National Institute for Health and Care Excellence

Draft for consultation

Antenatal Care

[O] Monitoring fetal growth

NICE guideline <number>

Evidence reviews underpinning recommendations 1.2.26 to 1.2.29

February 2021

Draft for consultation

These evidence reviews were developed by the National Guideline Alliance which is a part of the Royal College of Obstetricians and Gynaecologists

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN:

Contents

Monitoring fetal growth

Review question

What is the best method using third trimester measurements to predict birth weight?

Introduction

In the UK, it is current practice for women with low risk pregnancies to have symphysis-fundal height (SFH) measurements during the third trimester to monitor growth of the baby. Routine ultrasound is not current practice. This question aims to compare which technique is most accurate in monitoring fetal growth.

Summary of the protocol

See Table 1 for a summary of the Population, Index test, Reference standard and Outcomes and prioritisations of this review.

Table 1: Summary of the protocol

Population	All women with unselected or low-risk pregnancies
Index tests	The use of the following third trimester measurements, individually or in combination, to predict birth weight at birth: • Fetal biometry using ultrasound • Symphysis-fundal height (SFH) measurement
Reference standard	Reference standard is either • Relevant birth weight centile Or, if no chart is used, reference standard is: • Actual absolute birth weight threshold
Outcomes and prioritisation	Critical • Sensitivity for detecting SGA and LGA • Specificity for detecting SGA and LGA Important • Positive predictive value for SGA and LGA • Negative predictive value for SGA and LGA Predictive values calculated following meta-analysis of sensitivity and specificity using prevalences from most representative single studies

LGA: large for gestational age; SGA: small for gestational age.

For further details see the review protocol in appendix A.

Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual 2014. Methods specific to this review question are described in the review protocol in appendix A.

Declarations of interest were recorded according to NICE's conflicts of interest policy.

Clinical evidence

Included studies

Nineteen studies were included in this review, 11 retrospective cohort studies (Aviram 2017, Barel 2016, Ben-Haroush 2007, Blue 2018, Blue 2019, Callec 2015, Gabbay-Benziv 2016, Khan 2019, Lin 1990, Rad 2018, Turitz 2014); 6 prospective cohort studies (Akolekar 2019, Erkamp 2020, Sekar 2016, Skovron 1991, Sovio 2015, Sovio 2018); 1 nested case-control study (Harding 1995); and 1 population based study (Monier 2015).

The included studies are summarised in Table 2.

Two studies were conducted in Australia (Harding 1995, Sekar 2016); 1 study was conducted in France (Callec 2015); 3 studies were conducted in Israel (Aviram 2017, Barel 2016, Ben-Haroush 2007); 4 studies were conducted in the UK (Akolekar 2019, Khan 2019, Sovio 2015, Sovio 2018); 1 study was conducted in the Netherlands (Erkamp 2020), 6 studies were conducted in USA (Blue 2018, Blue 2019, Lin 1990, Rad 2018, Skovron 1991, Turitz 2014). One study did not mention which country it was conducted in (Gabbay-Benziv 2016).

Two additional studies (Bardin 2020, Duncan 2020) were identified in final update searches for the review that met the protocol inclusion criteria but did not affect the evidence base or draft recommendations. The searches were initially updated in May 2020 but due to the atypical prolongation of guideline development due to COVID-19 pandemic, the searches were updated again in September 2020. New evidence identified in this final update search which did not impact on the conclusions were not fully included in the report but are referenced in appendix M.

See the literature search strategy in appendix B and study selection flow chart in appendix C.

Excluded studies

Studies not included in this review are listed, and reasons for their exclusion are provided in appendix K.

Summary of studies included in the evidence review

Summaries of the studies that were included in this review are presented in Table 2.

Table 2: Summary of included studies

Study	Population	Index test	Reference standard	Outcomes
Akolekar 2019 Prospective cohort study UK	N=45 847 singleton pregnancies	Ultrasound estimated fetal weight <10th percentile (Hadlock formula) >7d from delivery	Birth weight <10th percentile for gestational age based on the fetal medicine foundation, fetal and neonatal population weight charts	• Sensitivity for SGA • Specificity for SGA
Aviram 2017 Retrospective cohort study Israel	N=7 996 singleton pregnancies	Different ultrasound tests (20 variations) <7d from delivery	Birth weight >90th percentile for gestational age	• Sensitivity for LGA • Specificity for LGA

Study	Population	Index test	Reference standard	Outcomes
Barel 2016 Retrospective cohort study Israel	N=14 089 singleton pregnancies	Ultrasound estimated fetal weight <10th percentile (Hadlock formula) <7d from delivery	Birth weight <10th percentile for gestational age based on actual birth weight from departmental computerised database	• Sensitivity for SGA • Specificity for SGA
Ben-Haroush 2007 Retrospective cohort study Israel	N=259 women	Ultrasound estimated fetal weight ≤10th and ≥90th percentile (Hadlock formula) >7d from delivery	Birth weight ≤10th and ≥90th percentile for gestational age	• Sensitivity for SGA • Specificity for SGA • Sensitivity for LGA • Specificity for LGA
Blue 2018 Retrospective cohort study US	N=1 704 singleton pregnancies	Ultrasound estimated fetal weight <10th percentile (Hadlock formula) >7d from delivery	Birth weight <10th percentile for gestational age	• Sensitivity for SGA • Specificity for SGA
Blue 2019 Retrospective cohort study US	N=831 singleton pregnancies	Ultrasound estimated fetal weight <10th percentile (Hadlock formula) <7d from delivery	Birth weight <10th percentile for gestational age	• Sensitivity for SGA • Specificity for SGA
Callec 2015 Prospective cohort study France	N=1 897 pregnant women	Ultrasound estimated fetal weight <10th percentile (Hadlock formula) >7d from delivery	Birth weight <10th percentile for gestational age	• Sensitivity for SGA • Specificity for SGA
Erkamp 2020 Prospective cohort study Netherlands	N=7 677 pregnant women	Ultrasounds estimated fetal weight <10th percentile, >90th percentile >7 days from delivery	Birth weight <10th percentile for gestational age, >90th percentile for gestational age	• Sensitivity for SGA • Specificity for SGA • Sensitivity for LGA • Specificity for LGA
Gabbay-Benziv 2016 Retrospective cohort study Country not reported	N=6 126 pregnant women	Different ultrasound tests (20 variations) <7d from delivery	Birth weight >90th percentile for gestational age	• Sensitivity for LGA • Specificity for LGA

Study	Population	Index test	Reference standard	Outcomes
Harding 1995 Nested case-control study Australia	N=1 135 pregnant women	SH <10th percentile for GA >7d from delivery	Birth weight <10th percentile for GA using charts constructed from the Western Australia population	• Sensitivity for SGA • Specificity for SGA
Khan 2019 Retrospective cohort study UK	N=67 836 singleton pregnancies	Ultrasound estimated fetal weight >90th percentile (Hadlock formula) >7d from delivery	Birth weight >90th percentile for gestational age based on the fetal medicine foundation fetal and neonatal population weight charts	• Sensitivity for LGA • Specificity for LGA
Lin 1990 Retrospective cohort study US	N=463 pregnant women	Ultrasound AC <10th percentile (Shepards equation) >7d from delivery	Birth weight <10th percentile for gestational age	• Sensitivity for SGA • Specificity for SGA
Monier 2015 Population-based study France	N=14 100 live and stillbirths	Ultrasound (defined as suspicion of FGR during pregnancy in the medical notes) >7d from delivery	Birthweight <10th centile for gestational age	• Sensitivity for SGA • Specificity for SGA
Rad 2018 Retrospective cohort study US	N=1 594 pregnancies	Ultrasound estimated fetal weight <10th percentile (Hadlock formula) >7d from delivery	Birth weight <10th percentile for gestational age	• Sensitivity for SGA • Specificity for SGA
Sekar 2016 Prospective cohort study Australia	N=150 pregnant women	Ultrasound estimated fetal weight <10th percentile or >90th percentile (Hadlock) <7d from delivery	Birth weight <10th percentile or >90th percentile	• Sensitivity for SGA • Specificity for SGA • Sensitivity for LGA • Specificity for LGA
Skovron 1991 Prospective cohort study US	N=768 pregnant women	Ultrasound estimated fetal weight (Shepards formula) and AC <10th percentile for GA >7d from delivery	Birthweight <10th percentile	• Sensitivity for SGA • Specificity for SGA
Sovio 2015	N=4 512 pregnant women	Ultrasound estimated fetal weight <10th	Birth weight <10th percentile for gestational age	• Sensitivity for SGA • Specificity for SGA

Study	Population	Index test	Reference standard	Outcomes
Prospective cohort study UK		percentile (Hadlock formula) >7d from delivery		
Sovio 2018 Prospective cohort study UK	N=4 512 pregnant women	Ultrasound estimated fetal weight >90th percentile (Hadlock formula) >7d from delivery	Birth weight >90th percentile for gestational age	• Sensitivity for LGA • Specificity for LGA
Turitz 2014 Retrospective cohort study US	N=10 642 singleton pregnancies	Ultrasound estimated fetal weight <10th percentile (Hadlock formula) >7d from delivery	Birth weight <10th percentile for gestational age based on the Alexander curve (a national reference nomogram)	• Sensitivity for SGA • Specificity for SGA

AC: abdominal circumference; FGR: fetal growth restriction; GA: gestational age; SH: symphysis-fundal height.

See the full evidence tables in appendix D and the forest plots in appendix E.

Quality assessment of studies included in the evidence review

See the evidence profiles in appendix F.

Included studies

A systematic review of the economic literature was conducted but no economic studies were identified which were applicable to this review question.

A single economic search was undertaken for all topics included in the scope of this guideline. See supplementary material 2 for details.

Excluded studies

There was no economic evidence identified for this review question and therefore there is no excluded studies list in appendix K.

Summary of included economic evidence

No economic studies were identified which were applicable to this review question.

Economic model

No economic modelling was undertaken for this review because the committee agreed that other topics were higher priorities for economic evaluation.

Evidence statements

Clinical evidence statements

Ultrasound done more than 7 days before delivery had poor sensitivity for small-for-gestational age (SGA; 11 studies, low quality evidence) and very good specificity for SGA

(12 studies, moderate quality evidence). When ultrasound was done fewer than 7 days before delivery the sensitivity for SGA remained poor but was slightly improved (4 studies, low quality evidence) and the specificity for SGA was very good (4 studies, high quality evidence).

Ultrasound done more than 7 days before delivery had poor sensitivity for large-for-gestational age (LGA; 4 studies, moderate quality evidence) and very good specificity for LGA (4 studies, very low quality evidence). When ultrasound was done fewer than 7 days before delivery the sensitivity for LGA remained poor but was improved (2 studies, low quality evidence) and the specificity for LGA was very good (2 studies, low quality evidence).

Symphysis-fundal height measurements done more than 7 days before delivery had very poor sensitivity for SGA (1 study, high quality evidence) and moderate specificity for SGA (1 study, high quality evidence).

The committee's discussion of the evidence

Interpreting the evidence

The outcomes that matter most

The committee agreed that they would prioritise sensitivity over specificity for this diagnostic test accuracy review. They considered the impact of true positives (correctly identifying SGA/LGA babies and allowing for appropriate management to be in place for their birth), true negatives (reassuring mothers of babies who are appropriate-for-gestational age), false positives (potentially promoting definitive interventions that are unnecessary – for example earlier induction for LGA babies) and false negatives (failing to identify babies that may require more intensive monitoring and peripartum care) and noted that false negatives could be particularly impactful – hence a particular need to focus on the sensitivity of tests. The committee considered the positive and negative predictive values as additional information alongside sensitivity and specificity in order to allow them to understand what the impact of a system that recommended a certain action for all positive or negative test results would have.

The quality of the evidence

The quality of the evidence ranged from high to very low, typically evidence was downgraded for issues relating to inconsistency and imprecision. The inconsistency may have been driven in parts by subtly different approaches to imaging (for example using different variants of formulae available for estimating size on ultrasound) although the criteria for the review were chosen to minimise this.

The committee noted that there was very little evidence available on the accuracy of SFH measurements (one small study looking at SGA and no studies looking at LGA) and particularly on the accuracy of repeated measurements as opposed to a one-off assessment.

Benefits and harms

The evidence showed that ultrasound is not very sensitive for SGA or LGA, though it is more sensitive when done closer to delivery and it is more sensitive than SFH measurement – although this outcome is based on a single small study. Ultrasound was quite specific for both SGA and LGA, again this was improved if done closer to delivery and again more accurate than for SFH measurement.

Overall the evidence suggested that neither of the main modalities for assessing fetal growth were particularly accurate, with sensitivity being particularly poor (at best a point estimate of 70% for ultrasound done for LGA less than 7 days from delivery).

The results of this review were interpreted alongside evidence review Q on routine third trimester ultrasound for fetal growth. That review broadly concluded that routinely ultrasound scanning all women in the third trimester (as opposed to selectively scanning those in whom there were concerns or clinical suspicions of adverse outcomes) did not convey a clinically important benefit. Selective scanning criteria in low risk pregnancies vary but typically were at least partially informed by SFH measurements. While the evidence in this review suggested that SFH measurement is not very sensitive, SFH measurement is easily performed with little resource implications and essentially no adverse effects (in terms of the test itself, inaccurate results will still have adverse effects). If SFH measurement was not done routinely, it would make the selective choice of who should receive an ultrasound scan more challenging. Therefore overall the committee agreed, despite the (limited) evidence of low sensitivity, it was appropriate to offer SFH measurement at each antenatal appointment after 24+0 weeks unless the woman is already undergoing regular growth scans (in which case there would be no additional benefit).

The committee were aware that there are some risk factors for fetal growth restriction and agreed that a risk assessment should be done in early pregnancy (at booking appointment) when all pre and early pregnancy risk factors could be considered and again in the second trimester, when other risk factors may have become apparent (for example gestational hypertension). The committee were aware of available risk assessment tools, such as those in the Saving Babies Lives Care Bundle version 2 (2019) and RCOG Green-Top guideline on investigation and management of small-for-gestational age fetus (2013).

The committee also made informal consensus based recommendations about the response to concerns about babies being either SGA or LGA as per SFH measurement. For babies possibly being SGA, the committee agreed an ultrasound was required as being SGA may be associated with critical adverse outcomes including stillbirth that could require intervention of some kind. The urgency of this ultrasound would be dictated by the overall clinical findings and whether or not there were other reasons to be concerned about the wellbeing of the baby (for example a reduction in fetal movements) or mother (for example raised blood pressure or proteinuria). If there were concerns about babies being LGA, the committee made a weaker recommendation to consider an ultrasound (for example to check for volume of amniotic fluid), however, LGA is less commonly associated with critical adverse outcomes such as stillbirth and may not warrant further investigation or intervention (particularly if the baby has been consistently LGA as opposed to changing growth trajectories), although LGA increases the risk of for example shoulder dystocia.

Cost effectiveness and resource use

It was noted that diagnostic outcomes would not in themselves lend to recommendations for routine ultrasound scanning as the committee were of the view that such recommendations should be made in conjunction with clinical outcomes such as stillbirth and NICU admission. Therefore, the committee discussed this topic in conjunction with the clinical and economic evidence included in Evidence Review Q (Wastlund 2019). This study included diagnostic outcomes which were identified in this clinical review (Sovio 2018) and assessed the sensitivity and specificity of routine and selective ultrasound for identifying LGA and SGA.

The sensitivity of ultrasound was generally poor from studies included in the accompanying clinical review. The implications for cost effectiveness should be viewed in the context of the consequential management strategies of the diagnostic outcomes. Increasing true positives or reducing false negatives will impact on costs and effects if women receive appropriate treatment. For example, a lower false positive would reduce the costs and harms associated with unnecessary emergency caesarean sections. As the clinical review conducted in Evidence Review Q demonstrated no important differences between routine and selective ultrasound, the committee did not recommend routine ultrasound testing as it would likely not be a cost effective use of resources given the increase in cost and no improvement in clinical

outcomes. This conclusion was supported by economic evidence (Wastlund 2019) in evidence review Q.

Mindful of the substantial costs in routine provision, the committee's recommendation to offer symphysis fundal height measurement to all women reflects current practice. The committee highlighted that many trusts in England currently offer routine ultrasound for SGA detection. Where these recommendations lead to a reduction in routine ultrasound testing then a significant cost saving will be achieved.

Other factors the committee took into account

Evidence review Q showed that there was no important benefit of routine ultrasound assessment as opposed to selective assessment on clinically important outcomes. This review assessed accuracy of tests including ultrasound. The more accurate ultrasound is, the more likely that its routine use could have benefits. However if the subsequent management of the various possible diagnoses does not result in benefit then simply having an accurate test will not lead to better downstream outcomes.

The committee also noted that, while this is not an outcome in the protocol of evidence review Q, studies included in that review showed a higher detection rate of SGA/LGA cases in the routine arm compared with the selective arm. This is logical as unless ultrasound had 0% sensitivity or the choice of who should receive selective ultrasound was perfect (in other words all cases received ultrasound) detection rate will always be greater to some degree. The fact that the overall conclusion of review Q was that routine ultrasound did not convey a clinical benefit over selective ultrasound, shows that regardless of what the precise increase in detection rate is, it did not translate into a clinically meaningful benefit (or that any benefits were offset by possible harms of false positives).

References

Akolekar 2019

Akolekar, R., Panaitescu, A. M., Ciobanu, A., Syngelaki, A., Nicolaides, K. H., 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 in obstetrics & gynecology, 04, 04, 2019

Aviram 2017

Aviram, A., Yogev, Y., Ashwal, E., Hiersch, L., Hadar, E., Gabbay-Benziv, R., Prediction of large for gestational age by various sonographic fetal weight estimation formulas-which should we use?, Journal of Perinatology, 37, 513-517, 2017

Barel 2016

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 in der Medizin, 37, 283-9, 2016

Barden 2020

Bardin, R., Aviram, A., Hiersch, L., Hadar, E., Gabbay-Benziv, R., False diagnosis of small for gestational age and macrosomia-clinical and sonographic predictors, Journal of Maternal Fetal and Neonatal Medicine., 2020

Ben-Haroush 2007

Ben-Haroush, A., Yogev, Y., Hod, M., Bar, J., Predictive value of a single early fetal weight estimate in normal pregnancies, European Journal of Obstetrics, Gynecology, & Reproductive BiologyEur J Obstet Gynecol Reprod Biol, 130, 187-92, 2007

Blue 2018

Blue, N. R., Beddow, M. E., Savabi, M., Katukuri, V. R., Mozurkewich, E. L., Chao, C. R., 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, Obstetrics & GynecologyObstet Gynecol, 131, 835-841, 2018

Blue 2019

Blue, N. R., Savabi, M., Beddow, M. E., Katukuri, V. R., Fritts, C. M., Izquierdo, L. A., Chao, C. R., The Hadlock Method Is Superior to Newer Methods for the Prediction of the Birth Weight Percentile, Journal of ultrasound in medicine, 38, 587-596, 2019

Callec 2015

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 in obstetrics & gynecology, 46, 216-20, 2015

Duncan 2020

Duncan, J. R., Odibo, L., Hoover, E. A., Odibo, A. O., Prediction of Large for Gestational Age Neonates by Different Growth Standards, Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine., 29, 2020

Erkamp 2020

Erkamp, J. S., Voerman, E., Steegers, E. A. P., Mulders, Agmgj, Reiss, I. K. M., Duijts, L., Jaddoe, V. W. V., Gaillard, R., Second and third trimester fetal ultrasound population

screening for risks of preterm birth and small-size and large-size for gestational age at birth: a population-based prospective cohort study, BMC Medicine, 18, 63, 2020

Gabbay-Benziv 2016

Gabbay-Benziv, R., Aviram, A., Bardin, R., Ashwal, E., Melamed, N., Hiersch, L., Wiznitzer, A., Yogev, Y., Hadar, E., Prediction of Small for Gestational Age: Accuracy of Different Sonographic Fetal Weight Estimation Formulas, Fetal Diagnosis and Therapy, 40, 205-213, 2016

Harding 1995

Harding, K., Evans, S., Newnham, J., Screening for the small fetus: a study of the relative efficacies of ultrasound biometry and symphysiofundal height, Australian & New Zealand Journal of Obstetrics & Gynaecology, 35, 160-4, 1995

Khan 2019

Khan, N., Ciobanu, A., Karampitsakos, T., Akolekar, R., Nicolaides, K. H., Prediction of large-for-gestational-age neonate by routine third-trimester ultrasound, Ultrasound in obstetrics & gynecology, 54, 326-333, 2019

Lin 1990

Lin, C. C., Sheikh, Z., Lopata, R., The association between oligohydramnios and intrauterine growth retardation, 76, 1100-4, 1990

Monier 2015

Monier, I., Blondel, B., Ego, A., Kaminiski, M., Goffinet, F., Zeitlin, J., Poor effectiveness of antenatal detection of fetal growth restriction and consequences for obstetric management and neonatal outcomes: a French national study, 122, 518-27, 2015

Rad 2018

Rad, S., Beauchamp, S., Morales, C., Mirocha, J., Esakoff, T. F., Defining fetal growth restriction: abdominal circumference as an alternative criterion, Journal of Maternal-Fetal and Neonatal Medicine, 31, 3089-3094, 2018

Sekar 2016

Sekar, R., Khatun, M., Barrett, H. L., Duncombe, G., A prospective pilot study in assessing the accuracy of ultrasound estimated fetal weight prior to delivery, Australian and New Zealand Journal of Obstetrics and Gynaecology, 56, 49-53, 2016

Skovron 1991

Skovron, M. L., Berkowitz, G. S., Lapinski, R. H., Kim, J. M., Chitkara, U., Evaluation of early third-trimester ultrasound screening for intrauterine growth retardation, J Ultrasound MedJournal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine, 10, 153-9, 1991

Turitz 2014

Turitz, A. L., Quant, H., Schwartz, N., Elovitz, M., Bastek, J. A., Isolated abdominal circumference < 5% or estimated fetal weight 10 to 19% as predictors of small for gestational age infants, American Journal of Perinatology, 31, 469-476, 2014

Appendices

Appendix A – Review protocol

Review protocol for review question: What is the best method using third trimester measurements to predict birth weight?

Table 3: Review protocol

Field	Content
Review question	What is the best method using third trimester measurements to predict birth weight?
Type of review question	• Diagnostic test (predictive) accuracy review
Objective of the review	Monitoring fetal growth is essential to planning the care of pregnant women. In particular, it is important to make alternative provisions for babies that are thought to be small or large compared to normal size babies (e.g. alternative place of delivery to enable prompt admission to neonatal unit in the case of maternal or fetal complications). The aim of this review is to establish what techniques, or combination of techniques, using third trimester measurements, are effective in accurately categorising babies by birth weight and identifying small or large babies at delivery.
Eligibility criteria – population	• All women with unselected or low-risk pregnancies
Eligibility criteria –Index test(s)	The use of the following third trimester measurements, individually or in combination, to predict birth weight at delivery (either using specific birth weight centile (e.g. <10th centile) or absolute birth weight thresholds (e.g. ≤2500 g or ≥4000 g) will be examined: ○ Fetal biometry using ultrasound (e.g. abdominal circumference <10th centile; estimated fetal weight <3rd centile) ○ Symphysis-fundal height (SFH) measurement ○ Without growth chart ○ With customised growth charts ○ With non-customised growth chart
Eligibility criteria –Reference standard	Reference standard is either • Relevant birth weight centile as appropriate for test (e.g. birth weight <10th percentile for index test to identify small babies, or birth weight>90th centile for index test to identify large babies) using ○ Customised birth-weight chart, or ○ Non-customised birth-weight chart, Or,if no chart is used, reference standard is: • Actual absolute birth weight threshold (e.g. small ≤2500 g or large ≥4000 g) Note: studies may use more than one (threshold) definition of SGA or LGA. Definitions of these thresholds and related data will be extracted and data presented separately for each reference standard. Results for SFH measurement using birth-weight chart will be pooled unless there is serious or very serious heterogeneity.
Outcomes and prioritisation	**Critical outcome** Diagnostic test accuracy data (i.e. TP, FP, TN, FN) that allows calculation of • Sensitivity and specificity **Important outcome** Diagnostic test accuracy data (i.e. TP, FP, TN, FN) that allows calculation of

Field	Content
	• Positive and negative predictive values Note: Raw data will be extracted from studies and the relevant diagnostic accuracy pair measures calculated if not otherwise reported. Results will be presented separately by definition of reference standard (e.g. birth weight < 3rd, 5th or 10th percentile).

Eligibility criteria – study design	INCLUDE:
	• Systematic reviews of diagnostic test accuracy studies
	• Diagnostic test accuracy studies
	  ○ Prospective cohort studies
	  ○ Retrospective cohort studies
	  ○ Nested case-control studies within a cohort of known size
	The committee will prioritise direct comparison (i.e. fully paired or partially paired) development or validation studies of diagnostic tests (e.g. a study comparing the performance of both SFH measurement with customised chart to AC measurement using ultrasound relative to reference standard of BW<10th centile) and will prioritise prospective cohort studies over both retrospective studies and nested case-control studies when making recommendations. Note: For further details, see the algorithm in appendix H, Developing NICE guidelines: the manual.

Other inclusion exclusion criteria	**Exclusion** POPULATION: • High-risk pregnancies • Multiple pregnancies • Pregnancy with known or pre-existing congenital anomalies STUDY DESIGN: • Cross-sectional studies • Epidemiological review or review on associations • Experimental studies • Non-nested case control studies PUBLICATION STATUS: • Conference abstract LANGUAGE: • Non-English **Inclusion** COUNTRY: • High-income countries only (as defined by the World Bank; see https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups for classification of countries). Note that the use of the World Bank definitions of low-, middle- and high-income countries in this guideline is consistent with its use in the Postnatal care up to 8 weeks after birth (update) NICE guideline CG37.
Proposed sensitivity/sub-group analysis, or meta-regression	Sensitivity analysis according to study design will be conducted. In the presence of heterogeneity in the results for SFH measurement with a customised or non-customised chart, the following subgroup analysis will be conducted: • Type of chart: customised, non-customised For meta-analyses using hierarchical bivariate models, statistical heterogeneity will be assessed by visually examining the diagnostic accuracy plots and by examining the I^2 inconsistency statistic (with an I^2 value≥50% indicating serious heterogeneity, and ≥80% indicating very serious heterogeneity).
Selection process – duplicate screening/selection/analysis	Studies included in the Antenatal care for uncomplicated pregnancies guideline (CG62) that satisfy the review protocol will be included in this review. Review questions selected as high priorities for health economic analysis (and those selected as medium priorities and where health economic analysis could influence recommendations) will be subject to dual weeding and study selection; any discrepancies above 10% of the dual weeded resources will be

Field	Content
	resolved through discussion between the first and second reviewers or by reference to a third person. All data extraction will quality assured by a senior reviewer. Draft excluded studies and evidence tables will be circulated to the Topic Group for their comments. Resolution of disputes will be by discussion between the senior reviewer, Topic Advisor and Chair.
Data management (software)	NGA STAR software will be used to generate bibliographies/citations, and conduct study sifting and data extraction. RevMan v.5, STATA and WinBUGS software will be used to conduct multivariate meta-analysis and construct summary ROC curves as appropriate.
Information sources – databases and dates	Sources to be searched: Medline, Medline In-Process, CCTR, CDSR, DARE, HTA, Embase. Limits (e.g. date, study design): • Date limit: 2006 • Apply standard animal/non-English language exclusion
Identify if an update	This antenatal care update will replace the 2008 NICE guideline on antenatal care for uncomplicated pregnancies (CG62). The following relevant recommendations in CG62 regarding fetal growth and well-being were made: 1.10.1 Symphysis–fundal height should be measured and recorded at each antenatal appointment from 24 weeks. [2008] 1.10.2 Ultrasound estimation of fetal size for suspected large-for-gestational-age unborn babies should not be undertaken in a low-risk population. [2008] 1.10.3 Routine Doppler ultrasound should not be used in low-risk pregnancies. [2008]
Author contacts	Developer: National Guideline Alliance.
Highlight if amendment to previous protocol	For details please see section 4.5 of Developing NICE guidelines: the manual.
Search strategy – for one database	For details please see appendix B of the evidence report
Data collection process – forms/duplicate	A standardised evidence table format will be used, and published as appendix D (clinical evidence tables) or G (economic evidence tables) of the evidence report
Data items – define all variables to be collected	For details please see evidence tables in appendix D (clinical evidence tables) or G (economic evidence tables) of the evidence report
Methods for assessing bias at outcome/study level	Quality assessment of individual studies will be performed using the following checklists: • ROBIS tool for systematic reviews of diagnostic test accuracy studies • QUADAS-2 for diagnostic test accuracy studies For details please see section 6.2 of Developing NICE guidelines: the manual. The risk of bias across all available evidence will be evaluated for each member of paired accuracy measures (e.g. GRADE evaluation for both sensitivity and specificity of test) using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group: http://www.gradeworkinggroup.org/
Criteria for quantitative synthesis (where suitable)	**Meta-analysis** Depending on the availability of the evidence, the findings will be summarised narratively or quantitatively. Sensitivity and specificity, with 95% CIs will be used as primary outcomes for diagnostic test accuracy. These diagnostic accuracy parameters will be obtained from the studies or calculated by the technical team using data from the studies. Where 4 or more unbiased studies are included (e.g. there is no suggestion that the estimates of accuracy are systematically incorrect) then diagnostic meta-analysis will be conducted using the bivariate random effects model. Where fewer than 4 studies are included a univariate model will be used. **Interpretation of diagnostic test performance/patient-important outcomes** The committee discussed the patient-important outcomes associated with testing and each of the 4 diagnostic test outcomes. They agreed that routine fetal growth monitoring may lead to an increase in maternal anxiety. True positives and true negatives, which are both desirable outcomes, both lead to correct prediction of birth centile and allows appropriate planning for labour relative to this centile. False positives are not desirable as they lead to incorrect prediction of adverse birth centile and inappropriate active management of labour (e.g. induction), and therefore also wasted resources. However, management of labour is in a controlled environment (e.g. hospital), therefore allowing appropriate management when the incorrect prediction is discovered. False negatives are not desirable as the incorrect prediction regarding their adverse birth centile can lead to inappropriate planning by health services and adverse long-term health outcomes. For example, small babies/mothers have a higher risk of experiencing complications during labour and being admitted to the neonatal unit, whilst large babies have a higher risk of shoulder dystocia and are more likely to require the use of Caesarean Section. The committee

Field	Content
	agreed that in terms of incorrect diagnoses, the consequences of a false negative result are likely to be more serious for the mother/baby (especially for small babies who are not identified as such) and the healthcare system than a false positive result. **Clinical decision thresholds** Given the seriousness of false negatives, the committee agreed that, in principle, the clinical decision threshold for sensitivity below which a test would not be recommended is 0.8, and the clinical decision threshold above which a test would be recommended is 0.95. The committee agreed that, in principle, the clinical decision threshold for specificity below which a test would not be recommended is 0.75, and the clinical decision threshold above which a test would be recommended is 0.9. However, the committee recognised that there is substantive clinical uncertainty as to the diagnostic accuracy of SFH measurement with or without charts, which is current UK standard practice, with an estimated sensitivity between 0.17 and 0.93. These thresholds will be used to guide decision making and imprecision judgements.
Methods for analysis – combining studies and exploring (in)consistency	For details please see Supplement 1: methods and section 6.2 of Developing NICE guidelines: the manual.
Meta-bias assessment – publication bias, selective reporting bias	For details please see Supplement 1: methods and section 6.2 of Developing NICE guidelines: the manual. If sufficient relevant evidence is available, publication bias will be explored using RevMan software to examine funnel plots. Trial registries will be examined to identify missing evidence: Clinical trials.gov, NIHR Clinical Trials Gateway.
Assessment of confidence in cumulative evidence	For details please see sections 6.4 and 9.1 of Developing NICE guidelines: the manual.
Rationale/context – Current management	For details please see the introduction to the evidence review in the evidence report
Describe contributions of authors and guarantor	A multidisciplinary committee developed the guideline. The committee was convened by the National Guideline Alliance and chaired by Kate Harding in line with section 3 of Developing NICE guidelines: the manual. Staff from the National Guideline Alliance undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the guideline in collaboration with the committee. For details please see Supplement 1: methods.
Sources of funding/support	The National Guideline Alliance is funded by NICE and hosted by the Royal College of Obstetricians and Gynaecologists.
Name of sponsor	The National Guideline Alliance is funded by NICE and hosted by the Royal College of Obstetricians and Gynaecologists.
Roles of sponsor	NICE funds the National Guideline Alliance to develop guidelines for those working in the NHS, public health, and social care in England.
PROSPERO registration number	This protocol is not registered with PROSPERO.

CDSR: Cochrane Database of Systematic Reviews; CENTRAL: Cochrane Central Register of Controlled Trials; DARE: Database of Abstracts of Reviews of Effects; GRADE: Grading of Recommendations Assessment, Development and Evaluation; HTA: Health Technology Assessment; MID: minimally important difference; NGA: National Guideline Alliance; NHS: National health service; NICE: National Institute for Health and Care Excellence; RCT: randomised controlled trial; RoB: risk of bias; SD: standard deviation

Appendix B – Literature search strategies

Literature search strategies for review question: What is the best method using third trimester measurements to predict birth weight?

Database(s): Medline & Embase (Multifile)
Last searched on **Embase Classic+Embase** 1947 to 2020 September 04, **Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily** 1946 to September 04, 2020
Date of last search: 8th September 2020
Multifile database codes: emczd = Embase Classic+Embase; ppez= MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily

#	Searches
1	Fetal Growth Retardation/ use ppez
2	Infant, Small for Gestational Age/ use ppez
3	Infant, Low Birth Weight/ use ppez
4	"Embryonic and Fetal Development"/ use ppez
5	intrauterine growth retardation/ use emczd
6	*growth retardation/ use emczd
7	small for date infant/ use emczd
8	large for gestational age/ use emczd
9	low birth weight/ use emczd
10	fetus development/ use emczd
11	((intrauterine or fetal or foetal or fetus or foetus) adj growth adj (restrict$ or retard$)).tw,kw.
12	((IUGR or FGR) adj5 (growth$ or restrict$ or retard$)).tw,kw.
13	((small$ or large$) adj2 gestation$ age$).tw,kw.
14	((LGA or SGA) adj10 (gestation$ or age or large or small)).tw,kw.
15	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
16	Pubic Symphysis/ use ppez
17	pubis symphysis/ use emczd
18	((symphysisfundus or symphys$-fundus or symphysis-to-fundus or symphysisfundal or symphys$-fundal or symphysis-to-fundal or fundus-symphys$ or fundal-symphys$) adj3 (method$ or rule$ or measure$ or distance$ or evaluation$ or growth$ or increment$)).tw.
19	((symphys$-fundus or symphysis-to-fundus or symphys$-fundal or symphysis-to-fundal or fundus-symphys$ or fundal-symphys$) adj height).tw.
20	((distance$ or height$ or measur$) adj4 (fundus or fundal or symphys$) adj4 (fundus or fundal or symphys$)).mp.
21	(SFH$ adj measur$).tw.
22	Growth Charts/ use ppez
23	growth chart/ use emczd
24	((growth or height or weight) adj (chart$ or curve$ or centile$)).mp.
25	((conditional or customi?ed or non-customi?ed) adj2 growth).mp.
26	size chart$.tw.
27	((reference$ or centile$) adj chart$).tw.
28	abdominal circumference/ use emczd
29	(abdom$ adj3 (circumference$ or diamet$)).mp.
30	((fetal or foetal or fetus or foetus or birth) adj weight$ adj5 estimat$).mp.
31	*Anthropometry/ use ppez
32	*Biometry/ use ppez
33	*anthropometry/ use emczd
34	*biometry/ use emczd
35	((fetal or foetal or fetus or foetus) adj (anthropometr$ or biometr$)).mp.
36	((ultrasound$ or ultrasonogra$ or sonogra$) adj3 (anthropometr$ or biometr$)).mp.
37	(third$ trimester$ adj3 (ultrasound$ or ultrasonograph$ or sonograph$ or doppler$ or echograph$ or screening)).mp.
38	uter$ arter$ Doppler$.mp.
39	Cephalometry/ use ppez
40	head circumference/ use emczd
41	cephalometry/ use emczd
42	(head$ adj2 circumference$).tw,kw.
43	39 or 40 or 41 or 42
44	Fetus/
45	Fetal Diseases/ use ppez
46	fetus disease/ use emczd
47	44 or 45 or 46

#	Searches
48	43 and 47
49	*Prenatal Diagnosis/ use ppez
50	*prenatal diagnosis/ use emczd
51	(diagnos$ adj (ultrasound$ or ultrasonograph$ or sonograph$ or doppler$ or echograph$)).mp.
52	16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 48 or 49 or 50 or 51
53	(predict or prediction).ti.
54	(validat$ or rule$).ti,ab.
55	(predict$ and (outcome$ or risk$ or model$)).ti,ab.
56	((history or variable$ or criteria or scor$ or characteristic$ or finding$ or factor$) and (predict$ or model$ or decision$ or identif$ or prognos$)).ti,ab.
57	(decision$.ti,ab. and Logistic models/) use ppez
58	(decision$.ti,ab. and Statistical model/) use emczd
59	(decision$ and (model$ or clinical$)).ti,ab.
60	(prognostic and (history or variable$ or criteria or scor$ or characteristic$ or finding$ or factor$ or model$)).ti,ab.
61	(stratification or discrimination or discriminate or c statistic or "area under the curve" or AUC or calibration or indices or algorithm or multivariable).ti,ab.
62	ROC curve/ use ppez
63	Receiver operating characteristic/ use emczd
64	"Sensitivity and Specificity"/ use ppez
65	"sensitivity and specificity"/ use emczd
66	(sensitivity or specificity).ti,ab.
67	(predictive value$ or PPV or NPV).ti,ab.
68	likelihood ratio$.ti,ab.
69	53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68
70	15 and 52 and 69
71	meta-analysis/
72	meta-analysis as topic/
73	systematic review/
74	meta-analysis/
75	(meta analy* or metanaly* or metaanaly*).ti,ab.
76	((systematic or evidence) adj2 (review* or overview*)).ti,ab.
77	((systematic* or evidence*) adj2 (review* or overview*)).ti,ab.
78	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
79	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
80	(search* adj4 literature).ab.
81	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
82	cochrane.jw.
83	((pool* or combined) adj2 (data or trials or studies or results)).ab.
84	letter/
85	editorial/
86	news/
87	exp historical article/
88	Anecdotes as Topic/
89	comment/
90	case report/
91	(letter or comment*).ti.
92	84 or 85 or 86 or 87 or 88 or 89 or 90 or 91
93	randomized controlled trial/ or random*.ti,ab.
94	92 not 93
95	animals/ not humans/
96	exp Animals, Laboratory/
97	exp Animal Experimentation/
98	exp Models, Animal/
99	exp Rodentia/
100	(rat or rats or mouse or mice).ti.
101	94 or 95 or 96 or 97 or 98 or 99 or 100
102	letter.pt. or letter/
103	note.pt.
104	editorial.pt.
105	case report/ or case study/
106	(letter or comment*).ti.
107	102 or 103 or 104 or 105 or 106
108	randomized controlled trial/ or random*.ti,ab.
109	107 not 108
110	animal/ not human/
111	nonhuman/

#	Searches
112	exp Animal Experiment/
113	exp Experimental Animal/
114	animal model/
115	exp Rodent/
116	(rat or rats or mouse or mice).ti.
117	109 or 110 or 111 or 112 or 113 or 114 or 115 or 116
118	101 use ppez
119	117 use emczd
120	118 or 119
121	(or/71-72,75,77-82) use ppez
122	(or/73-76,78-83) use emczd
123	121 or 122
124	15 and 52 and 123
125	70 or 124
126	120 and 125
127	125 not 126
128	limit 127 to english language
129	limit 128 to yr="2006 -Current"

Database(s): Cochrane Library

Last searched on **Cochrane Database of Systematic Reviews**, Issue 9 of 12, September 2020, **Cochrane Central Register of Controlled Trials**, Issue 9 of 12, September 2020

Date of last search: 8th September 2020

#	Searches
#1	MeSH descriptor: [Fetal Growth Retardation] this term only
#2	MeSH descriptor: [Infant, Small for Gestational Age] this term only
#3	MeSH descriptor: [Infant, Low Birth Weight] this term only
#4	MeSH descriptor: [Embryonic and Fetal Development] this term only
#5	(((intrauterine or fetal or foetal or fetus or foetus) NEXT growth NEXT (restrict* or retard*))):ti,ab,kw (Word variations have been searched)
#6	(((IUGR or FGR) NEAR/5 (growth* or restrict* or retard*))):ti,ab,kw
#7	((small* or large*) NEAR/2 gestation* NEXT age*):ti,ab,kw (Word variations have been searched)
#8	(((LGA or SGA) NEAR/10 (gestation* or age or large or small))):ti,ab,kw (Word variations have been searched)
#9	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8
#10	MeSH descriptor: [Pubic Symphysis] this term only
#11	(((symphysisfundus or symphys-fundus or symphysis-to-fundus or symphysisfundal or symphys-fundal or symphysis-to-fundal or fundus-symphys* or fundal-symphys*) NEAR/3 (method* or rule* or measure* or distance* or evaluation* or growth* or increment*))):ti,ab,kw
#12	(((symphys-fundus or symphysis-to-fundus or symphys-fundal or symphysis-to-fundal or fundus-symphys* or fundal-symphys*) NEXT height)):ti,ab,kw
#13	(((distance* or height* or measur*) NEAR/4 (fundus or fundal or symphys*) NEAR/4 (fundus or fundal or symphys*))):ti,ab,kw
#14	((SFH* NEXT measur*)):ti,ab,kw
#15	MeSH descriptor: [Growth Charts] this term only
#16	(((growth or height or weight) NEXT (chart* or curve* or centile*))):ti,ab,kw
#17	(((conditional or customised or non-customised or customized or non-customized) NEAR/2 growth)):ti,ab,kw
#18	(size NEXT chart*):ti,ab,kw
#19	(((reference* or centile*) NEXT chart*)):ti,ab,kw
#20	((abdom* NEAR/3 (circumference* or diamet*))):ti,ab,kw
#21	(((fetal or foetal or fetus or foetus or birth) NEXT weight* NEAR/5 estimat*)):ti,ab,kw
#22	MeSH descriptor: [Anthropometry] this term only
#23	MeSH descriptor: [Biometry] this term only
#24	(((fetal or foetal or fetus or foetus) NEXT (anthropometr* or biometr*))):ti,ab,kw
#25	(((ultrasound* or ultrasonogra* or sonogra*) NEAR/3 (anthropometr* or biometr*))):ti,ab,kw
#26	((third* NEXT trimester* NEAR/3 (ultrasound* or ultrasonograph* or sonograph* or doppler* or echograph* or screening))):ti,ab,kw
#27	(uter* NEXT arter* NEXT Doppler*):ti,ab,kw
#28	MeSH descriptor: [Cephalometry] this term only
#29	((head* NEAR/2 circumference*)):ti,ab,kw
#30	MeSH descriptor: [Fetus] this term only
#31	MeSH descriptor: [Fetal Diseases] this term only
#32	(#28 OR #29) AND (#30 OR #31)
#33	MeSH descriptor: [Prenatal Diagnosis] this term only
#34	((diagnos* NEXT (ultrasound* or ultrasonograph* or sonograph* or doppler* or echograph*))):ti,ab,kw
#35	#10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #32 OR #33 OR #34

#	Searches
#36	#9 AND #35 Publication Year from 2006 to current

Database(s): CRD: Database of Abstracts of Reviews of Effects (DARE), HTA Database

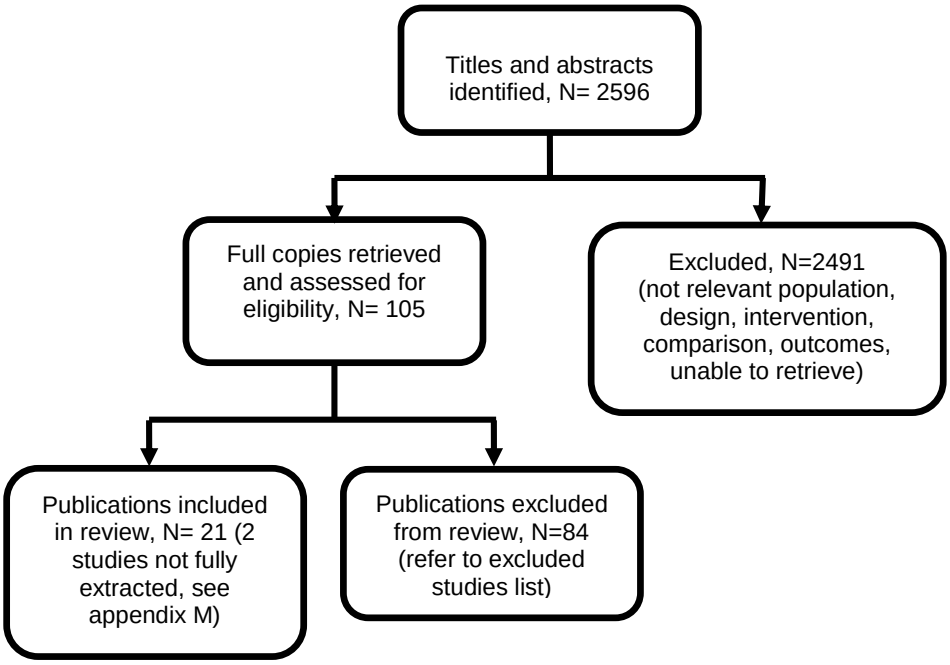
Date of last search: 8th September 2020

#	Searches
1	MeSH DESCRIPTOR Fetal Growth Retardation IN DARE,HTA
2	MeSH DESCRIPTOR Infant, Small for Gestational Age IN DARE,HTA
3	MeSH DESCRIPTOR Infant, Low Birth Weight IN DARE,HTA
4	MeSH DESCRIPTOR Embryonic and Fetal Development IN DARE,HTA
5	(((intrauterine or fetal or foetal or fetus or foetus) NEXT growth NEXT (restrict* or retard*))) IN DARE, HTA
6	(((IUGR or FGR) NEAR5 (growth* or restrict* or retard*))) IN DARE, HTA
7	(((small* or large*) NEAR2 gestation* age*)) IN DARE, HTA
8	(((LGA or SGA) NEAR10 (gestation* or age or large or small))) IN DARE, HTA
9	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8
10	MeSH DESCRIPTOR Pubic Symphysis IN DARE,HTA
11	(((symphysisfundus or symphys*-fundus or symphysis-to-fundus or symphysisfundal or symphys*-fundal or symphysis-to-fundal or fundus-symphys* or fundal-symphys*) NEAR3 (method* or rule* or measure* or distance* or evaluation* or growth* or increment*))) IN DARE, HTA
12	(((((symphys*-fundus or symphysis-to-fundus or symphys*-fundal or symphysis-to-fundal or fundus-symphys* or fundal-symphys*) NEXT height))) IN DARE, HTA
13	(((distance* or height* or measur*) NEAR4 (fundus or fundal or symphys*) NEAR4 (fundus or fundal or symphys*))) IN DARE, HTA
14	(((SFH* NEXT measur*))) IN DARE, HTA
15	MeSH DESCRIPTOR Growth Charts IN DARE,HTA
16	(((growth or height or weight) NEXT (chart* or curve* or centile*))) IN DARE, HTA
17	(((((conditional or customised or non-customised or customized or non-customized) NEAR2 growth))) IN DARE, HTA
18	((size NEXT chart*)) IN DARE, HTA
19	(((reference* or centile*) NEXT chart*)) IN DARE, HTA
20	(((abdom* NEAR3 (circumference* or diamet*)))) IN DARE, HTA
21	((((fetal or foetal or fetus or foetus or birth) NEXT weight* NEAR5 estimat*))) IN DARE, HTA
22	MeSH DESCRIPTOR Anthropometry IN DARE,HTA
23	MeSH DESCRIPTOR Biometry IN DARE,HTA
24	((((fetal or foetal or fetus or foetus) NEXT (anthropometr* or biometr*)))) IN DARE, HTA
25	((((ultrasound* or ultrasonogra* or sonogra*) NEAR3 (anthropometr* or biometr*)))) IN DARE, HTA
26	(((third* NEXT trimester* NEAR3 (ultrasound* or ultrasonograph* or sonograph* or doppler* or echograph* or screening)))) IN DARE, HTA
27	((uter* NEXT arter* NEXT Doppler*)) IN DARE, HTA
28	MeSH DESCRIPTOR Cephalometry IN DARE,HTA
29	(((head* NEAR2 circumference*))) IN DARE, HTA
30	MeSH DESCRIPTOR Fetus IN DARE,HTA
31	MeSH DESCRIPTOR Fetal Diseases IN DARE,HTA
32	#28 OR #29
33	#30 OR #31
34	#32 AND #33
35	MeSH DESCRIPTOR Prenatal Diagnosis IN DARE,HTA
36	(((diagnos* NEXT (ultrasound* or ultrasonograph* or sonograph* or doppler* or echograph*)))) IN DARE, HTA
37	#10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #34 OR #35 OR #36
38	#9 AND #37 Publication Year from 2006 to current

Appendix C – Clinical evidence study selection

Study selection for: What is the best method using third trimester measurements to predict birth weight?

Figure 1: Study selection flow chart



1 Appendix D – Clinical evidence tables

2 Evidence tables for review question: What is the best method using third trimester measurements to predict birth weight?

3 Table 4: Evidence tables

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes</th><th>Comments</th></tr>
<tr>
<td>Full citation

Akolekar, R., Panaitescu, A. M., Ciobanu, A., Syngelaki, A., Nicolaides, K. H., 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 in obstetrics & gynecology, 04, 04, 2019

Ref Id

1112815

Country/ies where the study was carried out

UK

Study type

Prospective cohort study

Aim of the study
Examine the contribution of small for gestational age (SGA) fetuses to the overall rate of adverse perinatal outcome and, to</td>
<td>Sample size
n=45,847 singleton pregnancies

Characteristics
Age (years): non-SGA 31.7 (27.4 to 35.4); SGA 30.9 (26.2 to 35)**, overall mean 31.3
Weight (kg): non-SGA 79.7 (71.5 to 91.10); SGA 73.4 (65.5 to 83.2)**
Height (cm): non-SGA 165 (161 to 170); SGA 163 (158 to 167)**
Racial origin:
white: non-SGA 30,812 (76%); SGA 3,348 (63.4%)**
Black: non-SGA 6,065 (15%); SGA 1,131 (21.4%)**
South Asian: non-SGA 1,697 (4.2%); SGA 488 (9.2%)**</td>
<td>Tests
Index test: US estimated fetal weight <10th percentile (Hadlock formula)
Reference standard: Birth weight <10th percentile for gestational age based on the fetal medicine foundation fetal and neonatal population weight charts

Timing
>7d from delivery (US done between 35+0 and 36+6 weeks gestation)</td>
<td>Methods
Visit included ultrasound examination for fetal anatomy and measurement of fetal head circumference, abdominal circumference and femur length for calculation of estimated fetal weight (EFW) using Hadlock formula, and trans abdominal colour doppler ultrasound measurement of mean Ut-PI, UA-PI and MCA-PI. Gestational age was determined by the measurements of metal crown-rump length at 11-13 weeks or fetal head circumference at 19-24 weeks. The ultrasound measurements were carried out by sonographers who had obtained the fetal medicine foundation certificate of competence in ultrasound examination. Statistical software package SPSS version 24.0 for windows and</td>
<td>Results
<table>
<tr><th></th><th>Reference test +</th><th>Reference test -</th><th>Total</th></tr>
<tr><td>Index test +</td><td>2420</td><td>1689</td><td>4109</td></tr>
<tr><td>Index test -</td><td>2860</td><td>38878</td><td>41738</td></tr>
<tr><td>Total</td><td>5280</td><td>40567</td><td>45847</td></tr>
</table>
<u>SGA</u>
Sensitivity= 45.83% (44.48% to 47.19%)*
Specificity= 95.84% (95.64% to 96.03%)*
Positive predictive value= 58.86% (57.51% to 60.18%)*
Negative predictive value= 93.16% (93.00% to 93.32%)*

Prevalence of SGA= 11.5%*

*Calculated by the NGA technical team</td>
<td>Limitations
Risk of bias assessed using QUADAS-II
DOMAIN 1: PATIENT SELECTION
A. RISK OF BIAS
1. Was a consecutive or random sample of patients enrolled? yes
2. Was a case-control design avoided? yes
3. Did the study avoid inappropriate exclusions? yes
Could the selection of patients have introduced bias? RISK: LOW
B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: LOW

DOMAIN 2: INDEX TESTS
A. RISK OF BIAS
1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes</td>
</tr>
</table>

Study details	Participants	Interventions	Methods	Outcomes	Comments
propose a 2-stage approach for prediction of a SGA neonate at routine ultrasound examination at 35 + 0 to 36 + 6 weeks **Study dates** March 2014-September 2018 **Source of funding** The Fetal Medicine Foundation (Charity No: 1037116)	• East Asian: non-SGA 813 (2%); SGA 126 (2.4%) • Mixed: non-SGA 1,180 (2.9%); SGA 187 (3.5%)* Cigarette smoker: non-SGA 2,961 (7.3%); SGA 762 (14.4%)** Conception: • natural: non-SGA 39,190 (96.6); SGA 5080 (96.2%) • ovulation drugs: non-SGA 223 (0.5%); SGA 34 (0.6%) • IVF:non-SGA 1154 (2.8%); SGA 166 (3.1%) Medical conditions: • Chronic hypertension: non-SGA 490 (1.2%); SGA 90 (1.7%)** • Type I Diabetes Mellitus: non-SGA 162 (0.4%); SGA 5 (0.1%)* • Type II Diabetes Mellitus: non-SGA 189 (0.5%); SGA 19 (0.4%)		MedCalc were used for data analysis		2. If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? unclear Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review

Study details	Participants	Interventions	Methods	Outcomes	Comments
	Obstetric history: • nulliparous: non-SGA 17,911 (44.2%); SGA 2,949 (55.9%) • Parous: - prior SGA: non-SGA 3,112 (7.75); SGA 964 (18.3%)** • Parous: - no prior SGA: non-SGA 19,544 (48.2%); SGA 1,367 (25.9%)** Gestational age at screening (weeks): non-SGA 36.1 (35.9 to 36.4); SGA 36.1 (35.9 to 36.4) Estimated fetal weight (z-score): non-SGA 0.01 (-0.59 to 0.60); SGA -1.39 (-2.08 to -0.85)** Gestational age at delivery (weeks): non-SGA 40.0 (39.1 to 40.9); SGA 39.4 (38.2 to 40.3)** Birth weight (z-score): non-SGA 0.13 (-0.45 to 0.75); SGA -1.72 (-2.14 to -1.48)** Birth weight (g): non-SGA 3,490 (3220 to 3790); SGA 2,715 (2510 to 2860)** **p-value <0.001; *p-value <0.01				question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? yes 2. Did all patients receive a reference standard? yes 3. Did patients receive the same reference standard? Yes 4. Were all patients included in the analysis? Yes Could the patient flow have introduced bias? RISK: LOW

Study details	Participants	Interventions	Methods	Outcomes	Comments
	Inclusion Criteria singleton pregnancies undergoing routine ultrasound examination at 35 + 0 to 36 + 6 weeks gestation and delivery of non-malformed liveborn or stillborn neonate. **Exclusion Criteria** Aneuploidy or major fetal abnormality				
Full citation Aviram, A., Yogev, Y., Ashwal, E., Hiersch, L., Hadar, E., Gabbay-Benziv, R., Prediction of large for gestational age by various sonographic fetal weight estimation formulas-which should we use?, Journal of Perinatology, 37, 513-517, 2017 **Ref Id** 1121728 **Country/ies where the study was carried out** Israel **Study type**	**Sample size** n=7996 singleton pregnancies **Characteristics** Maternal age (years): LGA 32.0±4.9; AGA 31.5±5.3** Medical conditions: • Gestational hypertension: LGA 18 (1.1); AGA 97 (1.5) • Preeclampsia without severe features: LGA 28 (1.7); AGA 184 (2.9)* • Preeclampsia with severe	**Tests** Index test: Different US tests (20 variations) Reference standard: Birth weight >90th percentile for gestational age **Timing** <7d from delivery (US done up to 3 days before delivery)	**Methods** For each sonographic fetal weight estimation (sEFW) examination, estimated fetal weight was calculated using 20 sonographic fetal weight estimation formulas published in literature. The formulas were subdivided into groups according to the combination of the fetal biometric indices incorporated in their equations. Gestational age was calculated by the LMP or by the first trimester ultrasound if discrepancy between them exceeded 7 days. sEFW included all standard fetal biometry measurements (AC, FL,	**Results** |	Reference test +	Reference test -	Total
---	---	---	---
Index test +	1215	574	1789
Index test -	402	5804	6207
Total	1618	6378	7996	 LGA Sensitivity= 75.1% Specificity= 91.0% Positive predictive value= 69.9% Negative predictive value= 92.9% Overall accuracy= 87.5	**Limitations** **Risk of bias assessed using QUADAS-II** DOMAIN 1: PATIENT SELECTION A. RISK OF BIAS 1. Was a consecutive or random sample of patients enrolled? Yes 2. Was a case-control design avoided? Yes 3. Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review

Study details	Participants	Interventions	Methods	Outcomes	Comments
Retrospective cohort study **Aim of the study** To compare the accuracy of various formulas for prediction of LGA neonates in pregnancies in which sonographic weight estimation was performed within 7 days of delivery, and rank the formulas by their LGA prediction performance. **Study dates** 1st July 2007 to 31st December 2014. **Source of funding** Not reported	features: LGA 4 (0.2); AGA 106 (1.7)** • Type 1 Diabetes Mellitus: LGA 30 (1.9); AGA 35 (0.5)** • Type 2 Diabetes Mellitus: LGA 23 (1.4); AGA 39 (0.6) • Gestational diabetes mellitus: LGA 286 (17.7); AGA 807 (12.7)** Obstetric history: • Parity: LGA 2.5±1.5; AGA 2.2±1.4** • Nulliparity: LGA 442 (27.3); AGA 2577 (40.4)** Fetal weight estimation performed up to 3 days before delivery: LGA 1256 (77.6); AGA 4715 (73.9) Gestational age at delivery (weeks): LGA 39.4±1.4; AGA 38.3±2.5** **p-value <0.001; *p-value 0.01 **Inclusion Criteria** Live-birth singleton pregnancy, gestational age at 37+0/7 to 42+0/7		BPD, HC), presenting part, placental location, and amniotic fluid estimation. The examinations were performed transabdominally using a high-quality ultrasound system by senior physicians who are ultrasound specialists or by experienced ultrasound technicians. Data analysis was performed using SPSS v 23.0. A p-value <0.05 was considered significant. Fisher's exact test and Mann-Whitney-Wilcoxon tests were used where appropriate.	Prevalence of LGA= 20.2%	question? CONCERN: LOW DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Unclear 2. If a threshold was used, was it pre-specified? Unclear Could the conduct or interpretation of the index test have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? Yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? Unclear Could the reference standard, its conduct, or its interpretation have

Study details	Participants	Interventions	Methods	Outcomes	Comments
	gestational weeks, and absence of major malformations or chromosomal abnormalities. **Exclusion Criteria** Women without documentation of biometric measurements (BPD, HC, AC, and FL) Women who delivered SGA neonates Women who were in active labour or with ruptured membranes at the time of sonographic assessment				introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? Yes 2. Did all patients receive a reference standard? Yes 3. Did patients receive the same reference standard? Yes 4. Were all patients included in the analysis? Yes Could the patient flow have introduced bias? RISK: LOW
Full citation Barel, O., Maymon, R., Elovits, M., Smorgick, N.,	**Sample size** N=14 089 SFWE estimations	**Tests** Index test: US estimated fetal weight <10th	**Methods** Sonographic fetal measurements were taken according to formal	**Results**	**Limitations** **Risk of bias assessed using QUADAS-II**

Study details	Participants	Interventions	Methods	Outcomes	Comments
Tovbin, J., Vaknin, Z., Evaluation of Fetal Weight Estimation Formulas in Assessing Small-for-Gestational-Age Fetuses, Ultraschall in der Medizin, 37, 283-9, 2016 **Ref Id** 756959 **Country/ies where the study was carried out** Israel **Study type** Retrospective cohort study **Aim of the study** To compare the accuracy of multiple sonographic fetal weight estimation models in assessing small-for-gestational age fetuses. **Study dates** January 2004 to September 2011 **Source of funding** Not reported	**Characteristics** Maternal age (years): 30.4±5.1 (range 16-53) Parity: 2.1±1.3 (1-13) Maternal weight (kg): 78±14.6 (range 33-175) Maternal height (m) 1.63±0.08 (range 1.34-1.86) Maternal BMI (kg/m2): 29.2±4.9 (range 12.8-68.3) Maternal gestational diabetes: 1293 (9.1%) Maternal pre-gestational diabetes: 60 (0.4%) **Inclusion Criteria** A live birth singleton pregnancy Birth weight >500g Gestational age >24 weeks **Exclusion Criteria** Detection of a fetal abnormality or a major malformation Active labour at the time of SFWE Ruptured membranes	percentile (Hadlock formula) Reference standard: Birth weight <10th percentile for gestational age based on actual birth weight from departmental computerised database **Timing** <7d from delivery (US done up to one week before delivery)	standards. The BPD, HC, AC, and FL were measured up to one week before delivery and expected birth weight was calculated using 26 different formulas. SFWEs were performed in obstetric ultrasound units by ultrasound technicians and by physicians trained in obstetrics and gynaecology. Statistical analysis was performed by SPSS. Fetal ultrasound measurements were used in the calculations of the formulas for the models analysed. The analysis was performed in several ways.	<table><tr><td></td><td>Reference test +</td><td>Reference test -</td><td>Total</td></tr><tr><td>Index test +</td><td>392</td><td>103</td><td>495</td></tr><tr><td>Index test -</td><td>826</td><td>12768</td><td>13594</td></tr><tr><td>Total</td><td>1218</td><td>12871</td><td>14089</td></tr></table> SGA Sensitivity= 32.20% Specificity= 99.20%	DOMAIN 1: PATIENT SELECTION A. RISK OF BIAS 1. Was a consecutive or random sample of patients enrolled? Unclear 2. Was a case-control design avoided? Yes 3. Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: LOW DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes 2. If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from

Study details	Participants	Interventions	Methods	Outcomes	Comments
					the review question? CONCERN: LOW DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? Yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? Yes 2. Did all patients receive a reference standard? Yes

Study details	Participants	Interventions	Methods	Outcomes	Comments
					3. Did patients receive the same reference standard? Yes 4. Were all patients included in the analysis? Yes Could the patient flow have introduced bias? RISK: LOW
Full citation Ben-Haroush, A., Yogev, Y., Hod, M., Bar, J., Predictive value of a single early fetal weight estimate in normal pregnancies, European Journal of Obstetrics, Gynecology, & Reproductive BiologyEur J Obstet Gynecol Reprod Biol, 130, 187-92, 2007 **Ref Id** 1121753 **Country/ies where the study was carried out** Israel **Study type** Retrospective cohort study	**Sample size** n=259 women **Characteristics** Maternal age (years): 28.5 ± 5.2 (17-42) Nulliparity: 35% Maternal weight before pregnancy (kg): 63 ± 12 (41-120) Maternal weight gain in pregnancy (kg): 13 ± 6 (-7 to +36) Maternal weight at delivery (kg): 76 ± 13 (50-144) Maternal BMI (kg/m2): 23.8 ± 4.4 (16.8-35.2) Gestational age of ultrasound (weeks): 32 ± 1.6 (28-34) Interval between EFW and delivery (weeks): 7 ± 2.2 (2-12)	**Tests** Index test: US estimated fetal weight ≤10th and ≥90th percentile (Hadlock formula) Reference standard: Birth weight ≤10th and ≥90th percentile for gestational age **Timing** >7d from delivery (US done at 28 to 34 weeks gestation)	**Methods** All ultrasound examinations were routinely performed at several outpatient clinics by experienced ultrasound technicians or physicians, and were covered by medical insurance. Since the sonographic measurements were performed prior to the admission of the patients at delivery - it was performed by different performers on different scanners. Gestational age was determined by last menstrual period and by ultrasonographic measurements of the crown-rump length before 12 weeks gestation. Fetal weight at 28-34 weeks gestation was estimated on the basis of	**Results** Reported results EFW SGA, BW SGA = 4 EFW SGA, BW LGA = 1 EFW SGA, BW AGA = 7 EFW LGA, BW SGA = 0 EFW LGA, BW LGA = 13 EFW LGA, BW AGA = 43 EFW AGA, BW SGA = 15 EFW AGA, BW LGA = 9 EFW AGA, BW AGA = 167 SGA Sensitivity=21% Specificity=96.6% Positive predictive value=33.3% Negative predictive value=93.9% Prevalence=7.3%* *Calculated by the NGA technical team LGA Sensitivity=56.5%	**Limitations** **Risk of bias assessed using QUADAS-II** DOMAIN 1: PATIENT SELECTION A. RISK OF BIAS 1. Was a consecutive or random sample of patients enrolled? yes 2. Was a case-control design avoided? yes 3. Did the study avoid inappropriate exclusions? No, women with any medical/obstetric problems in particular diabetes and hypertension, which in reality is a proportion of women who are at risk of having larger or smaller babies and of particular interest Could the selection of patients have introduced bias? RISK: MODERATE

Study details	Participants	Interventions	Methods	Outcomes	Comments
Aim of the study Evaluate the accuracy of ultrasound-based estimated fetal weight (EFW) at 28-34 weeks gestation in predicting small- and large-for-gestational age (SGA, LGA) infants at term **Study dates** Not reported **Source of funding** Not reported	Birth weight (g): 3230 ± 475 (1610-4360) Birth weight percentile: 49 ± 23* Birthweight percentile minus EFW percentile: -13 ± 26 (-75 to +80) *p=<0.001 compared with EFW percentile **Inclusion Criteria** Healthy, singleton pregnancy and ultrasound documentation of the fetal biparietal diameter, head circumference, and femur length performed at 28 and 34 weeks gestation **Exclusion Criteria** Hypertensive and diabetic pregnancies No smokers or medical/obstetrical problems		the biometry data using Hadlocks formula that used the fetal biparietal diameter, abdominal circumference, and femur length, and converted in percentiles according to locally developed growth charts for comparison with the birth weight percentiles. The birth weight percentiles were derived manually from the charts by one performer. A multivariate linear regression model was fitted to the data to predict the birth weight and the birth weight percentile. The resulting equation (projectile formula) of the stepwise analysis, which included the significant variables, was used to calculate the projected birth weight. The calculated birth weight was compared with the actual birth weight at delivery. Fetuses or infants with an estimated fetal or birth weight of ≤10 and ≥90 percentile, were categorised as SGA or LGA,respectively. For statistical analysis, SPSS statistical package was used, version 10.0. Analyses included paired Students t-test, receiver operating characteristic (ROC) curves, linear regression analysis, and	Specificity=81.8% Positive predictive value=23.2% Negative predictive value=95% Prevalence=8.8%* *Calculated by the NGA technical team	B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: MODERATE DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Unclear 2. If a threshold was used, was it pre-specified? Unclear, different sonographers using different ultrasound equipment to estimate fetal weight, no mention of protocol to follow Could the conduct or interpretation of the index test have introduced bias? RISK: MODERATE B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: MODERATE DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly

Study details	Participants	Interventions	Methods	Outcomes	Comments
			Kappa measure of agreement. For categorical analysis, chi-square test was used. The Bland and Altman plot and passing and bablock regression were used to compare between the calculated birth weight and the actual birth weight. Correlations and differences were considered significant with p was less than 0.05. Logathirmic transformation was used for skewed data. A group sample size of 168 subjects is sufficient to achieve 80% power to detect a difference of 10 percentiles between the EFW and birth weight percentiles, assuming a mean percentile of 50, wit known group standard deviations of 23 and an alpha value of 0.05		classify the target condition? yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? unclear Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? yes 2. Did all patients receive a reference standard? yes 3. Did patients receive the same reference standard? Yes 4. Were all patients included in the analysis? Yes Could the patient flow have introduced bias? RISK: LOW

Study details	Participants	Interventions	Methods	Outcomes	Comments
Full citation Blue, N. R., Beddow, M. E., Savabi, M., Katukuri, V. R., Mozurkewich, E. L., Chao, C. R., 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, Obstetrics & GynecologyObstet Gynecol, 131, 835-841, 2018 **Ref Id** 961485 **Country/ies where the study was carried out** USA **Study type** Retrospective cohort study **Aim of the study** To compare the RCOG and ACOG methods' ability to predict small for gestational age at birth.	**Sample size** N=1704 pregnancies **Characteristics** Maternal age (years): 28.8 (±6.5) Ethnicity: • White: 406 (23.8) • Hispanic: 844 (49.5) • Native American: 184 (10.8) • Black: 32 (1.9) • Asian: 41 (2.4) • Other or missing: 195 (11.4) Obstetric history: • Nulliparous: 461 (27.1) • Parous: 1243 (72.9) • Grand multiparous: 56 (3.3) Medical conditions:	**Tests** Index test: US estimated fetal weight <10th percentile (Hadlock formula) Reference test: Birth weight <10th percentile for gestational age **Timing** >7d from delivery (US done at mean of 14 days from delivery)	**Methods** Estimated fetal weights were calculated and estimated fetal weight or abdominal circumference percentiles assigned using the Hadlock estimated fetal weight and z score formulas. Statistical significance was by chi-squared, paired t test, or analysis of variance, depending on the type of variable.	**Results** | | Reference test + | Reference test - | Total | | --- | --- | --- | --- | | Index test + | 138 | 97 | 235 | | Index test - | 94 | 1375 | 1469 | | Total | 232 | 1472 | 1704 | SGA Sensitivity= 58.7% (52.1%-65.1%) Specificity= 93.6% (92.2%-94.8%) Positive likelihood ratio= 9.2 (7.3-11.5) Negative likelihood ratio= 0.44 (0.38-0.51)	**Limitations** **Risk of bias assessed using QUADAS-II** DOMAIN 1: PATIENT SELECTION A. RISK OF BIAS 1. Was a consecutive or random sample of patients enrolled? Yes 2. Was a case-control design avoided? Yes 3. Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: LOW DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes 2. If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index

Study details	Participants	Interventions	Methods	Outcomes	Comments
Study dates January 1st 2013 to March 31st 2017 **Source of funding** Not reported.	• Diabetes mellitus: 360 (21.1) • Pre-gestational diabetes mellitus: 81 (4.8) • Gestational diabetes mellitus: 279 (16.4) • Hypertensive disorder: 293 (17.2) • Chronic hypertension: 62 (3.6) • Preeclampsia: 223 (13.1) • Hemolysis, elevated liver enzymes, and low platelet count: 6 (0.4) • Eclampsia: 2 (0.1) Tobacco use: 239 (14) Illicit drug use: 275 (16.1) Gestational age at delivery (weeks): 37.7±2.8 Mean birth weight (grams): 2960 ± 865 **Inclusion Criteria** Neonates who both had an ultrasonographic estimated fetal weight performed within 30 days				test have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? Yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? Yes Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes</th><th>Comments</th></tr>
<tr><td></td><td>before (mean 14d from US to delivery) and were delivered at the study institution.

Exclusion Criteria
Multiple gestations, fetal hydrops, intrauterine fetal demise, inconsistent gestational age documentation, missing ultrasound or birth weight data, and congenital anomalies not allowing for accurate assessment of the biparietal diameter, head circumference, abdominal circumference, or femur length.</td><td></td><td></td><td></td><td>1. Was there appropriate interval between index tests and reference standard? Yes
2. Did all patients receive a reference standard? Yes
3. Did patients receive the same reference standard? Yes
4. Were all patients included in the analysis? Yes
Could the patient flow have introduced bias? RISK: LOW</td></tr>
<tr><td>Full citation

Blue, N. R., Savabi, M., Beddow, M. E., Katukuri, V. R., Fritts, C. M., Izquierdo, L. A., Chao, C. R., The Hadlock Method Is Superior to Newer Methods for the Prediction of the Birth Weight Percentile, Journal of ultrasound in medicine, 38, 587-596, 2019

Ref Id

1121773

Country/ies where the study was carried out</td><td>Sample size
N=831

Characteristics
Maternal age (years): SGA 27.7 (±6.5); non-SGA 28.9 (±6.5)
Ethnicity:

• Hispanic: SGA 66 (47.8); non-SGA 397 (47.8)
• White: SGA 35 (25.4); non-SGA 189 (22.7)
• Native American: SGA 10 (7.2);</td><td>Tests
Index test: US estimated fetal weight <10th percentile (Hadlock formula)
Reference standard: Birth weight <10th percentile for gestational age

Timing
<7d from delivery (US done at median 6.5 days from delivery)</td><td>Methods
Estimated fetal weights and percentiles were calculated by the Hadlock method, the Intergrowth-21st method, and the Salomon method. Each method's test characteristics to predict SGA were calculated. Ultrasound examinations were performed with Voluson E8, E10, or S10 machines by certified sonographers. Statistical analyses was by analysis of variance or chi-squared test, as appropriate. All analyses were performed with NCSS v.11.</td><td>Results
<table><tr><td></td><td>Reference test +</td><td>Reference test -</td><td>Total</td></tr><tr><td>Index test +</td><td>98</td><td>56</td><td>154</td></tr><tr><td>Index test -</td><td>40</td><td>637</td><td>677</td></tr><tr><td>Total</td><td>138</td><td>693</td><td>831</td></tr></table>SGA
Sensitivity= 71% (62.7%-78.4%)*
Specificity= 91.9% (88.5%-92.9%)*</td><td>Limitations
Risk of bias assessed using QUADAS-II
DOMAIN 1: PATIENT SELECTION
A. RISK OF BIAS
1. Was a consecutive or random sample of patients enrolled? Yes
2. Was a case-control design avoided? Yes
3. Did the study avoid inappropriate exclusions? Yes
Could the selection of patients have introduced bias? RISK: LOW
B. CONCERNS REGARDING APPLICABILITY Is there</td></tr>
</table>

Study details	Participants	Interventions	Methods	Outcomes	Comments
USA **Study type** Retrospective cohort study **Aim of the study** To compare a traditional ultrasound method for estimated fetal weight (EFW) calculation and fetal growth restriction diagnosis with 2 newer methods for the prediction of small for gestational age (SGA) at birth. **Study dates** January 1st 2013 to March 31st 2017 **Source of funding** Not reported	non-SGA 109 (13.1) • African American: SGA 4 (2.9); non-SGA 18 (2.2) • Asian: SGA 4 (2.9); non-SGA 18 (2.2) • Other/missing: SGA 17 (12.3); non-SGA 87 (10.5) Obstetric history: • Nulliparous: SGA 55 (39.9); non-SGA 229 (27.6) • Parous: SGA 83 (60.1); non-SGA 602 (72.4) • Grand multiparous: SGA 2 (1.4); non-SGA 36 (4.3) Medical conditions: • Diabetes: SGA 12 (8.7); non-SGA 203 (24.4) • Pre-gestational diabetes: SGA 3 (2.2); non-SGA 56 (6.7) • Gestational diabetes: SGA 9			Positive predictive value= 60.9% (54.6%-66.8%) Negative predictive value= 94% (92.4%-95.3%) *p≤0.03	concern that the included patients do not match the review question? CONCERN: LOW DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Unclear 2. If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? Yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? Unclear Could the reference standard, its conduct, or its

Study details	Participants	Interventions	Methods	Outcomes	Comments
	(6.5); non-SGA 149 (17.9) • Hypertensive disorder: SGA 46 (33.3); non-SGA 44 (5.3) • Chronic hypertension: SGA 9 (6.5); non-SGA 44 (5.3) • Preeclampsia: SGA 34 (24.6); non-SGA 151 (18.2) • Hemolysis, elevated liver enzymes, and low platelets: SGA 3 (2.2); non-SGA 6 (0.7) • Eclampsia: SGA 0 (0); non-SGA 1 (0.1) Tobacco use: SGA 27 (19.6); non-SGA 113 (13.6) Illicit drug use: SGA 30 (21.7); non-SGA 131 (15.8) Gestational age at delivery (weeks): SGA 36.6 (±3.4); non-SGA 37.0 (±3.0) Birth weight (grams): SGA 2069±610 (range 470 to 2919); non-SGA 2815±787 (range 470-4945)				interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? Yes 2. Did all patients receive a reference standard? Yes 3. Did patients receive the same reference standard? Yes 4. Were all patients included in the analysis? Yes Could the patient flow have introduced bias? RISK: LOW **Other information** US EFW <10th percentile (INTG and Salomon method) also studied.

Study details	Participants	Interventions	Methods	Outcomes	Comments
	Inclusion Criteria Singleton, live births delivered at the study Institute within 2 weeks (median 6.5d) of a US-derived estimated fetal weight. **Exclusion Criteria** Multiple gestations, discrepant gestational age (GA) documentation and fetal anomalies or conditions precluding accurate assessment of the biparietal diameter, head circumference, abdominal circumference, or femur length				
Full citation 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 in obstetrics & gynecology, 46, 216-20, 2015 **Ref Id** 1121804	**Sample size** n=1897 pregnant women **Characteristics** Age (years): 29.2 ± 4.9 Height (cm): 163 ± 6 Weight (kg): 62.2 ± 12.8 BMI (kg/m2): 23.3 ± 4.6 Chronic hypertension: 92 (4.8%) Gestational hypertension: 37 (2%) **Inclusion Criteria**	**Tests** Index test: US estimated fetal weight <10th centile (Hadlock formula) Reference standard: Birth weight <10th centile for gestational age **Timing** >7d from delivery (US done at 30-35 weeks gestation)	**Methods** Gestational age was determined from the date of the last menstrual period in women with a regular cycle, or by ultrasound assessment of crown-rump length or biparietal diameter. When there was a discrepancy of >7 days between age deducted from the last menstrual period and sonographic age, the sonographic age was used. SGA was defined as an estimated fetal weight (EFW) below the 10th	**Results** |	Reference test +	Reference test -	Total
---	---	---	---
Index test +	45	101	146
Index test -	110	1641	1751
Total	155	1742	1897	 SGA Sensitivity=29.0% (22.5%-36.6%)	**Limitations** **Risk of bias assessed using QUADAS-II** DOMAIN 1: PATIENT SELECTION A. RISK OF BIAS 1. Was a consecutive or random sample of patients enrolled? yes 2. Was a case-control design avoided? yes 3. Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? RISK: LOW B. CONCERNS REGARDING

Study details	Participants	Interventions	Methods	Outcomes	Comments
Country/ies where the study was carried out France **Study type** Prospective cohort study **Aim of the study** To evaluate the performance of screening for small-for gestational-age (SGA) fetuses by ultrasound biometry at 30-35 weeks gestation, and to determine the impact of screening on obstetric and neonatal outcomes **Study dates** 2003-2006 **Source of funding** Fondation pour la recherche médicale (FRM), french ministry of research: IFR program, INSERM Human Nutrition National Research Program, Diabetes National Research Program, French Ministry of Health Perinatality program, French agency for environment security, french national institute for population health surveillance, paris-sud	Pregnant and recruited prior to 24 weeks gestation **Exclusion Criteria** multiple pregnancy, known diabetes mellitus, illiteracy, and intention to deliver outside university hospital or move outside the region within 3 years of examination		percentile, according to the formula of Hadlock, where AC is abdominal circumference, FL is femur length, HC is head circumference and BPD is biparietal diameter. All ultrasound examinations were performed by one of five specialists who agreed on standardised procedures before the study commenced. Furthermore, the first five measurements made by each examiner were reviewed by another examiner to check for consistency with the protocol. The chi-square test or fishers exact test was used to compare qualitative variables and students t-test was used to compare continuous variables. Statistical analyses were carried out using SAS 9.3, p=<0.05 was considered to indicate statistical significance.	Specificity=94.2% (93%-95.2%) Positive predictive value=30.8% (23.9%-38.7%) Negative predictive value=93.7% (92.5%-94.8%) Prevalence=8.2%	APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: LOW DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes 2. If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? unclear

Study details	Participants	Interventions	Methods	Outcomes	Comments
university, french national institute for health education, nestle, mutually general de l'educations national, French-soeaking association for the study of diabetes and metabolism, national agency for research and the national institute for research in public health					Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? yes 2. Did all patients receive a reference standard? yes 3. Did patients receive the same reference standard? Yes 4. Were all patients included in the analysis? Yes Could the patient flow have introduced bias? RISK: LOW
Full citation	**Sample size**	**Tests**	**Methods**	**Results**	**Limitations Limitations**

Study details	Participants	Interventions	Methods	Outcomes	Comments
Erkamp, J. S., Voerman, E., Steegers, E. A. P., Mulders, Agmgj, Reiss, I. K. M., Duijts, L., Jaddoe, V. W. V., Gaillard, R., Second and third trimester fetal ultrasound population screening for risks of preterm birth and small-size and large-size for gestational age at birth: a population-based prospective cohort study, BMC Medicine, 18, 63, 2020 **Ref Id** 1241622 **Country/ies where the study was carried out** Netherlands **Study type** Prospective cohort study **Aim of the study** To examine whether single or combined second or third trimester fetal and placental ultrasound exminations are optimal to detect fetuses at risk for preterm birth, SGA and LGA **Study dates**	N = 7677 **Characteristics** Median age 30.3 (25.9 to 33.4 IQR), mean BMI 24.8, 58% Dutch/European, 56% nulliparous, 72% non-smokers **Inclusion Criteria** All pregnant women in a population based cohort study in Rotterdam (Generation R study) **Exclusion Criteria** Women without second and third trimester US data, non-singleton live births, missing outcome data	Index test: EFW was calculated using the Hadlock formula. EFW in lowest or highest decile was considered screen positive. Reference standard: information about birth weight was obtained from medical records, SGA and LGA at birth were defined as a gestational age adjusted birth weight <10th and >90th percentile in the study cohort respectively **Timing** >7d from delivery (US at median 30.4 weeks gestation)	Gestational age was established using data from the first ultrasound. All US were carried out in two dedicated research centres.	<table><tr><td>SGA</td><td>R+</td><td>R-</td></tr><tr><td>I+</td><td>331</td><td>436</td></tr><tr><td>I-</td><td>437</td><td>6466</td></tr></table>Sensitivity (calculated by NGA): 43% (95% CI 40 to 47%) Specificity (calculated by NGA): 94% (95% CI 93 to 94%) <table><tr><td>LGA</td><td>R+</td><td>R-</td></tr><tr><td>I+</td><td>273</td><td>494</td></tr><tr><td>I-</td><td>494</td><td>6409</td></tr></table>Sensitivity (calculated by NGA): 36% (95% CI 32 to 39%) Specificity (calculated by NGA): 93% (95% CI 92 to 93%)	**Risk of bias assessed using QUADAS-II** DOMAIN 1: PATIENT SELECTION A. RISK OF BIAS 1. Was a consecutive or random sample of patients enrolled? Yes 2. Was a case-control design avoided? Yes 3. Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: LOW DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes 2. If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? RISK: LOW

Study details	Participants	Interventions	Methods	Outcomes	Comments
Enrolled 2001-2005 **Source of funding** Academic/charitable organisations					B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? Yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>between index tests and reference standard? Yes
2. Did all patients receive a reference standard? Yes
3. Did patients receive the same reference standard? Yes
4. Were all patients included in the analysis? Yes
Could the patient flow have introduced bias? RISK: LOW</td></tr>
<tr><td>Full citation

Gabbay-Benziv, R., Aviram, A., Bardin, R., Ashwal, E., Melamed, N., Hiersch, L., Wiznitzer, A., Yogev, Y., Hadar, E., Prediction of Small for Gestational Age: Accuracy of Different Sonographic Fetal Weight Estimation Formulas, Fetal Diagnosis and Therapy, 40, 205-213, 2016

Ref Id

961951

Country/ies where the study was carried out</td><td>Sample size
N=6126 women with fetal weight estimation performed within 3 days of delivery

Characteristics
Maternal age (years): SGA 30.6 (±5.3); non-SGA 31.4 (±5.2) p=0.002
Nulliparity: SGA 329 (51.5); non-SGA 1976 (36) p=0.000
Gestational age at delivery (weeks): SGA 37.5 (±2.4); non-SGA 38.8 (±2.2)
Birth weight (grams): SGA 2190 (±426); non-SGA 3310 (±596) p=0.000</td><td>Tests
Index test: Different ultrasound tests (20 variations)
Reference test: birth weight >90th percentile for gestational age

Timing
<7d from delivery (US within 3 days of delivery)</td><td>Methods
Sonographic fetal weight estimations included all standard fetal biometry measurements (AC, FL, BPD and HC), presenting part, placental location and amniotic fluid estimation. The examinations were performed trans-abdominally using a high quality ultrasound system by senior physicians who are ultrasound specialists or by experienced ultrasound technicians. In the latter case, examinations were reviewed by a specialised physician.</td><td>Results
<table><tr><td></td><td>Reference test +</td><td>Reference test -</td><td>Total</td></tr><tr><td>Index test +</td><td>441</td><td>159</td><td>601</td></tr><tr><td>Index test -</td><td>197</td><td>5329</td><td>5525</td></tr><tr><td>Total</td><td>638</td><td>5488</td><td>6126</td></tr></table><u>SGA</u>
Sensitivity= 69.2%
Specificity= 97.1%
Positive predictive value= 73.7%
Negative predictive value= 96.4%</td><td>Limitations
Risk of bias assessed using QUADAS-II
DOMAIN 1: PATIENT SELECTION
A. RISK OF BIAS
1. Was a consecutive or random sample of patients enrolled? Yes
2. Was a case-control design avoided? Yes
3. Did the study avoid inappropriate exclusions? Yes
Could the selection of patients have introduced bias? RISK: LOW
B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the</td></tr>
</table>

Study details	Participants	Interventions	Methods	Outcomes	Comments
Not reported **Study type** Retrospective cohort study **Aim of the study** To compare the accuracy of various sonographic estimated fetal weight (sEFW) formulas for the prediction of small for gestational age (SGA) neonates. **Study dates** July 1st 2007 to 31st December 2014 **Source of funding** Not reported	**Inclusion Criteria** Live birth, singleton pregnancy, birth weight >500g, gestational age >24 weeks, and absence of major malformations or chromosomal abnormalities. **Exclusion Criteria** Women without documentation of all biometric measurements or women who were in active labour or with ruptured membranes at the time of sonographic assessment.		Data analysis by SPSS. A value of p<0.05 was considered significant. Fisher's exact test and Mann-Whitney-Wilcoxon test used as appropriate, depending on variable.	Positive likelihood ratio= 24.0 Negative likelihood ratio= 0.30	review question? CONCERN: LOW DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes 2. If a threshold was used, was it pre-specified? Unclear Could the conduct or interpretation of the index test have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? Yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? Unclear Could the reference standard, its conduct, or its interpretation have

Study details	Participants	Interventions	Methods	Outcomes	Comments
					introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? Yes 2. Did all patients receive a reference standard? Yes 3. Did patients receive the same reference standard? Yes 4. Were all patients included in the analysis? No. 196 participants excluded due to major anomaly or chromosomal abnormality identified or missing biometrical information. Could the patient flow have introduced bias? RISK: LOW

Study details	Participants	Interventions	Methods	Outcomes	Comments
Full citation Harding, K., Evans, S., Newnham, J., Screening for the small fetus: a study of the relative efficacies of ultrasound biometry and symphysiofundal height, Australian & New Zealand Journal of Obstetrics & Gynaecology, 35, 160-4, 1995 **Ref Id** 659829 **Country/ies where the study was carried out** Australia **Study type** Nested case-control study **Aim of the study** To investigate the most appropriate cut-off values for detecting birthweight <10th percentile at various GA using SH measurements, and ultrasound measurement of fetal abdominal circumference, and to compare these cut-off values with those in common practice. **Study dates** not reported	**Sample size** n=1,135 women in the 'intensive group' of an RCT **Characteristics** Maternal age (years): 28 (23-32) Race: • Caucasian: 88.3% • Aboriginal: 2.3% • Other: 9.4% Parity: • 0: 47.4% • 1: 29.9% • 2: 14.6% • >2: 8.1% Smoking: • Nil: 72.6% • 1-10/day: 17.1% • 10-20/day: 7.6% • >20/day: 2.7% Pregnancy and induced hypertension: 16.7% Antepartum haemorrhage: 8.5% Induction of labour: 32.8% Mode of delivery:	**Tests** Index test: SH <10th percentile for GA Reference standard: birth weight <10th percentile for GA using charts constructed from the Western Australia population **Timing** >7d from delivery (SH measured at 34 weeks analysed)	**Methods** GA was calculated from the last menstrual period unless it differed by more than 7 days from that predicted by ultrasound biometry. The women were scanned again at 24, 28, 34 and 38 weeks At each of the visits the SH was measured by a research midwife. These midwives were not involved in the clinical care of these women and were blinded to the hospital records and previous SH measurements. Each woman was asked to empty her bladder prior to the measurement. The measurement was made with the blank side of the tape measure facing upwards and extended from the uterine fundus to the upper border of the symphis pubis. All results were recorded to nearest 0.5cm. Receiver operator characteristics (ROC) curves were produced for each test as described by Sackett et al. Sensitivity was plotted against 1-specificity for a range of possible cut-offs. Using these curves it is possible to visually compare the efficacy of different cut-off	**Results** | | Reference +ve | Reference -ve | total | | --- | --- | --- | --- | | Index +ve | 33 | 105 | 138 | | Index -ve | 75 | 700 | 775 | | total | 108 | 805 | 913 | SGA (34 weeks) Sensitivity= 30.56% (22.05% to 40.16%)* Specificity= 86.96% (84.43 to 89.21%)* Positive predictive value= 23.86% (18.30% to 30.48%)* Negative predictive value= 90.35% (89.17% to 91.41%)* Prevalence of SGA= 11.8%* *Calculated by the NGA technical team	**Limitations** **Risk of bias assessed using QUADAS-II** DOMAIN 1: PATIENT SELECTION A. RISK OF BIAS 1. Was a consecutive or random sample of patients enrolled? yes 2. Was a case-control design avoided? yes 3. Did the study avoid inappropriate exclusions? No, all women with conditions predisposing to SGA or LGA were excluded Could the selection of patients have introduced bias? RISK: HIGH B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: HIGH DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes 2. If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? RISK: LOW

Study details	Participants	Interventions	Methods	Outcomes	Comments
Source of funding Raine research foundation of the University of Western Australia, National Health and Medical Research Council of Australia, and the Foundation of Women's and Infant's Health, King Edwards Memorial Hospital	• Elective CS: 8.2% • Non-elective CS: 8% • Spontaneous vaginal delivery: 65.2% • Instrumental vaginal delivery: 18.6% Neonatal: • Male sex: 51.4% • GA at delivery (weeks): 39 + 5 (38 + 4 to 40 + 4) • Preterm delivery (<37 weeks): 4.8% • Birthweight (g): 3373 • Apgar <7 at 5 min: 13% **Inclusion Criteria** Pregnant and recruited at 16-20 weeks gestation, who were receiving 5 scans between 18 weeks and 38 weeks gestation as part of an RCT **Exclusion Criteria** All women with an antenatal condition known		values for individual tests in addition to the overall efficacy of each test. The best cut off value was defined as the point which was furthest from the line of equality, and the best test as the one with the smallest area under the curve. SGA was defined as birth weight <10th percentile using charts constructed from the Western Australia population. These charts account for maternal height, parity, and infant sex although infant sex was not included in this study for the reason that the gender is usually unknown at the time of prenatal examination		B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? yes Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes</th><th>Comments</th></tr>
<tr><td></td><td>to affect fetal growth, including diabetes, preexisting hypertension, maternal renal disease, fetal congenital abnormalities and multiple pregnancies</td><td></td><td></td><td></td><td>between index tests and reference standard? yes
2. Did all patients receive a reference standard? yes
3. Did patients receive the same reference standard? Yes
4. Were all patients included in the analysis? No, 913/1,135=80% included in SH analysis
Could the patient flow have introduced bias? RISK: LOW

Other information
34 weeks measurements taken to represent 3rd trimester</td></tr>
<tr><td>Full citation

Khan, N., Ciobanu, A., Karampitsakos, T., Akolekar, R., Nicolaides, K. H., Prediction of large-for-gestational-age neonate by routine third-trimester ultrasound, Ultrasound in obstetrics & gynecology, 54, 326-333, 2019

Ref Id

1122199</td><td>Sample size
n=67,836 total population, of which:
n=21,989 women screened at 31 + 0 to 33 + 6 weeks
n=45,847 women screened at 35 + 0 to 36 + 6 weeks (diagnostic accuracy data for this population only)

Characteristics</td><td>Tests
Index test: US estimated fetal weight >90th percentile (Hadlock formula)
Reference standard: Birth weight >90th percentile for gestational age based on the fetal medicine foundation fetal and neonatal population weight charts

Timing
>7d from delivery (US at 35+0 to 36+6 weeks)</td><td>Methods
At third trimester visits , an ultrasound examination for fetal anatomy and measurement of fetal HC, AC, and FL for calculation of EFW using the Hadlock formula. Gestational age was determined by the measurement of fetal crown-rump length at 11-14 weeks or fetal HC at 19-24 weeks. The ultrasound examinations were carried out by our examiners who had obtained the fetal</td><td>Results
Birth ≥37 weeks
<table><tr><td></td><td>Reference test +</td><td>Reference test -</td><td>Total</td></tr><tr><td>Index test +</td><td>1944</td><td>2559</td><td>4503</td></tr><tr><td>Index test -</td><td>2285</td><td>39059</td><td>41344</td></tr><tr><td>Total</td><td>4229</td><td>41618</td><td>45847</td></tr></table>LGA</td><td>Limitations
Risk of bias assessed using QUADAS-II
DOMAIN 1: PATIENT SELECTION
A. RISK OF BIAS
1. Was a consecutive or random sample of patients enrolled? yes
2. Was a case-control design avoided? yes
3. Did the study avoid inappropriate exclusions? yes
Could the selection of patients have introduced bias? RISK: LOW</td></tr>
</table>

Study details	Participants	Interventions	Methods	Outcomes	Comments
Country/ies where the study was carried out UK **Study type** Retrospective cohort study **Aim of the study** To evaluate and compare the performance of routing ultrasonographic EFW and fetal AC at 31 + 0 to 33 + 6 and 35 + 0 to 36 + 6 weeks gestation in the prediction of a LGA neonate born at ≥37 weeks gestation, second to assess the additive value of fetal growth velocity between 32 and 36 weeks gestation to the performance of EFW at 35 + 0 to 36 + 6 weeks gestation for prediction of a LGA neonate, third, to define the predictive performance of a LGA neonate of different EFW cut-offs on routine ultrasound examination at 35 + 0 to 36 + 6 weeks gestation, and fourth, to propose a 2-stage strategy for identifying pregnancies with a LGA fetus that may benefit from iatrogenic delivery during the 38th gestational week	Age (years): non-LGA 31.5 (27.2 to 35.3); LGA 32.2 (28.3 to 35.8)** Weight (kg): non-LGA 79.2 (70.0 to 89.0); LGA 88.0 (78.5 to 100.0)** Height (cm): non-LGA 165 (160 to 169); LGA 167 (163 to 171)** Racial origin: • white: non-LGA 30,677 (73.7%); LGA 3,483 (82.4%)** • Black: non-LGA 6,708 (16.1%); LGA 488 (11.5%)** • South Asian: non-LGA 2,085 (5%); LGA 100 (2.4%)* • East Asian: non-LGA 882 (2.1%); LGA 57 (1.3%)* • Mixed: non-LGA 1,266 (3%); LGA 101 (2.4%)*** Cigarette smoker: non-LGA 3,565 (8.6%); LGA 158 (3.7%)** Conception: • natural: non-LGA 40,205 (96.6%); LGA 4,065 (96.1%) • ovulation drugs: non-LGA 228		medicine foundation certificate of competence in ultrasound examination for fetal abnormalities. The outcome measures of the study were birth weight >90th and >97th percentiles born at ≥37 weeks gestation, based on the fetal medicine foundation fetal and neonatal population charts. The windows statistical software package SPSS statistics for windows version 24 were used for data analysis	Sensitivity= 45.97% (44.46% to 47.48%)* Specificity= 93.85% (93.62% to 94.08%)* Positive predictive value= 43.10% (41.88% to 44.32%)* Negative predictive value= 94.49% (94.34% to 94.63%)* Prevalence of LGA= 9.2%*	B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: LOW DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? unclear 2. If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? yes 2. Were the reference standard results interpreted without

Study details	Participants	Interventions	Methods	Outcomes	Comments
Study dates May 2011 to September 2018 **Source of funding** The fetal medicine foundation (Charity no: 1037116)	(0.5%); LGA 29 (0.7%) • IVF:non-LGA 1185 (2.8%); LGA 135 (3.2%) Medical conditions: • Chronic hypertension: non-LGA 530 (1.3%); LGA 50 (1.2%) • Type I Diabetes Mellitus: non-LGA 118 (0.3%); LGA 49 (1.2%)** • Type II Diabetes Mellitus: non-LGA 169 (0.4%); LGA 39 (0.9%)** Obstetric history: • nulliparous: non-LGA 19,456 (46.7%); LGA 1,404 (33.2%) • Parous: - prior LGA: non-LGA 1,825 (4.4%); LGA 956 (22.6%)** • Parous: - no prior LGA: non-LGA 20,337 (48.9%); LGA 1,869 (44.2%)**				knowledge of the results of the index test? unclear Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? yes 2. Did all patients receive a reference standard? yes 3. Did patients receive the same reference standard? Yes 4. Were all patients included in the analysis? Yes Could the patient flow have introduced bias? RISK: LOW

Study details	Participants	Interventions	Methods	Outcomes	Comments
	Gestational age at screening (weeks): non-LGA 36.1 (35.9 to 36.4); LGA 36.1 (35.9 to 36.4) Estimated fetal weight (z-score): non-LGA -0.03 (-0.66 to 0.57); LGA 1.21 (0.71 to 1.75)** Gestational age at delivery (weeks): non-LGA 39.9 (39.0 to 40.8); LGA 40.0 (39.1 to 40.9)** Birth weight (z-score): non-LGA: -0.13 (-0.79 to 0.45); LGA 1.63 (1.44 to 1.93)** Birth weight (g): non-LGA 3,365 (3070 to 3645); LGA 4,240 (4065 to 4400)** ***p-value <0.05 **p-value <0.001; *p-value <0.01 **Inclusion Criteria** Singleton pregnancy delivering a non-malformed liveborn or stillborn neonate. **Exclusion Criteria** aneuploidy and/or major fetal abnormality				
Full citation	**Sample size** n=463	**Tests**	**Methods** Serial ultrasound examinations were	**Results**	**Limitations** **Risk of bias assessed using QUADAS-II**

Study details	Participants	Interventions	Methods	Outcomes	Comments
Lin, C. C., Sheikh, Z., Lopata, R., The association between oligohydramnios and intrauterine growth retardation, 76, 1100-4, 1990 **Ref Id** 1172087 **Country/ies where the study was carried out** USA **Study type** Retrospective cohort study **Aim of the study** To determine if oligohydramnios increases the accuracy of prenatal diagnosis of FGR. **Study dates** September 1985 to April 1988 **Source of funding** Not reported	**Characteristics** Maternal risk factors: • Hypertension: IUGR with OH 25%; IUGR without OH 28% • Smoking: IUGR with OH 56%; IUGR without OH 63% • Ethanol and other substance abuse: IUGR with OH 25%; IUGR without OH 20% • Misc medical problems: IUGR with OH 13%; IUGR without OH 20% Fetal outcome: • GA at birth (weeks): IUGR with OH 37.8 (1.7); IUGR without OH 37.7 (2) • Mean birth weight (g): IUGR with OH 2242 (273); IUGR without OH 2145 (377)	Index test: US AC <10th percentile (Shepards equation) Reference test: birth weight <10th percentile for GA **Timing** >7d from delivery (latest US could be 34 weeks)	performed by a single experienced ultrasonographer using a 3.5 MHz curvilinear real-time scanner and/or a 5-MHz sector scanner. Multiple static axial and longitudinal images of the fetus and uterus were taken during each sonographic examination. Each examination included fetal measurements such as biparietal diameter, head circumference, abdominal circumference, femur length, estimated fetal weight, and calculation of the ratios of head circumference to abdominal circumferences and femur length to abdominal circumference. Shepards equation for predicting fetal weight by ultrasound was used. At least 3 ultrasounds were performed for each case of IUGR. The first was done before 26 weeks gestation and was followed by 2 additional examinations in the 3rd trimester, separated by an interval of 2-4 weeks. An initial ultrasound examination was performed to confirm the GA. If subsequent US abdominal circumference measurements were below the 10th percentile	<table><tr><td></td><td>Reference +ve</td><td>Reference -ve</td><td>total</td></tr><tr><td>Index +ve</td><td>56</td><td>91</td><td>147</td></tr><tr><td>Index -ve</td><td>8</td><td>308</td><td>316</td></tr><tr><td>total</td><td>64</td><td>399</td><td>463</td></tr></table> SGA Sensitivity= 87.50% (76.85% to 94.45%)* Specificity= 77.19% (72.76% to 81.22%)* Positive predictive value= 38.05% (33.40% to 42.93%)* Negative predictive value= 97.47% (95.27% to 98.67%)* Prevalence of SGA= 13.8%* *Calculated by the NGA technical team	DOMAIN 1: PATIENT SELECTION A. RISK OF BIAS 1. Was a consecutive or random sample of patients enrolled? yes 2. Was a case-control design avoided? yes 3. Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: LOW DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Unclear 2. If a threshold was used, was it pre-specified? Unclear Could the conduct or interpretation of the index test have introduced bias? RISK: MODERATE B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from

Study details	Participants	Interventions	Methods	Outcomes	Comments
	• Preterm delivery (<37 weeks): IUGR with OH 19%; IUGR without OH 18% • Meconium-stained fluid: IUGR with OH 25%; IUGR without OH 33% • Intrapartum FHR decelerations: IUGR with OH 93%; IUGR without OH 69% • Intrapartum fetal acidosis (pH <7.20): IUGR with OH 20%; IUGR without OH 30% • Apgar score < 7: @1 min IUGR with OH 13%; IUGR without OH 18%; @5 min IUGR with OH 0%; IUGR without OH 3% **Inclusion Criteria** Pregnant women with an obstetric ultrasound examination performed at the Chicago Lying-In hospital **Exclusion Criteria**		according to the table of Tamura and Sabbagha, the case was classified as "suspected IUGR". Medical records of the study subjects were obtained for information on associated maternal risk factors and neonatal factors. Several analyses were performed to compare the first and second groups. Statistical analyses were performed using the chi squared test, two-sample student t test, and fisher exact test where appropriate. The level of significance was p <0.5 (two-tailed)		the review question? CONCERN: MODERATE DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? yes 2. Did all patients receive a reference standard? yes

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes</th><th>Comments</th></tr>
<tr><td></td><td>Multiple gestations, ruptured membranes, fetal malformations, or uncertain dates (cases lacking either an early ultrasound or a first - trimester clinical assessment).</td><td></td><td></td><td></td><td>3. Did patients receive the same reference standard? Yes
4. Were all patients included in the analysis? Yes
Could the patient flow have introduced bias?
RISK: LOW</td></tr>
<tr><td>Full citation

Monier, I., Blondel, B., Ego, A., Kaminiski, M., Goffinet, F., Zeitlin, J., Poor effectiveness of antenatal detection of fetal growth restriction and consequences for obstetric management and neonatal outcomes: a French national study, 122, 518-27, 2015

Ref Id

1122375

Country/ies where the study was carried out

France

Study type

Population-based study

Aim of the study</td><td>Sample size
n=14,404 population meeting inclusion criteria
n=14,100 included in analyses

Characteristics
Maternal characteristics
Maternal age (years): TP 28.8; FN 29.3; FP 28.3; TN 29.7
Nulliparous: TP 54.4%; FN 56.5%; FP 44.4%; TN 42.1%
Medical/obstetric factors:

• risk factors for FGR: TP 35.7%; FN 13.8%; FP 27.5%; TN 10.7%
• other risk factors: TP 12.1%; FN 8.8%; FP 27.9%; TN 13%</td><td>Tests
Index tests: US (defined as suspicion of FGR during pregnancy in the medical notes)
Reference standard: birthweight <10th centile for gestational age

Timing

>7d from delivery (US done at 30-35 weeks)</td><td>Methods
Suspicion of FGR was determined by whether there was mention of suspected growth retardation during pregnancy in the medical records. According to French recommendations, prenatal care should include a minimum of 7 prenatal visits and 3 US for a term birth. An ultrasound is recommended for each trimester of pregnancy and the 3rd trimester ultrasound is performed 30-35 weeks of gestation. Its main objective is to detect abnormalities of fetal growth and congenital abnormalities which cannot be diagnosed earlier. Suspicion of FGR should be based on an estimated fetal weight or</td><td>Results
<table><tr><td></td><td>Reference +ve</td><td>Reference -ve</td><td>total</td></tr><tr><td>Index +ve</td><td>265</td><td>271</td><td>536</td></tr><tr><td>Index -ve</td><td>954</td><td>12,610</td><td>13,564</td></tr><tr><td>total</td><td>1219</td><td>12,881</td><td>14,100</td></tr></table>
<u>SGA</u>

Sensitivity= 21.74% (19.45% to 24.16%)*

Specificity= 97.90% (97.63% to 98.14%)*

Positive predictive value= 49.30% (45.34% to 53.26%)*</td><td>Limitations
Risk of bias assessed using QUADAS-II
DOMAIN 1: PATIENT SELECTION
A. RISK OF BIAS
1. Was a consecutive or random sample of patients enrolled? yes
2. Was a case-control design avoided? yes
3. Did the study avoid inappropriate exclusions? yes
Could the selection of patients have introduced bias? RISK: LOW
B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question?
CONCERN: LOW

DOMAIN 2: INDEX TESTS
A. RISK OF BIAS</td></tr>
</table>

Study details	Participants	Interventions	Methods	Outcomes	Comments
To assess the proportion of SGA and normal birthweight infants suspected of fetal growth restriction during pregnancy, and to investigate obstetric and neonatal outcomes by suspicion of FGR and SGA status at birth **Study dates** 2010 **Source of funding** The 2010 French Perinatal Survey was funded by the Ministry of Health. Inserm unit 1153 received a grant from the Bettencourt foundation (coups d'elan pour la recherché francaise) in support of its research activities. One author was supported by a research grant from the assistance publique hopitaux de paris.	• no risk factor (low risk): TP 52.2%; FN 77.4%; FP 44.6%; TN 76.3% History of still birth: TP 5.1%; FN 1.3%; FP 2.7; TN 1.9 History of an SGA infant: TP 12.9%; FN 4.6%; FP 11.6%; TN 2.6% Pre-eclampsia: TP 8%; FN 2.6%; FP 10%; TN 1.8% BMI (kg/m2): TP 22.2; FN 22.4; FP 22.4; TN 23.5 Smoke in 3rd trimester: TP 33.6%; FN 32.6%; FP 23.6; TN 15.4% **Neonatal characteristics** Male sex: TP 42.3%; FN 52.1%; FP 41%; TN 53% Gestational age at birth (weeks): TP 37.4; FN 39.3; FP 37.1; TN 39.1 Birthweight (g): TP 2195; FN 2639; FP 2635; TN 3375 Birthweight percentile: • <3rd: TP 56.2%; FN 31.7% • 3rd-9th: TP 43.8%; FN 68.3% • 10th-25th: FP 47.6%; TN 13.8%		other biometric measurement under the 10th percentile for gestational age. In the study, information was noted on whether the medical term suspected FGR, but further details were not available on ultrasounds or doppler velocimetry. SGA was defined as a birthweight below the 10th percentile for gestational age and sex using the French reference standards. The population was divided into 4 groups on the basis of SGA status at birth and antenatal suspicion of FGR. Maternal and neonatal characteristics were described using chi-squared test or fishers exact test, as appropriate.	Negative predictive value= 93.00% (92.81% to 93.20%)* Prevalence of SGA= 8.6%* *Calculated by the NGA technical team	1. Were the index test results interpreted without knowledge of the results of the reference standard? Unclear 2. If a threshold was used, was it pre-specified? Unclear, diagnosis of FGR was based on documentation in the medical notes. Furthermore, different sonographers using different ultrasound equipment to estimate fetal weight, no mention of protocol to follow Could the conduct or interpretation of the index test have introduced bias? RISK: HIGH B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: HIGH DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? unclear

Study details	Participants	Interventions	Methods	Outcomes	Comments
	• >25th: FP 52.4%; 86.2% TP - true positive; FN - false negative; FP - false positive **Inclusion Criteria** All live birth and stillbirths with at least a birthweight of 500g were included **Exclusion Criteria** Births outside of continental France, medical terminations of pregnancy, multiple pregnancies, cases with missing data on gestational age, birthweight, and fetal sex were excluded				Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? yes 2. Did all patients receive a reference standard? yes 3. Did patients receive the same reference standard? Yes 4. Were all patients included in the analysis? no, 304 were excluded due to insufficient documentation Could the patient flow have introduced bias? RISK: LOW

Study details	Participants	Interventions	Methods	Outcomes	Comments
Full citation Rad, S., Beauchamp, S., Morales, C., Mirocha, J., Esakoff, T. F., Defining fetal growth restriction: abdominal circumference as an alternative criterion, Journal of Maternal-Fetal and Neonatal Medicine, 31, 3089-3094, 2018 **Ref Id** 963091 **Country/ies where the study was carried out** USA **Study type** Retrospective cohort study **Aim of the study** Compare EFW and AC percentiles as screening tests near term for SGA newborns in an effort to determine the best screening test for FGR. **Study dates** December 2008 to May 2014 **Source of funding** None	**Sample size** n=1594 **Characteristics** Maternal age (years): not SGA 33.9 (30.2-37.7); SGA 32.2 (29.1-36.3) Parity: not SGA 0 (0-1); SGA 0 (0-1) BMI (kg/m2): not SGA 29.4 (26.3-34.0); SGA 27.5 (24.9-30.8) Race/ethnicity: • Caucasian: not SGA 69.2%; SGA 58.5% • Black: not SGA 14.2%; SGA 19.3% • Asian: not SGA 14.8%; SGA 20.4% • Hispanic: not SGA 2%; SGA 1.9% GA at ultrasound (weeks): not SGA 37.3 (36.6-38.1); SGA 37.3 (36.7-37.1) GA at delivery (weeks): not SGA 39.2 (36.6-39.9); SGA 38.9 (38.1-39.5) Days from ultrasound of delivery: not SGA 11 (5.4-19.3); SGA 8.3 (3.2-14.5)	**Tests** Index test: US estimated fetal weight <10th percentile (Hadlock formula) Reference standard: birthweight <10th percentile **Timing** >7d from delivery (median 10.6 days)	**Methods** All ultrasound biometric measurements were performed by 1 of 12 experienced certified ultrasound technologists and/or maternal-fetal medicine physician specialists using GE Voluson or Phillips iU22 ultrasound machines. If more than one ultrasound were performed during the study period, only data from the ultrasound performed closest to delivery were included. EFW and AC percentile were calculated for each fetus using Hadlocks formula and standard (composite measurement of the fetal head, abdomen and femur) and categorised as <3, <5, <10< and/or >10 for GA. Newborn birthweight percentiles for GA were calculated using the Alexander et al standard. Newborns were classified as SGA or not, SGA was defined as a birthweight ≤10th percentile for GA. FGR was defined in 4 different ways: 1) AC percentile <10; EFW <10; both AC and EFW <10, either AC or EFW <10. Primary outcome was SGA birthweight.	**Results** | | Reference +ve | Reference -ve | total | |---|---|---|---| | Index +ve | 134 | 27 | 161 | | Index -ve | 131 | 1302 | 1433 | | total | 265 | 1329 | 1594 | <u>SGA</u> Sensitivity= 50.57% (44.38% to 56.74%)* Specificity= 97.97% (97.06% to 98.66%)* Positive predictive value= 83.20% (77.0% to 88.0%)* Negative predictive value= 90.10% (88.53% to 91.52%)* Prevalence of SGA= 16.6%* *Calculated by the NGA technical team	**Limitations** **Risk of bias assessed using QUADAS-II** DOMAIN 1: PATIENT SELECTION A. RISK OF BIAS 1. Was a consecutive or random sample of patients enrolled? yes 2. Was a case-control design avoided? yes 3. Did the study avoid inappropriate exclusions? yes Could the selection of patients have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: LOW DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Unclear 2. If a threshold was used, was it pre-specified? Unclear Could the conduct or interpretation of the index test have introduced bias? RISK: LOW B. CONCERNS REGARDING

Study details	Participants	Interventions	Methods	Outcomes	Comments
	Inclusion Criteria All singleton non-anomalous pregnancies undergoing ultrasound for fetal growth at >36 weeks (median 10.6d to delivery from US) gestation for any indication who delivered at the studies institution **Exclusion Criteria** Unknown or inaccurate GA dating, multiple gestations, major structural and/or chromosomal abnormalities, fetal demise, and delivery at a different institution		Sensitivity, specificity, false positive rate (FPR), positive-predictive value (PPV), and negative predictive value (NPV) of the various FGR definitions for SGA were calculated. Fishers exact and Wilcoxon Rank-sum tests were were used to compare variables. A p value of ≤0.5 was considered significant. The power analysis indicated that at least 140 SGA newborns would be needed to detect a 15% difference in sensitivity among FGR definitions, with 80% power. two-sided, and 5% level of significance (exact sign test of equality of paired proportions). All statistical analyses were conducted using SAS version 9.2		APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? unclear Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? yes

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>2. Did all patients receive a reference standard? yes
3. Did patients receive the same reference standard? Yes
4. Were all patients included in the analysis? yes
Could the patient flow have introduced bias? RISK: LOW</td></tr>
<tr><td>Full citation

Sekar, R., Khatun, M., Barrett, H. L., Duncombe, G., A prospective pilot study in assessing the accuracy of ultrasound estimated fetal weight prior to delivery, Australian and New Zealand Journal of Obstetrics and Gynaecology, 56, 49-53, 2016

Ref Id

446616

Country/ies where the study was carried out

Australia

Study type</td><td>Sample size
n=150

Characteristics
Maternal age (years): 31.1 (5.6)
Pregnancy BMI (kg/m2):

• normal 18.5-24.9: 30%
• overweight 25-29.9: 28.7%
• obese ≥ 30: 38.7%

Ethnicity:

• caucasian: 86.7%

Parity:</td><td>Tests
Index test: US estimated fetal weight <10th percentile or >90th percentile (Hadlock)
Reference standard: Birth weight <10th percentile or >90th percentile

Timing
<7d from delivery (all US done within 7 days of delivery)</td><td>Methods
All participants in the study were consecutively enrolled and allocated a study number. Women were allocated an odd or even number to one of the departments where ultrasound scans in the pregnancy were performed in the hospital, namely the department of medical imaging where general and obstetric-related examinations were performed and the centre for advanced prenatal care (where high risk pregnancies are scanned and monitored). To assess the inter observer reliability, 2 obstetric ultrasound operators, either medical practitioners or</td><td>Results

<table><tr><td></td><td>Reference +ve</td><td>Reference -ve</td><td>total</td></tr><tr><td>Index +ve</td><td>14</td><td>1</td><td>15</td></tr><tr><td>Index -ve</td><td>1</td><td>134</td><td>135</td></tr><tr><td>total</td><td>15</td><td>135</td><td>150</td></tr></table>
SGA
Sensitivity= 93.33% (68.05% to 99.83%)*
Specificity= 99.26% (99.94% to 99.98%)*
Positive predictive value= 99.33% (66.41% to 99%)*</td><td>Limitations
Risk of bias assessed using QUADAS-II
DOMAIN 1: PATIENT SELECTION
A. RISK OF BIAS
1. Was a consecutive or random sample of patients enrolled? yes
2. Was a case-control design avoided? yes
3. Did the study avoid inappropriate exclusions? No, only women who were scheduled for a induction or planned caesarean were included in this study
Could the selection of patients have introduced bias? RISK: HIGH
B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the</td></tr>
</table>

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes</th><th>Comments</th></tr>
<tr>
<td>Prospective cohort study

Aim of the study
To assess the accuracy of EFW measured by 2 monographers within 1 week of delivery using hadlock formula

Study dates
February to December 2013

Source of funding
Not reported</td>
<td>• 0: 46.6%
• 1: 53.4%

Nonsmoker: 92%
Actual birthweight (g): 3373
gestational week at birth:

• 28 + 5 to 35 + 6 weeks: 12.7%
• 37 to 41 weeks: 87.3%

Male: 76%

Inclusion Criteria
Pregnant women with singleton pregnancies who were either booked for induction of labour or elective caesarean section.

Exclusion Criteria
Multiple pregnancies and known fetal abnormalities</td>
<td></td>
<td>sonographers scanned each woman. Some of the sonographers and one of the medical practitioners worked within both departments. There were 15 obstetric ultrasound operators who scanned the women and 5 of them scanned more frequently than others. All sonographers had at least 1-year experience in obstetric scanning. The first sonographer would perform an EFW, amniotic fluid index and doppler studies assessing fetal well-being. Care was taken to delete the biometric measurements from the ultrasound screen, after a hard copy was made, before the 2nd sonographer entered the room. Subsequently, the 2nd sonographer, was blinded to the results of the 1st sonographer, performed the same measurements. Ultrasonography was performed using curvilinear 3.5-5, voluson E platforms BT10. Each monographer performed a total of 1-3 sets of measurements for biparietal diameter, abdominal circumference, head circumference, and femur length (BPD, AC, HC and FL) recorded in</td>
<td>Negative predictive value= 99.26% (95.28% to 99.84%)*

Prevalence of SGA= 10%*

<table>
<tr><th></th><th>Reference +ve</th><th>Reference -ve</th><th>total</th></tr>
<tr><td>Index +ve</td><td>9</td><td>6</td><td>15</td></tr>
<tr><td>Index -ve</td><td>6</td><td>129</td><td>135</td></tr>
<tr><td>total</td><td>15</td><td>135</td><td>150</td></tr>
</table>
<u>LGA</u>
Sensitivity= 60% (32.29% to 83.66%)*
Specificity= 95.56% (90.58% to 98.35%)*
Positive predictive value= 60% (38.25% to 78.42%)*
Negative predictive value= 95.56% (92.04% to 97.56%)*

Prevalence of LGA= 10%*

*Calculated by the NGA technical team</td>
<td>review question?
CONCERN: HIGH

DOMAIN 2: INDEX TESTS
A. RISK OF BIAS
1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes
2. If a threshold was used, was it pre-specified? Yes
Could the conduct or interpretation of the index test have introduced bias?
RISK: LOW

B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question?
CONCERN: LOW

DOMAIN 3: REFERENCE STANDARD
A. RISK OF BIAS
1. Is the reference standard likely to correctly classify the target condition? yes
2. Were the reference standard results interpreted without knowledge of the results of the index test? yes
Could the reference standard, its conduct, or its interpretation have</td>
</tr>
</table>

Study details	Participants	Interventions	Methods	Outcomes	Comments
			mm on each woman using standard views. The women and treating Drs were aware of the EFW. EFW was calculated according to the formula by Hadlock et al. Estimation of gestational age was by mothers last normal menstrual period or by ultrasound scanning before 20 weeks gestation. Fetuses with an EFW <10th percentile of birthweight for gestational age were classified as SGA and EFW >90th percentile of birthweight for gestational age were classified as LGA. Postdelivery, all babies were weighed on the day of birth consistently on Seca model 727 birth scales. The accuracy of fetal weight was examined by calculating the mean % difference using the formula (EFW-BW/BW) x 100. Cronbachs alpha measured the inter observer reliability between the 2 trained sonographers. Reliability coefficients were also measured for these 4 parameters individually. F-test was used to compare the biometric measurements and EFW within the sonographers.		introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? yes 2. Did all patients receive a reference standard? yes 3. Did patients receive the same reference standard? Yes 4. Were all patients included in the analysis? Yes Could the patient flow have introduced bias? RISK: LOW , **Risk of bias assessed using QUADAS-II** DOMAIN 1: PATIENT SELECTION A. RISK OF BIAS 1. Was a consecutive or random sample of patients enrolled? yes 2. Was a case-control design avoided? yes

Study details	Participants	Interventions	Methods	Outcomes	Comments
			The paired t-test was used to estimate the mean differences in individual biometric parameters measured by 2 sonographers and to test the mean percentage differences of EFWs. Sensitivity and specificity were calculated for the diagnostic assessment of SGA and LGA foetuses.		3. Did the study avoid inappropriate exclusions? No, only women who were scheduled for a induction or planned caesarean were included in this study Could the selection of patients have introduced bias? RISK: HIGH B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: HIGH DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes 2. If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW DOMAIN 3: REFERENCE STANDARD

Study details	Participants	Interventions	Methods	Outcomes	Comments
					A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? yes Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? yes 2. Did all patients receive a reference standard? yes 3. Did patients receive the same reference standard? Yes 4. Were all patients included in the analysis? Yes

Study details	Participants	Interventions	Methods	Outcomes	Comments
					Could the patient flow have introduced bias? RISK: LOW
Full citation Skovron, M. L., Berkowitz, G. S., Lapinski, R. H., Kim, J. M., Chitkara, U., Evaluation of early third-trimester ultrasound screening for intrauterine growth retardation, J Ultrasound MedJournal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine, 10, 153-9, 1991 **Ref Id** 1172088 **Country/ies where the study was carried out** USA **Study type** Prospective cohort study **Aim of the study** Evaluate the usefulness of early third trimester ultrasound fetal biometry for detecting IUGR and to compare the efficacy of	**Sample size** n=768 **Characteristics** Maternal: • age (mean, years): SGA 28; non-SGA 27 • Parity (multiparous): SGA 46%; non-SGA 46% • Ethnicity (non-white): SGA 71%; non-SGA 68% • Provider (private): SGA 30%; non-SGA 34% • Medical conditions (noted): SGA 38%; non-SGA 21% • Medications (noted): SGA	**Tests** Index test: US estimated fetal weight (Shepards formula) and AC <10th percentile for GA Reference standard: Birthweight <10th percentile **Timing** >7d from delivery (all early third trimester ultrasounds)	**Methods** Ultrasound examination for determination of fetal size between 26 and 34 weeks gestation. Data abstracted from the US examination records included BPD, HC, AC, FL, EFW, sonographic gestational age, and number of previous sonographic examinations. Ultrasound measurements were obtained in the standard manner using a linear-array, real-time system with a 3.5-MHz focused transducer. EFW was calculated from BPD and AC measurements, using the equation of Shepard et al. In 627 pregnancies, GA was based on the date of LMP, which was within 2 weeks of that determined by sonography. In 129 pregnancies, GA was determined by a previous dating scan, and in 12 pregnancies, GA was based on the physicians clinical judgement.	**Results** | | Reference +ve | Reference -ve | total | | --- | --- | --- | --- | | Index +ve | 17 | 21 | 38 | | Index -ve | 52 | 678 | 730 | | total | 69 | 699 | 768 | SGA Sensitivity= 24.64% (15.05% to 36.49%)* Specificity= 97.00% (95.44% to 98.13%)* Positive predictive value= 44.78% (31.02% to 59.39%)* Negative predictive value= 92.86% (91.91% to 93.71%)* Prevalence of SGA= 9.0%*	**Limitations** **Risk of bias assessed using QUADAS-II** DOMAIN 1: PATIENT SELECTION A. RISK OF BIAS 1. Was a consecutive or random sample of patients enrolled? yes 2. Was a case-control design avoided? yes 3. Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: LOW DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Unclear

Study details	Participants	Interventions	Methods	Outcomes	Comments
several recommended parameters **Study dates** 1985-1987 **Source of funding** Health services improvement fund of empire state blue cross and blue shield	20%; non-SGA 9% Ultrasound: • Gestation at first study examination (median, weeks): SGA 30; non-SGA 30 • Subsequent examinations: SGA 51%; non-SGA 35% • HC (mean): SGA 26.5; non-SGA 27.3 • AC (mean): SGA 23.7; non-SGA 25.3 • EFW (mean, g): SGA 1261; non-SGA 1468 • FL/AC (mean): SGA 2.3; non-SGA 2.2 Neonatal: • Preterm (<37 weeks): SGA 13%; non-SGA 10% • Sex of infant (male): SGA 46%; non-SGA 51% • VLBW infant (<1500g): SGA		Infants falling below the 10th percentile of birth weight for GA and sex, according to the normogram developed by Brenner et al, were categorised as SGA. Percentile and deviation of fetal ultrasound measurements for GA was assigned with reference to normograms for AC, HC, EFW and FL/AC ratio. The performance of the four ultrasound parameters in detecting IUGR was examined by ROC curve analysis.	*Calculated by the NGA technical team	2. If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? RISK: MODERATE B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: MODERATE DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes</th><th>Comments</th></tr>
<tr><td></td><td>9%; non-SGA 0.4%

Inclusion Criteria
Singleton pregnancies

Exclusion Criteria
Gestational diabetes, placenta praevia, preterm labour, Rh sensitisation, fetal anomalies</td><td></td><td></td><td></td><td>question? CONCERN: LOW

DOMAIN 4: FLOW AND TIMING
A. RISK OF BIAS
1. Was there appropriate interval between index tests and reference standard? yes
2. Did all patients receive a reference standard? yes
3. Did patients receive the same reference standard? Yes
4. Were all patients included in the analysis? Yes
Could the patient flow have introduced bias?
RISK: LOW</td></tr>
<tr><td>Full citation

Sovio, U., White, I. R., Dacey, A., Pasupathy, D., Smith, G. C. S., Screening for fetal growth restriction with universal third trimester ultrasonography in nulliparous women in the Pregnancy Outcome Prediction (POP) study: a prospective cohort study [Erratum: Lancet 2015; 386(10008): 2058],</td><td>Sample size
n=4512 women

Characteristics
Maternal age (years):

• <20 years: 139 (4%)
• 20-24.9 years: 520 (13%)
• 25-29.9 years: 1225 (31%)</td><td>Tests
Index test: US estimated fetal weight <10th percentile (Hadlock)
Reference test: Birth weight <10th percentile (calculated from a UK reference).

Timing
>7d from delivery (36 week appointment)</td><td>Methods
All research scans after the dating scan were done by one of a team of six sonographers, all of whom received standard training. All ultrasound examinations followed the same protocols as those used in the clinical service. At the 28 and 36 week research appointments, umbilical and uterine artery Doppler flow velocimetry were</td><td>Results
<table><tr><td></td><td>Reference +ve</td><td>Reference -ve</td><td>total</td></tr><tr><td>Index +ve</td><td>199</td><td>363</td><td>562</td></tr><tr><td>Index -ve</td><td>153</td><td>3262</td><td>3415</td></tr><tr><td>total</td><td>352</td><td>3625</td><td>3977</td></tr></table></td><td>Limitations
Risk of bias assessed using QUADAS-II
DOMAIN 1: PATIENT SELECTION
A. RISK OF BIAS
1. Was a consecutive or random sample of patients enrolled? Yes
2. Was a case-control design avoided? Yes
3. Did the study avoid inappropriate exclusions? Yes</td></tr>
</table>

Study details	Participants	Interventions	Methods	Outcomes	Comments
Lancet, 386, 2089-2097, 2015 **Ref Id** 1122666 **Country/ies where the study was carried out** United Kingdom **Study type** Prospective cohort study **Aim of the study** To determine the diagnostic effectiveness of universal ultrasonic fetal biometry in the third trimester as a screening test for small-for-gestational-age (SGA) infants, and whether the risk of morbidity associated with being small differed in the presence or absence of ultrasonic markers of fetal growth restriction. **Study dates** 14th January 2008 to 31st July 2012 **Source of funding** Medical Research Council, National Institute for Health Research, Cambridge	• 30-34.9 years: 1485 (37%) • 35-39.9 years: 534 (13%) • ≥40 years: 74 (2%) Ethnicity: • White: 3696 (93%) • Missing: 69 (2%) Married: 2727 (69%) Smokers: 185 (5%) Alcohol consumption: • Any: 183 (5%) • Missing: 69 (2%) BMI (kg/m^2): • <25: 2325 (58%) • 25-29.9: 1117 (28%) • 30-34.9: 377 (9%) • 35-39.9: 110 (3%) • ≥40: 47 (1%) • Missing: 1 (<1%) Diabetes: • Type 1 or type 2: 12 (<1%)		repeated, and ultrasonographic measurement of fetal biparietal diameter, head circumference, abdominal circumference, and femur length were also done using standard techniques. Gestational age was defined on the basis of ultrasonographic estimation at the time of the first scan, as recommended.	<u>SGA</u> Sensitivity= 57% (95% CI 51 to 62)* Specificity= 90% (89 to 91)* Positive predictive value= 35% (31 to 39)** Negative predictive value= 96% (95 to 96)* *p<0.0001 **p=0.0001	Could the selection of patients have introduced bias? RISK: Low B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN:Low DOMAIN 2: INDEX TESTS A. RISK OF BIAS 1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes 2. If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? RISK: Low B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: Low DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? Yes

Study details	Participants	Interventions	Methods	Outcomes	Comments
Comprehensive Biomedical Research Centre, and the Stillbirth and Neonatal Death Society.	• Gestational: 162 (4%) • Missing: 5 (<1%) **Inclusion Criteria** • Primiparous women with a singleton pregnancy; • Women who attended research scans booked before delivery; • Women who had a live birth at the Rosie Hospital. **Exclusion Criteria** • Multiple pregnancy; • Women who delivered before their 28 week scan appointment				2. Were the reference standard results interpreted without knowledge of the results of the index test? Yes Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: Low B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN:Low DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? Yes 2. Did all patients receive a reference standard? Yes 3. Did patients receive the same reference standard? Yes 4. Were all patients included in the analysis? No (5.4% lost to follow up) Could the patient flow have introduced bias? RISK: Low

Study details	Participants	Interventions	Methods	Outcomes	Comments
Full citation Sovio, U., Moraitis, A. A., Wong, H. S., Smith, G. C. S., Universal vs selective ultrasonography to screen for large-for-gestational-age infants and associated morbidity, Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 51, 783-791, 2018 **Ref Id** 963458 **Country/ies where the study was carried out** United Kingdom **Study type** Prospective cohort study **Aim of the study** To compare the diagnostic effectiveness of selective vs universal ultrasonography as a screening test for large-for-gestational age (LGA) infants, and to determine whether previously described ultrasound markers of excessive fetal	**Sample size** See Sovio 2015 **Characteristics** See Sovio 2015 **Inclusion Criteria** See Sovio 2015 **Exclusion Criteria** See Sovio 2015	**Tests** Index test: US estimated fetal weight >90th percentile (Hadlock formula) Reference standard: EFW>90th percentile (using an externally derived reference range) >7 days from birth	**Methods** See Sovio 2015	**Results** | | Reference +ve | Reference -ve | total | | --- | --- | --- | --- | | Index +ve | 67 | 110 | 177 | | Index -ve | 127 | 3562 | 3689 | | total | 194 | 3672 | 3866 | LGA Sensitivity= 38% (95% CI 31 to 45) p=0.005 Specificity= 97% (95% CI 96 to 97) p<0.0001 Positive predictive value= 35 (95% CI 28 to 41) p<0.002 Negative predictive value= 97 (95% CI 96 to 98) p=0.01	**Limitations** See Sovio 2015

Study details	Participants	Interventions	Methods	Outcomes	Comments
growth could identify suspected LGA fetuses that are at increased risk of adverse neonatal outcome. **Study dates** See Sovio 2015 **Source of funding** See Sovio 2015					
Full citation Turitz, A. L., Quant, H., Schwartz, N., Elovitz, M., Bastek, J. A., Isolated abdominal circumference < 5% or estimated fetal weight 10 to 19% as predictors of small for gestational age infants, American Journal of Perinatology, 31, 469-476, 2014 **Ref Id** 963604 **Country/ies where the study was carried out** USA **Study type** Retrospective cohort study **Aim of the study**	**Sample size** N=10 642 pregnancies **Characteristics** Not reported **Inclusion Criteria** Women with singleton pregnancies who presented for at least one growth ultrasound between 26 and 36 weeks gestational age. **Exclusion Criteria** Fetal anomalies, stillbirths, and twins with one fetal loss.	**Tests** Index test: US estimated fetal weight <10th percentile (Hadlock formula) Reference standard: Birth weight <10th percentile for gestational age based on the Alexander curve (a national reference nomogram) **Timing** >7 days from delivery (US done between 26 and 36 weeks gestation)	**Methods** All growth ultrasounds were performed by skilled ultrasound personnel, under the supervision of attending Maternal Fetal Medicine physicians. Fetuses with growth restriction <10% underwent antepartum surveillance with twice weekly modified biophysical profile and weekly umbilical artery Doppler. Calculated percentiles were applied to the Alexander curve (national reference nomogram) to generate fetal weight percentages. Associations between categorical variables were compared with chi square analyses. Multivariable logistic regression equations were used, where appropriate, controlling for	**Results** |	Reference test +	Reference test -	Total
---	---	---	---
Index test +	267	284	551
Index test -	579	9512	10091
Total	846	9796	10642	 <u>SGA</u> Sensitivity= 31.6% Specificity= 97.1% Positive predictive value= 70.2% Negative predictive value= 86.9%	**Limitations** **Risk of bias assessed using QUADAS-II** DOMAIN 1: PATIENT SELECTION A. RISK OF BIAS 1. Was a consecutive or random sample of patients enrolled? Yes 2. Was a case-control design avoided? Yes 3. Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: LOW DOMAIN 2: INDEX TESTS A. RISK OF BIAS

Study details	Participants	Interventions	Methods	Outcomes	Comments
To determine whether: • Isolated fetal abdominal circumference <5% (AC5) in absence of growth restriction (estimated fetal weight <10% [EFW10]) • Or borderline fetal growth 10 to 19% (EFW10–19) predicts subsequent fetal and/or neonatal growth restriction. **Study dates** January 2008 to December 2011 **Source of funding** Grant Number K12HD001265 (PI Driscoll; Scholar Bastek) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development.			confounders. Analysis was done by STATA.		1. Were the index test results interpreted without knowledge of the results of the reference standard? Unclear 2. If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW DOMAIN 3: REFERENCE STANDARD A. RISK OF BIAS 1. Is the reference standard likely to correctly classify the target condition? Yes 2. Were the reference standard results interpreted without knowledge of the results of the index test? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW B. CONCERNS REGARDING APPLICABILITY Is there

Study details	Participants	Interventions	Methods	Outcomes	Comments
					concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW DOMAIN 4: FLOW AND TIMING A. RISK OF BIAS 1. Was there appropriate interval between index tests and reference standard? Yes 2. Did all patients receive a reference standard? Yes 3. Did patients receive the same reference standard? Yes 4. Were all patients included in the analysis? Yes Could the patient flow have introduced bias? RISK: LOW **Other information** Women were included if they had an ultrasound from 26 gestational weeks.

1

Appendix E – Forest plots

Forest plots for review question: What is the best method using third trimester measurements to predict birth weight?

Figure 2: Ultrasound for SGA, >7 days from delivery

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Akolekar 2019	2420	1689	2860	38878	0.46 [0.44, 0.47]	0.96 [0.96, 0.96]
Ben-Haroush 2007	4	8	15	232	0.21 [0.06, 0.46]	0.97 [0.94, 0.99]
Blue 2018	138	97	94	1375	0.59 [0.53, 0.66]	0.93 [0.92, 0.95]
Callec 2015	45	101	110	1641	0.29 [0.22, 0.37]	0.94 [0.93, 0.95]
Erkamp 2020	331	436	437	6466	0.43 [0.40, 0.47]	0.94 [0.93, 0.94]
Lin 1990	56	91	8	308	0.88 [0.77, 0.94]	0.77 [0.73, 0.81]
Monier 2015	265	271	954	12610	0.22 [0.19, 0.24]	0.98 [0.98, 0.98]
Rad 2018	134	27	131	1302	0.51 [0.44, 0.57]	0.98 [0.97, 0.99]
Skovron 1991	17	21	52	678	0.25 [0.15, 0.36]	0.97 [0.95, 0.98]
Sovio 2015	199	363	153	3262	0.57 [0.51, 0.62]	0.90 [0.89, 0.91]
Turitz 2014	267	284	579	9512	0.32 [0.28, 0.35]	0.97 [0.97, 0.97]

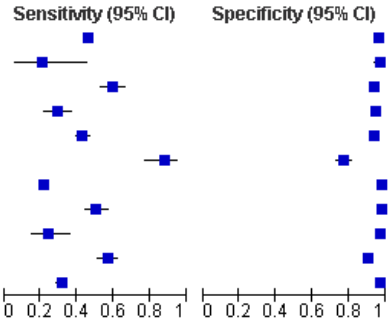


Figure 3: Ultrasound for SGA, <7 days from delivery

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Barel 2016	392	103	826	12768	0.32 [0.30, 0.35]	0.99 [0.99, 0.99]
Blue 2019	98	56	40	637	0.71 [0.63, 0.78]	0.92 [0.90, 0.94]
Gabbay-Benziv 2016	441	159	197	5329	0.69 [0.65, 0.73]	0.97 [0.97, 0.98]
Sekar 2016	14	1	1	134	0.93 [0.68, 1.00]	0.99 [0.96, 1.00]

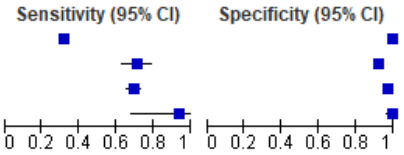


Figure 4: Ultrasound for LGA, >7 days from delivery

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ben-Haroush 2007	13	43	10	193	0.57 [0.34, 0.77]	0.82 [0.76, 0.86]
Erkamp 2020	273	494	494	6409	0.36 [0.32, 0.39]	0.93 [0.92, 0.93]
Khan 2019	1944	2559	2285	39059	0.46 [0.44, 0.47]	0.94 [0.94, 0.94]
Sovio 2018	67	127	110	3562	0.38 [0.31, 0.45]	0.97 [0.96, 0.97]

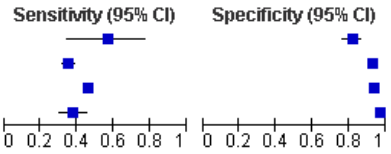


Figure 5: Ultrasound for LGA, <7 days from delivery

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Aviram 2017	1215	574	402	5804	0.75 [0.73, 0.77]	0.91 [0.90, 0.92]
Sekar 2016	9	6	6	129	0.60 [0.32, 0.84]	0.96 [0.91, 0.98]

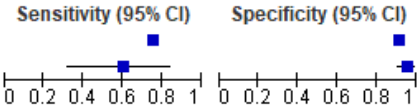
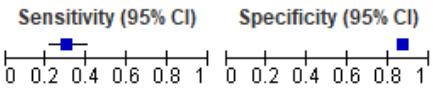


Figure 6: Symphysis-fundal height, >7 days from delivery

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Harding 1995	33	105	75	700	0.31 [0.22, 0.40]	0.87 [0.84, 0.89]



Appendix F – GRADE tables

GRADE tables for review question: What is the best method using third trimester measurements to predict birth weight?

Table 5: Fetal growth monitoring

Index test	No of studies	No of participants	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality of evidence (GRADE)	PPV[6]	NPV[6]
Ultrasound for SGA, >7 days from delivery	11[1]	88921	Sensitivity = 0.43 (0.29 to 0.58)	No serious	Very serious[2]	No serious	No serious	LOW	0.53	0.93
			Specificity = 0.95 (0.92 to 0.97)	No serious	Serious[3]	No serious	No serious	MODERATE		
Ultrasound for SGA, <7 days from delivery	4[1]	21196	Sensitivity = 0.66 (0.30 to 0.92)	No serious	Serious[3]	No serious	Serious[4]	LOW	0.81	0.96
			Specificity = 0.98 (0.91 to 0.99)	No serious	No serious	No serious	No serious	HIGH		
Ultrasound for LGA, >7 days from delivery	4[1]	57642	Sensitivity = 0.43 (0.24 to 0.65)	No serious	No serious	No serious	Serious[4]	MODERATE	0.38	0.94
			Specificity = 0.93 (0.80 to 0.98)	No serious	Serious[3]	No serious	Very serious[5]	VERY LOW		
Ultrasound for LGA, <7 days from delivery	2[1]	8145	Sensitivity = 0.70 (0.13 to 0.96)	No serious	No serious	No serious	Very serious[5]	LOW	0.50	0.97
			Specificity = 0.93 (0.58 to 0.99)	No serious	No serious	No serious	Very serious[5]	LOW		
		913	Sensitivity = 0.31 (0.22 to 0.40)	No serious	Not applicable	No serious	No serious	HIGH	0.24	0.91

SFH for SGA, >7 days from delivery	1 (Harding 1995)		Specificity = 0.87 (0.84 to 0.89)	No serious	Not applicable	No serious	No serious	HIGH		

SGA: small for gestational age; LGA: large for gestational age; CI: confidence interval; SFH: symphysis-fundal height, PPV: positive predictive value, NPV: negative predictive value

1 See corresponding forest plot for studies contributing to this outcome

2 Evidence downgraded by 2 levels due to considerable heterogeneity in individual study estimates across one decision making threshold (0.8 and 0.95 for sensitivity, 0.75 and 0.9 for specificity) and visual inspection of plot

3 Evidence downgraded by 1 level due to considerable heterogeneity in individual study estimates across one decision making threshold (0.8 and 0.95 for sensitivity, 0.75 and 0.9 for specificity)

4 Evidence downgraded by 1 level due confidence intervals crossing one decision making threshold (0.8 and 0.95 for sensitivity, 0.75 and 0.9 for specificity)

5 Evidence downgraded by 2 levels due confidence intervals crossing two decision making thresholds (0.8 and 0.95 for sensitivity, 0.75 and 0.9 for specificity)

6 Calculated by applying meta-analysed sensitivity and specificity to representative prevalence for SGA (11.5% from Akolekar 2019) and LGA (9.2% from Khan 2019)

Appendix G – Economic evidence study selection

Economic evidence study selection for review question: What is the best method using third trimester measurements to predict birth weight?

A single economic search was undertaken for all topics included in the scope of this guideline. No economic studies were identified which were applicable to this review question. See supplementary material 2 for details.

Appendix H – Economic evidence tables

Economic evidence tables for review question: What is the best method using third trimester measurements to predict birth weight?

No evidence was identified which was applicable to this review question.

Appendix I – Economic evidence profiles

Economic evidence profiles for review question: What is the best method using third trimester measurements to predict birth weight?

No evidence was identified which was applicable to this review question.

Appendix J – Economic analysis

Economic analysis for review question: What is the best method using third trimester measurements to predict birth weight?

No economic analysis was conducted for this review question.

Appendix K – Excluded studies

Excluded studies for review question: What is the best method using third trimester measurements to predict birth weight?

Clinical studies

Table 6: Excluded studies and reasons for their exclusion

Study	Reason for exclusion
Akolekar, R., Ciobanu, A., Zingler, E., Syngelaki, A., Nicolaides, K. H., Routine assessment of cerebroplacental ratio at 35-37 weeks' gestation in the prediction of adverse perinatal outcome, American Journal of Obstetrics & GynecologyAm J Obstet Gynecol, 221, 65.e1-65.e18, 2019	Did not assess accuracy of ultrasound or SFH measurement for predicting birth weight.
Akolekar, R., Syngelaki, A., Gallo, D. M., Poon, L. C., Nicolaides, K. H., Umbilical and fetal middle cerebral artery Doppler at 35-37 weeks' gestation in the prediction of adverse perinatal outcome, Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 46, 82-92, 2015	Did not assess accuracy of ultrasound or SFH measurement for predicting birth weight.
Atkinson,M.W., Maher,J.E., Owen,J., Hauth,J.C., Goldenberg,R.L., Copper,R.L., The predictive value of umbilical artery Doppler studies for preeclampsia or fetal growth retardation in a preeclampsia prevention trial, Obstetrics and Gynecology, 83, 609-612, 1994	Index test not of interest (Doppler ultrasound)
Baird, S. M., Davies-Tuck, M., Coombs, P., Knight, M., Wallace, E. M., Detection of the growth-restricted fetus: which centile charts?, Sonography, 3, 81-86, 2016	Did not assess accuracy of ultrasound or SFH measurement for predicting birth weight.
Bais, J. M., Eskes, M., Pel, M., Bonsel, G. J., Bleker, O. P., Effectiveness of detection of intrauterine growth retardation by abdominal palpation as screening test in a low risk population: an observational study, European Journal of Obstetrics, Gynecology, & Reproductive Biology, 116, 164-9, 2004	Index test not of interest for review: clinical examination only
Bakalis, S., Peeva, G., Gonzalez, R., Poon, L. C., Nicolaides, K. H., Prediction of small-for-gestational-age neonates: screening by biophysical and biochemical markers at 30-34 weeks, Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 46, 446-51, 2015	Incorrect index tests (only US in combination with non-protocol relevant tests).
Bakalis, S., Silva, M., Akolekar, R., Poon, L. C., Nicolaides, K. H., Prediction of small-for-gestational-age neonates: screening by fetal biometry at 30-34 weeks, Ultrasound in obstetrics & gynecology : the official journal of	Incorrect index tests (only US by Z-score).

Study	Reason for exclusion
the International Society of Ultrasound in Obstetrics and Gynecology, 45, 551-558, 2015	
Basuki, T. R., 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 Diagnosis & TherapyFetal Diagn Ther, 45, 230-237, 2019	Incorrect index tests (only US by Z-score)
Beattie, R. B., Dornan, J. C., Antenatal screening for intrauterine growth retardation with umbilical artery Doppler ultrasonography, British Medical Journal, 298, 631-635, 1989	Index test not of interest for review: doppler ultrasound (umbilical artery)
Bergman, E., Axelsson, O., Kieler, H., Sonesson, C., Petzold, M., Relative growth estimated from self-administered symphysis fundal measurements, Acta Obstetricia et Gynecologica Scandinavica, 90, 179-85, 2011	Testing began before third trimester.
Bergman, E., Axelsson, O., Petzold, M., Sonesson, C., Kieler, H., Self-administered symphysis-fundus measurements analyzed with a novel statistical method for detection of intrauterine growth restriction: A clinical evaluation, Acta Obstetricia et Gynecologica Scandinavica, 90, 890-896, 2011	Testing began before third trimester.
Bligh, L. N., Al Solai, A., Greer, R. M., Kumar, S., Diagnostic Performance of Cerebroplacental Ratio Thresholds at Term for Prediction of Low Birthweight and Adverse Intrapartum and Neonatal Outcomes in a Term, Low-Risk Population, Fetal Diagnosis and Therapy, 43, 191-198, 2018	Incorrect index tests (only CPR).
Blue, N. R., Beddow, M. E., Savabi, M., Katukuri, V. R., Chao, C. R., Comparing the Hadlock fetal growth standard to the Eunice Kennedy Shriver National Institute of Child Health and Human Development racial/ethnic standard for the prediction of neonatal morbidity and small for gestational age, American Journal of Obstetrics & GynecologyAm J Obstet Gynecol, 219, 474.e1-474.e12, 2018	Inclusion criteria matching Blue 2018 included in review, cohort likely significantly overlapping, included only the larger study to minimise risk for double counting and data loss
Broere-Brown, Z. A., Schalekamp-Timmermans, S., Jaddoe, V. W. V., Steegers, E. A. P., Deceleration of fetal growth rate as alternative predictor for childhood outcomes: a birth cohort study, BMC Pregnancy & ChildbirthBMC Pregnancy Childbirth, 19, 216, 2019	Testing began before third trimester.
Caradeux, J., Eixarch, E., Mazarico, E., Basuki, T. R., Gratacos, E., Figueras, F., Second- to third-trimester longitudinal growth assessment for prediction of small-for-gestational age and late fetal growth restriction, Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 51, 219-224, 2018	Insufficient data provided for calculation of accuracy outcomes.
Caradeux, J., Eixarch, E., Mazarico, E., Basuki, T. R., Gratacos, E., Figueras, F., Second- to	Testing began before third trimester.

Study	Reason for exclusion
Third-Trimester Longitudinal Growth Assessment for the Prediction of Largeness for Gestational Age and Macrosomia in an Unselected Population, Fetal Diagnosis & TherapyFetal Diagn Ther, 43, 284-290, 2018	
Caradeux, J., Eixarch, E., Mazarico, E., Basuki, T. R., Gratacos, E., Figueras, F., Longitudinal growth assessment for prediction of adverse perinatal outcome in fetuses suspected to be small-for-gestational age, Ultrasound in Obstetrics & GynecologyUltrasound Obstet Gynecol, 52, 325-331, 2018	Majority of population suspected FGR on inclusion.
Carberry, A. E., Gordon, A., Bond, D. M., Hyett, J., Raynesâ€• Greenow, C. H., Jeffery, H. E., Customised versus populationâ€• based growth charts as a screening tool for detecting small for gestational age infants in lowâ€• risk pregnant women, Cochrane Database of Systematic Reviews, 2014	Systematic review, checked for references.
Cavalcante, R. O., Caetano, A. C., Nacaratto, D. C., Helfer, T. M., Martins, W. P., Nardozza, L. M., Moron, A. F., Araujo Junior, E., Fetal thigh and upper-arm volumes by three-dimensional ultrasound to predict low postnatal body mass index, Journal of Maternal-Fetal & Neonatal Medicine, 28, 1047-52, 2015	Inappropriate reference standard (BMI).
Cavallaro, A., Ash, S. T., Napolitano, R., Wanyonyi, S., Ohuma, E. O., Molloholli, M., Sande, J., Sarris, I., Ioannou, C., Norris, T., Donadono, V., Carvalho, M., Purwar, M., Barros, F. C., Jaffer, Y. A., Bertino, E., Pang, R., Gravett, M. G., Salomon, L. J., Noble, J. A., Altman, D. G., Papageorghiou, A. T., Quality control of ultrasound for fetal biometry: results from the INTERGROWTH-21st Project, Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 52, 332-339, 2018	Reporting on quality control, no accuracy outcomes.
Chauhan, S. P., Scardo, J. A., Hendrix, N. W., Magann, E. F., Morrison, J. C., Accuracy of sonographically estimated fetal weight with and without oligohydramnios. A case-control study, J Reprod MedThe Journal of reproductive medicine, 44, 969-73, 1999	Index test not of interest for review: reduced amniotic fluid by ultrasound
Choi, S. K. Y., Gordon, A., Hilder, L., Henry, A., Hyett, J. A., Brew, B. K., Joseph, F., Jorm, L., Chambers, G. M., Performance of six birthweight and estimated fetal weight standards for predicting adverse perinatal outcomes: a 10-year nationwide population-based study, Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology., 16, 2020	No protocol relevant outcomes (association with adverse outcomes not diagnostic accuracy)

Study	Reason for exclusion
Ciobanu, A., Anthoulakis, C., Syngelaki, A., Akolekar, R., Nicolaides, K. H., Prediction of small-for-gestational-age neonates at 35-37 weeks' gestation: contribution of maternal factors and growth velocity between 32 and 36 weeks, Ultrasound in obstetrics & gynecology, 53, 630-637, 2019	Insufficient data provided for calculation of accuracy outcomes.
Ciobanu, A., Formuso, C., Syngelaki, A., Akolekar, R., Nicolaides, K. H., 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 in obstetrics & gynecology, 53, 488-495, 2019	Insufficient data provided for calculation of accuracy outcomes.
Ciobanu, A., Khan, N., Syngelaki, A., Akolekar, R., Nicolaides, K. H., Routine ultrasound at 32 vs 36 weeks' gestation: prediction of small-for-gestational-age neonates, Ultrasound in obstetrics & gynecology, 53, 761-768, 2019	Insufficient data provided for calculation of accuracy outcomes.
Ciobanu, A., Rouvali, A., Syngelaki, A., Akolekar, R., Nicolaides, K. H., Prediction of small for gestational age neonates: screening by maternal factors, fetal biometry, and biomarkers at 35-37 weeks' gestation, American journal of obstetrics and gynecology, 220, 486.e1-486.e11, 2019	Insufficient data provided for calculation of accuracy outcomes.
Dall'Asta, A., Rizzo, G., Kiener, A., Volpe, N., Di Pasquo, E., Roletti, E., Mappa, I., Makatsariya, A., Maruotti, G. M., Saccone, G., Sarno, L., Papaccio, M., Fichera, A., Prefumo, F., Ottaviani, C., Stampalija, T., Frusca, T., Ghi, T., Identification of large-for-gestational age fetuses using antenatal customized fetal growth charts: Can we improve the prediction of abnormal labor course?, European Journal of Obstetrics, Gynecology, & Reproductive BiologyEur J Obstet Gynecol Reprod Biol, 248, 81-88, 2020	Only included population of women with suspected high risk of macrosomia
De Reu,P.A., Smits,L.J., Oosterbaan,H.P., Nijhuis,J.G., Value of a single early third trimester fetal biometry for the prediction of birth weight deviations in a low risk population, Journal of Perinatal Medicine, 36, 324-329, 2008	Inappropriate index test (single metric not EFW).
Di Lorenzo, G., Monasta, L., Ceccarello, M., Cecotti, V., D'Ottavio, G., Third trimester abdominal circumference, estimated fetal weight and uterine artery doppler for the identification of newborns small and large for gestational age, European Journal of Obstetrics Gynecology and Reproductive Biology, 166, 133-138, 2013	Insufficient data provided for calculation of accuracy outcomes.
Ego, A., Prunet, C., Lebreton, E., Blondel, B., Kaminski, M., Goffinet, F., Zeitlin, J., Customized and non-customized French intrauterine growth curves. i - Methodology, Journal de Gynecologie Obstetrique et Biologie de la Reproduction, 45, 155â€• 164, 2016	Not in English.

Study	Reason for exclusion
Fadigas, C., Saiid, Y., Gonzalez, R., Poon, L. C., Nicolaides, K. H., Prediction of small-for-gestational-age neonates: screening by fetal biometry at 35-37 weeks, Ultrasound in Obstetrics & Gynecology, 45, 559-65, 2015	Insufficient data provided for calculation of accuracy outcomes.
Figueras,F., Figueras,J., Meler,E., Eixarch,E., Coll,O., Gratacos,E., Gardosi,J., Carbonell,X., Customised birthweight standards accurately predict perinatal morbidity, Archives of Disease in Childhood Fetal and Neonatal Edition, 92, F277-F280, 2007	Did not assess accuracy of ultrasound or SFH measurement for predicting birth weight.
Flatley, C., Kumar, S., Is the fetal cerebroplacental ratio better that the estimated fetal weight in predicting adverse perinatal outcomes in a low risk cohort?, Journal of maternal-fetal & neonatal medicine, 32, 2380-2386, 2019	Did not assess accuracy of ultrasound or SFH measurement for predicting birth weight.
Flatley, C., Kumar, S., Is the fetal cerebroplacental ratio better that the estimated fetal weight in predicting adverse perinatal outcomes in a low risk cohort?, BJOG: An International Journal of Obstetrics and Gynaecology, 125, 6, 2018	Duplicate.
Francis, A., Gardosi, J., Effectiveness of ultrasound biometry at 34-36 weeks in the detection of SGA at birth, BJOG: An International Journal of Obstetrics and Gynaecology, 123 (Supplement 2), 22, 2016	Duplicate.
Francis, A., Hugh, O., Gardosi, J., Customized vs INTERGROWTH-21st standards for the assessment of birthweight and stillbirth risk at term, American Journal of Obstetrics & GynecologyAm J Obstet Gynecol, 218, S692-S699, 2018	Did not assess accuracy of ultrasound or SFH measurement for predicting birth weight.
Frick, A. P., Syngelaki, A., Zheng, M., Poon, L. C., Nicolaides, K. H., Prediction of large-for-gestational-age neonates: screening by maternal factors and biomarkers in the three trimesters of pregnancy, Ultrasound in obstetrics & gynecology, 47, 332-9, 2016	Insufficient data provided for calculation of accuracy outcomes.
Gjessing, H. K., Grottum, P., Okland, I., Eik-Nes, S. H., Fetal size monitoring and birth-weight prediction: a new population-based approach, Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 49, 500-507, 2017	Insufficient data provided for calculation of accuracy outcomes.
Gonzalez Gonzalez, N. L., Gonzalez Davila, E., Cabrera, F., Padron, E., Castro, J. R., Garcia Hernandez, J. A., Customized weight curves for Spanish fetuses and newborns, Journal of maternal-fetal & neonatal medicine, 27, 1495-9, 2014	Assessed accuracy of birthweight charts.
Gonzalez Gonzalez, N. L., Plasencia, W., Gonzalez Davila, E., Padron, E., Garcia Hernandez, J. A., Di Renzo, G. C., Bartha, J. L.,	Assessed accuracy of birthweight charts.

Study	Reason for exclusion
The effect of customized growth charts on the identification of large for gestational age newborns, Journal of maternal-fetal & neonatal medicine, 26, 62-5, 2013	
Goto, E., Symphysis-fundal height to identify large-for-gestational-age and macrosomia: a meta-analysis, Journal of Obstetrics & GynaecologyJ Obstet Gynaecol, 1-7, 2019	Systematic review, references checked
Goto, E., Ultrasound fetal anthropometry to identify large-for-gestational-age: a meta-analysis, Minerva Ginecologica, 71, 467-474, 2019	Systematic review, references checked
Grover, V., Usha, R., Kalra, S., Sachdeva, S., Altered fetal growth: antenatal diagnosis by symphysis-fundal height in India and comparison with western charts, Int J Gynaecol ObstetInternational journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 35, 231-4, 1991	Country not of interest for review: India (not high income country)
Hansen, D. N., Odgaard, H. S., Uldbjerg, N., Sinding, M., Sorensen, A., Screening for small-for-gestational-age fetuses, Acta Obstetricia et Gynecologica Scandinavica, 99, 503-509, 2020	Reports accuracy of screening program as a whole but not US specifically
Haragan, A. F., Hulsey, T. C., Hawk, A. F., Newman, R. B., Chang, E. Y., Diagnostic accuracy of fundal height and handheld ultrasound-measured abdominal circumference to screen for fetal growth abnormalities, American Journal of Obstetrics and Gynecology, 212, 820.e1-820.e8, 2015	Testing began before third trimester.
Hargreaves,K., Cameron,M., Edwards,H., Gray,R., Deane,K., Is the use of symphysis-fundal height measurement and ultrasound examination effective in detecting small or large fetuses?, Journal of Obstetrics and Gynaecology, 31, 380-383, 2011	Insufficient data provided for calculation of accuracy outcomes.
Hedriana, H. L., Moore, T. R., A comparison of single versus multiple growth ultrasonographic examinations in predicting birth weight, American Journal of Obstetrics & Gynecology, 170, 1600-4; discussion 1604-6, 1994	Outcomes not reported as pert protocol (index test measurements in standard deviation)
Hoftiezer, L., Hof, M. H. P., Dijs-Elsinga, J., Hogeveen, M., Hukkelhoven, Cwpm, van Lingen, R. A., From population reference to national standard: new and improved birthweight charts, American Journal of Obstetrics & GynecologyAm J Obstet Gynecol, 220, 383.e1-383.e17, 2019	Assessed accuracy of birthweight charts.
Indraccolo,U., Chiocci,L., Rosenberg,P., Nappi,L., Greco,P., Usefulness of symphysis-fundal height in predicting fetal weight in healthy term pregnant women, Clinical and Experimental Obstetrics and Gynecology, 35, 205-207, 2008	Incorrect index tests (US results use 50th percentile).

Study	Reason for exclusion
Kase,B.A., Carreno,C.A., Blackwell,S.C., Customized estimated fetal weight: a novel antenatal tool to diagnose abnormal fetal growth, American Journal of Obstetrics and Gynecology, 207, 218-5, 2012	Testing began before third trimester.
Kayem,G., Grange,G., Breart,G., Goffinet,F., Comparison of fundal height measurement and sonographically measured fetal abdominal circumference in the prediction of high and low birth weight at term, Ultrasound in Obstetrics and Gynecology, 34, 566-571, 2009	Thresholds not chosen prospectively to identify SGA/LGA but picked to optimise sensitivity from ROC curve.
Khalifa, E. A., Hassanein, S. A., Eid, H. H., Ultrasound measurement of fetal abdominal subcutaneous tissue thickness as a predictor of large versus small fetuses for gestational age, Egyptian Journal of Radiology and Nuclear Medicine, 50 (1) (no pagination), 2019	Not in high income country
Kim, M. A., Han, G. H., Kim, Y. H., Prediction of small-for-gestational age by fetal growth rate according to gestational age, 14, e0215737, 2019	Thresholds not chosen prospectively to identify SGA/LGA but picked to optimise sensitivity from ROC curve.
Lalys,L., Pineau,J.C., Guihard-Costa,A.M., Small and large foetuses: Identification and estimation of foetal weight at delivery from third-trimester ultrasound data, Early Human Development, 86, 753-757, 2010	Insufficient data to construct 2 x 2 table and calculate diagnostic outcome accuracy measures
Lindell,G., Marsal,K., Kallen,K., Predicting risk for large-for-gestational age neonates at term: a population-based Bayesian theorem study, Ultrasound in Obstetrics and Gynecology, 41, 398-405, 2013	Threshold not of interest for review: z score
McCowan, L. M. E., Thompson, J. M. D., Taylor, R. S., Baker, P. N., North, R. A., Poston, L., Roberts, C. T., Simpson, N. A. B., Walker, J. J., Myers, J., Kenny, L. C., Healy, D., Briley, A., Murphy, N., Snapes, E., Chan, E., Black, M., Prediction of small for gestational age infants in healthy nulliparous women using clinical and ultrasound risk factors combined with early pregnancy biomarkers, PLoS ONE, 12 (1) (no pagination), 2017	Not third trimester ultrasound
Miranda, J., Rodriguez-Lopez, M., Triunfo, S., Sairanen, M., Kouru, H., Parra-Saavedra, M., Crovetto, F., Figueras, F., Crispi, F., Gratacos, E., Prediction of fetal growth restriction using estimated fetal weight vs a combined screening model in the third trimester, Ultrasound in obstetrics & gynecology, 50, 603-611, 2017	Insufficient data provided for calculation of accuracy outcomes.
Najafzadeh, A., Graves, A., Re: 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. Sovio U, White IR, Dacey A, Pasupathy D, Smith GC, Sonography, 3, 70-71, 2016	Commentary.

Study	Reason for exclusion
Newnham,J.P., Patterson,L.L., James,I.R., Diepeveen,D.A., Reid,S.E., An evaluation of the efficacy of Doppler flow velocity waveform analysis as a screening test in pregnancy, American Journal of Obstetrics and Gynecology, 162, 403-410, 1990	Cut off for index test US AC <5th percentile with the reference standard set at birth weight <10th percentile
Okonofua, F. E., Ayangade, S. O., Chan, R. C., O'Brien, P. M., A prospective comparison of clinical and ultrasonic methods of predicting normal and abnormal fetal growth, Int J Gynaecol ObstetInternational journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 24, 447-51, 1986	Outcomes not reported as pert protocol (insufficient information for calculation of accuracy outcomes)
Ott, W. J., Doyle, S., Ultrasonic diagnosis of altered fetal growth by use of a normal ultrasonic fetal weight curve, Obstetrics and Gynecology, 63, 201-204, 1984	Outcomes not reported as pert protocol (index test measurements in standard deviation)
Papastefanou, I., Pilalis, A., Chrelias, C., Kassanos, D., Souka, A. P., Screening for birth weight deviations by second and third trimester ultrasound scan, Prenatal diagnosis, 34, 759-64, 2014	Did not report results of third trimester scans separately.
Pay, A. S. D., Froen, J. F., Staff, A. C., Jacobsson, B., Gjessing, H. K., Symphysis-fundus measurement - the predictive value of a new reference curve, Tidsskrift for Den Norske LaegeforeningTidsskr Nor Laegeforen, 137, 717-720, 2017	Not in English.
Pay, A., Froen, J. F., Staff, A. C., Jacobsson, B., Gjessing, H. K., Prediction of small-for-gestational-age status by symphysis-fundus height: a registry-based population cohort study, BJOG: An International Journal of Obstetrics & GynaecologyBjog, 123, 1167-73, 2016	Accuracy data not reported for specific SFH measurements/strategies
Persson, B., Stangenberg, M., Lunell, N. O., Brodin, U., Holmberg, N. G., Vaclavinkova, V., Prediction of size of infants at birth by measurement of symphysis fundus height, Br J Obstet GynaecolBritish journal of obstetrics and gynaecology, 93, 206-11, 1986	Outcomes not reported as pert protocol (index test measurements in standard deviation)
Pilalis, A., Souka, A. P., Papastefanou, I., Michalitsi, V., Panagopoulos, P., Chrelias, C., Kassanos, D., Third trimester ultrasound for the prediction of the large for gestational age fetus in low-risk population and evaluation of contingency strategies, Prenatal Diagnosis, 32, 846-853, 2012	Insufficient data provided for calculation of accuracy outcomes.
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 in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 49, 493-499, 2017	Majority of population suspected SGA on inclusion.

Study	Reason for exclusion
Pritchard, N., Lindquist, A., Siqueira, I. D. A., Walker, S. P., Permezel, M., INTERGROWTH-21st compared with GROW customized centiles in the detection of adverse perinatal outcomes at term, Journal of Maternal-Fetal and Neonatal Medicine, 33, 961-966, 2020	Did not report accuracy outcomes
Reboul, Q., Delabaere, A., Luo, Z. C., Nuyt, A. M., Wu, Y., Chauleur, C., Fraser, W., Audibert, F., Prediction of small-for-gestational-age neonate by third-trimester fetal biometry and impact of ultrasound-delivery interval, Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 49, 372-378, 2017	Thresholds not chosen for prediction but selected without detail on index test intention.
Rial-Crestelo, M., Martinez-Portilla, R. J., Cancemi, A., Caradeux, J., Fernandez, L., Peguero, A., Gratacos, E., Figueras, F., Added value of cerebro-placental ratio and uterine artery Doppler at routine third trimester screening as a predictor of SGA and FGR in non-selected pregnancies, Journal of maternal-fetal & neonatal medicine, 32, 2554-2560, 2019	No outcomes on accuracy of US alone.
Ricchi, A., Pignatti, L., Bufalo, E., De Salvatore, C., Banchelli, F., Neri, I., Estimation of fetal weight near term: comparison between ultrasound and symphysis-fundus evaluation by Johnson's rule, Journal of maternal-fetal & neonatal medicine, 1-5, 2019	Insufficient data provided for calculation of accuracy outcomes.
Roeckner, J. T., Odibo, L., Odibo, A. O., The value of fetal growth biometry velocities to predict large for gestational age (LGA) infants, Journal of Maternal Fetal and Neonatal Medicine., 2020	Population only women referred for US for clinical suspicion of growth abnormality
Rogers, M. S., Needham, P. G., Evaluation of fundal height measurement in antenatal care, Aust N Z J Obstet GynaecolThe Australian & New Zealand journal of obstetrics & gynaecology, 25, 87-90, 1985	Outcomes not reported as pert protocol (index test measurements in standard deviation)
Rosenberg, K., Grant, J. M., Tweedie, I., Aitchison, T., Gallagher, F., Measurement of fundal height as a screening test for fetal growth retardation, Br J Obstet GynaecolBritish journal of obstetrics and gynaecology, 89, 447-50, 1982	Included 2nd trimester measurements
Sananes, N., Guigue, V., Kohler, M., Bouffet, N., Cancellier, M., Hornecker, F., Hunsinger, M. C., Kohler, A., Mager, C., Neumann, M., Schmerber, E., Tanghe, M., Nisand, I., Favre, R., Use of Z-scores to select a fetal biometric reference curve, Ultrasound in Obstetrics and Gynecology, 34, 404-409, 2009	Assessed accuracy of reference curves and included second trimester measurements.
Secher, N. J., Lundbye-Christensen, S., Qvist, I., Bagger, P., An evaluation of clinical estimation of fetal weight and symphysis fundal distance for detection of SGA infants, European Journal of	Index test not of interest for review: clinical examination/abdominal palpitation only

Study	Reason for exclusion
Obstetrics, Gynecology, & Reproductive Biology, 38, 91-6, 1991	
Sijmons, E. A., Reuwer, P. J., van Beek, E., Bruinse, H. W., The validity of screening for small-for-gestational-age and low-weight-for-length infants by Doppler ultrasound, Br J Obstet GynaecolBritish journal of obstetrics and gynaecology, 96, 557-61, 1989	Index test not of interest for review: doppler ultrasound (umbilical artery)
Souka, A. P., Papastefanou, I., Michalitsi, V., Pilalis, A., Kassanos, D., Specific formulas improve the estimation of fetal weight by ultrasound scan, Journal of Maternal-Fetal and Neonatal Medicine, 27, 737-742, 2014	Assessed overall accuracy of multiple formulae but without specific cut-off outcomes.
Souka, A. P., 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 in Obstetrics & GynecologyUltrasound Obstet Gynecol, 39, 535-42, 2012	Thresholds not chosen for prediction but selected without detail on index test intention.
Souka, A. P., Papastefanou, I., Pilalis, A., Michalitsi, V., Panagopoulos, P., Kassanos, D., Performance of the ultrasound examination in the early and late third trimester for the prediction of birth weight deviations, Prenatal diagnosis, 33, 915-20, 2013	Thresholds not chosen for prediction but selected without detail on index test intention.
Sovio, U., Smith, G. C. S., Comparison of estimated fetal weight percentiles near term for predicting extremes of birth weight percentile, American journal of obstetrics and gynecology., 21, 2020	Outcomes on this cohort already included from Sovio 2015
Sparks,T.N., Cheng,Y.W., McLaughlin,B., Esakoff,T.F., Caughey,A.B., Fundal height: a useful screening tool for fetal growth?, Journal of Maternal-Fetal and Neonatal Medicine, 24, 708-712, 2011	Majority of population suspected SGA/LGA on inclusion.
Todros, T., Ferrazzi, E., Arduini, D., Bastonero, S., Bezzeccheri, V., Biolcati, M., Bonazzi, B., Gabrielli, S., Pilu, G. L., Rizzo, G., et al.,, Performance of Doppler ultrasonography as a screening test in low risk pregnancies: results of a multicentric study, J Ultrasound MedJournal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine, 14, 343-8, 1995	Index test not of interest for review: doppler ultrasound (umbilical artery)
Warsof, S. L., Cooper, D. J., Little, D., Campbell, S., Routine ultrasound screening for antenatal detection of intrauterine growth retardation, Obstetrics and Gynecology, 67, 33-39, 1986	Index tests done before the third trimester

1

Economic studies

A single economic search was undertaken for all topics included in the scope of this guideline. No economic studies were identified which were applicable to this review question. See supplementary material 2 for details.

Appendix L – Research recommendations

Research recommendations for review question: What is the best method using third trimester measurements to predict birth weight?

No research recommendations were made for this review question.

Appendix M – Additional studies in update searches

Table 6 : Summary of studies identified but not extracted

Study	Why the study was not fully extracted and included
Bardin 2020	Reported accuracy of US <7 days from delivery consistent with meta-analysis and would not affect recommendations (SGA: sensitivity of 0.65, specificity of 0.97; LGA/macrosomia sensitivity of 0.68, specificity 0.94)
Duncan 2020	Reported accuracy of US <7 days from delivery generally consistent with meta-analysis and would not affect recommendations (LGA sensitivity of 0.30, specificity 0.98)