weeks) at a single community hospital	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** NR	**Singleton/multiple:** Twin
Country: USA and Australia	**Research question:** 3	**Parental age:** NR
Setting: Maternal and fetal center		**Ethnicity:** NR
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Retrospective		**Diabetes:** NR
Number of participants: 178 (89 pregnancies)		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Chauhan (2003)[248]	**Inclusion criteria:** Chronic HT according to ACOG criteria; reliable GA based on US before 22 weeks; EFW by US within 3 weeks of delivery	**Participant category:** High risk
Publication type: Full paper	**Exclusion criteria:** Known fetal abnormalities; multiple gestation; gestational HT; PET; DM	**Singleton/multiple:** Singleton
Country: USA	**Research question:** 3	**Parental age:** NR
Setting: NR		**Ethnicity:** NR
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Retrospective		**Diabetes (pre-gestational and gestational):** 0
Number of participants: 142		**Smoking:** NR

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
		PET/gestational HT: NR
		Female neonate: NR
Cowan (2021)[249]	**Inclusion criteria:** Pre-GAP cohort: all mothers who gave birth between 1 January and 31 December 2012 in CMH maternity facilities and received care by hospital-employed community midwives working in a continuity of care model. This was prior to widespread use of GROW charts and before the introduction of the NZMFM SGA guideline in 2013. Post-GAP cohort: comprised all mothers who gave birth between 1 April 2017 and 31 March 2018 in CMH maternity facilities and who received care by hospital-employed community midwives	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Pre-GAP and post-GAP cohorts: Births under the care of self-employed midwives were excluded as clinical records were not accessible. Further exclusions comprised booking > 20 weeks and 0 days, no maternal height or weight recorded, multiple pregnancy, baby born < 24 weeks and 0 days or with a major congenital anomaly	**Singleton/multiple:** Singleton
Country: New Zealand	**Research question:** 2	**Parental age:** NR
Setting: Tertiary obstetric facility		**Ethnicity (before GAP implementation):** Maori 16.8%; Pacific peoples 41%; Asian 25.9%; ME/LA/A 2.4%; European 13.6%; other 0.3%
Funding: NR		**Ethnicity (after GAP implementation):** Maori 17.3%; Pacific peoples 33.3%; Asian 33.3%; ME/LA/A 3.1%; European 13.1%; other 0.0%
Study design: Comparative observational		**Parity:** NR
Recruitment: Retrospective		**BMI(before GAP implementation), < 18.5; 18.5–24.9; 25–29.9; ≥ 30 :** 38; 306; 239; 522
Number of participants (before GAP implementation; after GAP implementation): 1105; 1082		**BMI(after GAP implementation), < 18.5; 18.5–24.9; 25–29.9; ≥ 30 :** 19; 342; 269; 452
		Diabetes: NR
		Smoking (before GAP implementation): 194
		Smoking (after GAP implementation): 29
		PET/gestational HT (before GAP implementation): 64
		PET/gestational HT (after GAP implementation): 77

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
		Female neonate: NR
De Jong (2000)[250]	**Inclusion criteria:** Primiparous women with one or more of the following: pre-existing HT, smoking > 15 cigarettes per day, age ≥ 35 years; multiparous women with an obstetric history of IUGR or pregnancy-induced hypertensive disorder	**Participant category:** High risk
Publication type: Full paper	**Exclusion criteria:** Multiple pregnancies; evidence of fetal congenital abnormality	**Singleton/multiple:** Singleton
Country: The Netherlands	**Research question:** 3	**Parental age, mean (SD) years:** 32.9 (4.5)
Setting: Department of Obstetrics and Gynecology, university hospital		**Ethnicity:** European 185
Funding: NR		**Parity (nulliparous):** 146
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Retrospective		**Diabetes:** NR
Number of participants: 215		**Smoking:** NR
		PET: 37
		Female neonate: 107
De Reu (2008)[251]	**Inclusion criteria:** Singleton pregnancies; Caucasian; no risk factors or pathology at the onset of the third trimester	**Participant category:** Low risk
Publication type: Full paper	**Exclusion criteria:** No scan performed at the end of the first trimester or at the onset of the third trimester; incomplete data (missing gender, AC or HC)	**Singleton/multiple:** Singleton
Country: The Netherlands	**Research question:** 3	**Parental age:** NR
Setting: Primary care midwifery practice		**Ethnicity (development cohort):** White/Caucasian 3499
Funding: NR		**Ethnicity (validation cohort):** White/Caucasian 725
Study design: Model development and validation (diagnostic accuracy data)		**Parity (≥ 1) (development cohort):** 3499
Recruitment: Retrospective		**Parity (≥ 1) (validation cohort):** 725
Number of participants: 4224		**BMI:** NR
		Diabetes: NR
		Smoking: NR
		PET/gestational HT: NR
		Female neonate (development cohort): 1677

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
		Female neonate (validation cohort): 364
De Silva (2021)[252]	**Inclusion criteria:** Singleton pregnancy; accurate dating scan prior to 20 weeks	**Participant category:** Low risk
De Silva (2017)[253]	**Exclusion criteria:** Medical disorders complicating pregnancy	**Singleton/multiple:** Singleton
Publication type: Full paper	**Research question:** 3	**Parental age, median (range) years:** 30 (18–41)
Country: Sri Lanka		**Ethnicity:** NR
Setting: General hospital		**Parity (nulliparous):** 243
Funding: NR		**BMI:** NR
Study design: Diagnostic accuracy		**Diabetes:** NR
Recruitment: Prospective		**Smoking:** NR
Number of participants: 508		**PET/gestational HT:** NR
		Female neonate: NR
Dude (2021)[254]	**Inclusion criteria:** Nulliparous; age ≥ 18 years; class II BMI 35.0–39.9 kg/m^2 or class III BMI ≥ 40.0 kg/m^2; term delivery (37 weeks or greater gestation); singleton gestation; US USEFW within 5 weeks of delivery, for any indication	**Participant category:** Obese (BMI ≥ 35)
Publication type: Full paper	**Exclusion criteria:** Fetus with a major anomaly as this may impair the accuracy of EFW	**Singleton/multiple:** Singleton
Country: USA	**Research question:** 4	**Parental age, mean (SD) years:** 29.1 (6.3)
Setting: NR		**Ethnicity:** Non-Hispanic White 192; non-Hispanic Black 207; Asian 25; Hispanic 160; other/unknown 102
Funding: Government		**Parity:** NR
Study design: Diagnostic accuracy		**BMI, < 18.5; 18.5–24.9; 25–29.9; ≥ 30 :** 0; 0; 0; 690
Recruitment: Retrospective		**Diabetes:** NR
Number of participants: 690		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Dude (2018)[255]	**Inclusion criteria:** Age 18 years or greater; gestational or pregestational diabetes; term delivery (37.0 WG or greater); singleton gestations; USEFW within 5 weeks of delivery	**Participant category:** Diabetes

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Publication type: Full paper	**Exclusion criteria:** Major fetal anomaly; informal 'bedside' USs not read by the Division of Maternal–Fetal Medicine	**Singleton/multiple:** Singleton
Country: USA	**Research question:** 4	**Parental age, mean (SD) years:** 31.4 (5.4)
Setting: NR		**Ethnicity:** Non-Hispanic White 168; Non-Hispanic Black 75; Hispanic 108; other 105
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI, < 18.5; 18.5–24.9; 25–29.9; ≥ 30:** NR; NR; NR; 309
Recruitment: Retrospective		**Diabetes, pre-gestational type I; pre-gestational; type II; gestational:** 54; 54; 413
Number of participants: 521		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Haragan (2015)[256]	**Inclusion criteria:** Between 24 and 40 weeks' GA; singleton pregnancy; undergoing a scheduled ultrasound to assess fetal growth	**Participant category:** High risk
Publication type: Full paper	**Exclusion criteria:** Known fetal congenital anomalies; aneuploidy; unsure estimated date of conception	**Singleton/multiple:** Singleton
Country: USA	**Research question:** 3	**Parental age, mean (range) years:** 28.3 (18–41)
Setting: American Institute of Ultrasound in Medicine and Prenatal Wellness Center at the Medical University of South Carolina		**Ethnicity:** Black 113; White 131; other 7
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI, mean (range):** 33.9 (18.1–71.6)
Recruitment: Prospective		**BMI, < 18.5; 18.5–24.9; 25–29.9; ≥ 30:** NR; NR; NR; 145
Number of participants: 251		**Diabetes (pre-gestational and gestational):** 52
		Smoking: 29
		PET/gestational HT: 65
		Female neonate: NR
Harper (2013)[257]	**Inclusion criteria:** Twin gestation; sonographically obtained biometric measurements (BPD, HC, AC and FL) available within 7 days of delivery; BW available	**Participant category:** General population

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Publication type: Full paper	**Exclusion criteria:** Intrauterine fetal death; pregnancies complicated by anomalies	**Singleton/multiple:** Twin
Country: USA	**Research question:** 3	**Parental age, median (IQR) years:** 30.9 (26.3 to 35.3)
Setting: Tertiary care		**Ethnicity:** White 159; Black 59
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI, median (IQR):** 25.6 (22.5 to 30.7)
Recruitment: Retrospective		**Diabetes:** NR
Number of participants: 540 (270 pregnancies)		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: 270
Henrichs (2019)[258]	**Inclusion criteria:** Antenatal care in a participating midwifery practice at enrolment; age ≥ 16 years; singleton pregnancy; no major obstetric or medical risk factors; reliable expected date of delivery based on a dating scan or a reliable first day of the LMP	**Participant category:** Low risk
van Roekel (2023)[259]	**Exclusion criteria:** NR	**Singleton/multiple:** Singleton
Publication type: Full paper	**Research question:** 4	**Parental age, mean years (intervention; control; total):** 31; 31; 31
Country: The Netherlands		**Ethnicity (intervention):** Dutch 5096; other Western 766; non-Western 1202; missing 3
Setting: Midwifery practices		**Ethnicity (control):** Dutch 4684; other Western 576; non-Western 714; missing 5
Funding: The Netherlands Organisation for Health Research and Development		**Ethnicity (total):** Dutch 9780; other Western 1342; non-Western 1916; missing 8
Study design: RCT		**Parity (nulliparous) (intervention):** 3368
Recruitment: Prospective		**Parity (nulliparous) (control):** 2928
Number of participants [intervention (routine third-trimester US at 28–30 and 34–36 weeks + usual care); control (usual care, serial SFH measurements + US when clinically indicated); total]: 7067; 5979; 13,046		**Parity (nulliparous) (total):** 6296 **BMI, < 18.5; 18.5–24.9; 25–29.9; ≥ 30; missing (intervention):** 232; 4583; NR; NR; 103
		BMI, < 18.5; 18.5–24.9; 25–29.9; ≥ 30; missing (control): 185; 4025; NR; NR; 62

continued

DOI: 10.3310/AJLK7403

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
		BMI, < 18.5; 18.5–24.9; 25–29.9; ≥ 30; missing (total): 417; 8608; NR; NR;165
		Diabetes: NR
		Smoking (intervention): 1003
		Smoking (control): 835
		Smoking (total): 1838
		PET/gestational HT: NR
		Female neonate (intervention): 3480
		Female neonate (control): 2923
		Female neonate (total): 6403
Hugh (2021)[260]	**Inclusion criteria:** All maternity units in England, classified by GAP status in 2017: not in GAP program; GAP partially implemented (incomplete adoption of protocol, no monitoring and audit); GAP implemented completely	**Participant category:** Low risk
Publication type: Full paper	**Exclusion criteria:** Births that did not occur in an NHS institution; births for which the place of birth was not recorded; births for which the postcode of the place of birth could not be mapped to a maternity unit	**Singleton/multiple:** Mixed
Country: UK	**Research question:** 2	**Parental age, median (IQR) years (non-GAP pre-implementation; non-GAP post-implementation; complete-GAP pre-implementation; complete-GAP post-implementation):** 30.3 (26–35); 31.4 (27–35); 29 (24–34); 30.2 (26–34)
Setting: NR		**Ethnicity:** NR
Funding: NR		**Parity (nulliparous) (non-GAP pre-implementation; non-GAP post-implementation; complete-GAP pre-implementation; complete-GAP post-implementation):** 53,523; 87,621; 63,241; 112,406
Study design: Comparative observational		**BMI:** NR
Recruitment: Retrospective analysis of ONS data		**Diabetes:** NR

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Number of participants (non-GAP pre-implementation; non-GAP post-implementation; complete-GAP pre-implementation; complete-GAP post-implementation): 204,811; 199,575; 289,539; 274,721		**Smoking:** NR
		PET/gestational HT: NR **Female neonate (non-GAP pre-implementation; non-GAP post-implementation; complete-GAP pre-implementation; complete-GAP post-implementation):** 99,967; 97,091; 140,911; 133,894
Humphries (2002)[261]	**Inclusion criteria:** Non-anomalous; singleton pregnancy; delivery within 14 days of US	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** NR	**Singleton/multiple:** Singleton
Country: USA	**Research question:** 4	**Parental age:** NR
Setting: Maternal and fetal development department, regional centre		**Ethnicity:** NR
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 238		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Jayawardena (2019)[262]	**Inclusion criteria:** General population	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Multiple pregnancies; known fetal anomalies	**Singleton/multiple:** Singleton
Country: Australia	**Research question:** 2	**Parental age, mean (range) years (before GAP implementation; after GAP implementation):** 31.2 (14–45); 31.3 (16–48)
Setting: Tertiary care		**Ethnicity (before GAP implementation):** Australian European 470; British European 17; Eastern European 4; Irish European 7; Central/Western European 30; Bangladeshi 10; Indian 47; Chinese 35; Other far Eastern 0; South East Asian 96; Other Asian 32; Middle Easter 22; North African 23; Sub-Saharan African 98; North American 6; South and Central American 10

continued

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Funding: NR		**Ethnicity (after GAP implementation):** Australian European 406; British European 17; Eastern European 20; Irish European 12; Central/Western European 11; Bangladeshi 7; Indian 67; Chinese 30; Other far Eastern 17; South East Asian 95; Other Asian 12; Middle Easter 24; North African 47; Sub-Saharan African 71; North American 11; South and Central American 14
Study design: Comparative observational		**Parity (nulliparous) (before GAP implementation; after GAP implementation):** 438; 438
Recruitment: Prospective		**BMI, mean (range) (before GAP implementation; after GAP implementation):** 24.8 (16–58); 24.6 (9–53)
Number of participants (before GAP implementation; after GAP implementation): 936; 882		**Diabetes:** NR
		Smoking: NR
		PET/gestational HT (before GAP implementation; after GAP implementation): 71; 46
		Female neonate: NR
Leung (2008)[263]	**Inclusion criteria:** Chinese women; singleton pregnancy	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Chromosomal abnormalities; structural abnormalities; intrauterine death	**Singleton/multiple:** Singleton
Country: China	**Research question:** 3	**Parental age, median:** 33
Setting: University fetal medicine unit		**Ethnicity:** Chinese 2760
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 2760		**Smoking:** 32
		PET/gestational HT: NR
		Female neonate: NR
Li (2021)[264]	**Inclusion criteria:** Singleton pregnancies managed at the authors' hospital	**Participant category:** General population

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Publication type: Full paper	**Exclusion criteria:** Structural abnormalities, liver or brain malformations; GA impossible to determine; too many motion-related artifacts during MRI, which made it impossible to distinguish organ boundaries	**Singleton/multiple:** Singleton
Country: China	**Research question:** 3	**Parental age, mean (SD):** 28.9 (4.86)
Setting: Womens' hospital, department of radiology		**Ethnicity:** NR
Funding: Public Project of Technology Research and Social Development in Zhejiang Province		**Parity, median:** 1.55
Study design: Diagnostic accuracy		**BMI, mean (SD):** 24.47 (1.63)
Recruitment: Retrospective		**Diabetes:** NR
Number of participants: 147		**Smoking:** NR
		PET/gestational HT:
		Female neonate: NR
Lindström (2023)[265]	**Inclusion criteria:** Women who gave birth at 22 + 0 gestational weeks or later; singleton pregnancy; US EFW estimation within 2 days of delivery	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Stillbirth; missing BW; missing or erroneous US measurements; pregnancies that were initially multiple with intrauterine death of at least one fetus	**Singleton/multiple:** Singleton
Country: Sweden	**Research question:** 3	**Parental age:** NR
Setting: Registry data		**Ethnicity (country of birth):** Nordic country 20,213; remaining Europe 1704; Asia and Oceania 3990; Africa 2038; North and south America 366; Missing 3210
Funding: Swedish Infant Death Foundation		**Parity (nulliparous):** 14,269
Study design: Diagnostic accuracy		**BMI, < 18.5; 18.5–24.9; 25–29.9; ≥ 30; missing:** 829; 14,894; 8103; 5869; 1826
Recruitment: Retrospective		**Diabetes:** NR
Number of participants: 31,521		**Smoking:** 1556
		PET/gestational HT: NR
		Female neonate: 15,397

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Marchand (2022)[266]	**Inclusion criteria:** Singleton live pregnancy; examination between 14 + 0 and 41 + 6 weeks with expected dates of birth determined by US using the crown–rump length; one fetal US per pregnancy	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Congenital fetal anomalies, and sonographic evidence of malformations, such as gastroschisis, skeletal dysplasia or hydrops fetalis; LGA fetus	**Singleton/multiple:** Singleton
Country: Germany	**Research question:** 4	**Parental age, median (range) years:** 33 (14–50)
Setting: Department of Gynecology and Obstetrics, St. Franziskus Hospital Muenster		**Ethnicity:** NR
Funding: None		**Parity:** NR
Study design: Diagnostic accuracy		**BMI, median (range):** 25 (15–65)
Recruitment: Retrospective		**Diabetes:** NR
Number of participants: 9292		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: 4527
Martin-Palumbo (2022)[267]	**Inclusion criteria:** Singleton pregnancy; absence of congenital anomalies and obstetric complications, such as premature rupture of membranes or those related with serious maternal diseases	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Maternal age under 18 years old; women unable to understand investigators' language; FGR due to severe placental vascular insufficiency	**Singleton/multiple:** Singleton
Country: Spain	**Research question:** 3	**Parental age, mean (range) years:** 34.6 (18.8 to 47.5)
Setting: Tertiary care		**Ethnicity:** Black 9; East Asian 12; Mixed 73; White 383
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 488		**Smoking:** 463
		PET/gestational HT: NR
		Female neonate: NR

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Mathewlynn (2022)[268]	**Inclusion criteria:** Singleton pregnancy; dating scan using crown–rump length measurement between 11 + 0 and 13 + 6 weeks; growth scan between 35 + 0 and 36 + 6 weeks; BW recorded at delivery pregnancy; dating scan using crown–rump length measurement between 11 + 0 and 13 + 6 weeks; growth scan between 35 + 0 and 36 + 6 weeks; BW recorded at delivery	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Fetal anomaly; stillbirth; delivery before 37 + 0 or after 42 + 6 weeks; incomplete biometry (either HC, AC or FL missing); delivery outside the study hospital	**Singleton/multiple:** Singleton
Country: UK	**Research question:** 4	**Parental age, mean (SD):** 31.12 (5.34)
Setting: Tertiary care obstetric unit		**Ethnicity:** Bangladeshi 78; British European 8575; Chinese 158; Indian 228; Irish European 111; mixed African-European 42; mixed Asian-European 53; mixed Caribbean-European 90; Pakistani 316; other 2340; unclassified 5687
Funding: NR		**Parity (nulliparous):** 7779
Study design: Diagnostic accuracy		**BMI, mean:** 25.54
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 17,678		**Smoking:** 1700
		PET/gestational HT: NR
		Female neonate: NR
Michaeli (2022)[269]	**Inclusion criteria:** Singleton; term neonates (i.e. > 37 weeks and 0 days of gestation); without anomalies	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Neonates with fetal weight estimation and BW above the 75th percentile	**Singleton/multiple:** Singleton
Country: Israel	**Research question:** 1 and 3	**Parental age, mean (SD):** 28.7 (5.7)
Setting: Tertiary care		**Ethnicity:** NR
Funding: NR		**Parity (nulliparous):** 25,637
Study design: Comparative observational		**BMI:** NR
Recruitment: Retrospective		**Diabetes (gestational):** 3040
Number of participants: 100,198		**Smoking:** NR

continued

DOI: 10.3310/AJLK7403

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
		PET/gestational HT: NR
		Female neonate: NR
[a]**Monier (2022a)**[270]	**Inclusion criteria:** Singleton live births; information available on 3rd trimester US (30–35 WG)	**Participant category:** General population
Monier (2022b)[271]	**Exclusion criteria:** Mothers < 18 years; missing data on gestation al age and EFW at 3rd trimester US; missing data on sex, BW and GA at delivery	**Singleton/multiple:** Singleton
Publication type: Full paper	**Research question:** 4	**Parental age:** NR
Country: France		**Ethnicity:** NR
Setting: Public and private maternity units		**Parity (nulliparous):** 4274
Funding: NR		**BMI, < 18.5; 18.5–24.9; 25–29.9; ≥ 30 :** 674; 5848; 1761; 1096
Study design: Diagnostic accuracy		**Diabetes (pre-gestational and gestational):** 1115
Recruitment: Retrospective		**Smoking:** NR
Number of participants: 9940		**PET/gestational HT:** 397
		Female neonate: 4809
[a]**Nwabuobi (2020)**[272]	**Inclusion criteria:** Women referred for fetal growth US from providers who practice within the centres or use the centres for the provision of prenatal diagnosis (common indications for growth USs in the study included: uterine size/date discrepancies, history of HT or diabetes, maternal history of chronic renal or vascular disease, and hemoglobinopathies)	**Participant category:** High risk
Pressman (2022)[273]	**Exclusion criteria:** Multiple gestations; chromosomal or congenital malformations	**Singleton/multiple:** Singleton
Odibo (2018a)[274]	**Research question:** 4	**Parental age, mean (SD) years:** 29.6 (6.1)
Roeckner (2021)[275]		**Ethnicity:** Asian 52; African American 325; Hispanic 148; European/White 512; Multiracial/other 17
Odibo (2018b)[156]		**Parity:** NR
Odibo (2018c)[155]		**BMI, mean (SD):** 28.9 (9.1)
Nwabuobi (2018)[276]		**Diabetes (pre-gestational):** 84
Publication type: Full paper		**Smoking:** 91
Country: USA		**PET/gestational HT:** NR

Study details	Selection criteria	Participant details
Setting: Division of Maternal–Fetal Medicine, Department of Obstetrics and Gynecology, university hospital		**Female neonate:** NR
Funding: Eunice Kennedy Shriver NICHD		
Study design: Diagnostic accuracy		
Recruitment: Prospective		
Number of participants: 1054		
Nymark Hansen (2019)[277]	**Inclusion criteria:** Singleton pregnancies	**Participant category:** High risk and general population
Publication type: Full paper	**Exclusion criteria:** Abortion/miscarriage before 22 weeks of gestation; delivery outside of North Denmark Region	**Singleton/multiple:** Singleton
Country: Denmark	**Research question:** 1 and 3	**Parental age, NR (range) years:** 29.5 (25 to 34)
Setting: Department of Obstetrics and Gynecology, Aalborg University Hospital		**Ethnicity:**
Funding: Research Foundation of North Denmark Region		**Parity (nulliparous):** 1361
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Retrospective		**Diabetes (pre-gestational and gestational):** 170
Number of participants: 2928		**Smoking:** 255
		PET/gestational HT: 138
		Female neonate: 1403
Poljak (2017)[278]	**Inclusion criteria:** Suspected SGA referred, after primary screening with SFH measurement and US examination (the indication for referral in all cases was US evidence of EFW < 10th percentile on the GROW chart)	**Participant category:** High risk
Publication type: Full paper	**Exclusion criteria:** Abnormal Doppler; oligohydramnios; structural anomalies identified at any point in their care	**Singleton/multiple:** Singleton
Country: UK	**Research question:** 4	**Parental age, median (range) years:** 28 (15–48)
Setting: Specialist fetal growth clinic		**Ethnicity:** Caucasian 92
Funding: NR		**Parity, median (range):** 1 (0 to 8)
Study design: Diagnostic accuracy		**BMI, median (range):** 24 (17 to 47)
Recruitment: Prospective		**Diabetes:** NR

continued

DOI: 10.3310/AJLK7403

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Number of participants: 105		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: 66
Price (2022)[279]	**Inclusion criteria:** Singleton pregnancy; GAs between 26/0 weeks and 36/6 weeks; US dating prior to 24/0 weeks; women referred for fetal growth USs (common indications for growth USs in the study included uterine size/date discrepancies, maternal history of HT or diabetes, maternal history of chronic renal or vascular disease, and hemoglobinopathies)	**Participant category:** High risk
Price (2020)[280]	**Exclusion criteria:** Chromosomal or congenital malformations	**Singleton/multiple:** Singleton
Publication type: Full paper	**Research question:** 3	**Parental age, NR (NR):** 30.5 (NR)
Country: USA		**Ethnicity:** NR
Setting: NR		**Parity (nulliparous):** 209
Funding: NR		**BMI, median:** 29.1
Study design: Diagnostic accuracy		**Diabetes (pre-gestational):** 56
Recruitment: Prospective		**Smoking:** 46
Number of participants: 612		**PET/gestational HT:** NR
		Female neonate: NR
Rad (2018)[281]	**Inclusion criteria:** Singleton pregnancy; non-anomalous; US for fetal growth at >= 36 gestational weeks for any indication; delivery at study institution	**Participant category:** General population
Rad (2015)[282]	**Exclusion criteria:** Unknown or inaccurate GA dating; fetal demise	**Singleton/multiple:** Singleton
Publication type: Full paper	**Research question:** 3	**Parental age, range years:** 29.1 to 36.3
Country: USA		**Ethnicity:** Caucasian 1074; Black 239; Asian 250; Hispanic 31
Setting: Tertiary care		**Parity, median (range):** 0 (0 to 1)
Funding: None		**BMI, median (range):** 29.4 (26.3 to 34.0)
Study design: Diagnostic accuracy		**Diabetes:** NR
Recruitment: Retrospective		**Smoking:** NR
Number of participants: 1594		**PET/gestational HT:** NR

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
		Female neonate: NR
Ridha (2022)[283]	**Inclusion criteria:** Singleton pregnancies = or > 20 weeks; delivery at the study unit within 7 days of US	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Fetal demise at time of US; major foetal anomalies that would potentially affect any of the four biometry measurements (AC), BPD, HC and FL such as anencephaly, achondroplasia or gastroschisis	**Singleton/multiple:** Singleton
Country: New Zealand	**Research question:** 3	**Parental age:** NR
Setting: National Women's Health unit		**Ethnicity:** African 35; Asian 453; Pacific people 397; European 567; Indian 323; Maori 211; Middle Eastern 41; Latin American 31; other 3
Funding: None		**Parity (nulliparous):** 1062
Study design: Diagnostic accuracy		**BMI, < 18.5; 18.5–24.9; 25–29.9; ≥ 30 :** 63; 884; 490; 592
Recruitment: Retrospective		**Diabetes:** NR
Number of participants: 2061		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Salomon (2005)[284]	**Inclusion criteria:** Singleton pregnancies undergoing first, second and third-trimester US examination at 11–14, 20–24 and 30–34 weeks of gestation, respectively, as part of routine antenatal care in France	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** NR	**Singleton/multiple:** Singleton
Country: France	**Research question:** 3	**Parental age, median (range) years:** 30 (18–42)
Setting: Service de Gynécologie-Obstétrique, Centre Hospitalier Intercommunal de Poissy-St Germain		**Ethnicity:** NR
Funding: NR		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Prospective		**Diabetes:** NR
Number of participants: 356		**Smoking:** NR
		PET/gestational HT: NR

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
		Female neonate: NR
Sklar (2017)[285]	**Inclusion criteria:** Triplet pregnancies	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** NR	**Singleton/multiple:** Triplet
Country: Canada	**Research question:** 3	**Parental age, median years:** 31
Setting: Tertiary care		**Ethnicity:** NR
Funding: The Lois Hole Hospital for Women through the Women and Children's Health Research Institute		**Parity (nulliparous):** 42
Study design: Diagnostic accuracy		**BMI:** NR
Recruitment: Retrospective		**Diabetes (gestational):** 12
Number of participants: 78 (234 fetuses)		**Smoking:** 6
		PET/gestational HT: 15
		Female neonate: NR
Skrastad (2013)[286]	**Inclusion criteria:** All pregnant women offered routine US examinations at 18 weeks of gestation	**Participant category:** General population
Skrastad (2014)[287]	**Exclusion criteria:** Estimated date of delivery was not predicted	**Singleton/multiple:** Singleton
Publication type: Full paper	**Research question:** 3	**Parental age, mean (SD) years (intervention; control):** 27 (5); 27 (5)
Country: Norway		**Ethnicity:** NR
Setting: National Center for Fetal Medicine, St Olav's Hospital		**Parity (nulliparous) (intervention; control):** 1448; 1501
Funding: National Center for Fetal Medicine, Department of Obstetrics and Gynecology, St Olav's Hospital, Trondheim University Hospital and Department of Laboratory Medicine Children's and Women's Health, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim		**BMI:** NR
Study design: RCT		**Diabetes:** NR
Recruitment: Prospective		**Smoking (intervention; control):** 852; 893
Number of participants [intervention (US at 18 and 33 weeks); control (US at 18 weeks only)]: 3190; 3236		**PET/gestational HT:** NR
		Female neonate: NR

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Sovio (2015)[288]	**Inclusion criteria:** Nulliparous women attending for their dating USS at the study hospital	**Participant category:** General population
Sovio (2014)[289]	**Exclusion criteria:** Multiple pregnancy	**Singleton/multiple:** Singleton
Publication type: Full paper	**Research question:** 3	**Parental age:** NR
Country: UK		**Ethnicity (no clinically indicated US ≥ 26 weeks):** White 2151; missing 40
Setting: Rosie Hospital, Cambridge		**Ethnicity (one or more clinically indicated US ≥ 26 weeks):** White 1545; missing 29
Funding: National Institute for Health Research, Medical Research Council, Sands, and GE Healthcare		**Ethnicity (all):** White 3696; missing 69
Study design: Diagnostic accuracy		**Parity (nulliparous) (no clinically indicated US ≥ 26 weeks; one or more clinically indicated US ≥ 26 weeks; total):** 2311; 1666; 3977
Recruitment: Prospective		**BMI, < 25; 25 to 29.9; ≥ 30; missing (no clinically indicated US ≥ 26 weeks):** 1416; 667; 228; 0
Number of participants (no clinically indicated US ≥ 26 weeks; one or more clinically indicated US ≥ 26 weeks; total): 2311; 1666; 3977		**BMI, < 25; 25 to 29.9; ≥ 30; missing (one or more clinically indicated US ≥ 26 weeks):** 909; 450; 306; 1
		BMI, < 25; 25 to 29.9; ≥ 30; missing (all): 2325; 1117; 534; 1
		Diabetes (pre-gestational; gestational) (no clinically indicated US ≥ 26 weeks): 0; 2
		Diabetes (pre-gestational; gestational) (one or more clinically indicated US ≥ 26 weeks): 14; 160
		Diabetes (pre-gestational; gestational) (all): 14; 162
		Smoking (no clinically indicated US ≥ 26 weeks; one or more clinically indicated US ≥ 26 weeks; total): 106; 79; 185
		PET/gestational HT: NR
		Female neonate: NR

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Temming (2017)[290]	**Inclusion criteria**: Singleton pregnancy; non-anomalous (pregnancies with ultrasonographic soft markers for aneuploidy, including echogenic intracardiac focus, single umbilical artery and pyelectasis were included); between 17 0/7 and 22 6/7 weeks of gestation; EFW performed at the time of ultrasonography	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Fetal structural malformations; chromosomal abnormalities; identified infectious aetiologies; pregnancy did not continue past 20 weeks of gestation; outcome data missing or not available	**Singleton/multiple:** Singleton
Country: USA	**Research question:** 4	**Parental age, mean (SD) years:** 28.7 (5.8)
Setting: NR		**Ethnicity:** African American 4280
Funding: NIH training grant; NIH National Centre for Advancing Translational Sciences grant		**Parity:** NR
Study design: Diagnostic accuracy		**BMI, mean (SD):** 26.3 (8.2)
Recruitment: Retrospective		**BMI ≥ 30 :** 3669
Number of participants: 12,783		**Diabetes (pre-gestational):** 632
		Smoking: 956
		PET/gestational HT: NR
		Female neonate: NR
Turitz (2014)[291]	**Inclusion criteria:** Singleton pregnancies who presented for at least one growth US between 26 and 36 weeks GA	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Fetal anomalies; stillbirths; twins with one fetal loss	**Singleton/multiple:** Singleton
Country: USA	**Research question:** 3	**Parental age:** NR
Setting: Department of Obstetrics and Gynecology, Hospital of the University of Pennsylvania		**Ethnicity:** NR
Funding: Eunice Kennedy Shriver National Institute of Child Hand Human Development		**Parity:** NR
Study design: Diagnostic accuracy		**BMI:** NR

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Recruitment: Retrospective		**Diabetes:** NR
Number of participants: 10,642		**Smoking:** NR
		PET/gestational HT: NR
		Female neonate: NR
Tuuli (2011a)[292]	**Inclusion criteria:** LMP pregnancy dating consistent with first trimester US; first and second trimester US examinations at the study unit	**Participant category:** General population
Tuuli (2011b)[293]	**Exclusion criteria:** Multiple gestations; major structural and chromosomal anomalies	**Singleton/multiple:** Singleton
Publication type: Full paper	**Research question:** 3	**Parental age, mean (SD) years:** 34.6 (0.1)
Country: USA		**Ethnicity:** African American 1080; White 5832; other 1521
Setting: Department of Obstetrics and Gynecology, Washington University School of Medicine		**Parity (nulliparous):** 2653
Funding: NR		**BMI, mean (SD):** 26.4 (0.1)
Study design: Diagnostic accuracy		**Diabetes (pre-gestational):** 184
Recruitment: Retrospective		**Smoking:** 651
Number of participants: 8433		**PET/gestational HT:** NR
		Female neonate: 4158
Vieira (2022)[294]	**Inclusion criteria:** All women who gave birth in participating clusters (maternity units) during the year prior to randomisation and during the trial (November 2016–February 2019)	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Multiple pregnancies; fetal abnormalities; births before 24 + 1 weeks	**Singleton/multiple:** Singleton
Country: UK	**Research question:** 2	**Parental age, median (IQR) (GAP implementation; SC):** 31.8 (27.9–35.5); 32 (27.9–35.4)
Setting: 13 Maternity units in England		**Ethnicity (GAP implementation):** White 6236; Black 1398; Asian 2252; mixed 178; other 1021
Funding: Guy's and St Thomas' Charity (MAJ150704), Stillbirth and Neonatal Death Charity – SANDS (RG1011/16) and Tommy's Charity		**Ethnicity (SC):** White 8659; Black 2085; Asian 1864; mixed 359; other 842

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Study design: Cluster RCT		**Parity (nulliparous) (GAP implementation; SC):** 5726; 6560
Recruitment: Prospective		**BMI, < 18.5; 18.5–24.9; 25–29.9; ≥ 30 (GAP implementation):** 377; 5726; 3018; 1975
Number of participants (GAP implementation; SC according to local guidelines): 11,096 (5 clusters); 13,810 (6 clusters)		**BMI, < 18.5; 18.5–24.9; 25–29.9; ≥ 30 (SC):** 470; 6518; 4074; 2748
		Diabetes (pre-gestational) (GAP implementation; SC): 299; 94
		Diabetes (gestational) (GAP implementation; SC): 707; 713
		Smoking (GAP implementation; SC): 569; 698
		PET/gestational HT (GAP implementation; SC): 216/219; 100/136
		Female neonate (GAP implementation; SC): 5433/11,023; 6745/13,798
Wan[295]	**Inclusion criteria:** Singleton pregnancy; delivered at > 20 WG	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Terminations of pregnancy for fetal structural or chromosomal abnormalities	**Singleton/multiple:** Singleton
Country: Australia	**Research question:** 3	**Parental age, median (IQR) years:** 30.9 (27.3 to 34.4)
Setting: Metropolitan maternity service		**Ethnicity:** NR
Funding: NR		**Parity (nulliparous):** 5317
Study design: Diagnostic accuracy		**BMI, median:** 25
Recruitment: Retrospective		**Diabetes (pre-gestational and gestational):** 748
Number of participants: 13,334		**Smoking:** 917
		PET/gestational HT: NR
		Female neonate: NR
Whitham (2023)[296]	**Inclusion criteria:** Singleton, non-anomalous pregnancy; ongoing beyond the second trimester anatomic survey; delivered at the study centre	**Participant category:** General population
Publication type: Full paper	**Exclusion criteria:** Missing imaging studies	**Singleton/multiple:** Singleton
Country: USA	**Research question:** 3	**Parental age, median (IQR) years:** 30 (26 to 33)

TABLE 17 Baseline study details (*continued*)

Study details	Selection criteria	Participant details
Setting: Tertiary care		**Ethnicity:** White/Caucasian 1084; Black 155; Asian 93; other 255
Funding: Virginia Engineering in Medicine grant		**Parity, median (IQR):** 1 (0–2)
Study design: Diagnostic accuracy		**BMI, median (IQR):** 25.4 (22.3–30.6)
Recruitment: Retrospective		**Diabetes (pre-gestational type I; pre-gestational type II; gestational):** 14; 33; 114
Number of participants: 1587		**Smoking:** current 104; former 222; never 1261
		PET/gestational HT: NR
		Female neonate: NR

CMH, Counties Manukau Health; IQR, interquartile range; LMP, last menstrual period; ME/LA/A, Middle Eastern/Latin American/African; NHMRC, National Health and Medical Research Council; NZMFM, New Zealand Maternal Fetal Medicine; SD, standard deviation; USEFW, US performed for fetal weight estimate; WG, weeks' gestation; yrs, years.
a Primary publication.

Note
Publications in bold have provided data for inclusion in this assessment

DOI: 10.3310/AJLK7403

TABLE 18 Fetal/neonatal and parental outcomes, research question 1

Study (year)	Index test result	*N*	Outcome	Outcome *n*	Mean (SD)
Michaeli (2022)[269]	FP	5671	Spontaneous onset of labour	4343	
Michaeli **PET/gestational HT:** NR **PET/gestational HT:** NR (2022)[269]	FP	5671	Induction of labour	649	
Michaeli (2022)[269]	FP	5671	Elective caesarean delivery	679	
Michaeli (2022)[269]	FP	5671	Spontaneous delivery	4508	
Michaeli (2022)[269]	FP	5671	Instrumental delivery	248	
Michaeli (2022)[269]	FP	5671	Caesarean delivery	915	
Michaeli (2022)[269]	FP	5671	Placental complications	148	
Michaeli (2022)[269]	FP	5671	Haemoglobin drop > 3 g/dl	305	
Michaeli (2022)[269]	FP	5671	Use of blood products	50	
Michaeli (2022)[269]	FP	5671	Puerperal fever	14	
Michaeli (2022)[269]	FP	5671	Maternal re-admission	20	
Michaeli (2022)[269]	FP	5671	BW (g)		3165 (359)
Michaeli (2022)[269]	FP	5671	Apgar 1 minute < 7	124	
Michaeli (2022)[269]	FP	5671	Apgar 5 minutes < 7	22	
Michaeli (2022)[269]	FP	5671	NICU admission	51	
Michaeli (2022)[269]	FP	5671	NICU for 72 hours	27	
Michaeli (2022)[269]	FP	5671	Prolonged neonatal hospitalisation (> 5 days)	592	
Michaeli (2022)[269]	TP	3040	Spontaneous onset of labour	1781	
Michaeli (2022)[269]	TP	3040	Induction of labour	969	
Michaeli (2022)[269]	TP	3040	Elective caesarean delivery	290	
Michaeli (2022)[269]	TP	3040	Spontaneous delivery	2243	
Michaeli (2022)[269]	TP	3040	Instrumental delivery	241	
Michaeli (2022)[269]	TP	3040	Caesarean delivery	556	
Michaeli (2022)[269]	TP	3040	Placental complications	104	

TABLE 18 Fetal/neonatal and parental outcomes, research question 1 (*continued*)

Study (year)	Index test result	N	Outcome	Outcome n	Mean (SD)
Michaeli (2022)[269]	TP	3040	Haemoglobin drop > 3 g/dl	189	
Michaeli (2022)[269]	TP	3040	Use of blood products	36	
Michaeli (2022)[269]	TP	3040	Puerperal fever	12	
Michaeli (2022)[269]	TP	3040	Maternal re-admission	11	
Michaeli (2022)[269]	TP	3040	BW (g)		2496.3 (277)
Michaeli (2022)[269]	TP	3040	Apgar 1 minute < 7	149	
Michaeli (2022)[269]	TP	3040	Apgar 5 minutes < 7	42	
Michaeli (2022)[269]	TP	3040	NICU admission	182	
Michaeli (2022)[269]	TP	3040	NICU for 72 hours	151	
Michaeli (2022)[269]	TP	3040	Prolonged neonatal hospitalisation (> 5 days)	564	
Michaeli (2022)[269]	FN	8508	Spontaneous onset of labour	6955	
Michaeli (2022)[269]	FN	8508	Induction of labour	1055	
Michaeli (2022)[269]	FN	8508	Elective caesarean delivery	498	
Michaeli (2022)[269]	FN	8508	Spontaneous delivery	6674	
Michaeli (2022)[269]	FN	8508	Instrumental delivery	728	
Michaeli (2022)[269]	FN	8508	Caesarean delivery	1106	
Michaeli (2022)[269]	FN	8508	Placental complications	295	
Michaeli (2022)[269]	FN	8508	Haemoglobin drop > 3 g/dl	595	
Michaeli (2022)[269]	FN	8508	Use of blood products	93	
Michaeli (2022)[269]	FN	8508	Puerperal fever	25	
Michaeli (2022)[269]	FN	8508	Maternal re-admission	51	
Michaeli (2022)[269]	FN	8508	BW (g)		2630.94 (234.64)
Michaeli (2022)[269]	FN	8508	Apgar 1 minute < 7	384	
Michaeli (2022)[269]	FN	8508	Apgar 5 minutes < 7	143	
Michaeli (2022)[269]	FN	8508	NICU admission	328	
Michaeli (2022)[269]	FN	8508	NICU for 72 hours	175	

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 18 Fetal/neonatal and parental outcomes, research question 1 (*continued*)

Study (year)	Index test result	*N*	Outcome	Outcome *n*	Mean (SD)
Michaeli (2022)[269]	FN	8508	Prolonged neonatal hospitalisation (> 5 days)	1010	
Michaeli (2022)[269]	TN	82,979	Spontaneous onset of labour	72,234	
Michaeli (2022)[269]	TN	82,979	Induction of labour	6964	
Michaeli (2022)[269]	TN	82,979	Elective caesarean delivery	3782	
Michaeli (2022)[269]	TN	82,979	Spontaneous delivery	71,845	
Michaeli (2022)[269]	TN	82,979	Instrumental delivery	4546	
Michaeli (2022)[269]	TN	82,979	Caesarean delivery	6588	
Michaeli (2022)[269]	TN	82,979	Placental complications	2358	
Michaeli (2022)[269]	TN	82,979	Haemoglobin drop > 3 g/dl	4906	
Michaeli (2022)[269]	TN	82,979	Use of blood products	701	
Michaeli (2022)[269]	TN	82,979	Puerperal fever	188	
Michaeli (2022)[269]	TN	82,979	Maternal re-admission	389	
Michaeli (2022)[269]	TN	82,979	BW (g)		3323.77 (158.66)
Michaeli (2022)[269]	TN	82,979	Apgar 1 minute < 7	1595	
Michaeli (2022)[269]	TN	82,979	Apgar 5 minutes < 7	383	
Michaeli (2022)[269]	TN	82,979	NICU admission	704	
Michaeli (2022)[269]	TN	82,979	NICU for 72 hours	321	
Michaeli (2022)[269]	TN	82,979	Prolonged neonatal hospitalisation (> 5 days)	5139	
Nymark Hansen (2019)[277]	TP	61	Caesarean delivery	26	
Nymark Hansen (2019)[277]	TP	26	Elective caesarean delivery, among all caesarean sections	7	
Nymark Hansen (2019)[277]	TP	61	Intended vaginal delivery	44	
Nymark Hansen (2019)[277]	TP	44	Induction among intended vaginal deliveries	34	
Nymark Hansen (2019)[277]	TP	35	Vacuum among vaginal deliveries	6	
Nymark Hansen (2019)[277]	TP	56	Umbilical artery pH < 7.1	4	

TABLE 18 Fetal/neonatal and parental outcomes, research question 1 (continued)

Study (year)	Index test result	*N*	Outcome	Outcome *n*	Mean (SD)
Nymark Hansen (2019)[277]	TP	60	Apgar score < 7 after 5 minutes	3	
Nymark Hansen (2019)[277]	TP	61	Stillborn	1	
Nymark Hansen (2019)[277]	TP	61	Neonatal death	1	
Nymark Hansen (2019)[277]	TP	61	Composite adverse neonatal outcome (umbilical artery pH < 7.1, Apgar score < 7 after 5 minutes and stillbirth or neonatal death)	8	
Nymark Hansen (2019)[277]	FP	158	Caesarean delivery	37	
Nymark Hansen (2019)[277]	FP	37	Elective caesarean delivery, among all caesarean sections	12	
Nymark Hansen (2019)[277]	FP	158	Intended vaginal delivery	128	
Nymark Hansen (2019)[277]	FP	128	Induction among intended vaginal deliveries	57	
Nymark Hansen (2019)[277]	FP	121	Vacuum among vaginal deliveries	8	
Nymark Hansen (2019)[277]	FP	147	Umbilical artery pH < 7.1	5	
Nymark Hansen (2019)[277]	FP	158	Apgar score < 7 after 5 minutes	1	
Nymark Hansen (2019)[277]	FP	158	Stillborn	0	
Nymark Hansen (2019)[277]	FP	158	Neonatal death	0	
Nymark Hansen (2019)[277]	FP	158	Composite adverse neonatal outcome (umbilical artery pH < 7.1, Apgar score < 7 after 5 minutes, stillbirth or neonatal death)	6	
Nymark Hansen (2019)[277]	FN	37	Caesarean delivery	9	
Nymark Hansen (2019)[277]	FN	9	Elective caesarean delivery, among all caesarean sections	0	
Nymark Hansen (2019)[277]	FN	37	Intended vaginal delivery	31	
Nymark Hansen (2019)[277]	FN	31	Induction among intended vaginal deliveries	10	
Nymark Hansen (2019)[277]	FN	28	Vacuum among vaginal deliveries	2	
Nymark Hansen (2019)[277]	FN	30	Umbilical artery pH < 7.1	2	

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 18 Fetal/neonatal and parental outcomes, research question 1 (*continued*)

Study (year)	Index test result	*N*	Outcome	Outcome *n*	Mean (SD)
Nymark Hansen (2019)[277]	FN	35	Apgar score < 7 after 5 minutes	1	
Nymark Hansen (2019)[277]	FN	37	Stillborn	1	
Nymark Hansen (2019)[277]	FN	37	Neonatal death	0	
Nymark Hansen (2019)[277]	FN	37	Composite adverse neonatal outcome (umbilical artery pH < 7.1, Apgar score< 7 after 5 minutes, stillbirth or neonatal death)	3	
Nymark Hansen (2019)[277]	TN	2672	Caesarean delivery	535	
Nymark Hansen (2019)[277]	TN	535	Elective caesarean delivery, among all caesarean sections	203	
Nymark Hansen (2019)[277]	TN	2672	Intended vaginal delivery	2221	
Nymark Hansen (2019)[277]	TN	2221	Induction among intended vaginal deliveries	631	
Nymark Hansen (2019)[277]	TN	2137	Vacuum among vaginal deliveries	179	
Nymark Hansen (2019)[277]	TN	2443	Umbilical artery pH < 7.1	113	
Nymark Hansen (2019)[277]	TN	2654	Apgar score < 7 after 5 minutes	21	
Nymark Hansen (2019)[277]	TN	2672	Stillborn	9	
Nymark Hansen (2019)[277]	TN	2672	Neonatal death	2	
Nymark Hansen (2019)[277]	TN	2672	Composite adverse neonatal outcome (umbilical artery pH < 7.1, Apgar score < 7 after 5 minutes, stillbirth or neonatal death)	138	

TABLE 19 Fetal/neonatal and parental outcomes, research question 2

Study	Participants/arm	N	Outcome	Outcome n	mean	Comparison measure	Comparison value
Chandra (2022)[238]	Before GAP implementation, 2011–4	26,295	Stillbirth from 28 weeks	162	–	–	–
Chandra (2022)[238]	After GAP implementation, 2015–8	31,624	Stillbirth from 28 weeks	147	–	RR (95% CI)	0.78 (0.60 to 1.02)
Chandra (2022)[238]	Before GAP implementation, 2011–4	26,295	Stillbirth at term	NR	–	–	–
Chandra (2022)[238]	After GAP implementation, 2015–8	31,624	Stillbirth at term	NR	–	RR (95% CI)	0.37 (0.20 to 0.66)
Cowan (2021)[249]	Pre-GAP	1105	Stillbirth	6	–	–	–
Cowan (2021)[249]	Post-GAP	1082	Stillbirth	7	–	–	–
Cowan (2021)[249]	Pre-GAP	1105	Neonatal death	1	–	–	–
Cowan (2021)[249]	Post-GAP	1082	Neonatal death	2	–	–	–
Cowan (2021)[249]	Pre-GAP	1105	Induction of labour	312	–	–	–
Cowan (2021)[249]	Post-GAP	1082	Induction of labour	368	–	–	–
Cowan (2021)[249]	Pre-GAP	1105	Caesarian birth	348	–	–	–
Cowan (2021)[249]	Post-GAP	1082	Caesarian birth	395	–	–	–
Cowan (2021)[249]	Pre-GAP	1105	Pre-term birth	82	–	–	–
Cowan (2021)[249]	Post-GAP	1082	Pre-term birth	104	–	–	–
Cowan (2021)[249]	Pre-GAP	1105	Post-term birth	296	–	–	–
Cowan (2021)[249]	Post-GAP	1082	Post-term birth	241	–	–	–
Hugh (2021)[260]	Non-GAP pre-implementation 2008 (39 units)	204,811	Stillbirths	1049	–	–	–
Hugh (2021)[260]	Non-GAP post-implementation 2017 (39 units)	199,575	Stillbirths	872	–	Difference in still birth rate	−14.60%
Hugh (2021)[260]	Complete GAP pre-implementation 2008 (65 units)	289,539	Stillbirths	1465	–	–	–
Hugh (2021)[260]	Complete GAP post-implementation 2017 (65 units)	274,721	Stillbirths	1095	–	Difference in still birth rate	−21.10%
Hugh (2021)[260]	Complete GAP, top 20 units for SGA detection rate, pre-implementation 2008 (20 units)	74,161	Stillbirths	374	–	–	–
Hugh (2021)[260]	Complete GAP, top 20 units for SGA detection rate, post-implementation 2017 (20 units)	67,315	Stillbirths	224	–	Difference in still birth rate	−24%[a]

continued

DOI: 10.3310/AJLK7403

TABLE 19 Fetal/neonatal and parental outcomes, research question 2 (*continued*)

Study	Participants/arm	N	Outcome	Outcome n	mean	Comparison measure	Comparison value
Jayawardena (2019)[262]	Before GAP implementation	936	Induction of labour	305	–	–	–
Jayawardena (2019)[262]	After GAP implementation	882	Induction of labour	324	–	–	–
Jayawardena (2019)[262]	Before GAP implementation, SGA infants	99	Induction of labour	37	–	–	–
Jayawardena (2019)[262]	After GAP implementation, SGA infants	83	Induction of labour	53	–	–	–
Jayawardena (2019)[262]	Before GAP implementation, non-SGA infants	837	Induction of labour	268	–	–	–
Jayawardena (2019)[262]	After GAP implementation, non-SGA infants	799	Induction of labour	271	–	–	–
Jayawardena (2019)[262]	Before GAP implementation	936	Emergency caesarean	125	–	–	–
Jayawardena (2019)[262]	After GAP implementation	882	Emergency caesarean	111	–	–	–
Jayawardena (2019)[262]	Before GAP implementation, SGA infants	99	Emergency caesarean	23	–	–	–
Jayawardena (2019)[262]	After GAP implementation, SGA infants	83	Emergency caesarean	15	–	–	–
Jayawardena (2019)[262]	Before GAP implementation, non-SGA infants	837	Emergency caesarean	102	–	–	–
Jayawardena (2019)[262]	After GAP implementation, non-SGA infants	799	Emergency caesarean	96	–	–	–
Jayawardena (2019)[262]	Before GAP implementation	936	Elective caesarean	140	–	–	–
Jayawardena (2019)[262]	After GAP implementation	882	Elective caesarean	139	–	–	–
Jayawardena (2019)[262]	Before GAP implementation, SGA infants	99	Elective caesarean	21	–	–	–
Jayawardena (2019)[262]	After GAP implementation, SGA infants	83	Elective caesarean	4	–	–	–
Jayawardena (2019)[262]	Before GAP implementation, non-SGA infants	837	Elective caesarean	119	–	–	–

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 19 Fetal/neonatal and parental outcomes, research question 2 (*continued*)

Study	Participants/arm	*N*	Outcome	Outcome *n*	mean	Comparison measure	Comparison value
Jayawardena (2019)[262]	After GAP implementation, non-SGA infants	799	Elective caesarean	135	–	–	–
Jayawardena (2019)[262]	Before GAP implementation	936	5-minutes Apgar score < 7	18	–	–	–
Jayawardena (2019)[262]	After GAP implementation	882	5-minutes Apgar score < 7	14	–	–	–
Jayawardena (2019)[262]	Before GAP implementation	936	Admission to NISC	100	–	–	–
Jayawardena (2019)[262]	After GAP implementation	882	Admission to NISC	75	–	–	–
Jayawardena (2019)[262]	Before GAP implementation, SGA infants	99	Admission to NISC	23	–	–	–
Jayawardena (2019)[262]	After GAP implementation, SGA infants	83	Admission to NISC	14	–	–	–
Jayawardena (2019)[262]	Before GAP implementation, non-SGA infants	837	Admission to NISC	77	–	–	–
Jayawardena (2019)[262]	After GAP implementation, non-SGA infants	799	Admission to NISC	61	–	–	–
Jayawardena (2019)[262]	Before GAP implementation	936	Perinatal death	1	–	–	–
Jayawardena (2019)[262]	After GAP implementation	882	Perinatal death	2	–	–	–
Jayawardena (2019)[262]	Before GAP implementation, SGA infants	99	Perinatal death	1	–	–	–
Jayawardena (2019)[262]	After GAP implementation, SGA infants	83	Perinatal death	1	–	–	–
Jayawardena (2019)[262]	Before GAP implementation, non-SGA infants	837	Perinatal death	0	–	–	–
Jayawardena (2019)[262]	After GAP implementation, non-SGA infants	799	Perinatal death	1	–	–	–
Jayawardena (2019)[262]	Before GAP implementation	936	GA at birth (weeks)	NR	39	–	–
Jayawardena (2019)[262]	After GAP implementation	882	GA at birth (weeks)	NR	39	–	–

continued

TABLE 19 Fetal/neonatal and parental outcomes, research question 2 (*continued*)

Study	Participants/arm	*N*	Outcome	Outcome *n*	mean	Comparison measure	Comparison value
Jayawardena (2019)[262]	Before GAP implementation, SGA infants	99	GA at birth (weeks)	NR	38.2	–	–
Jayawardena (2019)[262]	After GAP implementation, SGA infants	83	GA at birth (weeks)	NR	38.7	–	–
Jayawardena (2019)[262]	Before GAP implementation, non-SGA infants	837	GA at birth (weeks)	NR	39.1	–	–
Jayawardena (2019)[262]	After GAP implementation, non-SGA infants	799	GA at birth (weeks)	NR	39	–	–
Jayawardena (2019)[262]	Before GAP implementation	936	BW (g)	NR	3367.6	–	–
Jayawardena (2019)[262]	After GAP implementation	882	BW (g)	NR	3381.7	–	–
Jayawardena (2019)[262]	Before GAP implementation, non-SGA infants	837	BW (g)	NR	2554.7	–	–
Jayawardena (2019)[262]	After GAP implementation, non-SGA infants	799	BW (g)	NR	2649.73	–	–
Jayawardena (2019)[262]	Before GAP implementation	936	BW (g)	NR	3463.8	–	–
Jayawardena (2019)[262]	After GAP implementation	882	BW (g)	NR	3457.6	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Induction of labour	3273	–	Adjusted effect size (95% CI)	1.7 (−0.4 to 3.8)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Induction of labour	3715	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Spontaneous vaginal delivery	5992	–	Adjusted effect size (95% CI)	−0.1 (−2.6 to 2.4)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Spontaneous vaginal delivery	7526	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Operative vaginal delivery	1598	–	Adjusted effect size (95% CI)	−0.1 (−1.6 to 1.4)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Operative vaginal delivery	1947	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Elective caesarean section	1620	–	Adjusted effect size (95% CI)	−0.6 (−1.5 to 0.4)[b]

DOI: 10.3310/AJLK7403

TABLE 19 Fetal/neonatal and parental outcomes, research question 2 (*continued*)

Study	Participants/arm	*N*	Outcome	Outcome *n*	mean	Comparison measure	Comparison value
Vieira (2022)[294]	SC	13,810 (6 clusters)	Elective caesarean section	1920	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Emergency caesarean section	1853	–	Adjusted effect size (95% CI)	0.6 (−1.6 to 2.8)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Emergency caesarean section	2375	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Postpartum haemorrhage (> 1500 mls)	277	–	Adjusted effect size (95% CI)	−0.1 (−0.5 to 0.3)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Postpartum haemorrhage (> 1500 mls)	373	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Third-/fourth-degree tears	200	–	Adjusted effect size (95% CI)	−0.1 (−0.6 to 0.4)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Third-/fourth-degree tears	262	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Epidural	3129	–	Adjusted effect size (95% CI)	5.6 (−1.4 to 0.4)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Epidural	5027	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Episiotomy	2419	–	Adjusted effect size (95% CI)	−2.3 (−6.4 to 1.9)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Episiotomy	2431	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	GA at birth (weeks)	NR	39.4 (SD 2)	Adjusted effect size (95% CI)	0.0 (−0.1 to 0.1)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	GA at birth (weeks)	NR	39.4 (SC 1.9)	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Pre-term birth (< 37 weeks)	710	–	Adjusted effect size (95% CI)	0.0 (−0.8 to 0.9)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Pre-term birth (< 37 weeks)	842	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	BW (g)	NR	3297 (SD 567)	Adjusted effect size (95% CI)	−7.7 (−21.9 to 6.4)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	BW (g)	NR	3326 (SD 552)	–	–

continued

TABLE 19 Fetal/neonatal and parental outcomes, research question 2 (*continued*)

Study	Participants/arm	*N*	Outcome	Outcome *n*	mean	Comparison measure	Comparison value
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	5-minute Apgar score < 7	189	–	Adjusted effect size (95% CI)	−0.2 (−0.4 to 0.1)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	5-minute Apgar score < 7	304	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Arterial cord pH < 7.1	322	–	Adjusted effect size (95% CI)	0.3 (−0.4 to 1.0)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Arterial cord pH < 7.1	276	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Respiratory support at birth	533	–	Adjusted effect size (95% CI)	−1.0 (−2.7 to 0.7)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Respiratory support at birth	566	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Neonatal unit admission (including HDU and SCBU)	821	–	Adjusted effect size (95% CI)	0.4 (−0.8 to 1.7)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Neonatal unit admission (including HDU and SCBU)	2237	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Any major neonatal morbidity	522	–	Adjusted effect size (95% CI)	−1.2 (−3.4 to 1.0)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Any major neonatal morbidity	760	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Any neonatal brain injury (HIE + IVH)	38	–	–	–
Vieira (2022)[294]	SC	13,810 (6 clusters)	Any neonatal brain injury (HIE + IVH)	57	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Supplementary O2 > 28 days	17	–	–	–
Vieira (2022)[294]	SC	13,810 (6 clusters)	Supplementary O2 > 28 days	12	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Necrotising enterocolitis	9	–	–	–
Vieira (2022)[294]	SC	13,810 (6 clusters)	Necrotising enterocolitis	17	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Sepsis	510	–	–	–

TABLE 19 Fetal/neonatal and parental outcomes, research question 2 (*continued*)

Study	Participants/arm	*N*	Outcome	Outcome *n*	mean	Comparison measure	Comparison value
Vieira (2022)[294]	SC	13,810 (6 clusters)	Sepsis	742	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Retinopathy of prematurity	7	–	–	–
Vieira (2022)[294]	SC	13,810 (6 clusters)	Retinopathy of prematurity	23	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Any minor neonatal morbidity	333	–	Adjusted effect size (95% CI)	−0.0 (−1.6 to 1.5)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Any minor neonatal morbidity	359	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Hypothermia	16	–	–	–
Vieira (2022)[294]	SC	13,810 (6 clusters)	Hypothermia	23	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Hypoglycaemia	95	–	–	–
Vieira (2022)[294]	SC	13,810 (6 clusters)	Hypoglycaemia	164	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Nasogastric feeding	291	–	–	–
Vieira (2022)[294]	SC	13,810 (6 clusters)	Nasogastric feeding	273	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Stillbirth	34	–	Adjusted effect size (95% CI)	−0.07 (−0.14 to −0.01)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Stillbirth	50	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Neonatal death	8	–	Adjusted effect size (95% CI)	−0.02 (−0.08 to 0.04)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Neonatal death	6	–	–	–
Vieira (2022)[294]	GAP implementation	11,096 (5 clusters)	Perinatal mortality	41	–	Adjusted effect size (95% CI)	−0.09 (−0.17 to −0.004)[b]
Vieira (2022)[294]	SC	13,810 (6 clusters)	Perinatal mortality	57	–	–	–

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 19 Fetal/neonatal and parental outcomes, research question 2 (*continued*)

Study	Participants/arm	N	Outcome	Outcome n	mean	Comparison measure	Comparison value
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Induction of labour	297	–	Adjusted effect size (95% CI)	3.6 (−2.6 to 9.8)[c]
Vieira (2022)[294]	SC	995	Induction of labour	349	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Spontaneous vaginal delivery	393	–	Adjusted effect size (95% CI)	−3.3 (−7.4 to 0.7)[c]
Vieira (2022)[294]	SC	995	Spontaneous vaginal delivery	533	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Operative vaginal delivery	130	–	Adjusted effect size (95% CI)	0.6 (−3.0 to 4.2)[c]
Vieira (2022)[294]	SC	995	Operative vaginal delivery	131	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Elective Caesarean section	76	–	Adjusted effect size (95% CI)	−1.7 (−5.1 to 1.6)[c]
Vieira (2022)[294]	SC	995	Elective caesarean section	105	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Emergency caesarean section	204	–	Adjusted effect size (95% CI)	2.4 (−0.9 to 5.8)[c]
Vieira (2022)[294]	SC	995	Emergency caesarean section	213	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Postpartum haemorrhage (> 1500 ml)	14	–	Adjusted effect size (95% CI)	0.4 (−0.5 to 1.3)[c]
Vieira (2022)[294]	SC	995	Postpartum haemorrhage (> 1500 ml)	13	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Third-/fourth-degree tears	11	–	Adjusted effect size (95% CI)	1.1 (−0.1 to 2.2)[c]
Vieira (2022)[294]	SC	995	Third-/fourth-degree tears	8	–	–	–

TABLE 19 Fetal/neonatal and parental outcomes, research question 2 (*continued*)

Study	Participants/arm	*N*	Outcome	Outcome *n*	mean	Comparison measure	Comparison value
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Epidural	237	–	Adjusted effect size (95% CI)	−0.8 (−6.7 to 5.0)[c]
Vieira (2022)[294]	SC	995	Epidural	363	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Episiotomy	185	–	Adjusted effect size (95% CI)	−4.0 (−8.9 to 0.9)[c]
Vieira (2022)[294]	SC	995	Episiotomy	181	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	GA at birth (weeks)	NR	38.6 (SD 3.1)	Adjusted effect size (95% CI)	−0.3 (−0.5 to −0.1)[c]
Vieira (2022)[294]	SC	995	GA at birth (weeks)	NR	38.8 (SD 3)	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Pre-term birth (< 37 weeks)	133		Adjusted effect size (95% CI)	2.3 (−1.5 to 6.2)[c]
Vieira (2022)[294]	SC	995	Pre-term birth (< 37 weeks)	136	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	BW (g)	NR	2436 (SD 550)	Adjusted effect size (95% CI)	−58 (−99 to −18)[c]
Vieira (2022)[294]	SC	995	BW (g)	NR	2482 (SD 550)	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	5 minute Apgar score < 7	33	–	Adjusted effect size (95% CI)	−0.5 (−1.7 to 0.8)[c]
Vieira (2022)[294]	SC	995	5-minute Apgar score < 7	52	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Arterial cord pH < 7.1	30	–	Adjusted effect size (95% CI)	−0.3 (−1.4 to 0.8)[c]

continued

DOI: 10.3310/AJLK7403

TABLE 19 Fetal/neonatal and parental outcomes, research question 2 (*continued*)

Study	Participants/arm	*N*	Outcome	Outcome *n*	mean	Comparison measure	Comparison value
Vieira (2022)[294]	SC	995	Arterial cord pH < 7.1	28	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Respiratory support at birth	65	–	Adjusted effect size (95% CI)	−1.3 (−4.4 to 1.8)[c]
Vieira (2022)[294]	SC	995	Respiratory support at birth	71	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Neonatal unit admission (including HDU and SCBU)	121	–	Adjusted effect size (95% CI)	1.5 (−2.4 to 5.4)[c]
Vieira (2022)[294]	SC	995	Neonatal unit admission (including HDU and SCBU)	228	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Any major neonatal morbidity	74	–	Adjusted effect size (95% CI)	0.5 (−3.2 to 4.2)[c]
Vieira (2022)[294]	SC	995	Any major neonatal morbidity	84	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Any neonatal brain injury (HIE + IVH)	9	–	–	–
Vieira (2022)[294]	SC	995	Any neonatal brain injury (HIE + IVH)	9	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Supplementary O_2 > 28 days	4	–	–	–
Vieira (2022)[294]	SC	995	Supplementary O_2 > 28 days	4	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Necrotising enterocolitis	3	–	–	–
Vieira (2022)[294]	SC	995	Necrotising enterocolitis	2	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Sepsis	72	–	–	–
Vieira (2022)[294]	SC	995	Sepsis	82	–	–	–

TABLE 19 Fetal/neonatal and parental outcomes, research question 2 (*continued*)

Study	Participants/arm	N	Outcome	Outcome n	mean	Comparison measure	Comparison value
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Retinopathy of prematurity	1	–	–	–
Vieira (2022)[294]	SC	995	Retinopathy of prematurity	2	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Any minor neonatal morbidity	71	–	Adjusted effect size (95% CI)	1.9 (−3.1 to 6.9)[c]
Vieira (2022)[294]	SC	995	Any minor neonatal morbidity	63	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Hypothermia	8	–	–	–
Vieira (2022)[294]	SC	995	Hypothermia	9	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Hypoglycaemia	26	–	–	–
Vieira (2022)[294]	SC	995	Hypoglycaemia	34	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Nasogastric feeding	60	–	–	–
Vieira (2022)[294]	SC	995	Nasogastric feeding	52	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Stillbirth	11	–	Adjusted effect size (95% CI)	−0.76 (−1.50 to −0.03)[c]
Vieira (2022)[294]	SC	995	Stillbirth	22	–	–	–
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Neonatal death	3	–	Adjusted effect size (95% CI)	−0.11 (−0.60 to 0.38)[c]
Vieira (2022)[294]	SC	995	Neonatal death	2	–	–	–

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 19 Fetal/neonatal and parental outcomes, research question 2 (*continued*)

Study	Participants/arm	*N*	Outcome	Outcome *n*	mean	Comparison measure	Comparison value
Vieira (2022)[294]	GAP implementation; outcomes among SGA babies by population and customised according to treatment allocation (mITT analysis)	807	Perinatal mortality	14	–	Adjusted effect size (95% CI)	−0.69 (−1.47 to 0.09)[c]
Vieira (2022)[294]	SC	995	Perinatal mortality	24	–	–	–

a Mixed-effects logistic regression analysis, adjusted for age and depravation score, showed that GAP implementation was associated with a reduction in stillbirths [OR 0.89 (95% CI 0.79 to 1.00)], with a stronger association in units with the highest SGA detection rates [OR 0.76 (95% CI 0.64 to 0.90)].
b Adjusted for age, ethnicity, parity and stratification factor.
c Adjusted for baseline, age, ethnicity, parity, and stratification factor.

TABLE 20 Accuracy data

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Agrawal (2016)[223]	Singleton; Uns.	Single; 2nd or 3rd; TCD/AC US	> 90th PCTL	Local; Agrawal (2016)[223]	< 10th PCTL	General; Lubchenco (1963)[297]	9	5	6	80	60 (32 to 84)	94 (87 to 98)
Agrawal (2016)[223]	Singleton; Uns.	Single; 3rd; TCD/AC US	> 90th PCTL	Local; Agrawal (2016)[223]	< 10th PCTL	General; Lubchenco (1963)[297]	12	8	3	77	80 (52 to 96)	91 (82 to 96)
Badr (2023)[224]	Singleton; Uns.	Single; 3rd; EFW MRI	< 10th PCTL	General; FMF[298]	< 10th PCTL	General; FMF[298]	204	36	144	1994	59 (53 to 64)	98 (98 to 99)
Badr (2023)[224]	Singleton; Uns.	Single; 3rd; EFW MRI	< 10th PCTL	General; INTERGROWTH-21st[299]	< 10th PCTL	General; INTERGROWTH-21st to Villar (2014)[299]	27	0	85	2266	24 (17 to 33)	100 (100 to 100)
Badr (2023)[224]	Singleton; Uns.	Single; 3rd; EFW MRI	< 10th PCTL	General; WHO, Kiserud (2017)[4]	< 10th PCTL	General; WHO, Kiserud (2017)[4]	60	11	89	2218	40 (32 to 49)	100 (99 to 100)
Badr (2023)[224]	Singleton; Uns.	Single; 3rd; EFW MRI	< 5th PCTL	General; FMF[298]	< 5th PCTL	General; FMF[298]	124	36	74	2144	63 (55 to 69)	98 (98 to 99)
Badr (2023)[224]	Singleton; Uns.	Single; 3rd; EFW MRI	< 5th PCTL	General; INTERGROWTH-21st[299]	< 5th PCTL	General; INTERGROWTH-21st to Villar (2014)[299]	13	0	46	2319	22 (12 to 35)	100 (100 to 100)
Badr (2023)[224]	Singleton; Uns.	Single; 3rd; EFW MRI	< 5th PCTL	General; WHO, Kiserud (2017)[4]	< 5th PCTL	General; WHO, Kiserud (2017)[4]	18	4	76	2280	19 (12 to 29)	100 (100 to 100)
Badr (2023)[224]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; FMF[298]	< 10th PCTL	General; FMF[298]	244	229	104	1801	70 (65 to 75)	89 (87 to 90)
Badr (2023)[224]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; WHO, Kiserud (2017)[4]	< 10th PCTL	General; WHO, Kiserud (2017)[4]	42	42	70	2224	38 (29 to 47)	98 (98 to 99)
Badr (2023)[224]	Singleton; Uns.	Single; 3rd; EFW US	< 5th PCTL	General; FMF[298]	< 5th PCTL	General; FMF[298]	64	123	85	2106	43 (35 to 51)	94 (93 to 95)
Badr (2023)[224]	Singleton; Uns.	Single; 3rd; EFW US	< 5th PCTL	General; INTERGROWTH-21st[299]	< 5th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	140	175	58	2005	71 (64 to 77)	92 (91 to 93)
Badr (2023)[224]	Singleton; Uns.	Single; 3rd; EFW US	< 5th PCTL	General; WHO, Kiserud (2017)[4]	< 5th PCTL	General; WHO, Kiserud (2017)[4]	21	20	38	2299	36 (24 to 49)	99 (99 to 99)

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Badr (2023)[224]	Uns.	Single; 3rd; EFW US	< 10th PCTL	General; INTERGROWTH-21st[299]	< 10th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	28	36	66	2248	30 (21 to 40)	98 (98 to 99)
Baird (2016)[225]	Singleton; high risk	Single; 3rd; EFW US	< 10th PCTL	General ethnicity-specific; GROW (with ethnicity), Gardosi (2020)[300]	< 10th PCTL	Local; Australian reference chart, Dobbins (2012)[301]	68	21	10	8	87 (78 to 94)	28 (13 to 47)
Baird (2016)[225]	Singleton; high risk	Single; 3rd; EFW US	< 10th PCTL	General; GROW (no ethnicity), Gardosi (2020)[300]	< 10th PCTL	Local; Australian reference chart, Dobbins (2012)[301]	69	21	9	8	88 (79 to 95)	28 (13 to 47)
Baird (2016)[225]	Singleton; high risk	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; Australian reference chart, Dobbins (2012)[301]	60	16	18	13	77 (66 to 86)	45 (26 to 64)
Baird (2016)[225]	Singleton; high risk	Single; 3rd; EFW US	< 10th PCTL	Local; Dobbins (2012)[301]	< 10th PCTL	Local; Australian reference chart, Dobbins (2012)[301]	53	9	25	20	68 (56 to 78)	69 (49 to 85)
Bais (2004)[227]	Singleton; low risk	Single; 1st or 2nd; abdominal palpation	N/A	N/A	< 10th PCTL	Local; Amsterdam growth chart, Bleker (2006)[302]	114	238	422	5545	21 (18 to 25)	96 (95 to 96)
Bais (2004)[227]	Singleton; low risk	Single; 1st or 2nd; abdominal palpation	N/A	N/A	< 2.3rd PCTL	Local; Amsterdam growth chart, Bleker (2006)[302]	26	324	67	5901	28 (19 to 38)	95 (94 to 95)
Barreto (2004)[228]	Singleton; high risk	Single; unclear; TCD/AC US	≥ 16.1	N/A	< 10th PCTL	Sex-specific; Lubchenco (1963)[297]	24	38	7	181	77 (59 to 90)	83 (77 to 87)
Bastek (2009)[229]	NR; PET	NR; unclear; EFW US	< 10th PCTL	Local; Alexander (1996)[303]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	17	4	13	59	57 (37 to 75)	94 (85 to 98)
Ben-Haroush (2007)[230]	Singleton; low risk	Single; 3rd; EFW US	< 10th PCTL	Local; Leiberman (1993)[304]	< 10th PCTL	Local; Israeli growth chart, Leiberman (1993)[304]	4	8	15	232	21 (6 to 46)	97 (94 to 99)
Ben-Haroush (2007)[230]	Singleton; low risk	Single; 3rd; EFW US	< 35th PCTL	Local; Leiberman (1993)[304]	< 10th PCTL	Local; Israeli growth chart, Leiberman (1993)[304]	8	24	11	216	42 (20 to 67)	90 (85 to 93)

DOI: 10.3310/AJLK7403

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Bergman (2022)[231]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL on neonatal BW curve alone	Local; Dollberg (2005)[305]	< 10th PCTL	Local; Israeli growth curve, Dollberg (2005)[305]	74	5	151	89	33 (27 to 39)	95 (88 to 98)
Bergman (2022)[231]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL on US EFW curve or neonatal BW curve	Local; Sapir (2017), Dollberg (2005)[305,306]	< 10th PCTL	Local; Israeli growth curve, Dollberg (2005)[305]	159	27	185	248	46 (41 to 52)	90 (86 to 93)
Blue (2018)[234]	Singleton; Uns.	Serial; 2nd and unclear; EFW US	< 10th PCTL at 18–22 weeks and < 10th PCTL at next exam	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	8	11	128	1079	6 (3 to 11)	99 (98 to 99)
Blue (2018)[233]	Singleton; high risk	Single; 3rd; AC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local sex-specific; US growth curve, Alexander (1996)[303]	138	116	97	1353	59 (52 to 65)	92 (91 to 93)
Blue (2018)[233]	Singleton; high risk	Single; 3rd; AC US	< 5th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local sex-specific; US growth curve, Alexander (1996)[303]	110	68	125	1401	47 (40 to 53)	95 (94 to 96)
Blue (2018)[233]	Singleton; high risk	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local sex-specific; US growth curve, Alexander (1996)[303]	138	94	97	1375	59 (52 to 65)	94 (92 to 95)
Blue (2018)[233]	Singleton; high risk	Single; 3rd; EFW US and AC US	AC < 10th PCTL OR EFW < 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local sex-specific; US growth curve, Alexander (1996)[303]	155	134	80	1335	66 (60 to 72)	91 (89 to 92)
Blue (2018)[233]	Singleton; high risk	Single; 3rd; EFW US and AC US	AC < 5th PCTL OR EFW < 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local sex-specific; US growth curve, Alexander (1996)[303]	146	107	89	1362	62 (56 to 68)	93 (91 to 94)
Blue (2018)[233]	Singleton; high risk	Single; 3rd; EFW US and AC US	AC < 10th PCTL AND EFW < 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local sex-specific; US growth curve, Alexander (1996)[303]	121	76	114	1393	52 (45 to 58)	95 (94 to 96)

continued

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Blue (2018)[233]	Singleton; high risk	Single; 3rd; EFW US and AC US	AC < 10th PCTL AND normal EFW	General; Hadlock (1991)[30]	< 10th PCTL	Local sex-specific; US growth curve, Alexander (1996)[303]	17	40	218	1429	7 (4 to 11)	97 (96 to 98)
Blue (2018)[233]	Singleton; high risk	Single; 3rd; EFW US and AC US	AC < 5th PCTL AND normal EFW	General; Hadlock (1991)[30]	< 10th PCTL	Local sex-specific; US growth curve, Alexander (1996)[303]	8	13	227	1456	3 (1 to 7)	99 (98 to 100)
Blue (2019)[232]	Singleton; high risk	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	98	56	40	637	71 (63 to 78)	92 (90 to 94)
Blue (2019)[232]	Singleton; high risk	Single; 3rd; EFW US	< 10th PCTL	Local; INTERGROWTH-21st[23]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	80	41	58	652	58 (49 to 66)	94 (92 to 96)
Blue (2019)[232]	Singleton; high risk	Single; 3rd; EFW US	< 10th PCTL	Local; Salomon (2007)[20]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	59	26	79	667	43 (34 to 52)	96 (95 to 98)
Blue (2019)[232]	Singleton; high risk	Single; 3rd; EFW US	< 5th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	76	35	62	658	55 (46 to 64)	95 (93 to 97)
Blue (2019)[232]	Singleton; high risk	Single; 3rd; EFW US	< 5th PCTL	Local; INTERGROWTH-21st[23]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	64	28	74	665	46 (38 to 55)	96 (94 to 97)
Blue (2019)[232]	Singleton; high risk	Single; 3rd; EFW US	< 5th PCTL	Local; Salomon (2007)[20]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	47	15	91	678	34 (26 to 43)	98 (96 to 99)
Blue (2019)[232]	Singleton; high risk	Single; 3rd; EFW US	< 15th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	116	97	22	596	84 (77 to 90)	86 (83 to 88)
Blue (2019)[232]	Singleton; high risk	Single; 3rd; EFW US	< 22nd PCTL	Local; INTERGROWTH-21st[23]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	114	92	24	601	83 (75 to 89)	87 (84 to 89)
Blue (2019)[232]	Singleton; high risk	Single; 3rd; EFW US	< 53rd PCTL	Local; Salomon (2007)[20]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	102	173	36	520	74 (66 to 81)	75 (72 to 78)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score 0.25	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	1652	33,577	26	24,197	98 (98 to 99)	42 (41 to 42)

TABLE 20 Accuracy data (*continued*)

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score 0.00	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	1621	26,499	57	31,275	97 (96 to 97)	54 (54 to 55)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score −0.25	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	1574	21,547	104	36,227	94 (93 to 95)	63 (62 to 63)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score 0.50	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	1668	40,042	10	17,732	99 (99 to 100)	31 (30 to 31)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score −0.50	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	1463	14,584	215	43,190	87 (85 to 89)	75 (74 to 75)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score 0.75	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	1674	45,418	4	12,356	100 (99 to 100)	21 (21 to 22)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score −0.75	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	1322	9274	356	48,500	79 (77 to 81)	84 (84 to 84)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score 1.00	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	1676	48,322	2	9452	100 (100 to 100)	16 (16 to 17)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score −1.00	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	1129	5515	549	52,259	67 (65 to 70)	90 (90 to 91)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score 1.25	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	1676	51,713	2	6061	100 (100 to 100)	10 (10 to 11)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score −1.25	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	967	3610	711	54,164	58 (55 to 60)	94 (94 to 94)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score 1.50	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	1676	53,952	2	3822	100 (100 to 100)	7 (6 to 7)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score −1.50	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	728	1699	950	56,075	43 (41 to 46)	97 (97 to 97)

continued

DOI: 10.3310/AJLK7403

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score 1.75	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	1678	55,492	0	2282	100 (100 to 100)	4 (4 to 4)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score −1.75	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	508	720	1170	57,054	30 (28 to 33)	99 (99 to 99)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score 2.00	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	1678	56,219	0	1555	100 (100 to 100)	3 (3 to 3)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score −2.00	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	345	283	1333	57,491	21 (19 to 23)	100 (99 to 100)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score −2.25	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	250	142	1428	57,632	15 (13 to 17)	100 (100 to 100)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score −2.50	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	164	36	1514	57,738	10 (8 to 11)	100 (100 to 100)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score −2.75	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	98	11	1580	57,763	6 (5 to 7)	100 (100 to 100)
Bonnevier (2022)[235]	Singleton; Uns.	Single; 3rd; EFW US	*z*-score −3.00	Local; Marsál (1996)[307]	*z*-score < −2.0	Local; Swedish growth curve, Maršál (1996)[307]	61	2	1617	57,772	4 (3 to 5)	100 (100 to 100)
Carbone (2012)[236]	Singleton; Uns.	Single; 1st; CRL US	*z*-score < −1	N/A	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	44	453	141	2691	24 (18 to 31)	86 (84 to 87)
Carbone (2012)[236]	Singleton; Uns.	Single; 1st; CRL US	*z*-score < −1	N/A	< 5th PCTL	Local; US growth curve, Alexander (1996)[303]	58	459	127	2685	31 (25 to 39)	85 (84 to 87)
Chang (2005)[245]	Singleton; Uns.	Single; 3rd; UAV US	< 10th PCTL	Local; Chang (2002)[308]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	39	32	1	410	98 (87 to 100)	93 (90 to 95)
Chang (2005)[245]	Singleton; Uns.	Single; 3rd; UAV US	< 5th PCTL	Local; Chang (2002)[308]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	8	17	32	425	20 (9 to 36)	96 (94 to 98)

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Chang (2005)[245]	Singleton; Uns.	Single; 3rd; UAV US	< 25th PCTL	Local; Chang (2002)[308]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	40	59	0	383	100 (93 to 100)	87 (83 to 90)
Chang (2005)[245]	Singleton; Uns.	Single; 3rd; UAV US	< 50th PCTL	Local; Chang (2002)[308]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	40	159	0	283	100 (93 to 100)	64 (59 to 69)
Chang (2005)[244]	Singleton; Uns.	Single; 3rd; TV US	< 10th PCTL	Local; Chang (2003)[310]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	26	25	4	257	87 (69 to 96)	91 (87 to 94)
Chang (2005)[244]	Singleton; Uns.	Single; 3rd; TV US	< 5th PCTL	Local; Chang (2003)[310]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	4	16	26	266	13 (4 to 31)	94 (91 to 97)
Chang (2005)[244]	Singleton; Uns.	Single; 3rd; TV US	< 25th PCTL	Local; Chang (2003)[310]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	30	56	0	226	100 (90, 100)	80 (75 to 85)
Chang (2005)[244]	Singleton; Uns.	Single; 3rd; TV US	< 50th PCTL	Local; Chang (2003)[310]	< 10th PCTL	Local; Taiwanese growth curve to Hsieh (1991)[309]	30	129	0	153	100 (90 to 100)	54 (48 to 60)
Chang (2006)[243]	Singleton; Uns.	Single; 3rd; HV US	< 10th PCTL	Local; Chang (2003)[311]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	41	33	1	225	98 (87 to 100)	87 (83 to 91)
Chang (2006)[243]	Singleton; Uns.	Single; 3rd; HV US	< 5th PCTL	Local; Chang (2003)[311]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	15	17	27	241	36 (22 to 52)	93 (90 to 96)
Chang (2006)[243]	Singleton; Uns.	Single; 3rd; HV US	< 25th PCTL	Local; Chang (2003)[311]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	42	69	0	189	100 (93 to 100)	73 (67 to 79)
Chang (2006)[243]	Singleton; Uns.	Single; 3rd; HV US	< 50th PCTL	Local; Chang (2003)[311]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	42	121	0	137	100 (93 to 100)	53 (47 to 59)
Chang (2006)[242]	Singleton; Uns.	Single; 3rd; liver volume US	< 10th PCTL	Local; Chang (2003)[312]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	41	24	1	351	98 (87 to 100)	94 (91 to 96)
Chang (2006)[242]	Singleton; Uns.	Single; 3rd; liver volume US	< 5th PCTL	Local; Chang (2003)[312]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	36	5	6	378	86 (71 to 95)	99 (97 to 100)

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Chang (2006)[242]	Singleton; Uns.	Single; 3rd; liver volume US	< 25th PCTL	Local; Chang (2003)[312]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	42	85	0	290	100 (93 to 100)	77 (73 to 81)
Chang (2006)[242]	Singleton; Uns.	Single; 3rd; liver volume US	< 50th PCTL	Local; Chang (2003)[312]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	42	232	0	143	100 (93 to 100)	38 (33 to 43)
Chang (2007)[241]	Singleton; Uns.	Single; 3rd; FV US	< 10th PCTL	Local; Chang (2007)[241]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	30	18	12	286	71 (55 to 84)	94 (91 to 96)
Chang (2007)[241]	Singleton; Uns.	Single; 3rd; FV US	< 5th PCTL	Local; Chang (2007)[241]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	15	14	27	290	36 (22 to 52)	95 (92 to 97)
Chang (2007)[241]	Singleton; Uns.	Single; 3rd; FV US	< 25th PCTL	Local; Chang (2007)[241]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	40	67	2	237	95 (84 to 99)	78 (73 to 82)
Chang (2007)[241]	Singleton; Uns.	Single; 3rd; FV US	< 50th PCTL	Local; Chang (2007)[241]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	42	124	0	180	100 (93 to 100)	59 (53 to 65)
Chang (2008)[240]	Singleton; Uns.	Single; 3rd; RV US	< 10th PCTL	Local; Yu (2000)[313]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	27	9	1	212	96 (82 to 100)	96 (92 to 98)
Chang (2008)[240]	Singleton; Uns.	Single; 3rd; RV US	< 5th PCTL	Local; Yu (2000)[313]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	2	3	26	218	7 (1 to 24)	99 (96 to 100)
Chang (2008)[240]	Singleton; Uns.	Single; 3rd; RV US	< 25th PCTL	Local; Yu (2000)[313]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	28	41	0	180	100 (90 to 100)	81 (76 to 86)
Chang (2008)[240]	Singleton; Uns.	Single; 3rd; RV US	< 50th PCTL	Local; Yu (2000)[313]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	28	130	0	91	100 (90 to 100)	41 (35 to 48)
Chang (2011)[239]	Singleton; Uns.	Single; 3rd; st volume of upper arm US	< 10th PCTL	Local; Chang (2011)[239]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	37	23	7	196	84 (70 to 93)	89 (85 to 93)
Chang (2011)[239]	Singleton; Uns.	Single; 3rd; st volume of upper arm US	< 5th PCTL	Local; Chang (2011)[239]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	37	14	7	205	84 (70 to 93)	94 (90 to 96)

DOI: 10.3310/AJLK7403

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Chang (2011)[239]	Singleton; Uns.	Single; 3rd; st volume of upper arm US	< 25th PCTL	Local; Chang (2011)[239]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	39	54	5	165	89 (75 to 96)	75 (69 to 81)
Chang (2011)[239]	Singleton; Uns.	Single; 3rd; st volume of upper arm US	< 50th PCTL	Local; Chang (2011)[239]	< 10th PCTL	Local; Taiwanese growth curve, Hsieh (1991)[309]	42	111	2	108	95 (85 to 99)	49 (43 to 56)
Chauhan (2003)[248]	Singleton; high risk; USA, centre II	Single; 3rd; EFW US	< 10th PCTL	Local; Alexander (1996)[303]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	11	6	4	44	73 (48 to 91)	88 (77 to 95)
Chauhan (2003)[248]	Singleton; high risk; USA, centre III	Single; 3rd; EFW US	< 10th PCTL	Local; Alexander (1996)[303]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	8	23	2	44	80 (48 to 97)	65 (52 to 76)
Chauhan (2004)[247]	Twin; Uns.; dichorionic – twin A	Single; 3rd; EFW US	< 10th PCTL	General (twin-specific); Ananth (1998)[314]	< 10th PCTL	General (twin-specific); Ananth (1998)[314]	4	5	6	74	40 (12 to 74)	94 (86 to 98)
Chauhan (2004)[247]	Twin; Uns.; dichorionic – twin B	Single; 3rd; EFW US	< 10th PCTL	General (twin-specific); Ananth (1998)[314]	< 10th PCTL	General (twin-specific); Ananth (1998)[314]	6	4	7	72	46 (19 to 75)	95 (87 to 99)
Chauhan (2006)[246]	Singleton; high risk	Single; 3rd; EFW US	< 10th PCTL	Local; Alexander (1996)[303]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	337	86	190	1341	64 (60 to 68)	94 (93 to 95)
De Jong (2000)[250]	Singleton; high risk	Multiple; 3rd; EFW US	< 10th PCTL on one or more 3rd tr. scan	Individualised; individually customised growth chart	< 10th PCTL	Individualised; individually customised antenatal growth chart, adjusted for maternal height and weight in early pregnancy, parity, ethnic group and fetal gender	44	17	21	133	68 (55 to 79)	89 (82 to 93)
De Reu (2008)[251]	Singleton; low risk; derivation cohort	Single; 3rd; AC US	< 10th PCTL	Local; derivation cohort	< 10th PCTL	Local; Dutch growth chart, Visser (2009)[315]	108	184	279	2878	28 (23 to 33)	94 (93 to 95)
De Reu (2008)[251]	Singleton; low risk; derivation cohort	Single; 3rd; AC US	< 10th PCTL	Local; derivation cohort	< 2.3rd PCTL	Local; Dutch growth chart, Visser (2009)[315]	139	214	248	2848	36 (31 to 41)	93 (92 to 94)

continued

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
De Reu (2008)[251]	Singleton; low risk; derivation cohort	Single; 3rd; AC US	< 5th PCTL	Local; derivation cohort	< 10th PCTL	Local; Dutch growth chart, Visser (2009)[315]	66	61	321	3001	17 (13 to 21)	98 (97 to 98)
De Reu (2008)[251]	Singleton; low risk; derivation cohort	Single; 3rd; AC US	< 5th PCTL	Local; derivation cohort	< 2.3rd PCTL	Local; Dutch growth chart, Visser (2009)[315]	93	92	294	2970	24 (20 to 29)	97 (96 to 98)
De Reu (2008)[251]	Singleton; low risk; derivation cohort	Single; 3rd; AC US	< 25th PCTL	Local; derivation cohort	< 10th PCTL	Local; Dutch growth chart, Visser (2009)[315]	205	612	182	2450	53 (48 to 58)	80 (79 to 81)
De Reu (2008)[251]	Singleton; low risk; derivation cohort	Single; 3rd; AC US	< 25th PCTL	Local; derivation cohort	< 2.3rd PCTL	Local; Dutch growth chart, Visser (2009)[315]	240	674	147	2388	62 (57 to 67)	78 (76 to 79)
De Reu (2008)[251]	Singleton; low risk; derivation cohort	Single; 3rd; HC US	< 10th PCTL	Local; derivation cohort	< 10th PCTL	Local; Dutch growth chart, Visser (2009)[315]	89	214	298	2848	23 (19 to 28)	93 (92 to 94)
De Reu (2008)[251]	Singleton; low risk; derivation cohort	Single; 3rd; HC US	< 10th PCTL	Local; derivation cohort	< 2.3rd PCTL	Local; Dutch growth chart, Visser (2009)[315]	93	245	294	2817	24 (20 to 29)	92 (91 to 93)
De Reu (2008)[251]	Singleton; low risk; derivation cohort	Single; 3rd; HC US	< 5th PCTL	Local; derivation cohort	< 10th PCTL	Local; Dutch growth chart, Visser (2009)[315]	50	122	337	2940	13 (10 to 17)	96 (95 to 97)
De Reu (2008)[251]	Singleton; low risk; derivation cohort	Single; 3rd; HC US	< 5th PCTL	Local; derivation cohort	< 2.3rd PCTL	Local; Dutch growth chart, Visser (2009)[315]	50	122	337	2940	13 (10 to 17)	96 (95 to 97)
De Reu (2008)[251]	Singleton; low risk; derivation cohort	Single; 3rd; HC US	< 25th PCTL	Local; derivation cohort	< 10th PCTL	Local; Dutch growth chart, Visser (2009)[315]	163	612	224	2450	42 (37 to 47)	80 (79 to 81)
De Reu (2008)[251]	Singleton; low risk; derivation cohort	Single; 3rd; HC US	< 25th PCTL	Local; derivation cohort	< 2.3rd PCTL	Local; Dutch growth chart, Visser (2009)[315]	186	674	201	2388	48 (43 to 53)	78 (76 to 79)
De Reu (2008)[251]	Singleton; low risk; validation cohort	Single; 3rd; AC US	< 25th PCTL	Local; derivation cohort	< 10th PCTL	Local; Dutch growth chart, Visser (2009)[315]	40	85	28	572	59 (46 to 71)	87 (84 to 90)
De Reu (2008)[251]	Singleton; low risk; validation cohort	Single; 3rd; AC US	< 25th PCTL	Local; derivation cohort	< 2.3rd PCTL	Local; Dutch growth chart, Visser (2009)[315]	46	105	22	552	68 (55 to 78)	84 (81 to 87)

TABLE 20 Accuracy data (*continued*)

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
de Silva (2021)[252]	Singleton; low risk	Serial; multiple; AC alone or both AC and HC US	Visual asses[a]	General; Loughna (2009)[316]	< 5th PCTL	General; Landmann (2006)[317]	185	38	39	246	83 (77 to 87)	87 (82 to 90)
Dude (2018)[255]	Singleton; diabetes; GDM (diet-controlled)	Single; 3rd; EFW US	< 10th PCTL	Local; Brenner (1976)[318]	< 10th PCTL	Local; US growth curves to Oken (2003)[319]	2	0	43	234	4 (1 to 15)	100 (99 to 100)
Dude (2018)[255]	Singleton; diabetes; GDM (requiring medication)	Single; 3rd; EFW US	< 10th PCTL	Local; Brenner (1976)[318]	< 10th PCTL	Local; US growth curves, Oken (2003)[319]	0	0	15	119	0 (0 to 18)	100 (98 to 100)
Dude (2018)[255]	Singleton; diabetes (gestational or pre-gestational DM)	Single; 3rd; EFW US	< 10th PCTL	Local; Brenner (1976)[318]	< 10th PCTL	Local; US growth curves, Oken (2003)[319]	3	0	61	457	5 (1 to 13)	100 (99 to 100)
Dude (2018)[255]	Singleton; diabetes; type 2 DM	Single; 3rd; EFW US	< 10th PCTL	Local; Brenner (1976)[318]	< 10th PCTL	Local; US growth curves, Oken (2003)[319]	1	0	3	46	25 (1 to 81)	100 (94 to 100)
Dude (2021)[254]	Singleton; BMI; BMI ≥ 40	Single; 3rd; EFW US	< 10th PCTL	Local; Brenner (1976)[318]	< 10th PCTL	Local; Brenner (1976)[318]	2	5	7	309	22 (3 to 60)	98 (96 to 99)
Dude (2021)[254]	Singleton; BMI; BMI 35.0–39.9	Single; 3rd; EFW US	< 10th PCTL	Local; Brenner (1976)[318]	< 10th PCTL	Local; Brenner (1976)[318]	3	3	7	354	30 (7 to 65)	99 (98 to 100)
Dude (2021)[254]	Singleton; BMI (BMI ≥ 35)	Single; 3rd; EFW US	< 10th PCTL	Local; Brenner (1976)[318]	< 10th PCTL	Local; Brenner (1976)[318]	5	8	14	663	26 (9 to 52)	99 (98 to 99)
Haragan (2015)[256]	Singleton; high risk	Single; 3rd; AC hh US	< 5th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curves, Olsen (2010)[320]	20	23	7	201	74 (54 to 89)	90 (85 to 93)
Haragan (2015)[256]	Singleton; high risk	Single; 3rd; AC US	< 5th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curves, Olsen (2010)[320]	15	9	12	215	56 (35 to 75)	96 (93 to 98)
Haragan (2015)[256]	Singleton; high risk	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curves, Olsen (2010)[320]	6	1	21	223	22 (9 to 42)	100 (98 to 100)

continued

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Haragan (2015)[256]	Singleton; high risk	Single; 3rd; FH US	FH: size < dates	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curves, Olsen (2010)[320]	10	11	17	213	37 (19 to 58)	95 (91 to 98)
Harper (2013)[257]	Twin; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local; Alexander (1996)[303]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	105	42	52	341	67 (59 to 74)	89 (85 to 92)
Harper (2013)[257]	Twin; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local; Alexander (1996)[303]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	110	52	47	331	70 (62 to 77)	86 (83 to 90)
Harper (2013)[257]	Twin; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local; Alexander (1996)[303]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	111	56	46	327	71 (63 to 78)	85 (81 to 89)
Henrichs (2019)[258]	Singleton; low risk; control: usual care[b]	Multiple; 3rd; AC US	< 10th PCTL	Local; Verburg (2008)[321]	< 10th PCTL	Local; Dutch growth curve, Verburg (2008)[321]	53	60	354	5031	13 (10 to 17)	99 (98 to 99)
Henrichs (2019)[258]	Singleton; low risk; control: usual care[b]	Multiple; 3rd; AC US and fetal abdominal growth	AC < 10th PCTL and decrease in AC of at least 20 centiles	Local; Verburg (2008)[321]	< 10th PCTL	Local; Dutch growth curve, Verburg (2008)[321]	78	153	329	4938	19 (15 to 23)	97 (96 to 97)
Henrichs (2019)[258]	Singleton; low risk; control: usual care[b]	Multiple; 3rd; fetal abdominal growth	decrease in AC of at least 20 centiles	Local; Verburg (2008)[321]	< 10th PCTL	Local; Dutch growth curve, Verburg (2008)[321]	34	100	373	4991	8 (6 to 11)	98 (98 to 98)
Henrichs (2019)[258]	Singleton; low risk; interven-tion: 28–30 weeks	Single; 3rd; AC US	< 10th PCTL	Local; Verburg (2008)[321]	< 10th PCTL	Local; Dutch growth curve, Verburg (2008)[321]	74	126	482	6227	13 (11 to 16)	98 (98 to 98)
Henrichs (2019)[258]	Singleton; low risk; interven-tion: 34–36 weeks	Single; 3rd; AC US	< 10th PCTL	Local; Verburg (2008)[321]	< 10th PCTL	Local; Dutch growth curve, Verburg (2008)[321]	63	44	491	6290	11 (9 to 14)	99 (99 to 99)
Henrichs (2019)[258]	Singleton; low risk; intervention[c]	Multiple; 3rd; AC US	< 10th PCTL	Local; Verburg (2008)[321]	< 10th PCTL	Local; Dutch growth curve, Verburg (2008)[321]	122	156	434	6197	22 (19 to 26)	98 (97 to 98)
Henrichs (2019)[258]	Singleton; Low risk; intervention[c]	Multiple; 3rd; AC US and fetal abdominal growth	AC < 10th PCTL and decrease in AC of at least 20 centiles	Local; Verburg (2008)[321]	< 10th PCTL	Local; Dutch growth curve, Verburg (2008)[321]	179	651	377	5702	32 (28 to 36)	90 (89 to 90)

TABLE 20 Accuracy data (*continued*)

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Henrichs (2019)[258]	Singleton; Low risk; intervention[c]	Multiple; 3rd; fetal abdominal growth	Decrease in AC of at least 20 centiles	Local; Verburg (2008)[321]	< 10th PCTL	Local; Dutch growth curve, Verburg (2008)[321]	76	522	480	5381	14 (11 to 17)	91 (90 to 92)
Humphries (2002)[261]	Singleton; Uns.; board-certified maternal–fetal medicine specialists	Single; 3rd; EFW US	Under 2500 g[d]	General; N/A	< 2500g	General; an absolute value was used to define SGA	14	22	4	198	78 (52 to 94)	90 (85 to 94)
Humphries (2002)[261]	Singleton; Uns.; registered diagnostic medical sonographers	Single; 3rd; EFW US	Under 2500 g[d]	General; N/A	< 2500g	General; an absolute value was used to define SGA	13	2	5	218	72 (47 to 90)	99 (97 to 100)
Leung (2008)[263]	Singleton; Uns.	Single; 1st; CRL US	<Expected by 3–6 days	Fok (2003)[322]	< 10th PCTL	Local; GA-specific and gender-adjusted *z*-score based on a locally derived BW reference	25	183	216	2336	10 (7 to 15)	93 (92 to 94)
Li (2021)[264]	Singleton; Uns.	Single; 3rd; FBV/FLV US	4.1	N/A	< 10th PCTL	Local; Chinese growth chart, Dai (2014)[323]	39	5	3	100	93 (81 to 99)	95 (89 to 98)
Lindström (2023)[265]	Singleton; Uns.	Single; 2nd or 3rd; EFW US; Hadlock 2c[29]	< 3rd PCTL	Local; Lindstrom (2021)[324]	< 3rd PCTL	Local; Swedish reference range, Lindström (2021)[324]	4369	666	3433	23,044	56 (55 to 57)	97 (97 to 97)
Lindström (2023)[265]	Singleton; Uns.	Single; 2nd or 3rd; EFW US; Shepard (1982)[22]	< 3rd PCTL	Local; Lindstrom (2021)[324]	< 3rd PCTL	Local; Swedish reference range, Lindström (2021)[324]	4642	1016	3160	22,703	59 (58 to 61)	96 (95 to 96)
Lindström (2023)[265]	Singleton; Uns.	Single; 2nd or 3rd; EFW US; Persson and Weldner (1986)[325]	< 3rd PCTL	Local; Lindstrom (2021)[324]	< 3rd PCTL	Local; Swedish reference range, Lindström (2021)[324]	6132	2250	1670	21,469	79 (78 to 80)	91 (90 to 91)

continued

DOI: 10.3310/AJLK7403

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Marchand (2022)[266]	Singleton; Uns.	Single; any; AC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	785	823	276	7408	74 (71 to 77)	90 (89 to 91)
Marchand (2022)[266]	Singleton; Uns.	Single; any; AC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	222	271	33	8766	87 (82 to 91)	97 (97 to 97)
Marchand (2022)[266]	Singleton; Uns.	Single; any; BPD US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	541	988	520	7243	51 (48 to 54)	88 (87 to 89)
Marchand (2022)[266]	Singleton; Uns.	Single; any; BPD US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	115	181	140	8856	45 (39 to 51)	98 (98 to 98)
Marchand (2022)[266]	Singleton; Uns.	Single; any; FL US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	690	823	371	7408	65 (62 to 68)	90 (89 to 91)
Marchand (2022)[266]	Singleton; Uns.	Single; any; FL US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	158	90	97	8947	62 (56 to 68)	99 (99 to 99)
Marchand (2022)[266]	Singleton; Uns.	Single; any; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	424	1811	637	6420	40 (37 to 43)	78 (77 to 79)
Marchand (2022)[266]	Singleton; Uns.	Single; any; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	255	8947	0	90	100 (99 to 100)	1 (1 to 1)
Marchand (2022)[266]	Singleton; Uns.	Single; any; HC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	584	741	477	7490	55 (52 to 58)	91 (90 to 92)
Marchand (2022)[266]	Singleton; Uns.	Single; any; HC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	163	271	92	8766	64 (58 to 70)	97 (97 to 97)
Marchand (2022)[266]	Singleton; Uns.	Single; any; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	329	741	732	7490	31 (28 to 34)	91 (90 to 92)
Marchand (2022)[266]	Singleton; Uns.	Single; any; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	33	90	222	8947	13 (9 to 18)	99 (99 to 99)

TABLE 20 Accuracy data (*continued*)

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Marchand (2022)[266]	Singleton; Uns.	Single; any; TDC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	255	741	806	7490	24 (21 to 27)	91 (90 to 92)
Marchand (2022)[266]	Singleton; Uns.	Single; any; TDC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	51	271	204	8766	20 (15 to 25)	97 (97 to 97)
Marchand (2022)[266]	Singleton; Uns.	Single; any; TDC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	244	494	817	7737	23 (20 to 26)	94 (93 to 95)
Marchand (2022)[266]	Singleton; Uns.	Single; any; TDC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	77	271	179	8766	30 (24 to 36)	97 (97 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; AC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	38	134	42	1085	48 (37 to 59)	89 (87 to 91)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; AC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	258	452	139	4065	65 (60 to 70)	90 (89 to 91)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; AC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	296	586	181	5150	62 (58 to 67)	90 (89 to 91)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; AC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	5	39	3	1252	63 (24 to 93)	97 (96 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; AC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	47	146	12	4709	80 (67, 89)	97 (96 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; AC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	52	184	15	5962	78 (66 to 87)	97 (97 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; BPD US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	22	146	58	1073	28 (18 to 39)	88 (86 to 90)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; BPD US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	143	542	254	3975	36 (31 to 41)	88 (87 to 89)

continued

TABLE 20 Accuracy data (*continued*)

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; BPD US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	165	688	312	5048	35 (30 to 39)	88 (87 to 89)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; BPD US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	3	39	5	1252	38 (9 to 77)	97 (96 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; BPD US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	22	97	37	4758	37 (25 to 51)	98 (98 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; BPD US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	25	136	42	6010	37 (26 to 50)	98 (97 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; FL US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	45	134	35	1085	56 (44 to 67)	89 (87 to 91)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; FL US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	234	452	163	4065	59 (54 to 64)	90 (89 to 91)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; FL US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	279	586	198	5150	58 (54 to 63)	90 (89 to 91)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; FL US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	7	39	1	1252	88 (46 to 100)	97 (96 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; FL US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	42	146	17	4709	71 (58 to 82)	97 (96 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; FL US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	49	184	18	5962	73 (61 to 83)	97 (97 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	4	171	76	1048	5 (1 to 12)	86 (84 to 88)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	83	723	314	3794	21 (17 to 25)	84 (83 to 85)

DOI: 10.3310/AJLK7403

TABLE 20 Accuracy data (*continued*)

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	87	893	390	4843	18 (15 to 22)	84 (83 to 85)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	7	1175	1	116	88 (46 to 100)	9 (7 to 11)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	59	4855	0	0	100 (95 to 100)	0 (0 to 0)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	66	6030	1	116	99 (91 to 100)	2 (2 to 2)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; HC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	28	122	52	1097	35 (25 to 46)	90 (88 to 92)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; HC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	175	407	222	4110	44 (39 to 49)	91 (90 to 92)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; HC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	203	528	274	5208	42 (38 to 47)	91 (90 to 92)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; HC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	5	52	3	1239	63 (24 to 93)	96 (95 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; HC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	31	146	28	4709	53 (40 to 66)	97 (96 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; HC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	36	197	31	5949	54 (42 to 66)	97 (96 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	15	122	65	1097	19 (11 to 29)	90 (88 to 92)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	119	407	278	4110	30 (26 to 35)	91 (90 to 92)

continued

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	134	528	343	5208	28 (24 to 32)	91 (90 to 92)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	0	26	8	1265	0 (0 to 31)	98 (97 to 99)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	13	49	46	4806	22 (12 to 35)	99 (99 to 99)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	13	74	54	6072	19 (11 to 31)	99 (98 to 99)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; TCD/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	15	98	65	1121	19 (11 to 29)	92 (90 to 93)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; TCD/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	87	271	310	4246	22 (18 to 26)	94 (93 to 95)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; TCD/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	103	369	374	5367	21 (18 to 25)	94 (93 to 94)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; TCD/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	1	39	7	1252	13 (0 to 54)	97 (96 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; TCD/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	19	146	40	4709	32 (20 to 46)	97 (96 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; TCD/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	20	184	47	5962	30 (19 to 42)	97 (97 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; TDC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	15	98	65	1121	19 (11 to 29)	92 (90 to 93)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; TDC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	95	452	302	4065	24 (20 to 29)	90 (89 to 91)

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; TDC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	110	549	367	5187	23 (19 to 27)	90 (90 to 91)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; TDC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	2	39	6	1252	25 (3 to 65)	97 (96 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; TDC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	15	146	44	4709	25 (15 to 38)	97 (96 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 2nd tr.	Single; 2nd; TDC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	17	184	50	5962	25 (15 to 37)	97 (97 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; AC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	123	44	14	450	90 (84 to 94)	91 (88 to 93)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; AC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	194	108	40	870	83 (78 to 88)	89 (87 to 91)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; AC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	489	253	91	2231	84 (81 to 87)	90 (89 to 91)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; AC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	47	12	1	571	98 (88 to 100)	98 (96 to 99)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; AC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	65	34	10	1103	87 (77 to 94)	97 (96 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; AC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	171	69	16	2808	91 (86 to 95)	98 (97 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; BPD US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	92	69	45	425	67 (58 to 75)	86 (83 to 89)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; BPD US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	152	127	82	851	65 (59 to 71)	87 (85 to 89)

continued

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; BPD US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	378	328	202	2156	65 (61 to 69)	87 (85 to 88)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; BPD US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	20	12	28	571	42 (28 to 57)	98 (96 to 99)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; BPD US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	31	11	44	1126	41 (30 to 53)	99 (98 to 99)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; BPD US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	89	58	98	2819	47 (40 to 55)	98 (97 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; FL US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	92	54	45	440	67 (58, to 75)	89 (86 to 92)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; FL US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	164	98	70	880	70 (64 to 76)	90 (88 to 92)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; FL US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	406	274	174	2210	70 (66 to 74)	89 (88 to 90)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; FL US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	19	0	29	583	40 (26 to 55)	100 (99 to 100)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; FL US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	44	11	31	1126	59 (47 to 70)	99 (98 to 99)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; FL US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	110	23	77	2854	59 (51 to 66)	99 (99 to 99)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	121	296	16	198	88 (81 to 93)	40 (36 to 44)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	140	411	94	567	60 (53 to 66)	58 (55 to 61)

TABLE 20 Accuracy data (*continued*)

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	328	920	252	1564	57 (52 to 61)	63 (61 to 65)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	48	583	0	0	100 (94 to 100)	0 (0 to 1)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	75	1137	0	0	100 (96 to 100)	0 (0 to 0)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	187	2877	0	0	100 (98 to 100)	0 (0 to 0)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; HC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	88	49	49	445	64 (55 to 72)	90 (87 to 93)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; HC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	152	68	82	910	65 (59 to 71)	93 (91 to 95)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; HC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	376	219	204	2265	65 (61 to 69)	91 (90 to 92)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; HC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	32	17	16	566	67 (52 to 80)	97 (95 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; HC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	45	34	30	1103	60 (48 to 71)	97 (96 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; HC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	126	98	61	2779	68 (60 to 74)	97 (96 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	37	30	100	464	27 (20 to 35)	94 (92 to 96)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	80	88	154	890	34 (28 to 40)	91 (89 to 93)

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 20 Accuracy data (*continued*)

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	192	209	388	2275	33 (29 to 37)	92 (90 to 93)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	3	0	45	583	6 (1 to 17)	100 (99 to 100)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	5	0	70	1137	7 (2 to 15)	100 (100 to 100)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	19	0	168	2877	10 (6 to 15)	100 (100 to 100)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; TCD US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	34	44	103	450	25 (18 to 33)	91 (88 to 93)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; TCD US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	47	78	187	900	20 (15 to 26)	92 (90 to 94)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; TCD US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	144	214	436	2270	25 (21 to 29)	91 (90 to 92)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; TCD/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	18	20	119	474	13 (8 to 20)	96 (94 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; TCD/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	56	29	178	949	24 (19 to 30)	97 (96 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; TCD/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	139	100	441	2384	24 (21 to 28)	96 (95 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; TCD/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	12	12	36	571	25 (14 to 40)	98 (96 to 99)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; TCD/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	21	34	54	1103	28 (18 to 40)	97 (96 to 98)

TABLE 20 Accuracy data (*continued*)

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; TCD/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	55	80	132	2797	29 (23 to 36)	97 (97 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; TDC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	5	17	43	566	10 (3 to 22)	97 (95 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; TDC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	10	23	65	1114	13 (6 to 23)	98 (97 to 99)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; 3rd; TDC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	34	87	153	2790	18 (13 to 24)	97 (96 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; AC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	171	101	38	911	82 (76 to 87)	90 (88 to 92)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; AC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	58	23	6	1134	91 (81 to 97)	98 (97 to 99)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; BPD US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	134	132	75	880	64 (57 to 71)	87 (85 to 89)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; BPD US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	38	35	26	1122	59 (46 to 71)	97 (96 to 98)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; FL US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	150	121	59	891	72 (65 to 78)	88 (86 to 90)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; FL US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	46	12	18	1145	72 (59 to 83)	99 (98 to 99)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	67	213	142	799	32 (26 to 39)	79 (76 to 81)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; FL/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	64	1157	0	0	100 (95 to 100)	0 (0 to 0)

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; HC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	136	101	73	911	65 (58 to 71)	90 (88 to 92)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; HC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	49	46	15	1111	77 (65 to 87)	96 (95 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	75	91	134	921	36 (29 to 43)	91 (89 to 93)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; HC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	11	0	53	1157	17 (9 to 29)	100 (100 to 100)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; TDC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	63	91	146	921	30 (24 to 37)	91 (89 to 93)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; TDC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	19	46	45	1111	30 (19 to 43)	96 (95 to 97)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; TDC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; German growth chart, Voigt (2006)[326]	65	51	144	961	31 (25 to 38)	95 (93 to 96)
Marchand (2022)[266]	Singleton; Uns.; tested during 3rd tr.	Single; unclear; TDC/AC US	> 90th PCTL	General; Hadlock (1991)[30]	< 3rd PCTL	Local; German growth chart, Voigt (2006)[326]	22	35	42	1122	34 (23 to 47)	97 (96 to 98)
Martin-Palumbo (2022)[267]	Singleton; low risk; 36 weeks' US	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 28th PCTL	Local; Yudkin (1987)[327]	65	94	16	313	80 (70 to 88)	77 (72 to 81)
Martin-Palumbo (2022)[267]	Singleton; low risk; tested during 3rd tr.	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; Yudkin (1987)[327]	25	5	56	402	31 (21 to 42)	99 (97 to 100)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; GROW, Gardosi (2020)[300]	< 10th PCTL	General; GROW, Gardosi (2020)[300]	422	173	1504	15,579	22 (20 to 24)	99 (99 to 99)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; GROW, Gardosi (2020)[300]	< 10th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	238	355	515	16,570	32 (28 to 35)	98 (98 to 98)

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; GROW, Gardosi (2020)[300]	< 10th PCTL	Local; UK growth chart, Freeman (1995)[328]	281	312	980	16,105	22 (20 to 25)	98 (98 to 98)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; GROW, Gardosi (2020)[300]	< 10th PCTL	Local; UK local reference, Mathewlynn (2022)[268]	316	270	1452	15,640	18 (16 to 20)	98 (98 to 98)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; GROW, Gardosi (2020)[300]	< 10th PCTL	Local; WHO, SACN, RCPCH, UK growth chart[329]	275	322	473	16,608	37 (33 to 40)	98 (98 to 98)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	General; GROW, Gardosi (2020)[300]	404	205	1522	15,547	21 (19 to 23)	99 (99 to 99)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	284	322	469	16,603	38 (34 to 41)	98 (98 to 98)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; UK growth chart, Freeman (1995)[328]	340	263	921	16,154	27 (25 to 30)	98 (98 to 99)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; UK local reference, Mathewlynn (2022)[268]	375	239	1393	15,671	21 (19 to 23)	99 (98 to 99)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; WHO, SACN, RCPCH, UK growth chart[329]	331	271	417	16,659	44 (41 to 48)	98 (98 to 99)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; INTERGROWTH-21st[23]	< 10th PCTL	General; GROW, Gardosi (2020)[300]	210	63	1716	15,689	11 (10 to 12)	100 (99 to 100)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; INTERGROWTH-21st[23]	< 10th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	160	118	593	16,807	21 (18 to 24)	99 (99 to 99)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; INTERGROWTH-21st[23]	< 10th PCTL	Local; UK growth chart, Freeman (1995)[328]	177	99	1084	16,318	14 (12 to 16)	99 (99 to 100)

continued

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; INTERGROWTH-21st[23]	< 10th PCTL	Local; UK local reference, Mathewlynn (2022)[268]	191	80	1577	15,830	11 (9 to 12)	100 (99 to 100)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; INTERGROWTH-21st[23]	< 10th PCTL	Local; WHO, SACN, RCPCH, UK growth chart[329]	189	85	559	16,845	25 (22 to 29)	100 (99 to 100)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; WHO, Kiserud (2017)[4]	< 10th PCTL	General; GROW, Gardosi (2020)[300]	401	205	1525	15,547	21 (19 to 23)	99 (99 to 99)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; WHO, Kiserud (2017)[4]	< 10th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	282	322	471	16,603	38 (34 to 41)	98 (98 to 98)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; WHO, Kiserud (2017)[4]	< 10th PCTL	Local; UK growth chart, Freeman (1995)[328]	337	263	924	16,154	27 (24 to 29)	98 (98 to 99)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; WHO, Kiserud (2017)[4]	< 10th PCTL	Local; UK local reference, Mathewlynn (2022)[268]	370	239	1398	15,671	21 (19 to 23)	99 (98 to 99)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; WHO, Kiserud (2017)[4]	< 10th PCTL	Local; WHO, SACN, RCPCH, UK growth chart[329]	329	271	419	16,659	44 (40 to 48)	98 (98 to 99)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local; Mathewlynn (2022)[268]	< 10th PCTL	General; GROW, Gardosi (2020)[300]	792	914	1134	14,838	41 (39 to 43)	94 (94 to 95)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local; Mathewlynn (2022)[268]	< 10th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	478	1219	275	15,706	64 (60 to 67)	93 (92 to 93)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local; Mathewlynn (2022)[268]	< 10th PCTL	Local; UK growth chart, Freeman (1995)[328]	631	1067	631	15,350	50 (47 to 53)	94 (93 to 94)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local; Mathewlynn (2022)[268]	< 10th PCTL	Local; UK local reference, Mathewlynn (2022)[268]	773	923	995	14,987	44 (41 to 46)	94 (94 to 95)

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Mathewlynn (2022)[268]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local; Mathewlynn (2022)[268]	< 10th PCTL	Local; WHO, SACN, RCPCH, UK growth chart[329]	500	1202	248	15,728	67 (63 to 70)	93 (93 to 93)
Michaeli (2022)[269]	Singleton; Uns.	Single; unclear; EFW US	< 10th PCTL	Local; Dollberg (2005)[305]	< 10th PCTL	Local; Israeli growth curve, Dollberg (2005)[305]	3040	5671	8508	82,979	26 (26 to 27)	94 (93 to 94)
Monier (2022)[271]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; FMF[298]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	498	1238	402	7802	55 (52 to 59)	86 (86 to 87)
Monier (2022)[271]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	338	560	562	8480	38 (34 to 41)	94 (93 to 94)
Monier (2022)[271]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; INTERGROWTH-21st[23]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	178	208	722	8832	20 (17 to 23)	98 (97 to 98)
Monier (2022)[271]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; WHO, Kiserud (2017)[4]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	341	570	559	8470	38 (35 to 41)	94 (93 to 94)
Monier (2022)[271]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local ethnicity-specific; NICHD-White, Buck Louis (2015)[331]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	456	1022	444	8018	51 (47 to 54)	89 (88 to 89)
Monier (2022)[271]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local; CFEF, Massoud (2016)[332]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	314	497	586	8543	35 (32 to 38)	95 (94 to 95)
Monier (2022)[271]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local; Epopé, Ego (2016)[333]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	181	208	719	8832	20 (18 to 23)	98 (97 to 98)
Monier (2022)[271]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local; Salomon (2007)[20]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	429	850	471	8190	48 (44 to 51)	91 (90 to 91)
Monier (2022)[270]	Singleton; Uns.; females	Single; 3rd; EFW US	< 10th PCTL	General sex-specific; WHO, Kiserud (2017)[4]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	124	220	296	4169	30 (25 to 34)	95 (94 to 96)

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Monier (2022)[270]	Singleton; Uns.; females	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	170	341	250	4048	40 (36 to 45)	92 (91 to 93)
Monier (2022)[270]	Singleton; Uns.; females	Single; 3rd; EFW US	< 10th PCTL	General; WHO, Kiserud (2017)[4]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	162	313	258	4076	39 (34 to 43)	93 (92 to 94)
Monier (2022)[270]	Singleton; Uns.; females	Single; 3rd; EFW US	< 10th PCTL	Local sex-specific; Epopé, Ego (2016)[333]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	58	74	362	4315	14 (11 to 17)	98 (98 to 99)
Monier (2022)[270]	Singleton; Uns.; females	Single; 3rd; EFW US	< 10th PCTL	Local; Epopé, Ego (2016)[333]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	80	126	340	4263	19 (15 to 23)	97 (97 to 98)
Monier (2022)[270]	Singleton; Uns.; males	Single; 3rd; EFW US	< 10th PCTL	General sex-specific; WHO, Kiserud (2017)[4]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	200	310	280	4341	42 (37 to 46)	93 (93 to 94)
Monier (2022)[270]	Singleton; Uns.; males	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	168	218	312	4433	35 (31 to 39)	95 (95 to 96)
Monier (2022)[270]	Singleton; Uns.; males	Single; 3rd; EFW US	< 10th PCTL	General; WHO, Kiserud (2017)[4]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	156	197	324	4454	33 (28 to 37)	96 (95 to 96)
Monier (2022)[270]	Singleton; Uns.; males	Single; 3rd; EFW US	< 10th PCTL	Local sex-specific; Epopé, Ego (2016)[333]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	118	120	362	4531	25 (21 to 29)	97 (97 to 98)
Monier (2022)[270]	Singleton; Uns.; males	Single; 3rd; EFW US	< 10th PCTL	Local; Epopé, Ego (2016)[333]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	99	81	381	4570	21 (17 to 25)	98 (98 to 99)
Monier (2022)[270]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General sex-specific; WHO, Kiserud (2017)[4]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	324	530	576	8510	36 (33 to 39)	94 (94 to 95)
Monier (2022)[270]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	338	559	562	8481	38 (34 to 41)	94 (93 to 94)
Monier (2022)[270]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; WHO, Kiserud (2017)[4]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	318	510	582	8530	35 (32 to 39)	94 (94 to 95)

DOI: 10.3310/AJLK7403

TABLE 20 Accuracy data (*continued*)

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Monier (2022)[270]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local sex-specific; Epopé, Ego (2016)[333]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	176	194	724	8846	20 (17 to 22)	98 (98 to 98)
Monier (2022)[270]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local; Epopé, Ego (2016)[333]	< 10th PCTL	Local; French BW reference charts, Mamelle (1996)[330]	179	207	721	8833	20 (17 to 23)	98 (97 to 98)
Nwabuobi (2020)[272]	Singleton; high risk	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	58	33	81	882	42 (33 to 50)	96 (95 to 98)
Nwabuobi (2020)[272]	Singleton; high risk	Single; 3rd; EFW US	< 10th PCTL	General; INTERGROWTH-21st[299]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	34	30	105	905	24 (18 to 32)	97 (95 to 98)
Nymark Hansen (2019)[277]	Singleton; high risk	Unclear; 3rd; EFW US	≤ 15th PCTL (last US)	Local; Marsál (1996)[307]	≤ 22nd PCTL	Local; Swedish growth curve, Maršál (1996)[307]	61	158	16	1614	79 (68 to 88)	91 (90 to 92)
Nymark Hansen (2019)[277]	Singleton; high risk	Unclear; 3rd; EFW US	≤ 15th PCTL (last US)	Local; Marsál (1996)[307]	≤ 15th PCTL	Local; Swedish growth curve, Maršál (1996)[307]	126	93	89	1541	86 (76 to 93)	83 (81 to 85)
Nymark Hansen (2019)[277]	Singleton; Uns.	Unclear; 3rd; Danish screening program[e]	≤ 15th PCTL[f]	Local; Marsál (1996)[307]	≤ 22nd PCTL	Local; Swedish growth curve, Maršál (1996)[307]	61	158	37	2672	62 (52 to 72)	94 (94 to 95)
Nymark Hansen (2019)[277]	Singleton; Uns.	Unclear; 3rd; Danish screening program[e]	≤ 15th PCTL[f]	Local; Marsál (1996)[307]	≤ 15th PCTL	Local; Swedish growth curve, Maršál (1996)[307]	126	93	177	2532	42 (36 to 47)	96 (96 to 97)
Odibo (2018)[274]	Singleton; high risk	Single; 3rd; EFW US	< 10th PCTL	Individualised; Gardosi (2009)[334]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	54	38	85	877	39 (31 to 47)	96 (94 to 97)
Poljak (2017)[278]	Singleton; high risk	Serial; multiple; AC US	< 10th PCTL[g]	General; Sovio (2015)[288]	< 10th PCTL	General; GROW, Gardosi (1995)[335]	14	2	48	41	23 (13 to 35)	95 (84 to 99)
Poljak (2017)[278]	Singleton; high risk	Serial; multiple; AC US	< 10th PCTL[g]	General; Sovio (2015)[288]	< 10th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	13	3	42	47	24 (13 to 37)	94 (83 to 99)
Poljak (2017)[278]	Singleton; high risk	Serial; multiple; AC US	< 10th PCTL[g]	General; Sovio (2015)[288]	< 10th PCTL	General; Mikolajczyk (2011)[336]	12	4	45	44	21 (11 to 34)	92 (80 to 98)

continued

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Poljak (2017)[278]	Singleton; high risk	Serial; multiple; AC US	< 10th PCTL[g]	General; Sovio (2015)[288]	< 10th PCTL	Local; WHO, UK growth chart, Cole (1998)[337]	12	4	39	50	24 (13 to 37)	93 (82 to 98)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; AC US	< 10th PCTL[g]	General; Chitty (1994)[338]	< 10th PCTL	General; GROW, Gardosi (1995)[335]	27	4	35	39	44 (31 to 57)	91 (78 to 97)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; AC US	< 10th PCTL[g]	General; Chitty (1994)[338]	< 10th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	27	4	28	46	49 (35 to 63)	92 (81 to 98)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; AC US	< 10th PCTL[g]	General; Chitty (1994)[338]	< 10th PCTL	General; Mikolajczyk (2011)[336]	27	4	30	44	47 (34 to 61)	92 (80 to 98)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; AC US	< 10th PCTL[g]	General; Chitty (1994)[338]	< 10th PCTL	Local; WHO, UK growth chart, Cole (1998)[337]	26	5	25	49	51 (37 to 65)	91 (80 to 97)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; AC US	< 10th PCTL[g]	General; Hadlock (1991)[30]	< 10th PCTL	General; GROW, Gardosi (1995)[335]	46	13	16	30	74 (62 to 84)	70 (54 to 83)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; AC US	< 10th PCTL[g]	General; Hadlock (1991)[30]	< 10th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	43	1	12	49	78 (65 to 88)	98 (89 to 100)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; AC US	< 10th PCTL[g]	General; Hadlock (1991)[30]	< 10th PCTL	General; Mikolajczyk (2011)[336]	44	15	13	33	77 (64 to 87)	69 (54 to 81)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; AC US	< 10th PCTL[g]	General; Hadlock (1991)[30]	< 10th PCTL	Local; WHO, UK growth chart, Cole (1998)[337]	42	17	9	37	82 (69 to 92)	69 (54 to 80)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; AC US	< 10th PCTL[g]	General; INTERGROWTH-21st[339]	< 10th PCTL	General; GROW, Gardosi (1995)[335]	22	2	40	41	35 (24 to 49)	95 (84 to 99)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; AC US	< 10th PCTL[g]	General; INTERGROWTH-21st[339]	< 10th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	21	3	34	47	38 (25 to 52)	94 (83 to 99)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; AC US	< 10th PCTL[g]	General; INTERGROWTH-21st[339]	< 10th PCTL	General; Mikolajczyk (2011)[336]	21	3	36	45	37 (24 to 51)	94 (83 to 99)

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; AC US	< 10th PCTL[g]	General; INTERGROWTH-21st[339]	< 10th PCTL	Local; WHO, UK growth chart, Cole (1998)[337]	21	2	30	52	41 (28 to 56)	96 (87 to 100)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; EFW US	< 10th PCTL[g]	General; GROW, Gardosi (1995)[335]	< 10th PCTL	General; GROW, Gardosi (1995)[335]	35	8	27	35	56 (43 to 69)	81 (67 to 92)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; EFW US	< 10th PCTL[g]	General; GROW, Gardosi (1995)[335]	< 10th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	30	12	25	38	55 (41 to 68)	76 (62 to 87)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; EFW US	< 10th PCTL[g]	General; GROW, Gardosi (1995)[335]	< 10th PCTL	General; Mikolajczyk (2011)[336]	30	13	27	35	53 (39 to 66)	73 (58 to 85)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; EFW US	< 10th PCTL[g]	General; GROW, Gardosi (1995)[335]	< 10th PCTL	Local; WHO, UK growth chart, Cole (1998)[337]	28	15	23	39	55 (40 to 69)	72 (58 to 84)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; EFW US	< 10th PCTL[g]	General; Hadlock (1991)[30]	< 10th PCTL	General; GROW, Gardosi (1995)[335]	41	6	21	37	66 (53 to 78)	86 (72 to 95)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; EFW US	< 10th PCTL[g]	General; Hadlock (1991)[30]	< 10th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	38	8	17	42	69 (55 to 81)	84 (71 to 93)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; EFW US	< 10th PCTL[g]	General; Hadlock (1991)[30]	< 10th PCTL	General; Mikolajczyk (2011)[336]	39	8	18	40	68 (55 to 80)	83 (70 to 93)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; EFW US	< 10th PCTL[g]	General; Hadlock (1991)[30]	< 10th PCTL	Local; WHO, UK growth chart, Cole (1998)[337]	37	10	14	44	73 (58 to 84)	81 (69 to 91)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; EFW US	< 10th PCTL[g]	General; Mikolajczyk (2011)[336]	< 10th PCTL	General; GROW, Gardosi (1995)[335]	37	9	25	34	60 (46 to 72)	79 (64 to 90)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; EFW US	< 10th PCTL[g]	General; Mikolajczyk (2011)[336]	< 10th PCTL	General; INTERGROWTH-21st, Villar (2014)[299]	35	7	20	43	64 (50 to 76)	86 (73 to 94)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; EFW US	< 10th PCTL[g]	General; Mikolajczyk (2011)[336]	< 10th PCTL	General; Mikolajczyk (2011)[336]	35	7	22	41	61 (48 to 74)	85 (72 to 94)

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 20 Accuracy data (*continued*)

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Poljak (2017)[278]	Singleton; high risk	Single; unclear; EFW US	< 10th PCTL[g]	General; Mikolajczyk (2011)[336]	< 10th PCTL	Local; WHO, UK growth chart, Cole (1998)[337]	34	8	17	46	67 (52 to 79)	85 (73 to 93)
Pressman (2022)[273]	Singleton; high risk	Single; 3rd; AC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	70	52	69	863	50 (42 to 59)	94 (93 to 96)
Pressman (2022)[273]	Singleton; high risk	Single; 3rd; EFW US and AC US	AC < 10th PCTL OR EFW < 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	76	61	63	854	55 (46 to 63)	93 (92 to 95)
Price (2022)[279]	Singleton; high risk	Serial; 2nd and 3rd; growth velocity for AC US	Change in AC, growth velocity < 10th PCTL	N/A	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	12	49	56	495	18 (9 to 29)	91 (88 to 93)
Price (2022)[279]	Singleton; high risk	Serial; 2nd and 3rd; growth velocity for EFW US	Change in EFW, growth velocity < 10th PCTL	N/A	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	10	50	58	494	15 (7 to 26)	91 (88 to 93)
Rad (2018)[281]	Singleton; Uns.	Single; 3rd; AC US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	170	39	95	1290	64 (5 8 to 70)	97 (96 to 98)
Rad (2018)[281]	Singleton; Uns.	Single; 3rd; AC US	< 5th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	116	16	149	1313	44 (38 to 50)	99 (98 to 99)
Rad (2018)[281]	Singleton; Uns.	Single; 3rd; AC US	< 3rd PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	97	13	168	1316	37 (31 to 43)	99 (98 to 99)
Rad (2018)[281]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	134	27	131	1302	51 (44 to 57)	98 (97 to 99)
Rad (2018)[281]	Singleton; Uns.	Single; 3rd; EFW US	< 5th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	64	4	201	1325	24 (19 to 30)	100 (99 to 100)
Rad (2018)[281]	Singleton; Uns.	Single; 3rd; EFW US	< 3rd PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	42	3	223	1326	16 (12 to 21)	100 (99 to 100)
Rad (2018)[281]	Singleton; Uns.	Single; 3rd; EFW US and AC US	AC < 10th PCTL OR EFW < 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	179	44	86	1285	68 (62 to 73)	97 (96 to 98)

TABLE 20 Accuracy data (*continued*)

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Rad (2018)[281]	Singleton; Uns.	Single; 3rd; EFW US and AC US	AC < 10th PCTL AND EFW < 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	125	21	140	1308	47 (41 to 53)	98 (98 to 99)
Ridha (2022)[283]	Singleton; Uns.	Single; 2nd or 3rd; EFW US	< 10th PCTL	Individualised; customised growth chart	< 10th PCTL	Individualised; customised growth chart	266	41	258	1496	51 (46 to 55)	97 (96 to 98)
Ridha (2022)[283]	Singleton; Uns.	Single; 2nd or 3rd; EFW US	< 3rd PCTL	Individualised; customised growth chart	< 3rd PCTL	Individualised; customised growth chart	106	22	126	1829	46 (39 to 52)	99 (98 to 99)
Ridha (2022)[283]	Singleton; Uns.	Single; 2nd or 3rd; EFW US	3rd to 9th PCTL	Individualised; customised growth chart	3rd–9th PCTL	Individualised; customised growth chart	58	121	234	1648	20 (15 to 25)	93 (92 to 94)
Salomon (2005)[284]	Singleton; Uns.	Serial; any; FGP model	< 40	N/A	< 10th PCTL	Local; French growth curves, Leroy (1971)[340]	21	30	14	291	60 (42 to 76)	91 (87 to 94)
Sklar (2017)[285]	Triplet; Uns.	Single; 3rd; EFW US	< 10th PCTL (at least 1 foetus)	Local; Kramer (2001)[341]	< 10th PCTL, at least one neonate	Local; Canadian growth curves, Kramer (2001)[341]	15	0	12	51	56 (35 to 75)	100 (94 to 100)
Skrastad (2013)[286]	Singleton; Uns.; control group	Single; 2nd; BPD US and mean abdominal diameter	≤−5% growth deviation	Local; Eik-Nes (1982)(1983)[342,343]	≤−22% weight deviation	Local; Swedish growth curve, Maršál (1996)[307]	44	173	52	2967	46 (36, to 56)	94 (94 to 95)
Skrastad (2013)[286]	Singleton; Uns.; study group	Serial; multiple; BPD US and mean abdominal diameter	≤−5% growth deviation	Local; Eik-Nes (1982)(1983)[342,343]	≤−22% weight deviation	Local; Swedish growth curve, Maršál (1996)[307]	77	390	19	2704	80 (71 to 88)	87 (86 to 89)
Sovio (2015)[288]	Singleton; Uns.; selective US	Single; 3rd; EFW US	< 10th PCTL[h]	General; Hadlock (1991)[30]	< 10th PCTL	Local; UK growth chart, Freeman (1995)[328]	69	69	283	3556	20 (16 to 24)	98 (98 to 99)
Sovio (2015)[288]	Singleton; Uns.; selective US	Single; 3rd; EFW US	< 10th PCTL[h]	General; Hadlock (1991)[30]	< 3rd PCTL	Local; UK growth chart, Freeman (1995)UK growth chart, Freeman *et al.*, Arch Dis Child 1995; 73 : 17 – 24	28	110	59	3780	32 (23 to 43)	97 (97 to 98)
Sovio (2015)[288]	Singleton; Uns.; universal US	Single; 3rd; EFW US	< 10th PCTL (last US)	General; Hadlock (1991)[30]	< 10th PCTL	Local; UK growth chart, Freeman (1995)[328]	199	363	153	3262	57 (51 to 62)	90 (89 to 91)

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Sovio (2015)[288]	Singleton; Uns.; universal US	Single; 3rd; EFW US	< 10th PCTL (last US)	General; Hadlock (1991)[30]	< 3rd PCTL	Local; UK growth chart, Freeman (1995)UK growth chart, Freeman *et al.*, Arch Dis Child 1995; 73 : 17 – 24	67	495	20	3395	77 (67 to 85)	87 (86 to 88)
Temming (2017)[290]	Singleton; Uns.	Single; 2nd; EFW US	< 10th PCTL	General; Warsof (1977) before 20 0/7 weeks of ges-tation and Hadlock (1991) 20 0/7 weeks of gestation onward[30,25]	< 10th PCTL	General; Fenton (2013)[344]	109	246	975	11,453	10 (8 to 12)	98 (98 to 98)
Temming (2017)[290]	Singleton; Uns.	Single; 2nd; EFW US	< 10th PCTL	General; Warsof (1977) before 20 0/7 weeks of ges-tation and Hadlock (1991) 20 0/7 weeks of gestation onward[30,25]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	131	224	1132	11,296	10 (9 to 12)	98 (98 to 98)
Temming (2017)[290]	Singleton; Uns.	Single; 2nd; EFW US	< 10th PCTL	General; Warsof (1977) before 20 0/7 weeks of ges-tation and Hadlock (1991) 20 0/7 weeks of gestation onward[30,25]	< 10th PCTL	Local; US growth curves, Oken (2003)[319]	134	221	1268	11,160	10 (8 to 11)	98 (98, to 98)
Turitz (2014)[291]	Singleton; Uns.	Single; 3rd; EFW US and AC US	AC < 5th PCTL OR EFW < 10th PCTL	Local; Alexander (1996)[303]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	801	561	1075	8205	43 (40 to 45)	94 (93 to 94)
Turitz (2014)[291]	Singleton; Uns.	Single; 3rd; EFW US and AC US	AC < 5th PCTL OR EFW < 10th PCTL	Local; Alexander (1996)[303]	< 5th PCTL	Local; US growth curve, Alexander (1996)[303]	489	878	397	8878	55 (52 to 59)	91 (90 to 92)
Turitz (2014)[291]	Singleton; Uns.	Single; 3rd; EFW US and AC US	AC < 5th PCTL AND EFW < 10th PCTL	Local; Alexander (1996)[303]	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	510	184	1366	8582	27 (25 to 29)	98 (98 to 98)
Turitz (2014)[291]	Singleton; Uns.	Single; 3rd; EFW US and AC US	AC < 5th PCTL AND EFW < 10th PCTL	Local; Alexander (1996)[303]	< 5th PCTL	Local; US growth curve, Alexander (1996)[303]	338	351	548	9405	38 (35 to 41)	96 (96 to 97)

TABLE 20 Accuracy data *(continued)*

Study (year)	Pregnancy and participants	Testing strategy (configuration; timing; parameter)	Test threshold	Test reference chart	Threshold for neonate SGA	BW reference chart	TP	FP	FN	TN	Sens. (95% CI)	Spec. (95% CI)
Tuuli (2011)[292]	Singleton; Uns.	Multiple; 1st and 2nd; CRL US (1st tr.) and HC US (2nd tr.)	GA-adjusted 1st-to-2nd -tr. fetal growth *z*-score < −1	N/A	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	141	828	410	7054	26 (22 to 29)	89 (89 to 90)
Tuuli (2011)[292]	Singleton; Uns.	Single; 1st; CRL US	GA-adjusted *z*-score < −1	N/A	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	118	1308	433	6574	21 (18 to 25)	83 (83 to 84)
Tuuli (2011)[292]	Singleton; Uns.	Single; 2nd; EFW US	GA-adjusted *z*-score < −1	N/A	< 10th PCTL	Local; US growth curve, Alexander (1996)[303]	205	1143	346	6739	37 (33 to 41)	85 (85 to 86)
Wan (2022)[295]	Singleton; Uns.	Single; 3rd; EFW US and AC US	AC < 10th PCTL OR EFW < 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	Local; Australian reference chart, Dobbins (2012)[301]	471	492	859	11,512	35 (33 to 38)	96 (96 to 96)
Wan (2022)[295]	Singleton; Uns.	Single; 3rd; SFH	SFH 3 cm ≤ GA	N/A	< 10th PCTL	Local; Australian reference chart, Dobbins (2012)[301]	291	1020	1039	10,984	22 (20 to 24)	92 (91 to 92)
Wan (2022)[295]	Singleton; Uns.	Single; 3rd; SFH, EFW and AC US	SFH 3 cm ≤ GA OR EFW < 10th PCTL OR AC < 10th PCTL	Local; EFW: Hadlock (1991) customised for the Australian population, AC: Chitty (1994)[30,338]	< 10th PCTL	Local; Australian reference chart, Dobbins (2012)[301]	527	1224	803	10,780	40 (37 to 42)	90 (89 to 90)
Whitham (2023)[296]	Singleton; Uns.	Single; 3rd; EFW US	< 10th PCTL	Local; Duryea (2014)[345]	< 10th PCTL	General; Fenton (2013)[344]	21	7	69	1490	23 (15 to 33)	100 (99 to 100)
Whitham (2023)[296]	Singleton; Uns.	Single; 3rd; EFW US and AC US	AC < 10th PCTL OR EFW < 10th PCTL	General; Hadlock (1991)[30]	< 10th PCTL	General; Fenton (2013)[344]	27	38	63	1459	30 (21 to 41)	97 (97 to 98)

PCTL, percentile; Sens., sensitivity; Spec., specificity; Uns., unselected populations.
a The mothers whose trajectory line for AC alone or both AC and HC deviated to the right of the original centile line were assigned to the ‘antenatally growth-restricted’ group, and those whose trajectory line for AC alone or both AC and HC followed the original centile line or deviated to its left were assigned to the ‘antenatally non-growth restricted’ group.
b Serial FH measurements with clinically indicated ultrasonography.
c Routine ultrasonography in the 3rd trimester (28–30 and 34–36 weeks) combined with usual care (i.e. serial FH measurements with clinically indicated ultrasonography).
d Assumed but not reported.
e EFW US in high-risk pregnancies only, no US after 24 WG in low-risk pregnancies.
f Last US; test negative included no EFW after 24 weeks.
g Growth velocity, which triggered iatrogenic delivery (induction of labour or elective caesarean section)
h Last US, test negative includes no 3rd trimester US.

DOI: 10.3310/AJLK7403

Appendix 3 Details of excluded studies with rationale

To be included in the review, studies had to fulfil the following criteria:

Population	Pregnant people, with and without risk factors for a FGR/SGA fetus/neonate
Index test	(Q1) Iatrogenic delivery or other clinical intervention, following the identification of FGR by fetal growth monitoring
	(Q2) Fetal growth monitoring as published in practice guidelines
	(Q3 and Q4) Fetal growth monitoring methods which measure size (where the method of monitoring is defined as the complete process, including technology, timing, personnel and reference charts used) for the prediction of SGA/FGR
Reference standard	(Q3 and Q4) BW, with reference chart and threshold used to define SGA
Comparator	(Q1) No iatrogenic delivery or other clinical intervention for suspected FGR
	(Q2) SC before guideline implementation
Outcome	(Q1) Rates of stillbirth; rates of neonatal death; clinical outcomes (e.g. neonatal morbidity, brain injuries, unplanned NICU admission and parental morbidity) in relation to both the intervention (e.g. iatrogenic delivery) and the result of fetal growth monitoring (e.g. TP, FP, TN, FN for FGR); rates of pre-term iatrogenic delivery and GA at iatrogenic delivery
	(Q2) Rates of stillbirth, rates of neonatal death, clinical outcomes (e.g. neonatal morbidity, NICU admission and parental morbidity), rates of pre-term iatrogenic delivery and GA at iatrogenic delivery before and after guideline implementation
	(Q3 and Q4) Test accuracy (the numbers of TP, FN, FP and TN test results) for the target condition
	(Q4) Proportion of assessments where a measurement cannot be obtained or where the reported result is not considered adequate for clinical decision-making; reported reasons why a measurement cannot be obtained or where the reported result is not considered adequate for clinical decision-making
Study design	DTA study or any comparative study design (RCT, CCT or comparative observational study)

Appendix 3, *Table 21* summarises studies which were screened for inclusion based on full-text publication but did not fulfil one or more of the above criteria. The table shows which of the criteria each study fulfilled ('Y') and on which item it failed ('N') or was unclear.

TABLE 21 Details of excluded studies with reasons for exclusion

Study details	Study design	Population	Index test/intervention	Comparator	Reference standard	Outcome
Abulhaj Martínez (2012)[49]	Y	Y	N	N/A	N	N
ACTRN (2012)[50]	Y	Y	N	N	N	N
Adam (2014)[51]	Y	Y	Y	N/A	N	N
Adams (2020)[52]	Y	Y	N	N	N	N
Aderoba (2023)[53]	Y	Y	N	Y	N/A	N
Aiken (2019)[54]	Y	Y	Y	N/A	Y	N
Aiken (2019)[54]	N	N	N	N	N	N
Akolekar (2019)[55]	Y	Y	N	N/A	Y	N
Ali (2017)[56]	Y	Y	N	N/A	N	Y
Andreasen (2020)[57]	Y	Y	Y	Y	N	N
Andrews (2022)[58]	Y	Y	N	Y	N	N
Angarita (2022)[59]	Y	Y	Y	N/A	Y	N
Ashfaq (2022)[60]	Y	Y	Y	N/A	N	Y
Ashfaq (2022)[61]	Y	Y	N	N/A	N	Y
Bailey (2022)[62]	Y	Y	Y	N/A	Y	N
Barel (2016)[63]	Y	Y	N	N/A	N	N
Basuki (2019)[64]	Y	Y	N	N	N	N
Beddow (2018)[65]	Y	Y	Y	N/A	Y	N
Bellussi (2017)[66]	Y	Y	Y	N/A	Y	N
Bergman (2006)[67]	Y	Y	Y	N/A	Y	N
Bhimarao (2015)[68]	Y	Y	Y	N/A	N	N
Bikmetova (2013)[69]	Y	Y	Y	N/A	N	N
Blue (2018)[70]	Y	Y	Y	N/A	Y	N
Blue (2018)[71]	Y	Y	Y	N/A	Y	N
Bocchi (2011)[72]	Y	Y	N	N/A	N	N
Burd (2009)[73]	N	N	N	N	N	N

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 21 Details of excluded studies with reasons for exclusion *(continued)*

Study details	Study design	Population	Index test/intervention	Comparator	Reference standard	Outcome
Callec (2015)[74]	Y	Y	N	N/A	N	N
Caradeux (2018)[75]	N	Y	N	N/A	Y	N
Carlin (2019)[76]	Y	Y	Y	N/A	Y	N
Cassells (2021)[77]	Y	Y	Y	N/A	Y	N
Chauhan (2006)[78]	Y	Y	N	N/A	N	Y
Chauhan (2012)[79]	Y	Y	Y	Y	Y	N
Chauhan (2011)[80]	Y	Y	Y	N/A	Y	N
Chen (2022)[81]	Y	N	Y	N/A	Y	N
Choi (2021)[82]	Y	Y	Y	N/A	N	N
Ciobanu (2019)[83]	Y	Y	Y	N/A	Y	N
Ciobanu (2019)[84]	Y	Y	Y	N/A	Y	N
Cordiez (2017)[85]	Y	N	Y	N/A	Y	Y
Craig (2021)[86]	N	N	Y	N/A	Y	N
De Silva (2017)[87]	Y	Y	Y	N/A	Y	N
de Sousa Basso (2016)[88]	Y	Y	N	N/A	N	N
Doulaveris (2020)[89]	Y	Y	Y	N/A	N	N
Dowe (2016)[90]	N	Y	Y	N	N/A	N
Dromey (2019)[91]	N	Y	Y	N	N/A	N
Dude (2018)[92]	Y	Y	Y	N/A	Y	N
Dude (2017)[93]	Y	Y	Y	N/A	Y	N
Febres-Cordero (2022)[94]	Y	Y	Y	N/A	N	N
Figueras (2022)[95]	N	N	N	N/A	N	N
Finneran (2020)[96]	Y	N	N	N/A	N	N
Fitzgerald (2013)[97]	Y	Y	Y	N	Y	N
Francis (2016)[98]	Y	Y	Y	N/A	Y	N
Francis (2016)[99]	Y	Y	Y	N/A	Y	N
Fuchs (2012)[100]	Y	Y	Y	N	N/A	N

TABLE 21 Details of excluded studies with reasons for exclusion (*continued*)

Study details	Study design	Population	Index test/intervention	Comparator	Reference standard	Outcome
Fuchs (2012)[100]	N	N	N	N	N	N
Gabbay-Benziv (2017)[101]	N	N	N	N/A	Y	Y
Gabbay-Benziv (2016)[102]	Y	Y	N	N/A	N	N
Gabbay-Benziv (2016)[103]	Y	Y	Y	N/A	Y	N
Gabbay-Benziv (2016)[103]	N	N	N	N	N	N
Gardosi (2013)[104]	Y	Y	Y	Y	N	N
Gardosi (2023)[105]	Y	Y	N	N	Y	N
Gawie-Rotman (2023)[106]	Y	N	N	N	N	N
Gelman (2022)[107]	N	Y	Y	N/A	Y	N
Gilmore (2017)[108]	N	N	Y	N/A	Y	N
Goetzinger (2013)[109]	Y	Y	N	N/A	N	N
Goldshore (2013)[110]	Y	Y	Y	N/A	Y	N
Grantz (2023)[111]	Y	Y	N	N	N	N
Green (2017)[112]	Y	Y	Y	N/A	Y	N
Grzybek (2019)[113]	Y	Y	Y	N/A	Y	N
Gupta (2008)[114]	Y	Y	Y	N/A	Y	N
Hebbar (2018)[115]	Y	Y	Y	N/A	N	N
Hairston (2022)[116]	Y	Y	Y	N/A	Y	N
Haragan (2015)[117]	Y	Y	Y	N/A	Y	N
Honart (2021)[118]	Y	Y	Y	N/A	Y	N
Horwitz (2021)[119]	Y	Y	Y	N/A	Y	N
Iliodromiti (2020)[120]	Y	Y	Y	N	N	N
Ivars (2010)[121]	Y	Y	Y	N/A	Y	N
Kabiri (2020)[122]	Y	Y	Y	N/A	N	N
Khan (2004)[123]	Y	Y	N	N/A	Y	N
Kim (2019)[124]	Y	Y	N	N/A	N	N

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 21 Details of excluded studies with reasons for exclusion (*continued*)

Study details	Study design	Population	Index test/intervention	Comparator	Reference standard	Outcome
Konwar (2021)[125]	Y	Y	N	N/A	N	N
Kopparam (2019)[126]	Y	Y	Y	N/A	Y	N
Kose (2019)[9]	Y	N	N	N	N	N
Landres (2013)[127]	N	N	N	N	N	N
Landres (2011)[128]	Y	N	Y	N/A	N	N
Larkin (2017)[129]	N	N	Y	N/A	Y	N
Larkin (2015)[130]	Y	Y	Y	N/A	Y	N
Law (2009)[131]	N	Y	N	N/A	Y	N
Leon-Martinez (2023)[132]	Y	Y	Y	Y	Y	N
Lesmes (2015)[133]	N	Y	N	N/A	Y	N
Luo (2001)[134]	Y	Y	Y	N/A	Y	N
Ma (2023)[135]	Y	Y	Y	Y	N	N/A
Mailath-Pokorny (2015)[136]	Y	Y	N	N/A	N	N
Maines (2020)[137]	Y	N	N	N	N	N
Malin (2014)[138]	Y	Y	Y	N/A	Y	N
Matthews (2019)[139]	Y	Y	Y	N/A	Y	N
McKenna (2003)[140]	Y	Y	N	Y	N/A	Y
McKenna (2004)[141]	Y	Y	Y	Y	Y	N
McKeown (2018)[142]	Y	Y	Y	N/A	Y	N
Mendoza-Carrera (2021)[143]	Y	Y	Y	N/A	N	N
Merriam (2014)[144]	N	N	Y	N/A	N	N
Meyer (2020)[145]	Y	Y	Y	N/A	Y	N
Michaeli (2020)[146]	Y	Y	Y	N	Y	N
Michaeli (2020)[146]	Y	Y	Y	Y	Y	N
Miranda (2017)[147]	N	Y	N	N/A	Y	N
Miranda (2023)[148]	Y	Y	N	Y	N	Y
Monier (2017)[149]	Y	Y	N	N/A	Y	N

TABLE 21 Details of excluded studies with reasons for exclusion (*continued*)

Study details	Study design	Population	Index test/intervention	Comparator	Reference standard	Outcome
Mori (2020)[150]	Y	Y	N	N/A	N	N
Muller (2021)[151]	Y	Y	Y	N/A	N	N
Nadham (2019)[152]	Y	Y	Y	N/A	Y	N
Najafzadeh (2016)[153]	N	N	N	N	N	N
Neel (2021)[154]	Y	Y	N	N/A	N	N
Odibo (2018)[155]	Y	Y	N	N/A	N	N
Odibo (2018)[156]	Y	Y	N	N/A	N	N
Ott (2002)[157]	Y	Y	N	N/A	N	N
Owen (2003)[158]	Y	Y	Y	N/A	N	N
Owen (2000)[159]	Y	Y	Y	N/A	N	Y
Padilla-Amigo (2022)[160]	Y	Y	N	N/A	N	Y
Palaniappan (2013)[161]	Y	Y	Y	N/A	Y	N
Papastefanou (2022)[162]	N	Y	Y	N/A	Y	Y
Papastefanou (2021)[163]	N	Y	Y	N	Y	N
Papastefanou (2021)[164]	Y	Y	N	N/A	Y	N
Papastefanou (2021)[165]	N	Y	N	N/A	Y	N
Papastefanou (2015)[166]	Y	Y	Y	N/A	Y	N
Patel (2020)[167]	N	Y	Y	N/A	Y	N
Paulsen (2019)[168]	Y	Y	N	N	Y	Y
Pay (2017)[169]	Y	Y	Y	N/A	Y	N
Pay (2016)[170]	N	Y	N	N/A	Y	Y
Pils (2018)[171]	Y	Y	Y	N/A	Y	N
Policiano (2022)[172]	Y	Y	Y	Y	Y	Y
Prechapanich (2004)[173]	Y	Y	Y	N/A	Y	N
Premru-Srsen (2019)[174]	Y	Y	N	N/A	N	N

continued

DOI: 10.3310/AJLK7403

Health Technology Assessment 2025 Vol. 29 No. 62

TABLE 21 Details of excluded studies with reasons for exclusion (*continued*)

Study details	Study design	Population	Index test/intervention	Comparator	Reference standard	Outcome
Priya Appanamuthu Mari (2021)[175]	Y	Y	N	N/A	Y	N
Procas-Ramon (2021)[176]	N	N	Y	N/A	Y	Y
Quinton (2015)[177]	Y	Y	Y	N/A	Y	N
Radhika (2015)[178]	Y	Y	Y	N/A	Y	N
Ravangard (2014)[179]	Y	Y	Y	N/A	Y	N
Reboul (2017)[180]	N	Y	N	N/A	Y	N
Roma (2015)[181]	Y	Y	Y	Y	N	N
Safonova (2016)[182]	Y	Y	Y	N/A	Y	N
Sanchez Fernandez (2015)[183]	Y	Y	N	N/A	N	N
Sandlin (2012)[184]	N	N	N	N/A	Y	N
Saviron-Cornudella (2021)[185]	N	Y	Y	N/A	Y	N
Schreiber (2023)[186]	Y	Y	N	N/A	N	N
Schreiber (2022)[187]	Y	Y	Y	N/A	Y	N
Scott (2019)[188]	Y	Y	N	N/A	Y	N
Sekar (2016)[189]	Y	Y	Y	N/A	N	N
Severi (2011)[190]	N	Y	Y	N/A	Y	N
Shen (2017)[191]	Y	N	N	N/A	N	N
Shmueli (2017)[192]	Y	Y	N	N/A	N	N
Sim (1993)[193]	Y	Y	N	Y	N	N
Sinkey (2017)[194]	Y	Y	Y	N/A	Y	N
Sokol Karadjole (2017)[195]	Y	Y	Y	N/A	Y	N
Sotiriadis (2019)[196]	Y	Y	N	N/A	Y	N
Souka (2012)[197]	N	Y	N	N/A	Y	N
Sovio (2021)[198]	Y	Y	Y	N/A	Y	N
Sparks (2011)[199]	Y	Y	N	N/A	Y	Y
St (2016)[200]	N	N	N	N	N	N

TABLE 21 Details of excluded studies with reasons for exclusion (*continued*)

Study details	Study design	Population	Index test/intervention	Comparator	Reference standard	Outcome
Stanger (2013)[201]	Y	Y	N	N/A	N	N
Stirnemann (2014)[202]	Y	Y	Y	N/A	Y	N
Stout (2016[203]	Y	Y	Y	N/A	Y	N
Strassberg (2019[204]	N	N	Y	Y	N	Y
Swanzy (2017)[205]	Y	N	N	N/A	Y	N
Temming (2017)[206]	Y	N	Y	N/A	Y	N
Thiebaugeorges (2000)[207]	Y	Y	Y	N	Y	N
Trudell (2013)[208]	N	Y	N	N/A	N	N
Turitz (2013)[209]	Y	Y	Y	N/A	Y	N
Turitz (2013)[210]	Y	Y	Y	N/A	Y	N
Turner (2019)[211]	Y	Y	Y	Y	N/A	N
Turner (2016)[212]	Y	Y	Y	N/A	Y	N
Turner (2019)[213]	Y	Y	Y	Y	N/A	N
Turner (2016)[214]	Y	Y	Y	Y	N/A	N
Van De Kamp (2019)[215]	Y	Y	Y	Y	Y	N
Verger (2020)[216]	Y	Y	Y	N/A	Y	N
Vieira (2014)[217]	Y	Y	Y	N/A	Y	N
Walid (2009)[218]	Y	Y	Y	N/A	Y	N
Wanyonyi (2021)[219]	Y	Y	N	Y	N	N
Wright (2020)[220]	N	N	N	N/A	N	N
Yordan (2018)[221]	Y	Y	Y	N/A	Y	N
Yoshida (2000)[222]	Y	Y	Y	N/A	Y	Y

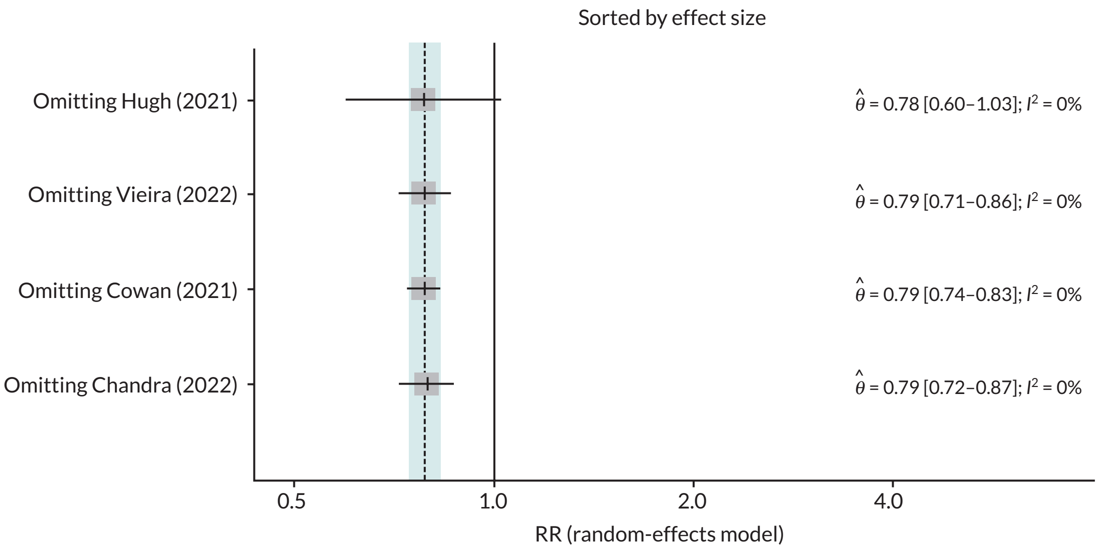
DOI: 10.3310/AJLK7403

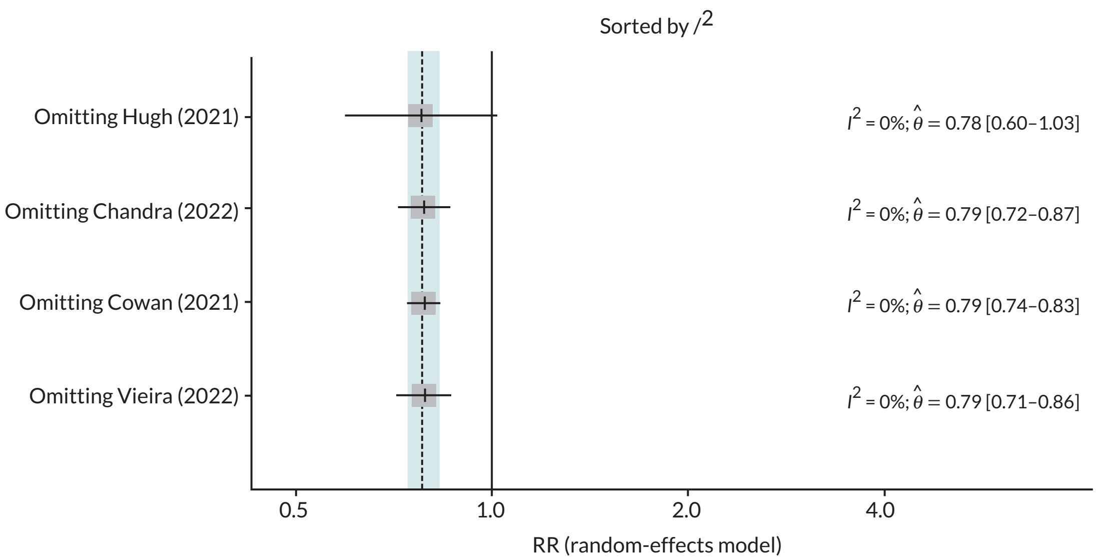
Health Technology Assessment 2025 Vol. 29 No. 62

Appendix 4 Influence analysis results

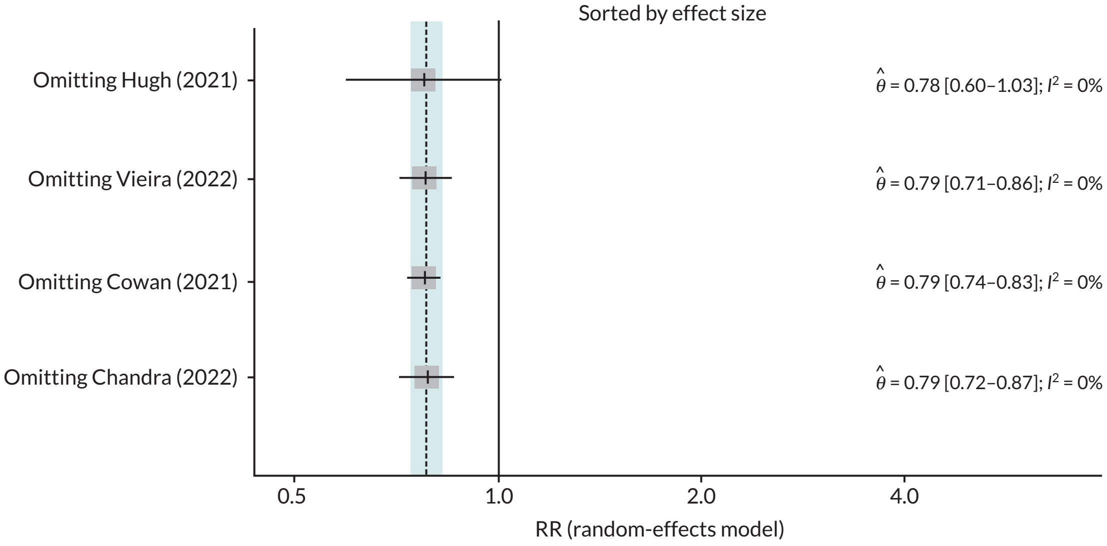
Influence analysis, leave-one-out method for the effect of GAP implementation on stillbirth outcomes; RR

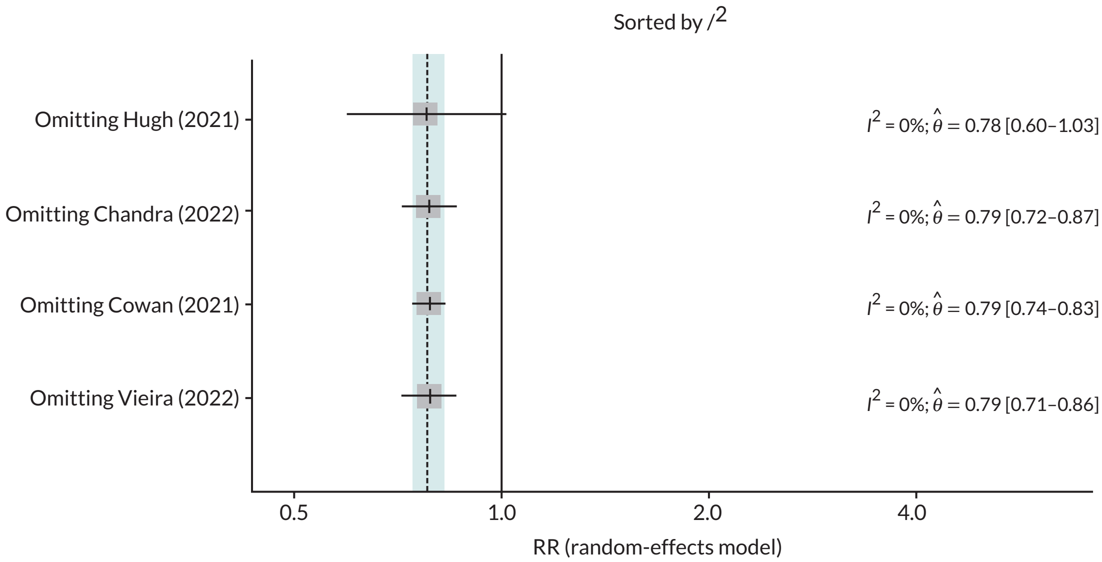
Influence analysis, leave-one-out method for the effect of GAP implementation on induction of labour outcomes; OR

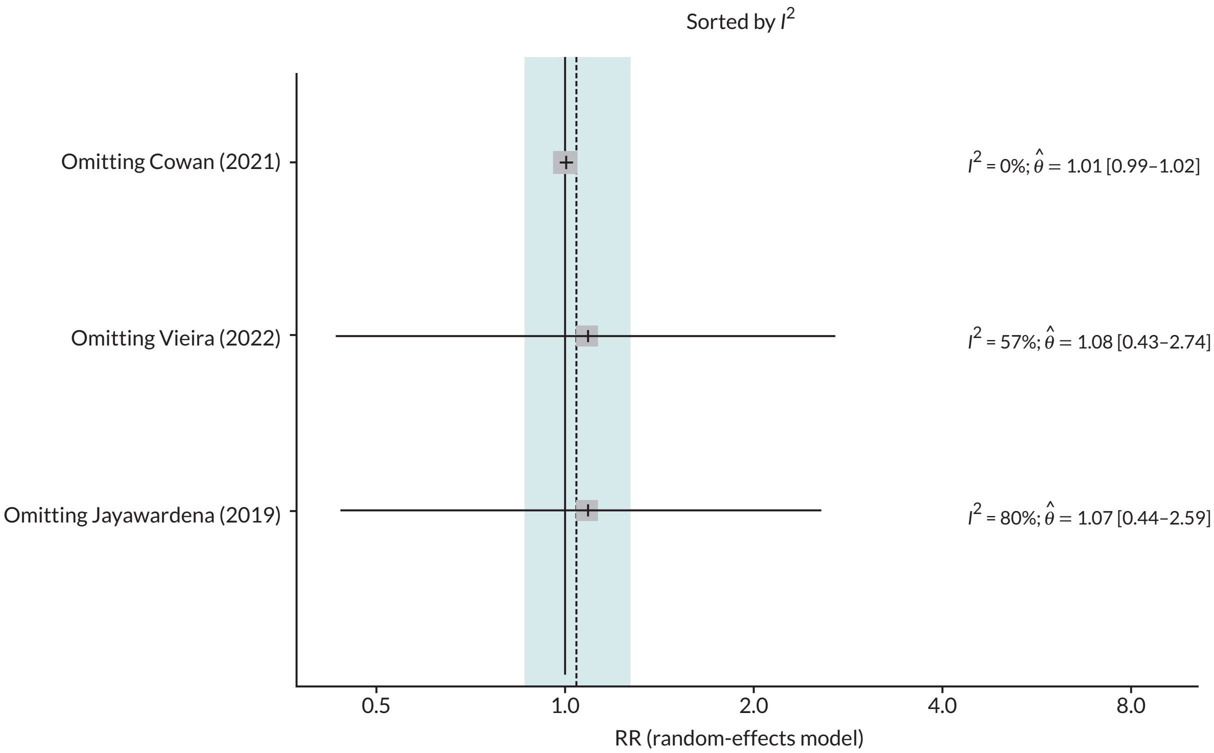


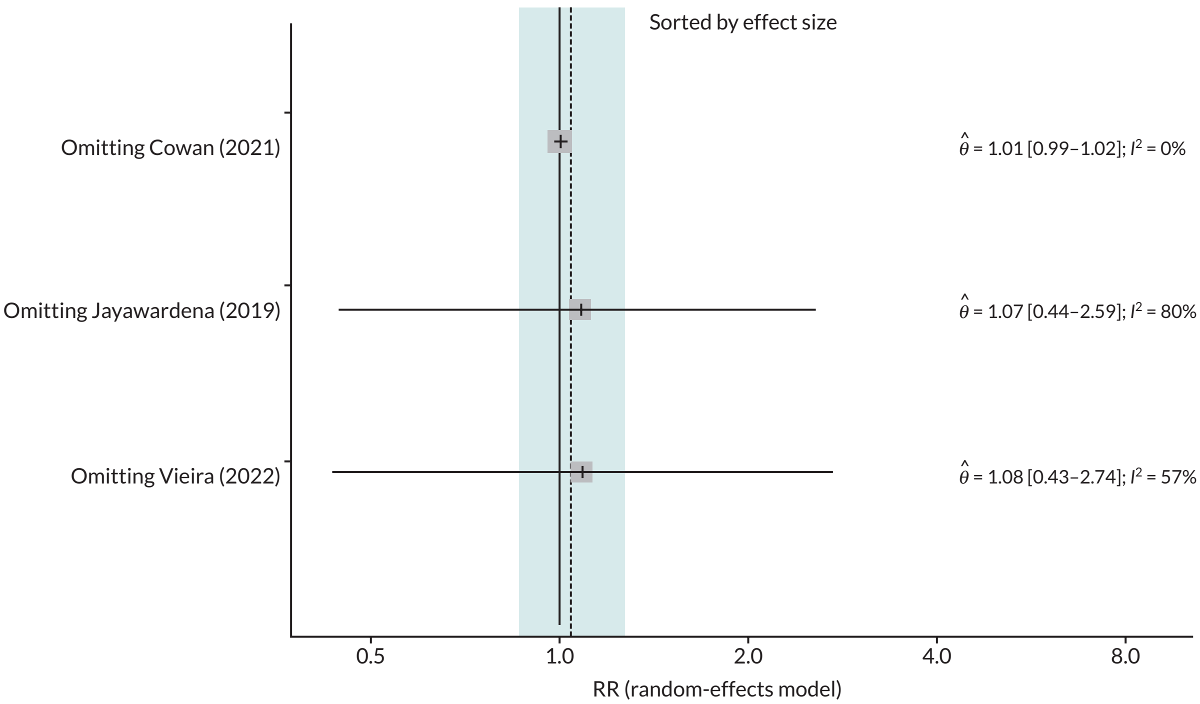


Influence analysis, leave-one-out method for the effect of GAP implementation on caesarean section operations outcomes; RR






Sorted by I^2
Omitting Cowan (2021)
Omitting Vieira (2022)
Omitting Jayawardena (2019)
I^2 = 0%; $\hat{\theta}$ = 1.01 [0.99–1.02]
I^2 = 57%; $\hat{\theta}$ = 1.08 [0.43–2.74]
I^2 = 80%; $\hat{\theta}$ = 1.07 [0.44–2.59]
0.5
1.0
2.0
4.0
8.0
RR (random-effects model)


Sorted by effect size
Omitting Cowan (2021)
Omitting Jayawardena (2019)
Omitting Vieira (2022)
$\hat{\theta}$ = 1.01 [0.99–1.02]; I^2 = 0%
$\hat{\theta}$ = 1.07 [0.44–2.59]; I^2 = 80%
$\hat{\theta}$ = 1.08 [0.43–2.74]; I^2 = 57%
0.5
1.0
2.0
4.0
8.0
RR (random-effects model)

EME
HSDR
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 and Care 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