National Institute for Health and Care Excellence

Final

Twin and Triplet Pregnancy

[B] Evidence review for the optimal screening programme to detect fetal growth restriction (intrauterine growth restriction) in twin and triplet pregnancy

NICE guideline NG137

Evidence review

September 2019

Final

This evidence review was developed by the National Guideline Alliance which is a part of the Royal College of Obstetricians and Gynaecologists

Disclaimer

Copyright

ISBN: 978-1-4731-3513-0

Contents

The optimal screening programme to detect intrauterine growth restriction in twin and triplet pregnancy

Review question

What is the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy?

Introduction

Twin pregnancies complicated by intrauterine growth restriction (IUGR) are at increased risk of perinatal mortality and morbidity. Inconsistencies in the diagnostic criteria; for example, birthweight below a certain threshold, the inclusion of size discordance, or which threshold to use, make the incidence of this condition and its prediction difficult to determine. Following a recent international consensus, one parameter was agreed to diagnose selective fetal growth restriction irrespective of chorionicity; i.e. estimated fetal weight of one of the twins less than the third centile. Alternatively, at least 2 out of 3 contributory parameters were also agreed: (1) estimated fetal weight less than the 10th centile of one of the twins; (2) estimated fetal weight discordance of 25% or more; and (3) umbilical artery pulsatility index of the smaller twin above the 95th centile. The current review aims to identify whether these consensus measurements or any other individual measure or combination of measurements could accurately identify IUGR during pregnancy.

Summary of the protocol

Table 1 summarises the Population, Index test, Reference standard and Outcome (PIRO) characteristics of this review. Even though it is a diagnostic test accuracy protocol and IUGR could develop in the first trimester, the terminology of screening is used in the first trimester because one ultrasound is used to screen for one or more complications whereas from the second trimester onwards this is referred to as diagnostic monitoring since it relates to regular monitoring to diagnose complications.

Table 1: Summary of protocol (Population, Index test, Reference standard and Outcome [PIRO] table)

Population	**For twin pregnancies:** • monochorionic diamniotic • monochorionic monoamniotic • dichorionic diamniotic **For triplet pregnancies:** • dichorionic triamniotic • monochorionic triamniotic • dichorionic, diamniotic (a monochorionic twins set) and monochorionic monoamniotic • trichorionic, triamniotic Setting: Secondary or tertiary care centres
Index test	**Estimated during ultrasound scan at 11 weeks 0 days (11^{+0}) to 13 weeks 6 days (13^{+6}):** • discrepant crown-rump length

	• discrepant nuchal translucency **Estimated during ultrasound scan at 14 weeks onwards:** • growth discordancy (fetal biometry including head circumference, abdominal circumference, femur length, biparietal diameter and estimated fetal weight based on formula of these parameters including difference in estimated fetal weight of each twin ≥15%) • amniotic fluid discordancy (amniotic fluid index or maximum pool depth, discordancy between twins in amniotic fluid volume) • doppler studies (umbilical artery and vein and middle cerebral artery doppler, ductus venosus doppler) • plotting symphysio-fundal height, estimated fetal weight and fetal biometric measurements on standard population or customised growth The diagnostic value of first and second trimester tests to detect IUGR will be examined. The above tests will be considered in isolation or in combination. Details regarding frequency and duration of testing throughout pregnancy presented in included studies will be recorded.
Reference standard	• Recognised reference standard for small for gestational age or intrauterine growth restriction including birthweight centiles by gestational age as reported in studies and standard deviation score (according to population or customised or twin specific growth charts) • Abdominal circumference, head circumference • Ponderal index and skinfold thickness • Intertwin weight discordance (any reported >15%) Analysis will be performed separately for the comparison of each index test to each reference standard test. A comparison of index tests to pooled reference standards will not be performed.
Outcomes	**Diagnostic value of first and second trimester tests** **Critical** • sensitivity (detection rate) and specificity **Important** • area under curve (AUC)

For the full review protocol see 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 and for a full description of the methods see supplementary document C.

Declaration of interests were recorded according to NICE's 2014 conflicts of interest policy from March 2017 until March 2018. From April 2018 onwards they were recorded according to NICE's 2018 conflicts of interest policy. Those interests declared until April 2018 were reclassified according to NICE's 2018 conflicts of interest policy (see Interests Register).

Clinical evidence

Included studies

One systematic review (Leombroni 2017), 4 prospective cohort studies (Fajardo-Exposito 2011; O'Connor 2013; Rodis 1990; Sayegh 1993), 15 retrospective cohort studies (Banks 2008; Blickstein 1996; Chamberlain 1991; Cordiez 2017; D'Antonio 2013; D'Antonio 2014; Dias 2010; Hill 1994; Jensen 1995; Johansen 2014; Neves 2017; Shah 1994; Sklar 2017; Storlazzi 1987; van de Waarsenburg 2015) and 1 cross-sectional study (Egan 1994) were included in the review. One further study for which the design could not be determined (Shahshahan 2011) was also included.

The systematic review (Leombroni 2017) included 20 studies (4 studies were prospective and 16 were retrospective cohort studies) which assessed the accuracy of ultrasonographic estimated fetal weight discordancy and ultrasonographic fetal abdominal-circumference discordancy, in predicting birth weight discordancy in women with twin pregnancy. Estimates from the relevant meta-analyses reported in Leombroni 2017 systematic review were included in the current evidence report. If studies included in the systematic review reported additional outcomes that were relevant to this review, then these studies were included independently. This resulted in 2 studies being included independently (O'Connor 2013; van de Waarsenburg 2015).

Four prospective cohort studies (Fajardo-Exposito 2011; O'Connor 2013; Rodis 1990; Sayegh 1993) used ultrasound and other measurements to assess twin discordancy or adverse outcomes during pregnancy.

Each of the retrospective cohort studies aimed to examine growth or weight discordancy in twin (Banks 2008; Chamberlain 1991; Cordiez 2017; D'Antonio 2013; D'Antonio 2013; Hill 1994; Jensen 1995; Johansen 2014; Neves 2017; Shah 1994; Storlazzi 1987; van de Waarsenburg 2015) or triplet (Sklar 2017) pregnancies using a variety of ultrasonographic indices.

The cross-sectional study (Egan 1994) sought to determine a nomogram for symphysio-fundal height measurements in twin pregnancies to screen for discordant growth in twins. The last study (Shahshahan 2011) evaluated discordancy in crown-rump length in the first trimester and its correlation with perinatal complications.

There was no evidence found for the following index tests: nuchal translucency and doppler studies.

The clinical studies included in this evidence review are summarised in Table 2.

See also the literature search strategy in appendix B, study selection flow chart in appendix C, study evidence tables in appendix D and GRADE profiles in appendix F.

Excluded studies

Studies excluded from this systematic review, with reasons for their exclusion, are listed in appendix K.

Summary of clinical studies included in the evidence review

Table 2 provides a brief summary of the included studies.

Table 2: Summary of included studies for twin and triplet pregnancy

Study	Population	Index test	Reference standard	Outcomes	Frequency and duration of screening for each study
Banks 2008 Retrospective cohort UK	N=108 dichorionic twin pregnancies	CRL discordancy ≥5% measured at 10 to 14 weeks' gestation	Intertwin BWD ≥20%	Diagnostic accuracy of CRL discordancy ≥5% to detect intertwin BWD (sensitivity and specificity; AUC)	1st trimester biometry was recorded at the time of the booking ultrasound scan at 10-14 weeks
Chamberlain 1991 Retrospective cohort Ireland	N=85 twin pregnancies Last USS performed within 7 or within 14 days of birth	EFWD ≥20% and ≥25% using 1) AC 2) FL and AC EFW calculation using FL and AC was based on Hadlock (1984)	Intertwin BWD ≥20% and ≥25%	Diagnostic accuracy of EFWD ≥20% estimated by AC and FL to detect BWD ≥20% when last USS to birth interval ≤7 days, and when last USS to birth interval ≤14 days Diagnostic accuracy of EFWD ≥25% estimated by AC and FL to detect BWD ≥25% when last USS to birth interval ≤7 days, and when last USS to birth interval ≤14 days (sensitivity and specificity)	In all twin pregnancies identified, sequential ultrasound examinations at 1-4 week intervals were performed. No other information regarding the frequency and duration of screening was reported
Cordiez 2017 Retrospective cohort France	N=236 twin pregnancies Sonographical data used in the analysis were collected during the	SGA was defined by EFW <10th percentile of the curve used. EFW was calculated using curves: 1) Hadlock's formula (1985), based on abdominal circumference, femo	GA defined as birth weight <10th percentile according to the French curves by Leroy and Lefort (Leroy 1971).	Diagnostic accuracy of EFW to detect SGA (defined as birth weight <10th percentile according to the French curves by Leroy and	Information regarding the frequency and duration of screening was not reported

Study	Population	Index test	Reference standard	Outcomes	Frequency and duration of screening for each study
	latest ultrasound performed less than 30 days before birth.	ral length, head circumference and biparietal diameter; 2) The customised curve (including maternal weight and height, parity and fetal sex) (Ego 2006); 3) The EPOPé unadjusted (Ego 2016); 4) Adjusted on the fetal sex (Ego 2016).	Sonographical data used in the analysis were collected during the latest ultrasound performed less than 30 days before birth.	Lefort (Leroy 1971)). Diagnostic accuracy of EFW to detect SGA (defined as birth weight <3rd percentile according to the French curves by Leroy and Lefort (Leroy 1971)) (sensitivity and specificity)	
D'Antonio 2013 Retrospective cohort UK	N=2155 women with twin pregnancies; n=1735 dichorionic, n=420 monochorionic Only ultrasound examinations just prior to birth were considered for the analysis	CRL discordancy measured at 11 to 14 weeks' gestation	1) Intertwin BWD 2) SGA <5th centile	Diagnostic accuracy of CRL discordancy to detect intertwin BWD (AUC)	A routine fetal structural survey was carried out at 20–22 weeks, and all monochorionic twins had 2 additional scans at around 17 and 19 weeks specifically to identify early features of FFTS
D'Antonio 2014 Retrospective cohort UK	N=2399 women with twin pregnancies; n=1942 dichorionic, n=457 monochorionic	1) AC discordancy 2) EFWD Hadlock's formula (1985) and measured between 20 and 22 weeks' gestation	Intertwin BWD ≥25%	Diagnostic accuracy of discordancy in AC or EFW to detect BWD ≥25% (AUC)	A routine fetal structural survey was carried out at 20–22 weeks, and all monochorionic twins had 2 additional scans

Study	Population	Index test	Reference standard	Outcomes	Frequency and duration of screening for each study
					at around 17 and 19 weeks specifically to identify early features of FFTS
Dias 2010 Retrospective cohort UK	N=660 women with twin pregnancies ; n=506 dichorionic, n=154 monochorionic	CRL discordance measured at 11 to 14 weeks' gestation	Intertwin BWD ≥15% and ≥25%	Diagnostic accuracy of CRL discordancy to detect intertwin BWD ≥15% and ≥25% (AUC)	Information regarding the frequency and duration of screening was not reported.
Egan 1994 Cross-sectional USA	N=160 women with twin pregnancies Using a cut-off of 20% difference for BWD, 143 of these were deemed normal and 17 discordant	SFH measurement	Intertwin BWD ≥20%	Diagnostic accuracy of SFH measurement to detect intertwin BWD ≥20% (sensitivity and specificity)	Information regarding the frequency and duration of screening was not reported. No US-to-birth interval provided
Fajardo-Exposito 2011 Retrospective cohort Spain	N=46 twin pregnancies ; n=35 dichorionic, n=11 monochorionic	CRL discordancy >15% measured at 11 to 14 weeks' gestation	1) Intertwin BWD >15% 2) SGA defined as BW <10th percentile (Santamaria 1998), at least 1 growth retarded neonate	Diagnostic accuracy of CRL discordancy to detect intertwin BWD and SGA (sensitivity and specificity)	Information regarding the frequency and duration of screening was not reported
Hill 1994 Retrospective cohort USA	N= 49 twin pregnancies scanned within 21 days of birth	Intertwin EFWD ≥20% EFW calculated from HC and AC according to Hadlock (1984)	Intertwin BWD ≥20%	Diagnostic accuracy of discrepancy in EFW ≥20% to detect BWD ≥20% (sensitivity and specificity)	Information regarding the frequency and duration of screening was not reported

Study	Population	Index test	Reference standard	Outcomes	Frequency and duration of screening for each study
Jensen 1995 Retrospective cohort Norway	N=73 twin pregnancies Last USS performed within 7 days of birth	1) EFW of an individual fetus ≤10th percentile 2) Intertwin EFWD d ≥20% EFW was calculated using Hadlock's formula (1984) based on BPD and AC.	1) IUGR at birth defined as birth weight <10th percentile 2) Intertwin BWD ≥20%	Diagnostic accuracy of EFW ≤10th centile to detect IUGR (fetal weight ≤10th centile) Diagnostic accuracy of EFWD ≥20% to detect BWD ≥20% (sensitivity and specificity)	Information regarding the frequency and duration of screening was not reported
Johansen 2014 Retrospective cohort Denmark	N=1993 (n=1,733 dichorionic and n=260 monochorionic) twin pregnancies	CRL discrepancy ≥10% measured at 11 to 14 weeks' gestation	Intertwin BWD ≥20%)	Diagnostic accuracy of CRL discordancy ≥10% to detect BWD ≥20% overall for dichorionic and monochorionic twins, for dichorionic twins only and for monochorionic twins only (sensitivity and specificity)	Information regarding the frequency and duration of screening was not reported
Leombroni 2017 Systematic review Italy and Norway Includes 20 studies *Prospective cohort study:* Klam 2005 O'Connor 2013	N=20 studies (4 prospective, 16 retrospective) N=5826 twin pregnancies	Studies used the index test that was represented by different thresholds of sonographic EFW discordancy (≥15%, ≥20%, ≥25%), calculated as ((larger EFW−smaller EFW)/larger EFW) ×100, or sonographic AC discordancy, calculated as ((larger AC−smaller AC)/larger AC) ×100.	Studies used the reference standard that was represented by the actual BWD, calculated as ((larger BW−smaller BW)/larger BW) ×100, as measured immediately after birth.	Diagnostic accuracy of EFWD ≥15% to detect BWD ≥15% Diagnostic accuracy of EFWD ≥20% to detect BWD ≥20% Diagnostic accuracy of EFWD ≥25% to detect BWD ≥25%	The interval between ultrasound and birth interval ranged between 2 and 59 days across studies, with one study not reporting this information

Study	Population	Index test	Reference standard	Outcomes	Frequency and duration of screening for each study
Simoes 2011 Van Mieghem 2009 *Retrospective cohort study:* Al Hassan 2012 Al-Obaidly 2015 Blickstein 1996 Caravello 1997 Chang 2006 Chittacharoen 2000 Danon 2008 Diaz-Garcia 2010 Fox 2011 Ghandi 2009 Gernt 2001 Hoopmann 2011 Khalil 2014 Ong 1999 Roberts 2001 Van de Waarsenburg 2015				Accuracy with EFW calculated using all fetal biometric parameters (head, abdomen, femur, 10 studies) Diagnostic accuracy of AC discordancy to detect BWD (≥15%, 3 studies) Diagnostic accuracy of AC discordancy to detect BWD (≥20%, 2 studies) Diagnostic accuracy of AC discordancy to detect BWD (≥25%, 6 studies) (sensitivity and specificity)	
Neves 2017 Retrospective cohort Portugal	N=176 twin pregnancies Data for the analyses used were measured at the last ultrasound; the median interval between the last ultrasound evaluation and birth	1) EFW discordancy ≥20% based on Hadlock's formula (1985) 2) Amniotic fluid amount (defined as olygoamnios = the deepest vertical pocket of amniotic fluid inferior to 2 cm)	Intertwin BWD ≥20%	Diagnostic accuracy of EFW ≥20% to detect intertwin BWD ≥20% Diagnostic accuracy of EFW (≥20%) to detect intertwin weight discordancy (≥20%) by chorionicity in	Information regarding the frequency and duration of screening was not reported

Study	Population	Index test	Reference standard	Outcomes	Frequency and duration of screening for each study
	was 2 weeks (IQR 0 - 3).			dichorionic twins and monochorionic twins Diagnostic accuracy of amniotic fluid to detect intertwin weight discordancy (≥20%) (sensitivity and specificity; AUC)	
O'Connor 2013 Prospective cohort Ireland	N=260 twin pregnancies	CRL discordancy >20% measured in the 1st trimester (11^{+0} to 14^{+0} weeks)	Intertwin BWD ≥18%	Diagnostic accuracy of CRL discordancy (>20%) to detect BWD (≥18%) (sensitivity and specificity)	Ultrasound examinations were made at enrolment (mean 16 weeks (range 13 - 19)) and again at 18-20 weeks for those enrolled prior to 18 weeks. CRL was recorded for each fetus in the 1st trimester. For monochorionic twins, two-weekly ultrasound surveillance was initiated at 16 weeks' gestation
Rodis 1990 Prospective cohort USA	N=25 women with twin pregnancy Last USS performed	1) EFW difference ≥20% using BPD and AC measurements	Intertwin BWD ≥20%	Diagnostic accuracy of EFWD ≥20%, when EFW calculated using BPD, AC (Shepard's formula), to	Information regarding the frequency and duration of screening was not reported

Study	Population	Index test	Reference standard	Outcomes	Frequency and duration of screening for each study
	within 7 days of birth	2) EFW difference ≥20% using FL and AC measurements EFW was calculated for each fetus using two formulae: one based on BPD and AC (Shepard's formula) and the other based on FL and AC (Hadlock's formula)		detect BWD ≥20% Diagnostic accuracy of EFWD ≥20%, when EFW calculated using FL and AC (Hadlock's formula), to detect BWD ≥20% (sensitivity and specificity)	
Sayegh 1993 Prospective cohort USA	N=78 women with twin pregnancy (including one with FFTS)	Intertwin EFW difference of ≥15%, ≥20% and ≥25%. Calculation of EFW was based on BPD and AC, according to Shepard's formula (1982).	Intertwin BWD ≥25%	Diagnostic accuracy of EFWD ≥25% to detect BWD ≥25% Diagnostic accuracy of EFWD ≥20% to detect BWD ≥25% Diagnostic accuracy of EFWD ≥15% to detect BWD ≥25% (sensitivity and specificity)	Information regarding the frequency and duration of screening was not reported
Shah 1994 Retrospective cohort USA	N=90 twin pregnancies but included in the analysis max=85 and min=54 Last USS performed within 7 days of birth	Intrapair differences in: 1) BPD 2) HC 3) AC 4) FL 5) EFW ≥20% EFW was computed by the method of Warsof et al. (1977) using FL and AC	Intertwin BWD ≥20%	Diagnostic accuracy of intertwin differences >5% (for BPD, HC, AC, FL) to detect BWD ≥20% Diagnostic accuracy of intertwin difference >10% (for BPD, HC, AC, FL) to detect BWD ≥20%	Information regarding the frequency and duration of screening was not reported

Study	Population	Index test	Reference standard	Outcomes	Frequency and duration of screening for each study
				Diagnostic accuracy of EFW difference ≥20% to detect BWD ≥20% (sensitivity and specificity)	
Shahshahan 2011 Unclear study design Iran	N=118 women with twin pregnancy	Discrepancy in CRL >11% measured at 7 to 14 weeks' gestation	Intertwin BWD >20%	Diagnostic accuracy of CRL discordancy >11% to detect SGA (defined as intertwin weight discordancy >20%) (sensitivity and specificity)	Information regarding the frequency and duration of screening was not reported
Sklar 2017 Retrospective cohort Canada	N=78 triplet pregnancies Data for the analyses used were measured closest to date of birth (median interval between last ultrasound and birth was 8 days (IQR 0 - 21); median 30.9 weeks' gestation).	1) FGR defined as EFW <10th percentile for gestational age using for reference the Canadian Perinatal Surveillance System singleton growth curves (Kramer 2001) 2) EFWD >25% which was calculated using Hadlock's formula, based on head circumference, abdominal circumference, femur length (Hadlock, 1985)	1) SGA defined as actual birth weight <10th percentile for gestational age using for reference the Canadian Perinatal Surveillance System singleton growth curves (Kramer, 2001) 2) Inter-triplet BWD > 25%	Diagnostic accuracy of EFW <10th percentile to detect SGA Diagnostic accuracy of EFW discordancy (>25%) to detect BWD >25% (sensitivity and specificity)	Information regarding the frequency and duration of screening was not reported
Storlazzi 1987 Retrospective cohort	N=43 twin pregnancy Last USS performed	Intertwin EFWD ≥20%. EFW calculation was based on BPD	Intertwin BWD ≥20	Diagnostic accuracy of EFWD ≥20% by BWD ≥20%	All participants had an ultrasound examination

Study	Population	Index test	Reference standard	Outcomes	Frequency and duration of screening for each study
USA	within 2 weeks of birth	and AC, using the formula of Shepard (1982) or on AC and FL using the formula of Hadlock (1984), when BPD was unobtainable.		(sensitivity and specificity)	upon admission to confirm the presence of twin gestation. The ultrasound evaluations were repeated every two weeks until birth. Only the results of the last scan were considered for analysis
Van de Waarsenburg 2015 Retrospective cohort The Netherlands	N=281 twin pregnancies	1) CRL discordancy (thresholds ≥11% and ≥20%) measured in the 1st trimester; 2) IUGR (at least 1 twin) defined as an EFW <10th percentile based on the last ultrasound before birth (median interval between the last ultrasound and birth was 8 days (IQR 0 - 59); 3) Amniotic fluid amount (oligohydramnios defined as the deepest vertical pocket of amniotic fluid of less than 2 cm), not reported when it was measured.	Intertwin BWD ≥20%	Diagnostic accuracy of CRL discordancy ≥11% to detect BWD ≥20% Diagnostic accuracy of CRL discordancy ≥20% to detect BWD ≥20% Diagnostic accuracy of IUGR (at least 1 twin, defined as EFW<10th percentile based on the last ultrasound before birth) to detect BWD (≥20%) Diagnostic accuracy of amniotic fluid amount (defined as oligohydramnios = the	Twin pregnancies were monitored according to a protocol based on chorionicity which included a 1st trimester determination of chorionicity, detailed anomaly scan at 20 weeks' gestation age and ultrasound assessment of growth and amniotic fluid volume at 20, 26, 30, 32, 34 and 36 weeks for dichorionic twin gestations and fortnightly from 14

Study	Population	Index test	Reference standard	Outcomes	Frequency and duration of screening for each study
				deepest vertical pocket of amniotic fluid of less than 2 cm) to detect BWD (≥20%) (sensitivity and specificity)	weeks onwards

AC: abdominal circumference; AUC: area under the curve (the curve represents different cut-off points); BPD: biparietal diameter; BW: birth weight; BWD: birth weight discordancy; CRL: crown-rump length; EFW: estimated fetal weight; EFWD: estimated fetal weight discordancy; FFTS: feto-fetal transfusion syndrome; FGR: fetal growth restriction; FL: femoral length; GA: gestational age; HC: head circumference; HC: AC ratio: head circumference: abdominal circumference ratio; IQR: interquartile range; IUGR: intrauterine growth restriction; N: number of participants included in the study; S:D ratio: peak systolic: end diastolic ratio; SFH: symphysio-fundal height; SGA: small for gestational age; US: ultrasound; USS: ultrasound screening

See appendix D for the full evidence tables.

Quality assessment of clinical studies included in the evidence review

Risk of bias was performed for all studies except those included in the Leombroni 2017 systematic review where the risk of bias assessment was taken from the review which provided a quality assessment for each individual study using the Quality of Diagnostic Accuracy Studies version 2 (QUADAS-II) checklist.

See appendix F for the full GRADE tables.

Economic evidence

Included studies

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

See the appendix B for the economic search strategy and appendix G for the economic evidence selection flow chart for further information.

Excluded studies

No full-text copies of articles were requested for this review and so there is no excluded studies list.

Summary of studies included in the economic evidence review

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

Only sensitivity and specificity values are provided in the evidence statements below. When assessing the diagnostic accuracy of sensitivity and specificity the following thresholds were used: high accuracy: more than 90%; moderate accuracy: 75% to 90%; and, low accuracy: less than 75%.

Area under the curve (AUC) measures are not reported in the evidence statements below as they are not related to a particular cut-off and are therefore difficult to interpret. AUC up to 70 are described as having poor ability to discriminate and AUC of 71 and above would be described as having moderate or good ability to discriminate. Estimates are reported for information in appendix D and appendix F (AUC estimates as reported in the included studies). For further details see the methods described in supplement document C.

Screening to identify a small for gestational age baby or intertwin birth weight discordancy in twin pregnancy in first trimester (11^{+0}-13^{+6} weeks' gestation)

Small for gestational age defined as <5th centile

Crown-rump length discordancy (continuous) – index test

Moderate quality evidence from 1 study (N=2155) showed that the overall crown-rump length discordancy for dichorionic and monochorionic twins measured using ultrasound had very poor ability to discriminate for the diagnosis of small for gestational age defined as birth weight <5th centile. Very low quality evidence from the same study (N=420) showed that the crown-rump length discordancy for monochorionic twins only had very poor ability to discriminate for the diagnosis of crown-rump length defined as birth weight <5th centile.

Small for gestational defined as <10th percentile

Crown-rump length discordancy >15% – index test

Moderate quality evidence from 1 study (N=46) showed that the sensitivity and specificity for crown-rump length discordancy >15% for dichorionic and monochorionic twins measured using ultrasound was 10% (0 to 45) and 94% (81 to 99) to detect small for gestational age defined as birth weight <10th percentile. Moderate quality evidence from the same study (N=35) showed that the sensitivity and specificity for crown-rump length discordancy >15% for dichorionic twins only was 13% (0 to 53) and 96% (81 to 100) to detect small for gestational defined as birth weight <10th percentile. Low quality evidence from the same study (N=11) showed that the sensitivity and specificity for crown-rump length discordancy >15% for monochorionic twins only was 0% (0 to 84) and 89% (52 to 100) to detect small for gestational age defined as birth weight <10th percentile.

Intertwin birth weight discordancy ≥15%

Crown rump length discordancy (continuous) – index test

Very low quality evidence from 1 study (N=660) showed that the overall CRL discordancy for dichorionic and monochorionic twins measured using ultrasound had very poor ability to discriminate for the diagnosis of intertwine birth weight discordancy ≥15%.

Intertwin birth weight discordancy >15%

Crown-rump length discordancy >15% – index test

Moderate quality evidence from 1 study (N=46) showed that the overall sensitivity and specificity for crown-rump length discordancy >15% for dichorionic and monochorionic twins measured using ultrasound was 13% (2 to 40) and 97% (83 to 100) to detect intertwin birthweight discordancy >15%. Moderate quality evidence from the same study (N=35) showed that the sensitivity and specificity for crown-rump length discordancy >15% for dichorionic twins only was 8% (0 to 38) and 96% (78 to 100) to detect intertwin birthweight

discordancy >15%. Very low quality evidence from the same study (N=11) showed that the sensitivity and specificity for crown-rump length discordancy >15% for monochorionic twins only was 33% (1 to 91) and 100% (63 to 100) to detect intertwin birthweight discordancy >15%.

Intertwin birth weight discordancy ≥18%

Crown-rump length discordancy >20% – index test

Moderate quality evidence from 1 study (N=260) showed that the sensitivity and specificity for crown-rump length discordancy >20% measured using ultrasound was 2% (0 to 11) and 100% (97 to 100) to detect birth weight discordancy ≥18%.

Intertwin birth weight discordancy ≥20%

Crown-rump length discordancy ≥5% – index test

Very low quality evidence from 1 study (N=108) showed that the sensitivity and specificity for crown-rump length discordancy ≥5% for dichorionic twins measured using ultrasound was 59% (36 to 79) and 60% (48 to 72) to detect intertwin birth weight discordancy ≥20%. Very low quality evidence from the same study (N=180) showed that crown-rump length discordancy for dichorionic twins only had very poor ability to discriminate for the intertwin birth weight discordancy ≥20%.

Crown-rump length discordancy ≥10% – index test

Low quality evidence from 1 study (N=1,993) showed that the overall sensitivity and specificity for crown-rump length discordancy ≥10% for dichorionic and monochorionic twins measured using ultrasound was 24% (19 to 31) and 87% (85 to 88) to detect intertwin birth weight discordancy ≥20%. Low quality evidence from the same study (N=1,733) showed that the sensitivity and specificity for crown-rump length discordancy ≥10% for dichorionic twins only was 24% (17 to 31) and 86% (85 to 88) to detect intertwin birth weight discordancy ≥20%. Low quality evidence from the same study (N=260) showed that the sensitivity and specificity for crown-rump length discordancy ≥10% for monochorionic twins only was 28% (14 to 47) and 89% (84 to 92) to detect intertwin birth weight discordancy ≥20%.

Crown-rump length discordancy ≥11% – index test

Moderate quality evidence from 1 study (N=281) showed that the sensitivity and specificity for crown-rump length discordancy ≥11% measured using ultrasound was 10% (3 to 23) and 95% (92 to 98) to detect intertwin birth weight discordancy ≥20%.

Crown-rump length discordancy ≥20% – index test

Moderate quality evidence from 1 study (N=281) showed that the sensitivity and specificity for crown-rump length discordancy ≥20% measured using ultrasound was 2% (0 to 13) and 99% (97 to 100) to detect intertwin birth weight discordancy ≥20%.

Intertwin birth weight discordancy >20%

Crown-rump length discordancy ≥11% – index test

Very low quality evidence from 1 study (N=118) showed that the sensitivity and specificity for crown-rump length discordancy ≥11% measured using ultrasound was 60% (32 to 84) and 87% (79 to 93) to detect birth weight discordancy >20%.

Intertwin birth weight discordancy ≥25%

Crown-rump length discordancy (continuous) – index test

Very low quality evidence from 1 study (N=660) showed that the overall crown-rump length discordancy for dichorionic and monochorionic twins measured using ultrasound had poor ability to discriminate for the diagnosis of intertwine birth weight discordancy ≥25%.

Diagnostic monitoring to identify intertwin birth weight discordancy ≥15% or more using fetal biometry discordancy in twin pregnancy in second trimester

Intertwin birth weight discordancy ≥15%

Overall performance of abdominal circumference discordancy – index test

Low quality evidence from 3 studies (N=1,090, systematic review) showed that the overall sensitivity and specificity for abdominal circumference discordancy measured using ultrasound was 27% (22 to 32) and 91% (89 to 92) to detect intertwin birth weight discordancy ≥15%. Ultrasound-to-birth interval was within 2 weeks in 2 studies, and not reported in one study. According to the review authors, due to "the multitude of cut-offs reported among studies, it was not possible to perform a comprehensive data synthesis for each [abdominal circumference] cut-off".

Intertwin birth weight discordancy ≥20%

Overall performance of abdominal circumference discordancy – index test

Low quality evidence from 2 studies (N=371, systematic review) showed that the overall sensitivity and specificity for abdominal circumference discordancy measured using ultrasound was 32% (21 to 45) and 91% (88 to 94) to detect intertwin birth weight discordancy ≥20%. Ultrasound-to-birth interval was 8 days (range 0-59) or within three weeks. According to the review authors, due to "the multitude of cut-offs reported among studies, it was not possible to perform a comprehensive data synthesis for each [abdominal circumference] cut-off".

Head circumference discordancy >5% – index test

Very low quality evidence from 1 study (N=54) showed that the sensitivity and specificity for head circumference discordancy >5% measured using ultrasound was 64% (31 to 89) and 74% (59 to 86) to detect intertwin birth weight discordancy ≥20%. Ultrasound-to-birth interval was within seven days.

Head circumference discordancy >10% – index test

Low quality evidence from 1 study (N=54) showed that the sensitivity and specificity for head circumference discordancy >10% measured using ultrasound was 18% (2 to 52) and 93% (81 to 99) to detect intertwin birth weight discordancy ≥20%. Ultrasound-to-birth interval was within seven days.

Abdominal circumference discordancy >5% – index test

Very low quality evidence from 1 study (N=85) showed that the sensitivity and specificity for abdominal circumference discordancy >5% measured using ultrasound was 89% (65 to 99) and 60% (47 to 72) to detect intertwin birth weight discordancy ≥20%. Ultrasound-to-birth interval was within 7 days.

Abdominal circumference discordancy >10% – index test

Very low quality evidence from 1 study (N=85) showed that the sensitivity and specificity for abdominal circumference discordancy >10% measured using ultrasound was 61% (36 to 83) and 90% (80 to 96) to detect intertwin birth weight discordancy ≥20%. Ultrasound-to-birth interval was within seven days.

Femur length discordancy >5% – index test

Low quality evidence from 1 study (N=79) showed that the sensitivity and specificity for femur length discordancy >5% measured using ultrasound was 47% (23 to 72) and 79% (67 to 88)

to detect intertwin birth weight discordancy ≥20%. Ultrasound-to-birth interval was within seven days.

Femur length discordancy >10% – index test

Low quality evidence from 1 study (N=79) showed that the sensitivity and specificity for femur length discordancy >10% measured using ultrasound was 18% (4 to 43) and 94% (84 to 98) to detect intertwin birth weight discordancy ≥20%. Ultrasound-to-birth interval was within seven days.

Biparietal diameter discordancy >5% – index test

Very low quality evidence from 1 study (N=64) showed that the sensitivity and specificity for biparietal diameter discordancy >5% measured using ultrasound was 57% (29 to 82) and 62% (47 to 75) to detect intertwin birth weight discordancy ≥20%. Ultrasound-to-birth interval was within 7 days.

Biparietal diameter discordancy >10% – index test

Low quality evidence from 1 study (N=64) showed that the sensitivity and specificity for biparietal diameter discordancy >10% measured using ultrasound was 36% (13 to 65) and 94% (83 to 99) to detect intertwin birth weight discordancy ≥20%. Ultrasound-to-birth interval was within seven days.

Intertwin birth weight discordancy ≥25%

Overall performance of abdominal circumference discordancy – index test

Very low quality evidence from 5 studies (N=1609, systematic review) showed that the overall sensitivity and specificity for head circumference discordancy measured using ultrasound was 71% (51 to 85) and 86% (62 to 96) to detect intertwin birth weight discordancy ≥25%. Ultrasound-to-birth interval was 3 days, 1.6±0.14 weeks, 2-4 weeks or within 2 or 3 weeks. According to the review authors, due to "the multitude of cut-offs reported among studies, it was not possible to perform a comprehensive data synthesis for each [abdominal circumference] cut-off".

Abdominal circumference discordancy (continuous) – index test

Low quality evidence from 1 study (N=2399) showed that the overall abdominal circumference discordancy for monochorionic and dichorionic twins measured using ultrasound had poor ability to discriminate for the diagnosis of intertwin birth weight discordancy ≥25%. Low quality evidence from the same study (N=457) showed that the abdominal circumference discordancy for monochorionic twins only had also poor ability to discriminate for the diagnosis of intertwin birth weight discordancy ≥25%.

Diagnostic monitoring to identify a small-for-gestational-age baby (defined as recognised reference standard for small for gestational age or intrauterine growth restriction) using estimated fetal birth weight discordancy in twin and triplet pregnancy in second trimester

Twin pregnancy

Growth curves (France)

Estimated fetal weight <3rd percentile based on Hadlock's 1985 curve (includes head circumference, abdominal circumference, femur length and biparietal diameter) – index test

Low quality evidence from 1 study (N=236) showed that the sensitivity and specificity for estimated fetal weight <3rd percentile measured using ultrasound and based on Hadlock's 1985 curve, which includes head circumference, abdominal circumference, femur length and biparietal diameter, was 64% (49 to 78) and 89% (86 to 92) to detect small for gestational age defined as birth weight <3rd percentile for gestational age using for reference the French

curves by Leroy and Lefort (1971). Ultrasound- to-birth interval was less than 30 days before birth.

Estimated fetal weight <3rd percentile according to the customized curve based on Ego 2006 (includes maternal weight and height, parity and fetal sex) – index test

Low quality evidence from 1 study (N=236) showed that the sensitivity and specificity for estimated fetal weight <3rd percentile measured using ultrasound and based on Ego's 2006 customized curve, which includes maternal weight and height, parity, and fetal sex, was 66% (50 to 80) and 86% (82 to 89) to detect small for gestational age defined as birth weight <3rd percentile for gestational age using for reference the French curves by Leroy and Lefort (1971). Ultrasound- to-birth interval was less than 30 days before birth.

Estimated fetal weight <3rd percentile according to the EPOPé unadjusted curve based on Ego 2016 – index test

Moderate quality evidence from 1 study (N=236) showed that the sensitivity and specificity for estimated fetal weight <3rd percentile measured using ultrasound and based on the EPOPé's unadjusted curve was 57% (42 to 71) and 89% (86 to 92) to detect small for gestational age defined as birth weight <3rd percentile for gestational age using for reference the French curves by Leroy and Lefort (1971). Ultrasound- to-birth interval was less than 30 days before birth.

Estimated fetal weight <3rd percentile according to the EPOPé adjusted (fetal sex) curve based on Ego 2016 – index test

Low quality evidence from 1 study (N=236) showed that the sensitivity and specificity for estimated fetal weight <3rd percentile measured using ultrasound and based on the EPOPé's adjusted curve was 64% (49 to 78) and 90% (87 to 93) to detect small for gestational age defined as birth weight <3rd percentile for gestational age using for reference the French curves by Leroy and Lefort (1971). Ultrasound- to-birth interval was less than 30 days before birth.

Estimated fetal weight <10th percentile based on Hadlock's 1985 curve (includes head circumference, abdominal circumference, femur length and biparietal diameter) – index test

Moderate quality evidence from 1 study (N=236) showed that the sensitivity and specificity for estimated fetal weight <10th percentile measured using ultrasound and based on Hadlock's 1985 curve, which includes head circumference, abdominal circumference, femur length and biparietal diameter, was 67% (60 to 74) and 80% (75 to 84) to detect small for gestational age defined as birth weight <10th percentile for gestational age using for reference the French curves by Leroy and Lefort (1971). Ultrasound to birth interval was less than 30 days before birth.

Estimated fetal weight <10th percentile according to the customized curve based on Ego 2006 (includes maternal weight and height, parity and fetal sex) – index test

Moderate quality evidence from 1 study (N=236) showed that the sensitivity and specificity for estimated fetal weight <10th percentile measured using ultrasound and based on Ego's customized curve, which includes maternal weight and height, parity, and fetal sex, was 63% (55 to 70) and 82% (76 to 86) to detect small for gestational age defined as birth weight <10th percentile for gestational age using for reference the French curves by Leroy and Lefort (1971). Ultrasound- to-birth interval was less than 30 days before birth.

Estimated fetal weight <10th percentile according to the EPOPé unadjusted curve based on Ego 2016 – index test

Moderate quality evidence from 1 study (N=236) showed that the sensitivity and specificity for estimated fetal weight <10th percentile measured using ultrasound and based on the

EPOPé's unadjusted curve was 60% (52 to 68) and 84% (79 to 88) to detect small for gestational age defined as birth weight <10th percentile for gestational age using for reference the French curves by Leroy and Lefort (1971). Ultrasound- to-birth interval was less than 30 days before birth.

Estimated fetal weight <10th percentile according to the EPOPé adjusted (fetal sex) curve based on Ego 2016 – index test

Moderate quality evidence from 1 study (N=236) showed that the sensitivity and specificity for estimated fetal weight <10th percentile measured using ultrasound and based on the EPOPé's adjusted curve was 57% (49 to 65) and 83% (79 to 87) to detect small for gestational age defined as birth weight <10th percentile for gestational age using for reference the French curves by Leroy and Lefort (1971). Ultrasound- to-birth interval was less than 30 days before birth.

Triplet pregnancy

Growth curves (Canada)

Estimated fetal weight <10th percentile – index test

Moderate quality evidence from 1 study (N=78) showed that the sensitivity and specificity for estimated fetal weight <10th centile measured using ultrasound was 56% (35 to 75) and 100% (93 to 100) to detect SGA defined as birth weight <10th percentile for gestational age using for reference the Canadian Perinatal Surveillance System singleton growth curves. Median ultrasound-to-birth interval was 8 days (range 0-21).

Estimated fetal weight discordancy >25% based on Hadlock's 1985 formula (includes head circumference, abdominal circumference and femur length) – index test

Very low quality evidence from 1 study (N=78) showed that the sensitivity and specificity for estimated fetal weight discordancy >25% measured using ultrasound and based on Hadlock's 1985 formula, which includes head circumference, abdominal circumference and femur length, was 80% (44 to 97) and 94% (86 to 98) to detect small for gestational age defined as birth weight <10th percentile for gestational age using for reference the Canadian Perinatal Surveillance System singleton growth curves. Median ultrasound-to-birth interval was 8 days (range 0-21).

Diagnostic monitoring to identify intrauterine growth restriction or intertwin birth weight discordancy ≥15% or more using growth discordancy in twin pregnancy in second trimester

Intertwin birth weight discordancy ≥15%

Estimated fetal weight discordancy ≥15% (overall performance) – index test

Low quality evidence from 6 studies (N=1477, systematic review) showed that the overall sensitivity and specificity for estimated fetal weight discordancy ≥15% measured using ultrasound was 68% (62 to 73) and 83% (79 to 87) to detect intertwin birth weight discordancy ≥15%. Ultrasound-to-birth interval was 48 h, 3 days (range 1-7), 15 days; within 28 days or 2 weeks.

Intertwin birth weight discordancy ≥20%

Estimated fetal weight discordancy ≥20% (overall performance) – index test

Low quality evidence from 7 studies (N=1780, systematic review) showed that the overall sensitivity and specificity for estimated fetal weight discordancy ≥20% measured using ultrasound was 65% (58 to 72) and 91% (87 to 94) to detect intertwin birth weight

discordancy ≥20%. Ultrasound-to-birth interval was 48 h; 3 days (range 1-7) or 8 days (range 0-59), 3, 10 or 15 days or within 28 days.

Estimated fetal weight discordancy ≥20% (based on abdominal circumference and femur length) – index test

Very low quality evidence from 3 studies (N=160) showed the overall sensitivity and specificity for estimated fetal weight discordancy ≥20% measured using ultrasound and based on abdominal circumference and femur length was 70% (34 to 93) and 89% 69 to 98) to detect intertwin birth weight discordancy ≥20%. Ultrasound-to-birth interval was within 7 days.

Estimated fetal weight discordancy ≥20% (overall performance) – index test

Very low quality evidence from 7 studies (N=491) showed the overall sensitivity and specificity for estimated fetal weight discordancy ≥20% measured using ultrasound was 71% (54 to 85) and 89% (83 to 94) to detect intertwin birth weight discordancy ≥20%. Ultrasound-to-birth interval was within 7 or 21 days, or within 2 weeks.

Estimated fetal weight discordancy ≥20% (based on abdominal circumference and femur length) – index test

Moderate quality evidence from 1 study (N=74) showed that the sensitivity and specificity for estimated fetal weight discordancy ≥20% measured using ultrasound and based on abdominal circumference and femur length was 46% (19 to 75) and 92% (82 to 97) to detect intertwin birth weight discordancy ≥20%. Ultrasound-to-birth interval was ≤14 days.

Estimated fetal weight discordancy ≥20% (based on Shepard's formula, includes abdominal circumference and biparietal diameter) – index test

Very low quality evidence from 1 study (N=29) showed that the sensitivity and specificity for estimated fetal weight discordancy ≥20% measured using ultrasound and based on abdominal circumference and biparietal diameter was 86% (57 to 98) and 80% (52 to 96) to detect intertwin birth weight discordancy ≥20%. Ultrasound-to-birth interval was within 7 days.

Estimated fetal weight discordancy ≥20% (based on Hadlock's formula) – index test

Very low quality evidence from 1 study (N=176) showed that estimated fetal weight discordancy ≥20% for monochorionic and dichorionic twins measured using ultrasound and based on Hadlock's formula had good ability to discriminate for the diagnosis of intertwin birth weight discordancy ≥20%. Very low quality evidence from the same study (N=123) showed estimated fetal weight discordancy ≥20% for dichorionic twins only had good ability to discriminate for the diagnosis of intertwin birth weight discordancy ≥20%. Very low quality evidence from the same study (N=53) showed estimated fetal weight discordancy ≥20% for monochorionic twins only had good ability to discriminate for the diagnosis of intertwin birth weight discordancy ≥20%. Median ultrasound-to-birth interval was 2 weeks (range 0-3).

Estimated fetal weight <10th percentile (at least 1 twin) based on Hadlock's 1991 formula (includes head circumference, abdominal circumference, femur length) – index test

Low quality evidence from 1 study (N=281) showed that the sensitivity and specificity for estimated fetal weight <10th percentile measured using ultrasound and based on Hadlock's 1991 formula, which includes head circumference, abdominal circumference and femur length, was 69% (53 to 82) and 80% (74 to 85) to detect intrauterine growth restriction defined as intertwin birth weight discordancy ≥20%. Median ultrasound-to-birth interval was 8 days (range 0-59).

Intertwin birth weight discordancy ≥25%

Estimated fetal weight discordancy ≥25% (overall performance) – index test

Very low quality evidence from 14 studies (N=3980, systematic review) showed that the overall sensitivity and specificity for estimated fetal weight discordancy ≥25% discordancy

measured using ultrasound was 58% (46 to 68) and 95% (93 to 97) to detect intertwin birth weight discordancy ≥25%. Ultrasound-to-birth interval was 48 h, 3, 14 or 15 days; within 3, 6 or 28 days; 1.6±0.14 wee; within 2, 2-4 or 3 weeks.

Estimated fetal weight discordancy ≥25% (based on abdominal circumference and femur length) – index test

Low quality evidence from 1 study (N=53) showed that the sensitivity and specificity for estimated fetal weight discordancy ≥25% measured using ultrasound and based on abdominal circumference and femur length was 50% (12 to 88) and 98% (89 to 100) to detect intertwin birth weight discordancy ≥25%. Ultrasound-to-birth interval was ≤7 days.

Estimated fetal weight discordancy ≥25% (based on abdominal circumference and femur length) – index test

Low quality evidence from 1 study (N=74) showed that the sensitivity and specificity for estimated fetal weight discordancy ≥25% measured using ultrasound and based on abdominal circumference and femur length was 38% (9 to 76) and 98% (92 to 100) to detect intertwin birth weight discordancy ≥25%. Ultrasound interval was ≤14 days.

Estimated fetal weight discordancy ≥25% (based on abdominal circumference and biparietal diameter) – index test

Very low quality evidence from 1 study (N=78) showed that the sensitivity and specificity for estimated fetal weight discordancy ≥25% measured using ultrasound and based on Shepard's formula, which includes abdominal circumference and biparietal diameter, was 77% (46 to 95) and 92% (83 to 97) to detect intertwin birth weight discordancy ≥25%. Ultrasound-to-birth interval was 1 to 6 weeks.

Estimated fetal weight discordancy (continuous, based on Hadlock's 1985 formula (includes head circumference, abdominal circumference, femur length) – index test

Low quality evidence from 1 study (N=2399) showed that the overall estimated fetal weight discordancy for monochorionic and dichorionic twins measured using ultrasound had poor ability to discriminate for the diagnosis of intertwin birth weight discordancy ≥25%. Low quality evidence from the same study (N=457) showed that the estimated fetal weight discordancy for monochorionic twins only had also poor ability to discriminate for the diagnosis of intertwin birth weight discordancy ≥25%.

Diagnostic monitoring to identify intertwin birth weight discordancy ≥20% using amniotic fluid discordancy in twin pregnancy in second trimester

Amniotic fluid discordancy – index test

Moderate quality evidence from 1 study (N=176) showed that the sensitivity and specificity for amniotic fluid discordancy (oligohydramnios) measured using ultrasound was 13% (4 to 28) and 97% (93 to 99) to detect intertwin birth weight discordancy ≥20%. Median ultrasound-to-birth interval was 2 weeks (range 0-3).

Another moderate quality evidence from 1 study (N=281) showed that the sensitivity and specificity for amniotic fluid discordancy (oligohydramnios) measured using ultrasound was 17% (7 to 31) and 85% (80 to 90) to detect intertwin birth weight discordancy ≥20%. Median ultrasound-to-birth interval was 8 days (range 0-59).

Diagnostic monitoring to identify intertwin birth weight discordancy ≥20% using symphysio-fundal height measurement in twin pregnancy in second trimester

Symphysio-fundal height measurement – index test

Low quality evidence from 1 study (N=160) showed that the sensitivity and specificity for symphysio-fundal height measured using ultrasound was 24% (7 to 55) and 83% (75 to 88) to detect intertwin birth weight discordancy ≥20%. No US-to-birth interval was reported; ultrasound was done between 16 and 36 weeks' gestation.

The committee's discussion of the evidence

Interpreting the evidence

The outcomes that matter most

The committee prioritised the diagnostic accuracy measure of sensitivity as a critical outcome in view of the increased perinatal mortality and morbidity associated with IUGR. In the second trimester or thereafter, detection of the presence or absence of IUGR is an important aim of each ultrasound assessment. The implications of a false negative test result would be increased risk of perinatal mortality while the implication of a false positive test result would be potentially increased risk of neonatal morbidity secondary to iatrogenic prematurity.

The committee discussed that a false negative test result can have a long term psychological impact on families because of the increased risk of perinatal mortality or neonatal morbidity and mortality associated with severe growth restriction. The consequences of the false negative test result can result in a range of disabilities.

The quality of the evidence

The quality of the accuracy of test results was assessed for the whole evidence base related to each index test using a modified GRADE approach (for a full description of methods see supplementary material C).

For the diagnostic accuracy measures in the first trimester the evidence was rated as very low to moderate quality. This was mainly due to risk of bias in the individual studies which often related to lack of clarity about whether the index test results were interpreted without knowledge of the results of the reference standard. There was also often imprecision in the evidence base with wide confidence intervals which indicated uncertainty about the quality of the accuracy measurement.

For the diagnostic accuracy measures in the second trimester, the evidence was rated as very low to moderate. This was also mainly due to risk of bias in the individual studies which often related to lack of clarity about whether the index test results were interpreted without knowledge of the results of the reference standard. There was also often variation in how and when the tests were performed (e.g. different ultrasound to birth interval) and imprecision in the evidence base with wide confidence intervals which indicated uncertainty about the quality of the accuracy measurement.

Benefits and harms

Even though the evidence was mixed and uncertain, the committee decided that strong recommendations are needed in the context of complications during pregnancy due to the increased risk of perinatal mortality and morbidity if such complications are not identified promptly.

The committee discussed and agreed that both terms 'intrauterine' and 'fetal' growth restriction are widely used and accepted among health professionals, but that 'fetal growth restriction' is now more commonly used (for instance it is the preferred terminology used by the Royal College of Obstetricians and Gynaecologists. They therefore used this in the guideline and the discussion section.

Screening for fetal growth restriction in the first trimester

Evidence in the first trimester related exclusively to the measurement of discrepant crown rump length. There was evidence that screening in the first trimester, using this measure, was not accurate in predicting growth discordance in the second and third trimester. The committee acknowledged, based on their expertise and experience, that other ultrasound measures would also not likely be accurate predictors due to uncertainties around these early measurements. They therefore decided not to recommend screening for fetal growth restriction for women with a twin or triplet pregnancy in the first trimester.

Diagnostic monitoring for fetal growth restriction in dichorionic twin and trichorionic triplet pregnancies

The committee decided to stratify recommendations into 2 sections according to 2 main risk groups. One set of recommendations relates to dichorionic twin and trichorionic triplet pregnancies and another set relates to complications of monochorionicity (because these pregnancies have a higher risk of severe growth discordance).

Based on their experience and expertise the committee agreed that abdominal palpation and symphysis fundal height measurements were inappropriate to measure growth of dichorionic twin and trichorionic triplet pregnancies as there was no way of calculating fetal growth discordance from them accurately. They therefore decided to retain the 2011 recommendation not to use these tests to detect fetal growth restriction.

After reviewing the evidence, the committee acknowledged the heterogeneity of the published studies, in particular the cut-off of the screening parameter measured and the birth weight cut-offs used to define its discordancy. They also discussed the lack of stratification by chorionicity in most of the studies. However, the committee focussed on particular cut-offs and used this evidence combined with their experience and expertise to agree that there were clear benefits for using ultrasound to detect selective fetal growth restriction. No harm could be identified for using ultrasound as the primary method of detecting fetal growth discordance since this is not an invasive measure. The committee agreed, based on experience, that there are benefits to using 2 or more biometric parameters to detect growth discordance and restriction because it would increase clinicians confidence in the interpretation of findings. Based on the evidence as well as experience, the committee agreed that the assessment of amniotic fluid levels helps to identify if a baby is constitutionally small or growth restricted. The amniotic fluid reflects the urine production of the baby and if a baby is growth restricted it will divert blood away from the kidneys towards the brain and the heart. This will reduce the urine production and hence lead to reduced amniotic fluid. Even though it has poor sensitivity the committee recommended this since it would be used in combination with the other growth parameters and is useful to build an overall clinical picture. Therefore the high specificity can help to eliminate false positive results from other measures.

The use of 2 or more biometric parameters would also allow a more accurate overall estimated fetal weight to be calculated. There were no obvious harms to using more than 1 biometric parameter since many different measurements are taken in each scan. Due to the risks of growth restrictions being lower in twins and triplet pregnancies that do not involve monochorionic babies, the committee decided that diagnostic monitoring could commence at the routine scan at 24 weeks rather than changing the schedules of appointment to accommodate a further earlier scan (for example at week 16 of pregnancy) which is not current practice (see the section ‘schedule of specialist appointments’ in the guideline). Adding an extra scan earlier would raise anxiety and add extra costs. The committee agreed that it would not lead to a significantly higher detection to justify this. Due to these reasons they recommended regular screening from week 24 of pregnancy for dichorionic twins and trichorionic triplets.

The evidence was limited on the frequency of ultrasound scanning for women with dichorionic twin pregnancies so the committee used their expertise and experience to make recommendations. The evidence of the scanning schedules that were used in the studies did not show a consistent pattern so the committee used their expertise and experience and agreed that women with a dichorionic twin pregnancy should have scans no more than 28 days apart because this would provide the best balance between the risks of a baby developing the condition, the woman's possible increased anxiety and additional costs. There was no evidence for different frequency of scanning of trichorionic pregnancies but the committee decided to recommend more frequent monitoring, at least every 14 days, because trichorionic triplets are at higher risk of developing growth restrictions than dichorionic twins.

Based on their expertise and experience the committee decided that it was important to calculate the estimated fetal weight (EFW) discordance to identify babies at high risk of having or developing growth restrictions. Based on their expertise and experience they provided a formula on how this would be calculated. This should then be documented so that any future trends can be identified (for example a discordance that is increasing over several scan appointments could raise concerns).

The committee decided to apply the same weight discordance cut-offs to all twin and triplet pregnancies, regardless of chorionicity. This was done because the definitions (as defined by the EFW discordance) are the same but the risks of complications, including fetal growth restriction, is higher in twin and triplet pregnancy that involve monochorionic babies. The reasons for the specific cut-offs are described in the section entitled 'weight discordance cut-offs for all twins and triplets' below.

Diagnostic monitoring for fetal growth restriction in twin and triplet pregnancy that involve monochorionic babies

Simultaneous monitoring of complications

There are several complications that are restricted to monochorionicity (feto-fetal transfusion syndrome (FFTS) and twin anaemia polycythaemia). Fetal growth restriction does not only occur in monochorionic babies but the risk of this complication is higher than in pregnancies not involving monochorionic babies. All of these are monitored by ultrasound. The committee highlighted that measurements from one ultrasound would be used to monitor for all complications simultaneously (such as FFTS, fetal growth restriction and twin anaemia polycythaemia sequence) rather than having separate ultrasound scans for each because they are not mutually exclusive conditions. An explanation about the relative likelihood of each complication and when they can occur during her pregnancy should be given to the woman so that she knows the reasons for the different ultrasound measurements that are taken.

Measurement parameters and frequency of monitoring

For the same reasons as in dichorionic twins and trichorionic triplet pregnancies (see above) the committee also recommended not to use abdominal palpation and symphysis fundal height measurements to monitor for fetal growth restriction. They also agreed that using ultrasound with 2 or more biometric parameters is equally appropriate for twin or triplet pregnancies involving monochorionic babies as it is for dichorionic twin and trichorionic triplet pregnancies (see above).

Based on their experience and expertise, the committee recommended that women with a monochorionic pregnancy need more frequent scans (at 14 day intervals) because these pregnancies have a higher risk of severe growth discordance. Scanning at 14 day intervals would allow the woman to be referred promptly to her specialist obstetrician for multiple pregnancy if concerns arise.

Based on their experience and expertise, the committee provided the calculation for fetal growth discordance in monochorionic twins to enable clinicians to calculate and assess the growth of each baby in relation to each other. They noted that triplet pregnancy involving a monochorionic set of babies may complicate calculations of growth discordance and they therefore recommended that a named specialist obstetrician should be involved in the assessment and calculation of triplets. The committee agreed that this is achievable because it would be a very small percentage of all pregnancies and that women with these pregnancies would already get more frequent specialist appointments.

Weight discordance cut-offs for all twins and triplets

The committee noted that there was evidence for both the 20% and 25% weight discordance cut-offs but that it was unclear whether one was better than the other and therefore based their decision on their experience and expertise. They agreed that the cut-offs should be the same regardless of chorionicity. A cut-off of 20% should raise concerns and therefore increase the frequency of monitoring (to weekly monitoring). This increased monitoring should also include doppler ultrasound assessment. This can measure whether the blood flow in the umbilical artery is normal in all fetuses which would be reassuring. If the doppler assessment indicates a high resistance in the umbilical artery it would be a sign of blood flow redistribution. In combination with the other measures this would be one indicator that may tip the balance between letting the pregnancy continue and intervening by offering an early caesarean section. Ongoing weekly monitoring for an estimated fetal weight discordance of 20% or above would then allow clinicians to assess whether the discordance increases over time. When the discordance is 25% or greater, it indicates fetal growth restriction of one baby. The committee noted that a growth discordance above this limit would lead to an increased risk of perinatal morbidity and mortality and may therefore need earlier intervention. When discussing the evidence, the committee acknowledged that it was mixed in terms of the measures used to estimate fetal weight, the different cut-offs for the screening parameters and for the definition of discordancy. However, even though the evidence was heterogeneous, based on their experience and knowledge the committee agreed that a combination of measures that are used at each scan would help build a general clinical picture as well as pick up any changes over time.

The committee also agreed that the estimated fetal weights themselves should be taken into account. Based on the evidence (where the 10th centile was used as a reference standard) they recommended the 10th centile for gestational age as a threshold for concern that should prompt increased monitoring

The Royal College of Obstetricians and Gynaecologists' Green Top guideline on monochorionic twin pregnancy recommends referring women 'for assessment and management in fetal medicine units with recognised relevant expertise' if there is an estimated fetal weight discordance of more than 20%. They therefore discussed whether their recommendations would conflict with the conclusions of the Green Top guideline. As described above the committee agreed with the Green Top guideline that this level should cause concern and prompt increased monitoring, but they recommended instead increasing to weekly monitoring and adding the extra parameter of a doppler assessment. This would be equivalent to the specialist assessment recommended by the Green Top guideline because it would need to be carried out by the specialist core team (in line with recommendation 1.3.1) who have experience and knowledge of managing twin and triplet pregnancies. The committee agreed that this would not be inconsistent with the Green Top guideline. An estimated fetal weight discordance of 25% or more (along with an EFW below the 10th centile) should warrant referral. At this level of discordance there would be an increased risk of perinatal morbidity and mortality that should prompt intervention rather than increased assessment. The tertiary level fetal medicine centre would have the expertise to weigh up the benefits and risks of conservative management, birth or invasive intrauterine therapy (in monochorionic pregnancies) to try to improve the chance of a positive pregnancy outcome.

Cost effectiveness and resource use

In the absence of any economic evidence or original analysis, the committee made a qualitative assessment about the cost effectiveness of screening and diagnostic monitoring to detect fetal growth restriction in twin and triplet pregnancy.

The committee considered that the perinatal mortality, morbidity and preterm birth associated with fetal growth restriction meant that monitoring for this in women with twin and triplet pregnancies was likely to be cost-effective because of the potential to reduce adverse outcomes by identifying high risk pregnancies.

The recommendations for twin pregnancies largely reinforce current practice. Therefore, the committee did not consider that their recommendations would have a significant impact on NHS resources or the provision of ultrasound scans at the local level.

The committee recognised that their recommendation to monitor triplet pregnancies at no more than 14-day intervals did represent a change in practice but that the number of pregnancies affected is small and is warranted because of the particularly high risk of fetal growth restriction in these pregnancies.

References

Al-Obaidly 2015

Al-Obaidly S, Parrish J, Murphy KE, Glanc P2, Maxwell C. The Accuracy of Estimating Fetal Weight and Inter-Twin Weight Discordance by Ultrasound in Twin Pregnancies in Women With Increased Body Mass Index. J Obstet Gynaecol Can, 37(8):696-701, 2015.

Al Hassan 2012

Al Hassan A, Al Ghany HA. Estimation of Fetal Body Weight in Twins: A New Mathematical Model. Iraqi Comm Med, 1:61-65, 2012.

Banks 2008

Banks CL, Nelson SM, Owen P. First and third trimester ultrasound in the prediction of birthweight discordance in dichorionic twins. Eur J Obstet Gynecol Reprod Biol, 38(1):34-8, 2008.

Caravello 1997

Caravello JW, Chauhan SP, Morrison JC, Magann EF, Martin JN Jr, Devoe LD. Sonographic examination does not predict twin growth discordance accurately. Obstet Gynecol, 89(4):529-33, 1997.

Chamberlain 1991

Chamberlain, P., Murphy, M., Comerford, F. R., How accurate is antenatal sonographic identification of discordant birthweight in twins? Eur J Obstet Gynecol Reprod BiolEuropean journal of obstetrics, gynecology, and reproductive biology, 40, 91-6, 1991.

Chang 2006

Chang YL, Chang TC, Chang SD, Cheng PJ, Chao AS, Hsieh PC, Soong YK. Sonographic prediction of significant intertwin birth weight discordance. Eur J Obstet Gynecol Reprod Biol, 127(1):35-40, 2006.

Chittacharoen 2000

Chittacharoen A, Leelapattana P, Rangsiprakarn R. Prediction of discordant twins by real-time ultrasonography combined with umbilical artery velocimetry. Ultrasound Obstet Gynecol, 15(2):118-21, 2000.

Cordiez 2017

Cordiez, S, Deruelle, P, Drumez, E, Bodart, S, Subtil, D, Houfflin-Debarge, V, Garabedian, C., Impact of customized growth curves on screening for small for gestational age twins, European Journal of Obstetrics, Gynecology, & Reproductive Biology, 215, 28-32, 2017.

Blickstein 1996

Blickstein I, Manor M, Levi R, Goldchmit R. Is intertwin birth weight discordance predictable? Gynecol Obstet Invest, 42(2):105-8, 1996.

Danon 2008

Danon D, Melamed N, Bardin R, Meizner I. Accuracy of ultrasonographic fetal weight estimation in twin pregnancies. Obstet Gynecol, 112(4):759-64, 2008.

D'Antonio 2013

D'Antonio F, Khalil A, Dias T, Thilaganathan B; Southwest Thames Obstetric Research Collaborative. Crown-rump length discordance and adverse perinatal outcome in twins: analysis of the Southwest Thames Obstetric Research Collaborative (STORK) multiple pregnancy cohort. Ultrasound Obstet Gynecol, 41(6):621-6, 2013

D'Antonio 2014

D'Antonio F, Khalil A, Thilaganathan B; Southwest Thames Obstetric Research Collaborative (STORK). Second-trimester discordance and adverse perinatal outcome in twins: the STORK multiple pregnancy cohort. BJOG, 121(4):422-9, 2014

Dias 2010

Dias T, Bhide A, Thilaganathan B. Early pregnancy growth and pregnancy outcome in twin pregnancies. Ceylon Med J. 2010 Sep;55(3):80-4.

Dias-Garcia 2010

Diaz-Garcia C, Bernard JP, Ville Y, Salomon LJ. Validity of sonographic prediction of fetal weight and weight discordance in twin pregnancies. Prenat Diagn, 30(4):361-7, 2010.

Egan 1994

Egan JFX, Vintzileos AM, Turner G, Fleming A, Scorza W, Wolf E etc. Correlation of Uterine Fundal Height with Ultrasonic Measurements in Twin Gestations, Journal of Maternal-Fetal Medicine, 3, 18-22, 1994.

Fajardo-Expósito 2011

Fajardo-Expósito MA, Hervías B, González FB, Melero-Jiménez V, Quintero-Prado R, Facio-Fernández MC, Bartha JL. First trimester fetal head and trunk volume predict growth disturbance in twin pregnancy. Prenat Diagn, 31(6):543-7, 2011.

Fox 2011

Fox NS, Saltzman DH, Schwartz R, Roman AS, Klauser CK, Rebarber A. Second-trimester estimated fetal weight and discordance in twin pregnancies: association with fetal growth restriction. J Ultrasound Med, 30(8):1095-101, 2011.

Gandhi 2009

Gandhi M, Fox NS, Russo-Stieglitz K, Hanley ME, Matthews G, Rebarber A. Effect of increased body mass index on first-trimester ultrasound examination for aneuploidy risk assessment. Obstet Gynecol, 114(4):856-9, 2009.

Gernt 2001

Gernt PR, Mauldin JG, Newman RB, Durkalski VL. Sonographic prediction of twin birth weight discordance. Obstet Gynecol, 97(1):53-6, 2001.

Hoopmann 2011

Hoopmann M, Kagan KO, Yazdi B, Grischke EM, Abele H Prediction of birth weight discordance in twin pregnancies by second- and third- trimester ultrasound. Fetal Diagn Ther, 30(1):29-34, 2011.

Hill 1994

Hill, L. M., Guzick, D., Chenevey, P., Boyles, D., Nedzesky, P., The sonographic assessment of twin growth discordancy, Obstetrics & Gynecology, 84, 501-4, 1994.

Jensen 1995

Jensen, O. H., Jenssen, H., Prediction of fetal weights in twins, Acta Obstet Gynecol ScandActa obstetricia et gynecologica Scandinavica, 74, 177-80, 1995.

Johansen 2014

Johansen, M. L, Oldenburg, A, Rosthoj, S, Cohn Maxild, J, Rode, L, Tabor, A., Crown-rump length discordance in the first trimester: a predictor of adverse outcome in twin pregnancies?, Ultrasound in Obstetrics & Gynecology, 43, 277-83, 2014.

Khalil 2014

Khalil A, D'Antonio F, Dias T, Cooper D, Thilaganathan B; Southwest Thames Obstetric Research Collaborative (STORK). Ultrasound estimation of birth weight in twin pregnancy: comparison of biometry algorithms in the STORK multiple pregnancy cohort. Ultrasound Obstet Gynecol, 44(2):210-20, 2014.

Klam 2005

Klam SL, Rinfret D, Leduc L. Prediction of growth discordance in twins with the use of abdominal circumference ratios. Am J Obstet Gynecol, 192(1):247-51, 2005.

Leombroni 2017

Leombroni, M, Liberati, M, Fanfani, F, Pagani, G, Familiari, A, Buca, D, Manzoli, L, Scambia, G, Rizzo, G, D'Antonio, F., Diagnostic accuracy of ultrasound in predicting birth-weight discordance in twin pregnancy: systematic review and meta-analysis, Ultrasound in Obstetrics & GynecologyUltrasound Obstet Gynecol, 50, 442-450, 2017.

Neves 2017

Neves, A. R, Nunes, F, Branco, M, Almeida, M. D. C, Santos Silva, I., The role of ultrasound in the prediction of birth weight discordance in twin pregnancies: are we there yet?, Journal of Perinatal Medicine, 29, 29, 2017.

O'Connor 2013

O'Connor, C, McAuliffe, F. M, Breathnach, F. M, Geary, M, Daly, S, Higgins, J. R, Dornan, J, Morrison, J. J, Burke, G, Higgins, S, Mooney, E, Dicker, P, Manning, F, McParland, P, Malone, F. D, Perinatal Ireland Research, Consortium, Prediction of outcome in twin

pregnancy with first and early second trimester ultrasound, Journal of Maternal-Fetal & Neonatal Medicine, 26, 1030-5, 2013.

Ong 1999

Ong S, Smith AP, Fitzmaurice A, Campbell D. Estimation of fetal weight in twins: a new mathematical model. Br J Obstet Gynaecol, 106(9):924-8, 1999.

Roberts 2001

Roberts WE, Gnam EC 3rd, Magann EF, Martin JN Jr, Morrison JC. Labor and membrane rupture in twin gestation. Can they affect the ability to estimate fetal weight? J Reprod Med, 46(5):462-6, 2001.

Rodis 1990

Rodis, J. F., Vintzileos, A. M., Campbell, W. A., Nochimson, D. J., Intrauterine fetal growth in discordant twin gestations, J Ultrasound MedJournal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine, 9, 443-8, 1990.

Sayegh 1993

Sayegh, S. K., Warsof, S. L., Ultrasonic prediction of discordant growth in twin pregnancies, Fetal Diagnosis & Therapy, 8, 241-6, 1993.

Shah 1994

Shah, Y. G., Sherer, D. M., Gragg, L. A., Casaceli, C. J., Woods, J. R., Jr., Diagnostic accuracy of different ultrasonographic growth parameters in predicting discordancy in twin gestation: a different approach, American Journal of Perinatology, 11, 199-204, 1994.

Shahshahan 2011

Shahshahan,Z, Hashemi,M., Crown-rump length discordance in twins in the first trimester and its correlation with perinatal complications, Journal of Research in Medical Sciences, 16, 1224-1227, 2011.

Simoes 2011

Simoes T, Julio C, Cordeiro A, Cohen A, Silva A, Blickstein I. Abdominal circumference ratio for the diagnosis of intertwin birth weight discordance. J Perinat Med, 39(1):43-6, 2011.

Sklar 2017

Sklar, C, Yaskina, M, Ross, S, Naud, K., Accuracy of Prenatal Ultrasound in Detecting Growth Abnormalities in Triplets: A Retrospective Cohort Study, Twin Research & Human Genetics: the Official Journal of the International Society for Twin Studies, 20, 84-89, 2017.

Storlazzi 1897

Storlazzi, E., Vintzileos, A. M., Campbell, W. A., Nochimson, D. J., Weinbaum, P. J., Ultrasonic diagnosis of discordant fetal growth in twin gestations, Obstetrics & Gynecology, 69, 363-7, 1987.

van de Waarsenburg 2015

van de Waarsenburg, M. K, Hack, K. E, Rijpma, R. J, Mulder, E. J, Pistorius, L, Derks, J. B., Ultrasonographic prediction of birth weight discordance in twin pregnancies, Prenatal Diagnosis, 35, 906-12, 2015.

van Mieghem 2009

van Mieghem T, Deprest J, Klaritsch P, Gucciardo L, Done' E, Verhaeghe J, Lewi L. Ultrasound prediction of intertwin birth weight discordance in monochorionic diamniotic twin pregnancies. Prenat Diagn, 29(3):240-4, 2009.

Appendices

Appendix A – Review protocol

1.2: Review protocol: What is the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy?

Table 3: Review protocol for the optimal screening programme to detect intrauterine growth restriction in twin and triplet pregnancy

ID	Field (based on PRISMA-P)	Content
I	Review question	What is the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy?
II	Type of review question	Diagnostic accuracy
III	Objective of the review	To determine what is the most accurate screening strategy for detecting IUGR in twin and triplet pregnancies, considering the optimum frequency and duration of ultrasound scans throughout pregnancy
IV	Eligibility criteria – population	**For twin pregnancies:** • monochorionic diamniotic • monochorionic monoamniotic • dichorionic diamniotic **For triplet pregnancies:** • dichorionic triamniotic • monochorionic triamniotic • dichorionic, diamniotic (a monochorionic twins set) and monochorionic monoamniotic • trichorionic, triamniotic Setting: Secondary or tertiary care centres
V	Eligibility criteria – diagnostic and prognostic factor(s)	**Index tests** Estimated during ultrasound scan at 11^{+0} to 13^{+6} weeks: • discrepant crown-rump length • discrepant nuchal translucency Estimated during ultrasound scan at 14 weeks onwards: • growth discordancy (fetal biometry including head circumference, abdominal circumference, femur length, biparietal diameter and estimated fetal weight based on formula of these parameters including difference in estimated fetal weight of each twin ≥ 15%) • amniotic fluid discordancy (amniotic fluid index or maximum pool depth, discordancy between twins in amniotic fluid volume)

ID	Field (based on PRISMA-P)	Content
		• Doppler studies (umbilical artery and vein and middle cerebral artery doppler, ductus venosus doppler) • plotting symphysio-fundal height, estimated fetal weight and fetal biometric measurements on standard population or customised growth charts, twin-specific charts, individual measurements or growth velocity The diagnostic value of first and second trimester tests to detect IUGR will be examined. The above tests will be considered in isolation or in combination. Details regarding frequency and duration of testing throughout pregnancy presented in included studies will be recorded
VI	Eligibility criteria – comparator(s)/control or reference (gold) standard	**Reference standard** • Recognised reference standard for small for gestational age or IUGR including birthweight centiles by gestational age as reported in studies and standard deviation score (according to population or customised or twin specific growth charts) • abdominal circumference, head circumference • Ponderal index and skinfold thickness. • Intertwin weight discordance (any reported >15%) Analysis will be performed separately for the comparison of each index test to each reference standard test. A comparison of index tests to pooled reference standards will not be performed
VII	Outcomes and prioritisation	Diagnostic value of first and second trimester tests **Critical:** • sensitivity • specificity Sensitivity was regarded as the more important measure for decision making as these are primarily screening diagnostic tests **Important:** • area under curve (AUC)
VIII	Eligibility criteria – study design	Systematic reviews of diagnostic accuracy studies Individual diagnostic accuracy studies including: • cross-sectional studies • cohort studies If insufficient data are available from prospective cohort studies, then retrospective cohort studies will be considered. Conference abstracts will not be considered
IX	Other inclusion exclusion criteria	Exclude:

ID	Field (based on PRISMA-P)	Content
		• Studies that report on quadruplet or higher-order multiple pregnancies as per scope • Studies that do not report results specifically for twin and/or triplet pregnancies • Studies that include <5 pregnant women • Structural or chromosomal anomalies • Intra-uterine death at study entry • Studies where 95% CIs for point estimates are not presented or where 2 x 2 contingency data are not presented or cannot be calculated
X	Proposed sensitivity/sub-group analysis, or meta-regression	Special consideration will be given to the following groups for which data will be reviewed and analysed separately if available: • twin pregnancies • triplet pregnancies • women with twin or triplet pregnancies who are aged 17 or less **For twin pregnancies:** • monochorionic diamniotic • monochorionic monoamniotic • dichorionic diamniotic **For triplet pregnancies:** • dichorionic triamniotic • monochorionic triamniotic • dichorionic, diamniotic (a monochorionic twins set) and monochorionic monoamniotic • trichorionic triamniotic Stratify by gestational age if data permit – stratifications as detailed in studies
XI	Selection process – duplicate screening/selection/analysis	Formal duplicate screening will not be undertaken for this question (as it has not been prioritised for economic analysis), although there will be senior supervision of the selection process. Hard copies of retrieved papers will be read by two reviewers and any disputes will be resolved in discussion with the Topic Advisor. Data extraction will be supervised by a senior reviewer. Draft excluded studies and evidence tables will be discussed with the Topic Advisor, prior to circulation to the Topic Group for their comments. Resolution of disputes will be by discussion between the senior reviewer, Topic Advisor and Chair
XII	Data management (software)	NGA STAR software will be used for generating bibliographies/citations, study sifting, data extraction and recording quality assessment using checklists. Meta-analyses will be performed using Cochrane Review Manager (RevMan5) and WinBUGS if available data permit

ID	Field (based on PRISMA-P)	Content
		A modified 'GRADE' method will be used to assess the quality of evidence for each index test
XIII	Information sources – databases and dates	Sources to be searched: Medline, Medline In-Process, CCTR, CDSR, DARE, HTA, Embase Search limits: • limit to English language • limit to human-only studies • no limit on study design • limit year of publication to 2010 for second trimester tests (date of previous guideline searches); no limits on year of publication for first trimester tests Supplementary search techniques: no supplementary search techniques will be used
XIV	Identify if an update	This is an update of a review performed in 2011 Question: What is the optimal screening programme to detect intrauterine growth restriction in multiple pregnancy? Chapter 6.4 of full guideline Recommendations **1.3.5 Monitoring for fetal growth restriction** 1.3.5.1 Do not use abdominal palpation or symphysis–fundal height measurements to predict intrauterine growth restriction in twin or triplet pregnancies. 1.3.5.2 Estimate fetal weight discordance using two or more biometric parameters at each ultrasound scan from 20 weeks. Aim to undertake scans at intervals of less than 28 days. Consider a 25% or greater difference in size between twins or triplets as a clinically important indicator of intrauterine growth restriction and offer referral to a tertiary level fetal medicine centre. 1.3.5.3 Do not use umbilical artery Doppler ultrasound to monitor for intrauterine growth restriction or birthweight differences in twin or triplet pregnancies. **Research recommendation** RR10 What is the pattern of fetal growth in health twin and triplet pregnancies, and how should intrauterine growth restriction be defined in twin and triplet pregnancies?
XV	Author contacts	Developer: National Guideline Alliance https://www.nice.org.uk/guidance/indevelopment/gid-ng10063

ID	Field (based on PRISMA-P)	Content
XVI	Highlight if amendment to previous protocol	For details please see section 4.5 of Developing NICE guidelines: the manual 2014
XVII	Search strategy – for one database	For details please see appendix B
XVIII	Data collection process – forms/duplicate	A standardised evidence table format will be used, and published as appendix G (clinical evidence tables) or H (economic evidence tables).
XIX	Data items – define all variables to be collected	For details please see evidence tables in appendix G (clinical evidence tables) or H (economic evidence tables)
XX	Methods for assessing bias at outcome/study level	Quality assessment of individual studies will be performed using the following checklists: • AMSTAR for systematic reviews • QUADAS II for cross sectional or cohort studies reporting diagnostic accuracy outcomes For details please see section 6.2 of Developing NICE guidelines: the manual 2014 The risk of bias across all available evidence will be evaluated for each outcome 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/
XXI	Criteria for quantitative synthesis (where suitable)	For details please see the methods chapter of the guideline and section 6.4 of Developing NICE guidelines: the manual 2014
XXII	Methods for analysis – combining studies and exploring (in)consistency	A full description of this is provided in the methods in supplementary material C
XXIII	Meta-bias assessment – publication bias, selective reporting bias	For details please see section 6.2 of Developing NICE guidelines: the manual 2014
XXIV	Assessment of confidence in cumulative evidence	For details please see sections 6.4 and 9.1 of Developing NICE guidelines: the manual 2014
XXV	Rationale/context – Current management	For details please see the introduction to the evidence review.
XXVI	Describe contributions of authors and guarantor	A multidisciplinary committee developed the guideline. The committee was convened by the National Guideline Alliance and chaired by Anthony Pearson in line with section 3 of Developing NICE guidelines: the manual 2014. 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. A full description of this is provided in the methods in supplementary material C

ID	Field (based on PRISMA-P)	Content
XXVII	Sources of funding/support	The National Guideline Alliance is funded by NICE and hosted by the Royal College of Obstetricians and Gynaecologists
XXVIII	Name of sponsor	The National Guideline Alliance is funded by NICE and hosted by the Royal College of Obstetricians and Gynaecologists
XXIX	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
XXX	PROSPERO registration number	Not registered with PROSPERO

AMSTAR: Assessing the Methodological Quality of Systematic Reviews; CCTR: Cochrane Central Register for Controlled Trials; CDSR: Cochrane Database of Systematic Reviews; CI: confidence interval; DARE: Database of Abstracts of Reviews of Effects; GRADE: Grading of Recommendations Assessment, Development and Evaluation; HTA: Health Technology Assessment; NGA: National Guideline Alliance; NICE: National Institute for Health and Care Excellence; QUADAS: Quality Assessment of Diagnostic Accuracy Studies

Appendix B – Literature search strategies

Literature search for review question: What is the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy?

Clinical Searches

Date of initial search: 20/02/2018

Database(s): Embase 1980 to 2018 Week 08, Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present

Date of updated search: 06/09/2018

Database(s): Embase 1980 to 2018 Week 36, Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present

#	Searches
1	exp Pregnancy, Multiple/ use ppez
2	exp multiple pregnancy/ use emez
3	((multiple* or twin* or triplet* or monozygotic or dizygotic or trizygotic) adj3 (birth* or pregnan* or gestation* or f?etus* or f?etal)).tw.
4	(chorionicity or monochorionic* or dichorionic* or trichorionic*).tw.
5	or/1-4
6	exp Medical Records/ use ppez
7	medical record/ use emez
8	exp Medical History Taking/ use ppez
9	exp anamnesis/ use emez
10	((medical or patient) adj2 (history or record*)).tw.
11	exp Palpation/ use ppez
12	palpation/ use emez
13	Crown-Rump Length/ use ppez
14	crown rump length/ use emez
15	(palpation adj3 abdom*).tw.
16	((crown rump or crown-rump) adj3 (length* or measur* or height* or estimat* or screen* or discord*)).tw.
17	(fundal height adj3 measur*).tw.
18	(symphysio?fundal adj3 measur*).tw.
19	(symphysio fundal adj3 measur*).tw.
20	((maternal or mother*) adj2 serum screening).tw.
21	exp Chorionic Gonadotropin/bl, di use ppez
22	chorionic gonadotropin/ use emez
23	(HCG or human chorionic gonadotropin).tw.
24	exp Fetal Proteins/ use ppez
25	fetoprotein/ use emez
26	alpha fetoprotein*.tw.
27	alpha feto protein*.tw.
28	AFP.tw.
29	Pregnancy-Associated Plasma Protein-A/ use ppez
30	pregnancy associated plasma protein A/ use emez
31	"PAPP A".tw.

#	Searches
32	PAPP alpha.tw.
33	((blood or alpha) adj2 (protein* or glycoprotein* or globulin* or macroglobulin*)).tw.
34	exp Inhibins/ use ppez
35	inhibin/ use emez
36	inhibin*.tw.
37	exp Estradiol/ use ppez
38	estradiol/ use emez
39	(estradiol or oestradiol).tw.
40	exp ultrasonography, doppler/ use ppez or exp ultrasonography, prenatal/ use ppez or exp ultrasonography/ use ppez
41	exp echograpy/ use emez or exp fetus echography/ use emez or doppler echography/ use emez
42	Nuchal Translucency Measurement/ use ppez
43	nuchal translucency measurement/ use emez
44	((antenatal* or prenatal* or fetal or foetal or fetus* or foetus*) adj3 (diagnos* or screen* or ultrason*)).tw.
45	((fetal or foetal or fetus* or foetus*) adj3 biomet*).tw.
46	(doppler adj3 ultraso*).tw.
47	(uterine artery adj3 doppler).tw.
48	(umbilical adj3 doppler).tw.
49	(middle cerebral artery adj3 doppler).tw.
50	(MCA adj3 doppler).tw.
51	(ductus venosus adj3 doppler).tw.
52	(descending aorta adj3 doppler).tw.
53	(inferior vena cava adj3 doppler).tw.
54	(IVC adj3 doppler).tw.
55	(nuchal adj3 (measur* or scan* or screen* or translucen* or test*)).tw.
56	Fetal Weight/ use ppez
57	fetus weight/ use emez
58	estimat* fetal weight.tw.
59	estimat* foetal weight.tw.
60	Amniotic Fluid/ use ppez
61	exp amnion fluid/ use emez
62	(amniotic fluid adj3 volume).tw.
63	or/6-62
64	Fetal Growth Retardation/bl, di, dg use ppez
65	exp intrauterine growth retardation/di use emez
66	(grow* adj3 (restrict* or retard* or discord*)).tw.
67	exp Infant, Low Birth Weight/bl, gd use ppez
68	exp low birth weight/di use emez
69	(intrauterine growth restrict* or intra-uterine growth restrict*).tw.
70	(small adj3 (gestation* or age)).tw.
71	IUGR.tw.
72	SGA.tw.
73	or/64-72
74	5 and (63 or 73)
75	limit 74 to (english language and yr="2010 -Current")
76	Letter/ use ppez
77	letter.pt. or letter/ use emez
78	note.pt.
79	editorial.pt.

#	Searches
80	Editorial/ use ppez
81	News/ use ppez
82	exp Historical Article/ use ppez
83	Anecdotes as Topic/ use ppez
84	Comment/ use ppez
85	Case Report/ use ppez
86	case report/ or case study/ use emez
87	(letter or comment*).ti.
88	or/76-87
89	randomized controlled trial/ use ppez
90	randomized controlled trial/ use emez
91	random*.ti,ab.
92	or/89-91
93	88 not 92
94	animals/ not humans/ use ppez
95	animal/ not human/ use emez
96	nonhuman/ use emez
97	exp Animals, Laboratory/ use ppez
98	exp Animal Experimentation/ use ppez
99	exp Animal Experiment/ use emez
100	exp Experimental Animal/ use emez
101	exp Models, Animal/ use ppez
102	animal model/ use emez
103	exp Rodentia/ use ppez
104	exp Rodent/ use emez
105	(rat or rats or mouse or mice).ti.
106	or/93-105
107	75 not 106

Date of initial search: 21/02/2018

Database(s): the Cochrane Library, issue 2 of 12, February 2018

Date of updated search: 06/09/2018

Database(s): the Cochrane Library, issue 9 of 12, September 2018

ID	Search
#1	MeSH descriptor: [Pregnancy, Multiple] explode all trees
#2	((multiple* or twin* or triplet* or monozygotic or dizygotic or trizygotic) near/3 (birth* or pregnan* or gestation* or foetus* or foetal or fetus* or fetal))
#3	(chorionicity or monochorionic or dichorionic or trichorionic)
#4	{or #1-#3}
#5	MeSH descriptor: [Fetal Development] this term only
#6	MeSH descriptor: [Fetus] explode all trees and with qualifier(s): [Abnormalities - AB, Blood supply - BS, Diagnostic imaging - DG]
#7	MeSH descriptor: [Infant, Low Birth Weight] explode all trees and with qualifier(s): [Blood - BL, Growth & development - GD]
#8	(grow* near/3 (restrict* or retard* or discord*))
#9	intrauterine growth restrict* or intra-uterine growth restrict*
#10	(small near/3 (gestation* or age))

ID	Search
#11	(IUGR or SGA)
#12	{or #5-#11}
#13	#4 and #12 Publication Year from 2010 to 2018

Health economics

(For the Cochrane Library, see above)

Date of initial search: 21/02/2018

Database(s): Embase 1980 to 2018 Week 08, Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present

Date of updated search: 06/09/2018

Database(s): Embase 1980 to 2018 Week 36, Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present

#	Searches
1	exp Pregnancy, Multiple/ use ppez
2	exp multiple pregnancy/ use emez
3	((multiple* or twin* or triplet* or monozygotic or dizygotic or trizygotic) adj3 (birth* or pregnan* or gestation* or f?etus* or f?etal)).tw.
4	(chorionicity or monochorionic* or dichorionic* or trichorionic*).tw.
5	or/1-4
6	exp Medical Records/ use ppez
7	medical record/ use emez
8	exp Medical History Taking/ use ppez
9	exp anamnesis/ use emez
10	((medical or patient) adj2 (history or record*)).tw.
11	exp Palpation/ use ppez
12	palpation/ use emez
13	Crown-Rump Length/ use ppez
14	crown rump length/ use emez
15	(palpation adj3 abdom*).tw.
16	((crown rump or crown-rump) adj3 (length* or measur* or height* or estimat* or screen* or discord*)).tw.
17	(fundal height adj3 measur*).tw.
18	(symphysio?fundal adj3 measur*).tw.
19	(symphysio fundal adj3 measur*).tw.
20	((maternal or mother*) adj2 serum screening).tw.
21	exp Chorionic Gonadotropin/bl, di use ppez
22	chorionic gonadotropin/ use emez
23	(HCG or human chorionic gonadotropin).tw.
24	exp Fetal Proteins/ use ppez
25	fetoprotein/ use emez
26	alpha fetoprotein*.tw.
27	alpha feto protein*.tw.
28	AFP.tw.
29	Pregnancy-Associated Plasma Protein-A/ use ppez
30	pregnancy associated plasma protein A/ use emez
31	"PAPP A".tw.
32	PAPP alpha.tw.

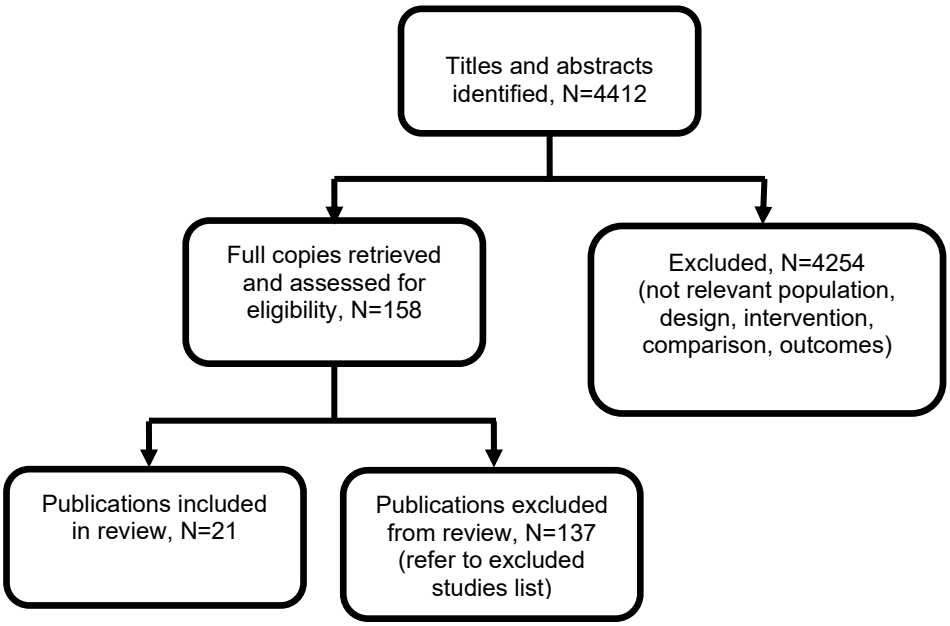
#	Searches
33	((blood or alpha) adj2 (protein* or glycoprotein* or globulin* or macroglobulin*)).tw.
34	exp Inhibins/ use ppez
35	inhibin/ use emez
36	inhibin*.tw.
37	exp Estradiol/ use ppez
38	estradiol/ use emez
39	(estradiol or oestradiol).tw.
40	exp ultrasonography, doppler/ use ppez or exp ultrasonography, prenatal/ use ppez or exp ultrasonography/ use ppez
41	exp echograpy/ use emez or exp fetus echography/ use emez or doppler echography/ use emez
42	Nuchal Translucency Measurement/ use ppez
43	nuchal translucency measurement/ use emez
44	((antenatal* or prenatal* or fetal or foetal or fetus* or foetus*) adj3 (diagnos* or screen* or ultrason*)).tw.
45	((fetal or foetal or fetus* or foetus*) adj3 biomet*).tw.
46	(doppler adj3 ultraso*).tw.
47	(uterine artery adj3 doppler).tw.
48	(umbilical adj3 doppler).tw.
49	(middle cerebral artery adj3 doppler).tw.
50	(MCA adj3 doppler).tw.
51	(ductus venosus adj3 doppler).tw.
52	(descending aorta adj3 doppler).tw.
53	(inferior vena cava adj3 doppler).tw.
54	(IVC adj3 doppler).tw.
55	(nuchal adj3 (measur* or scan* or screen* or translucen* or test*)).tw.
56	Fetal Weight/ use ppez
57	fetus weight/ use emez
58	estimat* fetal weight.tw.
59	estimat* foetal weight.tw.
60	Amniotic Fluid/ use ppez
61	exp amnion fluid/ use emez
62	(amniotic fluid adj3 volume).tw.
63	or/6-62
64	Fetal Growth Retardation/bl, di, dg use ppez
65	exp intrauterine growth retardation/di use emez
66	(grow* adj3 (restrict* or retard* or discord*)).tw.
67	exp Infant, Low Birth Weight/bl, gd use ppez
68	exp low birth weight/di use emez
69	(intrauterine growth restrict* or intra-uterine growth restrict*).tw.
70	(small adj3 (gestation* or age)).tw.
71	IUGR.tw.
72	SGA.tw.
73	or/64-72
74	5 and (63 or 73)
75	limit 74 to (english language and yr="2010 -Current")
76	Letter/ use ppez
77	letter.pt. or letter/ use emez
78	note.pt.
79	editorial.pt.
80	Editorial/ use ppez
81	News/ use ppez
82	exp Historical Article/ use ppez
83	Anecdotes as Topic/ use ppez
84	Comment/ use ppez
85	Case Report/ use ppez

#	Searches
86	case report/ or case study/ use emez
87	(letter or comment*).ti.
88	or/76-87
89	randomized controlled trial/ use ppez
90	randomized controlled trial/ use emez
91	random*.ti,ab.
92	or/89-91
93	88 not 92
94	animals/ not humans/ use ppez
95	animal/ not human/ use emez
96	nonhuman/ use emez
97	exp Animals, Laboratory/ use ppez
98	exp Animal Experimentation/ use ppez
99	exp Animal Experiment/ use emez
100	exp Experimental Animal/ use emez
101	exp Models, Animal/ use ppez
102	animal model/ use emez
103	exp Rodentia/ use ppez
104	exp Rodent/ use emez
105	(rat or rats or mouse or mice).ti.
106	or/93-105
107	75 not 106
108	Economics/
109	Value of life/
110	exp "Costs and Cost Analysis"/
111	exp Economics, Hospital/
112	exp Economics, Medical/
113	Economics, Nursing/
114	Economics, Pharmaceutical/
115	exp "Fees and Charges"/
116	exp Budgets/
117	(or/108-116) use ppez
118	health economics/
119	exp economic evaluation/
120	exp health care cost/
121	exp fee/
122	budget/
123	funding/
124	(or/118-123) use emez
125	budget*.ti,ab.
126	cost*.ti.
127	(economic* or pharmaco?economic*).ti.
128	(price* or pricing*).ti,ab.
129	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
130	(financ* or fee or fees).ti,ab.
131	(value adj2 (money or monetary)).ti,ab.
132	or/125-130
133	117 or 124 or 132
134	107 and 133

Appendix C – Clinical evidence study selection

Clinical evidence study selection for review question: What is the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy?

Figure 1: Flow diagram of clinical article selection for the optimal screening programme to detect intrauterine growth restriction in twin and triplet pregnancy



Appendix D – Clinical evidence tables

Clinical evidence tables for review question: What is the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy?

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Full citation Banks,C.L., Nelson,S.M., Owen,P., First and third trimester ultrasound in the prediction of birthweight discordance in dichorionic twins, European Journal of Obstetrics, Gynecology, and Reproductive Biology, 138, 34-38, 2008 **Ref Id** 97461 **Country/ies where the study was carried out** UK **Study type** Retrospective cohort study **Aim of the study** To test whether inter-twin disparity in 1st trimester biometry or the use of 3rd trimester fetal growth velocity or estimated	**Sample size** N=108 DC twin pregnancies. **Characteristics** Median maternal age (IQR): 35 (27-36); median gestation at birth (IQR): 36^{+3} weeks (35^{+1} to 37^{+5} weeks); median intertwin birth weight disparity (IQR): 0.29 kg (0.17 -0.54 kg); Median number of days from last scan to birth (IQR): 10 days (5-15). **Inclusion Criteria** Structurally and chromosomally normal twin gestations resulting in two live births after 24 weeks' gestation. **Exclusion Criteria** MC pregnancies.	**Tests** **Index test** CRL discordance >=5% measured at 10 to 14 weeks' gestation **Reference standard** Intertwin BWD >=20% Note: data for EFW and fetal growth velocity were not extracted as EFW was expressed as a standard deviation score, i.e. Z score.	**Methods** Data were collected from the perinatal database at the Princess Royal Maternity Unit, UK. The inter-twin disparities in CRL were calculated and expressed as a percentage of the larger twin (Kalish 2003). Birthweight disparity was calculated as the inter-twin BWD relative to the larger twin an expressed as a percentage. BWD was defined as ≥20% difference in birth weights.	**Results** Diagnostic accuracy of CRL (≥5%) discrepancy to predict BWD (≥20%): sensitivity (95% CI): 0.59 (0.36 to 0.79) specificity (95% CI): 0.60 (0.48 to 0.72) AUC (95% CI): 0.55 (0.44 to 0.66)	**Limitations** RoB was assessed using QUADAS-II **A. RoB** **Patient Sampling** Was a consecutive or random sample of patients enrolled? Unclear Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Unclear risk **B. Concerns regarding applicability:** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB**

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
fetal weight difference usefully predicts BWD. **Study dates** From September 2002 for 3 years. **Source of funding** Not reported.					Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? No Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standards likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					B. Concerns regarding applicability Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Unclear (CRL was measured at 10 to 14 weeks' gestation) Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Unclear concern **Other information** None
Full citation Chamberlain, P., Murphy, M., Comerford, F. R., How accurate is antenatal sonographic identification of discordant birthweight in twins?, Eur J Obstet	**Sample size** N=85 twin pregnancies with last USS performed within 7 days or within 14 days of birth. **Characteristics**	**Tests** **Screening tests** EFWD ≥20% and ≥25% using 1) AC only 2) FL and AC EFW calculation using FL and AC	**Methods** In all twin pregnancies identified, sequential ultrasound examinations at 1-4 week intervals were performed. No other information regarding the	**Results** Accuracy of EFW difference ≥20% estimated by AC and FL to determine BWD ≥20%: *Last USS to birth interval ≤7 days:*	**Limitations** RoB was assessed using QUADAS-II **Patient Sampling** **A. RoB** Was a consecutive or random sample of patients enrolled? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Gynecol Reprod BiolEuropean journal of obstetrics, gynecology, and reproductive biology, 40, 91-6, 1991 **Ref Id** 807890 **Country/ies where the study was carried out** Ireland **Study type** Retrospective cohort study **Aim of the study** To determine the accuracy of ultrasound determined interpair EFW percentage using EFW equations not dependent on BPD measurements in the antenatal identification of discordant birthweight in twins. **Study dates** January 1985 to December 1988. **Source of funding** Not reported.	All ultrasound examinations were performed by one examiner. Details of ethnicity and chorionicity not reported. **Inclusion Criteria** All twin pregnancies identified in the Fetal Assessment Unit, Department of Obstetrics and Gynaecology, Regional Hospital, Galway, Ireland, who underwent sequential USSs at 1-4 week intervals. **Exclusion Criteria** Interval between the last USS and birth of ≥14 days; intrauterine death in one fetus at referral or ≥ 14 days before birth; major congenital anomaly; failure to record birthweight within 6 hours of birth; AC and FL measurements too small for EFW determination.	was based on Hadlock (1984) **Reference test** Intertwin birthweight discordance ≥20% and ≥25%	frequency and duration of screening was reported. At each examination AC and, if possible, FL were measured and recorded. EFW for each fetus was determined from either AC measurement alone or from both AC and FL measurements. Details of equipment and method reported.	TP=6, FP=3, FN=5, TN=39 *Last USS to birth interval ≤14 days:* TP=6, FP=5, FN=7, TN=56 Accuracy of EFW difference ≥25% estimated by AC and FL to determine BWD ≥25%: *Last USS to birth interval ≤7 days:* TP=3, FP=1, FN=3, TN=46 *Last USS to birth interval ≤14 days:* TP=3, FP=1, FN=5, TN=6	Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Low risk **B. Concerns regarding applicability** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					the review question? Low concern **Reference Standard** **A. RoB** Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low risk **Other information** Study information transcribed from CG129 Multiple Pregnancy appendix H: Evidence tables. The RoB assessment was conducted by the NGA technical team
Full citation Cordiez, S, Deruelle, P, Drumez, E, Bodart, S, Subtil, D, Houfflin-Debarge, V, Garabedian, C., Impact of customized growth curves on screening for small for gestational age twins, European Journal of Obstetrics, Gynecology, & Reproductive Biology, 215, 28-32, 2017 **Ref Id** 794300	**Sample size** N=236 twin pregnancies **Characteristics** Maternal age (mean): SGA = 29.9 (5.6) GA at birth (median): 36 weeks (33 - 37) n=162 (34%) had a SGA<10th percentile at birth, among these n=44 (9%) were below the 3rd percentile. **Inclusion Criteria** All twin live births between 1 January 2010	**Tests** Ultrasound: **Index test** SGA was defined by EFW <10th percentile of the curve used. EFW was calculated using curves: 1) Hadlock's formula (1985), based on AC, FL, HC and BPD; 2) The customised curve (including maternal weight and height, parity and	**Methods** Sonographical data used in the analysis were collected during the latest ultrasound performed less than 30 days before birth. The fetal weight was calculated according to the formula of Hadlock (1985, based on AC, FL and HC). SGA was defined by EFW <10th percentile of the curve used (Gardosi 1992). The small weight for GA at birth was defined by	**Results** Diagnostic accuracy of EFW to predict SGA (defined as birth weight <10th percentile according to the French curves by Leroy and Lefort (Leroy 1971)). EFW based on curves: 1) Hadlock (1985) based on AC, FL, HC, BPD: sensitivity = 67.3% (59.5 - 74.4), specificity = 80% (75.1 - 84.3); 2) The customised curve (including	**Limitations** RoB was assessed using QUADAS-II **A. RoB** **Patient Sampling** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Low risk

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Country/ies where the study was carried out France **Study type** Retrospective cohort study **Aim of the study** To evaluate the SGA screening rates in twin pregnancies and to evaluate whether the use of adjusted or customized curves curve could help to better identify SGA fetuses. **Study dates** Between January 2010 and December 2013 **Source of funding** Not reported	and 31 December 2013 at the Jeanne de Flandre tertiary care maternity in Lille (France). **Exclusion Criteria** Those with FFTS during pregnancy or TAPS, malformation syndrome or intrauterine death in one twin.	fetal sex) (Ego 2006); 3) The EPOPé unadjusted (Ego 2016); 4) Adjusted on the fetal sex (Ego 2016). Sonographical data used in the analysis were collected during the latest ultrasound performed less than 30 days before birth. **Reference standard** SGA defined as birth weight <3rd or <10th percentile according to the French curves by Leroy and Lefort (Leroy 1971). Sonographical data used in the analysis were collected during the latest US performed less than 30 days before birth.	a weight <10th percentile according to the French curves by Leroy and Lefort (1971). The study used 4 growth curves: the Hadlock's curve (1985), the customised curve (including maternal weight and height, parity and fetal sex) (Ego 2006), the EPOPé unadjusted (M0) (Ego 2016) and adjusted on the fetal sex (M1) curves (Ego 2016). Information regarding the frequency and duration of screening was not reported.	maternal weight and height, parity and fetal sex) (Ego 2006): sensitivity = 63% (55 - 70.4), specificity = 82.3% (77.5 - 86.4); 3) The EPOPé unadjusted (Ego 2016): sensitivity = 59.9% (51.9 - 67.5), specificity = 83.5% (78.9 - 87.5); 4) Adjusted on the fetal sex (Ego 2016): sensitivity = 57.4% (49.4 - 65.1), specificity = 83.2% (78.6 - 87.2). Diagnostic accuracy of EFW to predict SGA (defined as birth weight <3rd percentile according to the French curves by Leroy and Lefort (Leroy 1971)). EFW based on curves: 1) Hadlock (1985) based on AC, femur length, HC, biparietal diameter: sensitivity = 63.6% (49.4 - 77.8), specificity = 89.2% (85.9 - 92); 2) The customised curve (including maternal	**B. Concerns regarding applicability:** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? One threshold - yes, another - no Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB**

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				weight and height, parity and fetal sex) (Ego 2006): sensitivity = 65.9% (50.1 - 79.5), specificity = 85.5% (81.8 - 88.7); 3) the EPOPé unadjusted (Ego 2016): sensitivity = 56.8% (42.2 - 71.4), specificity = 89.2% (85.9 - 92); 4) adjusted on the fetal sex (Ego 2016): sensitivity = 63.6% (49.4 - 77.8), specificity = 90.2% (87 - 92.8). No data were reported to calculate 2 by 2 table.	Is the reference standards likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Could the patient flow have introduced bias? Unclear concern **Other information** None

Full citation D'Antonio, F, Khalil, A, Dias, T, Thilaganathan, B, Southwest Thames Obstetric Research, Collaborative, Crown-rump length discordance and adverse perinatal outcome in twins: analysis of the Southwest Thames Obstetric Research Collaborative (STORK) multiple pregnancy cohort, Ultrasound in Obstetrics & Gynecology, 41, 621-6, 2013 **Ref Id** 794302 **Country/ies where the study was carried out** UK **Study type** Retrospective cohort study **Aim of the study** To ascertain the performance of 1st trimester CRL discordance in the prediction of adverse perinatal outcome in a large cohort of twin pregnancies.	**Sample size** N=2155 twin pregnancies; n=1735 DC, n=420 MC. **Characteristics** Median CRL discordancy (IQR): 3.5% (1.47 - 6.55); BW and EFW discordancy >25% was present in 11.8% and 12.8%. **Inclusion Criteria** All twin pregnancies booked for antenatal care in nine hospitals in the STORK over a period of 10 years. All women registering for routine antenatal care by 11 weeks' gestation were considered suitable for the analysis. **Exclusion Criteria** Termination of pregnancy, presence of fetal or chromosomal abnormalities, pregnancies of unknown chorionicity, MC monoamniotic pregnancies and high-order multiple gestations.	**Tests** **Index test** CRL discordance measured at 11 to 14 weeks' gestation (data used for the analysis were from the latest ultrasound) **Reference standard** 1) Intertwin BWD 2) SGA defined as the presence of at least one twin with BW<5th centile according to the singleton published reference ranges (Yudkin 1987)	**Methods** GA was determined by the CRL of the larger twin at the 11–14-week scan. A routine fetal structural survey was carried out at 20–22 weeks, and all MC twins had 2 additional scans at around 17 and 19 weeks specifically to identify early features of TTTS. CRL discordance (%) was calculated as 100*(larger CRL−smaller CRL)/ larger CRL. Ultrasound EFW was calculated using the Hadlock (1985) formula based on head circumference, abdominal circumference and femur length, while actual BW discordance (%) was calculated as 100*(larger BW−smaller BW)/larger BW. Only ultrasound examinations just prior to birth were considered for the analysis.	**Results** <u>Diagnostic accuracy of CRL discrepancy to predict BWD:</u> Overall: AUC (95% CI): 0.61 (0.56 to 0.65) MC twins only: AUC (95% CI): 0.61 (0.50 to 0.71) <u>Diagnostic accuracy of CRL discrepancy to predict SGA <5th centile:</u> Overall: AUC (95% CI): 0.56 (0.53 to 0.59) MC twins only: AUC (95% CI): 0.57 (0.49 to 0.66)	**Limitations** RoB was assessed using QUADAS-II **A. RoB** **Patient Sampling** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Low risk **B. Concerns regarding applicability:** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? No (continuous variable)

Study dates 10 years since 2000. **Source of funding** Not reported					Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standards likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing**

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					A. RoB Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low concern **Other information** None
Full citation D'Antonio, F, Khalil, A, Thilaganathan, B, Southwest Thames Obstetric Research, Collaborative, Second-trimester discordance and adverse perinatal outcome in twins: the STORK multiple pregnancy cohort, BJOG: An International Journal of Obstetrics & Gynaecology, 121, 422-9, 2014 **Ref Id** 794305	**Sample size** N=2399 twin pregnancies; n=1942 DC, n=457 MC. **Characteristics** Rate of BWD >=25% was 12.1% (10.7% in DC and MC twins) **Inclusion Criteria** All twin pregnancies booked for antenatal care in 9 hospitals in the STORK over a period of 10 years. All women registering for routine antenatal care by 11 weeks' gestation were considered suitable for the analysis.	**Tests** US **Index test** 1) AC discordancy 2) EFW discordance based on Hadlock's formula (1985); **Reference standard** BWD ≥25%	**Methods** GA was determined by the CRL of the larger twin at the 11–14-week scan. A routine fetal structural survey was carried out at 20–22 weeks, and all MC twins had two additional scans at around 17 and 19 weeks specifically to identify early features of TTTS. US EFW was calculated using the Hadlock (1985) formula based on HC, AC and FL. Ultrasound EFW discordance was calculated as 100*(larger EFW – smaller EFW)/larger EFW,	**Results** Diagnostic accuracy of EBWD to predict BWD ≥25%: Overall: AUC (95% CI): 0.63 (0.56 to 0.65) MC twins only: AUC (95% CI): 0.61 (0.50 to 0.71) Diagnostic accuracy of AC discordancy to predict BWD >=25%: Overall: AUC (95% CI): 0.61 (0.58 to 0.63) MC twins only: AUC (95% CI): 0.61 (0.58 to 0.63)	**Limitations** Risk of bias was assessed using QUADAS-II **A. RoB** **Patient Sampling** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Low risk **B. Concerns regarding applicability:** Patient characteristics and setting

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Country/ies where the study was carried out UK **Study type** Retrospective cohort study **Aim of the study** To ascertain the performance of 2nd trimester US biometry in the prediction of adverse perinatal outcomes in twin pregnancies. **Study dates** A period of 10 years from 2000. **Source of funding** Francesco D'Antonio is funded by University of Chieti, Italy, for a PhD in biomedical, clinical and experimental sciences.	**Exclusion Criteria** Termination of pregnancy, presence of fetal or chromosomal abnormalities, pregnancies of unknown chorionicity, MC monoamniotic pregnancies and high-order multiple gestations.		whereas actual birthweight discordancy was calculated as 100*(larger BW – smaller BW)/larger BW, and discordancy in abdominal circumference was calculated as 100*(larger AC – smaller AC)/larger AC. Only the ultrasound examinations at the time of the routine anomaly scan, between 20 and 22 weeks' gestation, were considered for the analysis.		Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? No (continuous variable) Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standards likely to correctly classify the target condition? Yes Were the reference standard results interpreted without

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low concern **Other information** None
Full citation	**Sample size**	**Tests** **Index test**	**Methods**	**Results**	**Limitations**

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Dias, T., Bhide, A., Thilaganathan, B., Early pregnancy growth and pregnancy outcome in twin pregnancies, Ceylon Medical JournalCeylon Med J, 55, 80-4, 2010 **Ref Id** 756371 **Country/ies where the study was carried out** UK **Study type** Retrospective cohort study **Aim of the study** To determine the association of CRL discrepancy and pregnancy outcome in monochorionic and dichorionic twins. **Study dates** Between December 1996 and September 2009. **Source of funding** Not reported.	N=660 twin pregnancies; n=506 DC, n=154 MC. **Characteristics** Median CRL discordancy: MC = 3.9% (+-8.34, range 0 to 59); DC = 3.2% (+-5.65, range 0-37.5); median BWD: MC = 8% (range 0-57); DC = 9.4% (range 0-69) No further description of the population. **Inclusion Criteria** Twin pregnancies with CRL between 45 mm and 84 mm. **Exclusion Criteria** Twin pregnancies referred from other hospitals.	CRL discordance (continuous) measured at 11 to 14 weeks' gestation. **Reference standard** 1) BWD ≥15% 2) BWD ≥25%	All the 1st trimester twin pregnancy data between 11 and 14 weeks were reviewed. The inter-twin CRL discrepancy was calculated by subtracting CRL of smaller twin (CRL-S) from the CRL of larger twin (CRL-L). A percentage difference of the CRL was computed by dividing CRL discrepancy by CRL of the large twin. BWD of twins calculated as the difference in the weights expressed as a % of that of the bigger twin. BWD of ≥15% was considered as a Grade I (mild) discordancy and ≥25% considered Grade II.	Diagnostic accuracy of CRL discrepancy to predict BWD >=15%: AUC (95% CI): 0.59 (0.54 to 0.65) Diagnostic accuracy of CRL discrepancy to predict BWD >=25%: AUC (95% CI): 0.63 (0.55 to 0.70)	RoB was assessed using QUADAS-II **A. RoB** **Patient Sampling** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Unclear as there is no description of the population Could the selection of patients have introduced bias? Unclear risk **B. Concerns regarding applicability:** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standards likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Flow and Timing **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low concern **Other information** None
Full citation Egan, James F. X., Vintzileos, Anthony M., Turner, Garry, Fleming, Alfred, Scorza, William, Wolf, Edward, Balducci, James, Correlation of Uterine Fundal Height with Ultrasonic Measurements in Twin Gestations, Journal of Maternal-Fetal Medicine, 3, 18-22, 1994 **Ref Id** 807896	**Sample size** N= 160 women with twin pregnancies. Using a cut-off of 20% difference for BWD, 143 of these were deemed normal and 17 discordant. **Characteristics** Women were 16 to 36 weeks pregnant at referral and had reliable menstrual dates that were confirmed by USS before the 20th week of pregnancy. 128 women (80%) were white; 20 (12.5%) Hispanic, 11 (7%) black, and 1 (0.5%).	**Tests** **Screening test** SFH measurement **Reference test** Intertwin birthweight discordancy ≥20%	**Methods** SFH and USS measurements (BPD, HC, AC, FL and amniotic fluid volume - single vertical pocket) were obtained in all women, at three different locations. EFW was derived using Hadlock formulae (BPD/AC and/or FL/AC). Using regression analysis, a normogram for SFH of the 143 normal twin pregnancies was obtained which was then used to determine the diagnostic	**Results** Diagnostic accuracy of SFH measurement in detecting intertwin weight discordance ≥20%: TP=4, FP=25, FN=13, TN=118	**Limitations** RoB was assessed using QUADAS-II **Patient Sampling** **A. RoB** Was a consecutive or random sample of patients enrolled? Unclear Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Unclear risk

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Country/ies where the study was carried out USA **Study type** Descriptive **Aim of the study** To establish a nomogram for SFH measurement in normal twin pregnancies and to determine whether twins with growth discordancy, as defined by US, can be detected by the nomogram. **Study dates** April 1987 – November 1991 **Source of funding** Not reported.	Details of chorionicity not reported. **Inclusion Criteria** Women with confirmed twin pregnancies, referred by physicians from the Division of at the University of Connecticut Health Centre, Farmington, USA, for further ultrasound evaluation, during April 1987 to November 1991. **Exclusion Criteria** Pregnancies with fetal anomalies or known medical or obstetrical complications.		accuracy of SFH measurement. Discordancy was confirmed at birth in all cases. Details of techniques and equipment reported.		**B. Concerns regarding applicability** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standard likely to correctly classify the target condition? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Unclear as no ultrasound-to-birth interval was provided. Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Unclear risk

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Other information Study information transcribed from CG129 Multiple Pregnancy appendix H: Evidence tables. The risk of bias assessment was conducted by the NGA technical team

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Full citation Fajardo-Exposito, M. A, Hervias, B, Gonzalez, F. B, Melero-Jimenez, V, Quintero-Prado, R, Facio-Fernandez, M. C, Bartha, J. L., First trimester fetal head and trunk volume predict growth disturbance in twin pregnancy, Prenatal Diagnosis, 31, 543-7, 2011 **Ref Id** 794311 **Country/ies where the study was carried out** Spain **Study type** Prospective cohort study **Aim of the study** To test the hypothesis that an inter-twin fetal head and trunk volume discrepancy determined during the 1st trimester of pregnancy detects a higher proportion of early growth discrepancies than traditional methods and would be an useful	**Sample size** N=46 twin pregnancies; n=35 DC, n=11 MC **Characteristics** Maternal age (mean): 31.4 (4.9); gestational age at birth (weeks): 35-37; birth weight (mean, g) of the 1st twin: 2357 (582.5), 2nd twin: 2310 (623.7); CRL (mean, mm) of the first twin: 70.7 (10.2), 2nd twin: 69.8 (9.6) **Inclusion Criteria** Twin pregnancy. **Exclusion Criteria** Not reported.	**Tests** **Index test** CRL discordance (threshold >15%) measured at 11 to 14 weeks' gestation. **Reference standard** 1) BWD >15% 2) Growth retardation at birth or SGA was defined as a BW <10th percentile at birth (Santamaria 1998), at least one growth retarded neonate.	**Methods** The study was conducted in a tertiary referral university hospital. Inter-twin differences were calculated with the formula [(larger − smaller twin)/larger twin] × 100. CRL discordancy was defined as a difference of 15% or more of CRL between twins using the larger twin as the index. If there was discordancy between twins the CRL used was the larger. All pregnancies were followed up until birth and none were excluded.	**Results** Diagnostic accuracy of CRL discrepancy (>15%) to predict BWD >15%: Overall: TP=2, FP=1, FN=13, TN=30 MC twins only: TP=1, FP=0, FN=2, TN=8 DC twins only: TP=1, FP=1, FN=11, TN=22 Diagnostic accuracy of CRL discordancy (>15%) to predict SGA defined as BW <10th percentile (Santamaria 1998), at least one growth retarded neonate: Overall: TP=1, FP=2, FN=9, TN=34 MC twins only: TP=0, FP=1, FN=2, TN=8 DC twins only: TP=1, FP=1, FN=7, TN=26	**Limitations** Risk of bias was assessed using QUADAS-II **A. RoB** **Patient Sampling** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Low risk **B. Concerns regarding applicability:** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
predictor of late growth disturbance in chromosomally normal twin pregnancies. **Study dates** Not reported **Source of funding** Not reported					Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standards likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Flow and Timing **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low concern **Other information** None
Full citation Hill, L. M., Guzick, D., Chenevey, P., Boyles, D., Nedzesky, P., The sonographic assessment of twin growth discordancy, Obstetrics & Gynecology, 84, 501-4, 1994 **Ref Id** 758263 **Country/ies where the study was carried out** USA	**Sample size** N= 49 twin pregnancies scanned within 21 days of birth. **Characteristics** Details of ethnicity or chorionicity not reported. **Inclusion Criteria** US examination at or after 15 weeks of pregnancy; last examination within 3 weeks of birth. **Exclusion Criteria** Late pregnancy test, first examination later than 10	**Tests** **Screening test** Intertwin EFW difference ≥20% EFW calculated from HC and AC according to Hadlock (1984) **Reference test** Intertwin BWD ≥20%	**Methods** All pregnancies underwent measurements of AC, FL, EFW, and TCD. Efficacies of the difference in AC (cut-off 20 mm), FL (cut-off 5mm), TCD (cut-off 4 mm) and EFW (cut-off 20%) in predicting twin discordancy was calculated. Details of equipment and method reported. Information regarding the frequency and duration of screening was not reported.	**Results** Prediction of fetal weight discordancy ≥20% using difference in EFW ≥20%: TP=13, FP=5, FN=1, TN=30	**Limitations** RoB was assessed using QUADAS-II **Patient Sampling** **A. RoB** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Low risk **B. Concerns regarding applicability**

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Study type Retrospective cohort study **Aim of the study** To evaluate the effectiveness of fetal biometry - AC, FL and TCD - for detecting twin growth discordancy. **Study dates** Not reported. **Source of funding** Not reported.	weeks' gestation, use of oral contraceptives up to 3 months before conception; irregular menses.				Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low risk **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low risk **Other information** Study information transcribed from CG129 Multiple Pregnancy

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					appendix H: Evidence tables. The RoB assessment was conducted by the NGA technical team
Full citation Jensen, O. H., Jenssen, H., Prediction of fetal weights in twins, Acta Obstet Gynecol ScandActa obstetricia et gynecologica Scandinavica, 74, 177-80, 1995 **Ref Id** 807899 **Country/ies where the study was carried out** Norway **Study type** Retrospective cohort study **Aim of the study** To determine the relative accuracy of US EFW in twin pregnancies and to assess the accuracy of identifying discordant twins. **Study dates**	**Sample size** N=73 twin pregnancies with last USS performed within 7 days of birth. **Characteristics** Details of ethnicity and chorionicity not reported. **Inclusion Criteria** All consecutive women with twin pregnancies who gave birth at Aker University Hospital between 1 January 1990 and 31 March 1993; EDD established by USS at 18 weeks of pregnancy; last USS performed within 7 days of birth. **Exclusion Criteria** None reported.	**Tests** **Screening tests** Intertwin EFW difference ≥20% EFW was calculated using Hadlock's formula (1984) based on BPD and AC. **Reference tests** Intertwin birthweight discordance ≥20%	**Methods** BPD and AC measurements were carried out in all women and EFW calculated from Hadlock's formula. Details of equipment/method reported. Information regarding the frequency and duration of screening was not reported.	**Results** Prediction of intertwin BWD ≥20% using EFW difference ≥20%: TP=9, FP=5, FN=5, TN=49	**Limitations** RoB was assessed using QUADAS-II **Patient Sampling** **A. RoB** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Unclear: no exclusion criteria were reported for this study Could the selection of patients have introduced bias? Unclear risk **B. Concerns regarding applicability** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
January 1990 to March 1993 **Source of funding** Not reported.					**Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low risk **Other information** Study information transcribed from CG129 Multiple Pregnancy appendix H: Evidence tables. The RoB assessment was conducted by the NGA technical team

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Full citation Johansen, M. L, Oldenburg, A, Rosthoj, S, Cohn Maxild, J, Rode, L, Tabor, A., Crown-rump length discordance in the first trimester: a predictor of adverse outcome in twin pregnancies?, Ultrasound in Obstetrics & Gynecology, 43, 277-83, 2014 **Ref Id** 795007 **Country/ies where the study was carried out** Denmark **Study type** Retrospective cohort study **Aim of the study** To evaluate outcome in twin pregnancies according to CRL discordance in the 1st trimester with the main focus on fetal loss and preterm birth before 34 weeks' gestation, in an attempt to assess the usefulness of CRL	**Sample size** N=1993 (n=1733 DC and n=260 MC) twin pregnancies. **Characteristics** Maternal age (median): DC: concordant = 31.8, DC discordant = 31.7 MC: concordant = 31.3, DC discordant = 28.2 CRL discordance (≥10%): DC: 156 (9%); MC: 32 (12%) **Inclusion Criteria** Diamniotic twin pregnancies with a chorionicity determination and 2 live fetuses identified at the time of the NT scan (at the 11−14-week). The earliest assessment of chorionicity and CRL was used. **Exclusion Criteria** Pregnancies with unknown chorionicity, MC monoamniotic pregnancies and pregnancies with a known reduction from a higher number of multiples.	**Tests** US **Index test** CRL discrepancy (threshold >=10%) measured at 11 to 14 weeks' gestation. **Reference standard** Intertwin weight discordance (threshold ≥20%).	**Methods** The cohort was identified by retrieving data on twin pregnancies with two live fetuses at the time of the NT scan from local Astraia servers in 14 of the 21 Departments of Obstetrics and Gynaecology in Denmark. The difference in CRL was calculated as the difference in the twin CRL measurements divided by the CRL of the larger twin and was expressed as a %. CRL discordance was defined as a discordance of ≥10% at the time of the NT scan. Intertwin weight discordance was calculated by dividing the difference in birth weight by the weight of the larger twin. BWD was defined as a discordance of ≥20%. Information regarding the frequency and duration of screening was not reported.	**Results** Diagnostic accuracy of CRL discordance (≥10%) to predict birth weight discordance (≥20%): Overall for DC and MC twins: TP=46, FP=242, FN=142, TN=1563 DC twins only: TP=37, FP=216, FN=119, TN=1361 MC twins only: TP=9, FP=26, FN=23, TN=202	**Limitations** RoB was assessed using QUADAS-II **A. RoB** **Patient Sampling** Was a consecutive or random sample of patients enrolled? Unclear Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Unclear risk **B. Concerns regarding applicability:** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
discordance as a predictor of adverse outcome. **Study dates** Between 2004 and 2006 **Source of funding** Not reported					Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standards likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Flow and Timing **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low concern **Other information** None
Full citation Leombroni, M, Liberati, M, Fanfani, F, Pagani, G, Familiari, A, Buca, D, Manzoli, L, Scambia, G, Rizzo, G, D'Antonio, F., Diagnostic accuracy of ultrasound in predicting birth-weight discordance in twin pregnancy: systematic review and meta-analysis, Ultrasound in Obstetrics & GynecologyUltrasound Obstet Gynecol, 50, 442-450, 2017	**Sample size** N=20 studies (4 prospective, 16 retrospective) N=5826 twin pregnancies The following data were extracted from the original papers Al Hassan (2012): n=107 Al-Obaidly (2015): n=300 Blickstein (1996): n=90 Caravello (1997): n=242 Chang (2006): n=575 Chittacharoen (2000): n=40 Danon (2008): n=278 Diaz-Garcia (2010): n=283 Fox (2011): n=306 Gandhi (2009): n=194 Gernt (2001): n=192	**Tests** **Index test - by each study** Al Hassan (2012): EFW (20%) discordance (weight formulae used: Campbell, Shepard, Hadlock) Al-Obaidly (2015): EFW (25%) discordance (weight formula used: Hadlock) Blickstein (1996): AC (18 mm) and EFW (15%, 20%, 25%) discordance	**Methods** For analysis, three different cut-offs of BW discordance were evaluated (≥15%, ≥20%, ≥25%). Two authors reviewed all abstracts independently and agreement regarding potential relevance was reached by consensus. The same two authors extracted relevant data and inconsistencies were discussed and consensus was reached, or the dispute was resolved by discussion with a third	**Results** <u>Diagnostic accuracy of EFW discordance (≥15%) to predict BW discordance (≥15%):</u> Overall accuracy (6 studies*, n=1477#): sensitivity (95% CI) = 67.9% (62.2 - 73.1) specificity (95% CI) = 83.3% (78.5 - 87.3) <u>Diagnostic accuracy of EFW discordance (≥20%) to predict BW discordance (≥20%):</u> Overall accuracy (7 studies*, n=1780#):	**Limitations** **AMSTAR** **Did the research questions and inclusion criteria for the review include the components of PICO?** Yes **Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?** Yes

Ref Id 794325 **Country/ies where the study was carried out** Italy and Norway **Study type** Systematic review **Aim of the study** To explore the accuracy of sonographic EFW discordance in predicting BW discordance and to ascertain the accuracy of sonographic fetal abdominal-circumference discordance in predicting BW discordance. **Study dates** Between 1996 and 28th July 2016 <u>Includes 20 studies</u> 1) Khalil A, D'Antonio F, Dias T, Cooper D, Thilaganathan B; Southwest Thames Obstetric Research Collaborative (STORK). Ultrasound estimation of birth weight in twin pregnancy: comparison of	Hoopmann (2011): n=196 Khalil (2014): n=293 Klam (2005): n=503 Roberts (2000): n=113 O'Connor (2013): n=960 Ong (1999): n=152 Simoes (2011): n=661 Van de Waarsenburg (2015): n=281 Van Mieghem (2009): n=60 **Characteristics** <u>Ultrasound to birth interval:</u> Al Hassan (2012): 3 days Al-Obaidly (2015): 14 days Blickstein (1996): within 2 weeks Caravello (1997): within 3 weeks Chang (2006): within 28 days Chittacharoen (2000): within 2 weeks Danon (2008): 3 days Diaz-Garcia (2010): 15 days Fox (2011): not stated Gandhi (2009): within 6 days Gernt (2001): within 16 days Hoopmann (2011): 3 days (range 1-7) Khalil (2014): 48 h Klam (2005): 2-4 weeks Roberts (2000): 3 days O'Connor (2013): not stated Ong (1999): 10 days Simoes (2011): 1.6 +-0.14 weeks	(weight formula used: Hadlock) Caravello (1997): AC (20 mm) and EFW (25%) discordance (weight formula used: Hadlock) Chang (2006): EFW (15%, 20%, 25%, 30%) discordance (weight formula: Hadlock) Chittacharoen (2000): AC (>20 mm) and EFW (>15%) discordance (weight formula used: Hadlock) Danon (2008): EFW (25%) discordance (weight formula used: Hadlock) Diaz-Garcia (2010): EFW (15%, 20%, 25%) discordance (weight formulae used: Ong, Warsof, Shepard, Hadlock) Fox (2011): EFW (15%) discordance (weight formula used: Hadlock) Gandhi (2009): EFW (25%) discordance (weight	author. If more than one study was published on the same cohort with identical endpoints, the report including the most comprehensive information on the population was included. The quality of the studies was assessed using the revised tool for the quality assessment of diagnostic accuracy studies (QUADAS-II) where each item was scored as having high, low or unclear risk if there was insufficient information to make an accurate judgment. Statistical analysis Summary estimates of sensitivity and specificity for EFW discordance in predicting actual BW discordance were calculated using the HSROC model. For meta-analyses that included less than 4 studies, the DerSimonian–Laird random-effects model was used. Publication bias was assessed using Deek's funnel plot asymmetry test (for ≥10 studies).	sensitivity (95% CI) = 65.4% (57.9 - 72.3) specificity (95% CI) = 90.8% (87.1 - 93.5) <u>Diagnostic accuracy of EFW discordance (≥25%) to predict BW discordance (≥25%):</u> Overall accuracy (14 studies*, n=3980#): sensitivity (95% CI) = 57.7% (46.3 - 68.3) specificity (95% CI) = 95.2% (92.8 - 96.9) *calculations based on hierarchical summary-operating characteristics model <u>Diagnostic accuracy of AC discordance to predict BW discordance (≥15%, 3** studies, n=1090#):</u> sensitivity (95% CI) = 26.5% (21.5 - 32) specificity (95% CI) = 90.6% (88.5 - 92.4) <u>Diagnostic accuracy of AC discordance to predict BW discordance (≥20%, 2** studies, n=371#):</u> sensitivity (95% CI) = 32.3% (20.9 - 45.3)	(registered on PROSPERO). **Did the review authors explain their selection of the study designs for inclusion in the review?** No **Did the review authors use a comprehensive literature search strategy?** Yes **Did the review authors perform study selection in duplicate?** Yes **Did the review authors perform data extraction in duplicate?** Yes **Did the review authors provide a list of excluded studies and justify the exclusions?** Yes **Did the review authors describe the included studies in adequate detail?** Partial **Did the review authors use a satisfactory technique for assessing the RoB in individual studies that were**

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
biometry algorithms in the STORK multiple pregnancy cohort. Ultrasound Obstet Gynecol 2014; 44: 210–220. 2) Van de Waarsenburg MK, Hack KE, Rijpma RJ, Mulder EJ, Pistorius L, Derks JB. Ultrasonographic prediction of birth weight discordance in twin pregnancies. Prenat Diagn 2015; 35: 906–912. 3) Al-Obaidly S, Parrish J, Murphy KE, Glanc P, Maxwell C. The Accuracy of Estimating Fetal Weight and Inter-Twin Weight Discordance by Ultrasound in Twin Pregnancies in Women With Increased Body Mass Index. J Obstet Gynaecol Can 2015; 37:696–701. 4) O'Connor C, McAuliffe FM, Breathnach FM, Geary M, Daly S, Higgins JR, Dornan J, Morrison JJ, Burke G, Higgins S, Mooney E, Dicker P, Manning F, McParland P, Malone FD; Perinatal Ireland Research Consortium. Prediction of outcome in twin pregnancy with first and early second trimester ultrasound. J	Van de Waarsenburg (2015): 8 days (range 0-59) Van Mieghem (2009): within 2 weeks **Inclusion Criteria** Studies that reported: - The index test that was represented by different thresholds of sonographic EFW discordance (≥15%, ≥20%, ≥25%), calculated as ((larger EFW−smaller EFW)/larger EFW) ×100, or sonographic AC discordance, calculated as ((larger AC−smaller AC)/larger AC)*100. - The reference standard that was represented by the actual BW discordance, calculated as ((larger BW−smaller BW)/larger BW)*100, as measured immediately after birth. **Exclusion Criteria** N=17 studies with the reasons for their exclusions were reported in the supplementary material.	formula used: Hadlock) Gernt (2001): EFW (25%) discordance (weight formula used: Hadlock) Hoopmann (2011): EFW (15%, 20%, 25%) discordance (weight formulae used: Shepard, Hadlock, Ferrero) Khalil (2014): AC and EFW (10%, 15%, 20%, 25%) discordance (weight formulae used: Combs, Hadlock, Hsieh, Ott, Roberts, Shinozuka, Woo, Jordan, Merz, Shepard, Vintzileos, Warsof, Ferrero, Ong, Campbell, Higginbottom, Honarvar) Klam (2005): AC (0.93) and EFW (25%) discordance (weight formula used: Hadlock) Roberts (2000): AC (20 mm) and EFW (25%) discordance (weight formulae		specificity (95% CI) = 91.2% (87.5 - 94.1) **calculations based on DerSimonian-Laird random-effects model <u>Diagnostic accuracy of AC discordance to predict BW discordance (≥25%, 6*** studies, n=1609#):</u> sensitivity (95% CI) = 70.8% (51.1 - 84.9) specificity (95% CI) = 86.4% (62.1 - 96.1) ***hierarchical summary-operating characteristics model Note: according to the authors, due to the multitude of AC cut-offs reported among studies, it was not possible to perform a comprehensive data synthesis for each cut-off. # the number of participants included in meta-analysis was not reported, it was calculated by the NGA 2019 technical team.	**included in the review?** Yes (QUADAS-II) **Did the review authors report on the sources of funding for the studies included in the review?** No **If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?** Yes **If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?** No; The evidence from this review was downgraded by the NGA 2019 technical team for heterogeneity between the included studies regarding the US-to-birth interval and for poor reporting as the included studies do not report the number of live birth or stillbirths.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Matern Fetal Neonatal Med 2013; 26: 1030–1035. 5) Al Hassan A, Al Ghany HA. Estimation of Fetal Body Weight in Twins: A New Mathematical Model. Iraqi J Comm Med 2012; 1: 61–65. 6) Hoopmann M, Kagan KO, Yazdi B, Grischke EM, Abele H. Prediction of birth weight discordance in twin pregnancies by second- and third-trimester ultrasound. Fetal Diagn Ther 2011; 30: 29–34. 7) Simoes T, Julio C, Cordeiro A, Cohen A, Silva A, Blickstein I. Abdominal circumference ratio for the diagnosis of intertwin birth weight discordance. J Perinat Med 2011; 39: 43–46. 8) Fox NS, Saltzman DH, Schwartz R, Roman AS, Klauser CK, Rebarber A. Second-trimester estimated fetal weight and discordance in twin pregnancies: association with fetal growth restriction. J Ultrasound Med 2011; 30:1095–1101. 9) Diaz-Garcia C, Bernard JP, Ville Y, Salomon LJ.		used: Shepard, Hadlock) O'Connor (2013): AC (10%) and EFW (10%) discordance (weight formula used: Hadlock) Ong (1999): EFW (20%) discordance (weight formulae used: Cambell, Ong, Shepard, Hadlock) Simoes (2011): AC (>10%, >20%, >30%) and EFW (25%) discordance (weight formula used: Hadlock) Van de Waarsenburg (2015): AC (1.2-1.3) and EFW (20%) discordance (weight formula used: Hadlock) Van Mieghem (2009): EFW (25%) discordance (weight formula used: Hadlock)			**Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?** No **Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?** Yes **If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?** No **Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?** No

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Validity of sonographic prediction of fetal weight and weight discordance in twin pregnancies. Prenat Diagn 2010; 30:361–367. 10) Van Mieghem T, Deprest J, Klaritsch P, Gucciardo L, Done E, Verhaeghe J, Lewi L. Ultrasound prediction of intertwin birth weight discordance in monochorionic diamniotic twin pregnancies. Prenat Diagn 2009; 29: 240–244. 11) Gandhi M, Ferrara L, Belogolovkin V, Moshier E, Rebaber A. Effect of increased body mass index on the accuracy of estimated fetal weight by sonography in twins. J Ultrasound Med 2009; 28: 301–308. 12) Danon D, Melamed N, Bardin R, Meizner I. Accuracy of ultrasonographic fetal weight estimation in twin pregnancies. Obstet Gynecol 2008; 112: 759–764. 13) Chang YL, Chang TC, Chang SD, Cheng PJ, Chao AS, Hsieh PC, Soong YK. Sonographic prediction					Note: according to the authors, the major limitations of the studies were the different GAs at the time of US assessment, heterogeneity in the time interval between the last US and birth, and lack of stratification by chorionicity for most of the studies. The assessment of RoB was taken from Leombroni 2017 review, based on QUADAS-II: Al Hassan (2012): **RoB** Patient selection: low risk Index test: low risk Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Al-Obaidly (2015): **RoB** Patient selection: low risk Index test: low risk Reference standard: low risk Flow and timing: low risk

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
of significant intertwin birth weight discordance. Eur J Obstet Gynecol Reprod Biol 2006; 127: 35–40. 14) Klam SL, Rinfret D, Leduc L. Prediction of growth discordance in twins with the use of abdominal circumference ratios. Am J Obstet Gynecol 2005; 192: 247–251. 15) Gernt PR, Mauldin JG, Newman RB, Durkalski VL. Sonographic prediction of twin birth weight discordance. Obstet Gynecol 2001; 97: 53–56. 16) Roberts WE, Gnam EC 3rd, Magann EF, Martin JN Jr, Morrison JC. Labor and membrane rupture in twin gestation. Can they affect the ability to estimate fetal weight? J Reprod Med 2001; 46: 462–466. 17) Chittacharoen A, Leelapattana P, Rangsiprakarn R. Prediction of discordant twins by real-time ultrasonography combined with umbilical artery velocimetry. Ultrasound Obstet Gynecol 2000; 15: 118–121. 18) Ong S, Smith AP, Fitzmaurice A, Campbell D.					**Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Blickstein (1996): **RoB** Patient selection: low risk Index test: low risk Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Caravello (1997): **RoB** Patient selection: low risk Index test: low risk Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Chang (2006): **RoB** Patient selection: low risk Index test: low risk Reference standard: low risk Flow and timing: low risk

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Estimation of fetal weight in twins: a new mathematical model. Br J Obstet Gynaecol 1999; 106: 924–928. 19) Caravello JW, Chauhan SP, Morrison JC, Magann EF, Martin JN Jr, Devoe LD. Sonographic examination does not predict twin growth discordance accurately. Obstet Gynecol 1997; 89: 529–533. 20) Blickstein I, Manor M, Levi R, Goldchmit R. Is intertwin birth weight discordance predictable? Gynecol Obstet Invest 1996; 42: 105–108. **Source of funding** Not reported					**Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Chittacharoen (2000): **RoB** Patient selection: low risk Index test: low risk Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Danon (2008): **RoB** Patient selection: unclear risk Index test: low risk Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: unclear risk Index test: low risk Reference standard: low risk Diaz-Garcia (2010): **RoB** Patient selection: low risk Index test: low risk

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Fox (2011): **RoB** Patient selection: high risk Index test: low risk Reference standard: low risk Flow and timing: high risk **Applicability concerns** Patient selection: high risk Index test: low risk Reference standard: low risk Gandhi (2009): **RoB** Patient selection: low risk Index test: low risk Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Gernt (2001): **RoB** Patient selection: low risk Index test: low risk

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Hoopmann (2011): **RoB** Patient selection: low risk Index test: low risk Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Khalil (2014): **RoB** Patient selection: low risk Index test: low risk Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Klam (2005): **RoB** Patient selection: low risk Index test: low risk

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Roberts (2000): **RoB** Patient selection: low risk Index test: low risk Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk O'Connor (2013): **RoB** Patient selection: high risk Index test: low risk Reference standard: low risk Flow and timing: high risk **Applicability concerns** Patient selection: high risk Index test: low risk Reference standard: low risk Ong (1999): **RoB** Patient selection: low risk Index test: low risk

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Simoes (2011): **RoB** Patient selection: low risk Index test: low risk Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Van de Waarsenburg (2015): **RoB** Patient selection: low risk Index test: low risk Reference standard: low risk Flow and timing: unclear risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk Van Mieghem (2009): **RoB**

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Patient selection: low risk Index test: low risk Reference standard: low risk Flow and timing: low risk **Applicability concerns** Patient selection: low risk Index test: low risk Reference standard: low risk **Other information** None
Full citation Neves, A. R, Nunes, F, Branco, M, Almeida, M. D. C, Santos Silva, I., The role of ultrasound in the prediction of birth weight discordance in twin pregnancies: are we there yet?, Journal of Perinatal Medicine, 29, 29, 2017 **Ref Id** 795319 **Country/ies where the study was carried out** Portugal **Study type** Retrospective cohort study	**Sample size** N=176 twin pregnancies **Characteristics** 69.9% were dichorionic, 30.1% MC twins. Maternal age (median): 33 (IQR 18 - 46); the interval between the last US evaluation and birth (median): 2 weeks (IQR 0 - 3); GA at birth (median): 35 weeks (IQR 26 - 38); EFW discordance (median): 8.9% (IQR 0.04 - 52.4); BWD (median): 10.2% (IQR 0 - 54.8); BWD ≥20% was present in 21.6% of the pregnancies. **Inclusion Criteria** Twin pregnancies	**Tests** US **Index test** 1) EFW discordance (≥20%) based on Hadlock's formula (1985); 2) Amniotic fluid amount (defined as olygoamnios = the deepest vertical pocket of amniotic fluid inferior to 2 cm) Data for the analyses used were measured at the last ultrasound; the median interval between the last ultrasound evaluation and birth	**Methods** Data were extracted from electronic patient records. The participants' records were reviewed and maternal characteristics, pregnancy and neonatal outcomes were registered; antenatal US records were collected from a computerized database. US parameters BWD were expressed in % and calculated as the difference between the measure in the larger and smaller twins, divided by the measure in the larger twin. Significant EBW and BW discordances were defined as the difference	**Results** Diagnostic accuracy of EFW (≥20%) to predict intertwin weight discordance (>=20%): TP=20, FP=19, FN=18, TN=119 AUC: 84% (76 - 92) Diagnostic accuracy of EFW (≥20%) to predict intertwin weight discordance (≥20%) by chorionicity: DC twins: AUC: 85% (76 - 95) MC twins: AUC: 82% (68 - 96) Diagnostic accuracy of amniotic fluid (defined as olygoamnios = the deepest vertical pocket of amniotic fluid inferior	**Limitations** RoB was assessed using QUADAS-II **A. RoB** **Patient Sampling** Was a consecutive or random sample of patients enrolled? Unclear Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Unclear risk **B. Concerns regarding applicability:** Patient characteristics and setting Are there concerns that the included patients and

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Aim of the study To analyse the accuracy of US biometry in the prediction of BWD in twin gestations and evaluate the influence of chorionicity and FGR on US performance. **Study dates** Between 2008 and 2014 **Source of funding** There were no funding sources for this study.	**Exclusion Criteria** Those with selective feticide or birth before 24 weeks, monoamnionicity, FFTS, fetal malformations and interval between US and birth >3 weeks.	was 2 weeks (IQR 0 - 3). **Reference standard** Intertwin weight discordance (>=20%)	between the two fetuses ≥20%. Information regarding the frequency and duration of screening was not reported.	to 2 cm) to predict intertwin weight discordance (>=20%): TP=5, FP=4, FN=33, TN=134	setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standards likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Unclear concern **Other information** None
Full citation O'Connor, C, McAuliffe, F. M, Breathnach, F. M, Geary, M, Daly, S, Higgins, J. R,	**Sample size** N=260 twin pregnancies **Characteristics**	**Tests** **Index test** CRL discordance (threshold >20%)	**Methods** This is a secondary analysis of the ESPRiT study which was	**Results** Diagnostic accuracy of CRL discordance (>20%) to predict birth	**Limitations** Risk of bias was assessed using QUADAS-II **A. RoB**

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Dornan, J, Morrison, J. J, Burke, G, Higgins, S, Mooney, E, Dicker, P, Manning, F, McParland, P, Malone, F. D, Perinatal Ireland Research, Consortium, Prediction of outcome in twin pregnancy with first and early second trimester ultrasound, Journal of Maternal-Fetal & Neonatal Medicine, 26, 1030-5, 2013 **Ref Id** 794330 **Country/ies where the study was carried out** Ireland **Study type** Prospective cohort study **Aim of the study** To determine the ultrasound biometric parameters in the 1st and early 2nd trimester that can predict adverse pregnancy outcome. **Study dates**	n=14 pregnancies were complicated by FFTS. Maternal age (mean): 32.7 (range 14-37), GA at birth (mean): 36 weeks **Inclusion Criteria** All twin pregnancies presenting to the study centres between 11 and 22 completed weeks' gestation, with both fetuses alive at the time of prelabour CS or of onset of labour. **Exclusion Criteria** Monoamnionicity, a major structural abnormality in either twin or fetal aneuploidy.	measured in the 1st trimester (11^{+0} to 14^{+0} weeks) **Reference standard** Intertwin weight discordance (threshold ≥18%).	a multicentre prospective study conducted at 8 academic perinatal centres. US examinations were made at enrolment (mean 16 weeks (range 13 - 19)) and again at 18-20 weeks for those enrolled prior to 18 weeks. CRL was recorded for each fetus in the 1st trimester. For MC twins, 2-weekly US surveillance was initiated at 16 weeks' gestation.	<u>weight discordance (≥18%):</u> TP=1, FP=1, FN=47, TN=211	**Patient Sampling** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Low risk **B. Concerns regarding applicability:** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? Unclear risk

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Between May 2007 and October 2009 **Source of funding** Study was supported by grant from Health Research Board of Ireland (Grant Code IMA/2005/3).					**B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standards likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB**

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low concern **Other information** None
Full citation Rodis, J. F., Vintzileos, A. M., Campbell, W. A., Nochimson, D. J., Intrauterine fetal growth in discordant twin gestations, J Ultrasound MedJournal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine, 9, 443-8, 1990 **Ref Id** 807903 **Country/ies where the study was carried out** USA	**Sample size** N=25 women with twin pregnancy who gave biwth within 7 days of the last USS. **Characteristics** Details of ethnicity or chorionicity not reported. **Inclusion Criteria** All women with twin pregnancies between 1985 and 1987 at the University of Connecticut Health Centre underwent serial USS if there was birthweight discordance ≥20%; confirmed dating and absence of major congenital	**Tests** **Screening tests** 1) EFW difference ≥20% using BPD and AC measurements 2) EFW difference ≥20% using FL and AC measurements EFW was calculated for each fetus using two formulae: one based on BPD and AC (Shepard's formula) and the other based on FL and AC (Hadlock's formula) **Reference test**	**Methods** 156 ultrasound examinations were performed and the mean discordancy was 27%. Details of equipment and methods reported. Information regarding the frequency and duration of screening was not reported.	**Results** Efficacy of predicting BWD ≥20% by EFWD ≥20%: when EFW calculated using BPD, AC (Shepard's formula): TP=12, FP=3, FN=2, TN=12 when EFW calculated using FL and AC (Hadlock's formula): TP=13, FP=4, FN=3, TN=25	**Limitations** RoB was assessed using QUADAS-II **Patient Sampling** **A. RoB** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Unclear: exclusion criteria were not reported for this study Could the selection of patients have introduced bias? Unclear risk **B. Concerns regarding applicability**

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Study type Prospective cohort study **Aim of the study** To assess longitudinal growth of twins who are ultimately discordant at birth and to see how they differ from the concordant group and to assess the accuracy of both Shepard's formula (using BPD and AC) and Hadlock's formula (employing FL and AC). **Study dates** 1985 to 1987 **Source of funding** Not reported.	anomalies in one or both fetuses. **Exclusion Criteria** None reported.	Intertwin BWD ≥20%			Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low risk **Other information** Study information transcribed from CG129

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Multiple Pregnancy appendix H: Evidence tables. The RoB assessment was conducted by the NGA technical team
Full citation Sayegh, S. K., Warsof, S. L., Ultrasonic prediction of discordant growth in twin pregnancies, Fetal Diagnosis & Therapy, 8, 241-6, 1993 **Ref Id** 758449 **Country/ies where the study was carried out** USA **Study type** Prospective cohort study **Aim of the study** To examine the ability of ultrasound to accurately predict discordant growth in twin pregnancies and to define the percent intertwin EFWD that best correlated with the previously	**Sample size** N=78 women with twin pregnancies (including one with FFTS). **Characteristics** When more than 1 scan was performed the most recent prior to birth was used and this varied from 1 day to 6 weeks and no standard interval was required to be included in the study. Details of chorionicity and ethnicity not reported. **Inclusion Criteria** All consecutive twin pregnancies at Sentara Norfolk General Hospital between 1 July 1984 and 20 June 1987 referred for targeted USS to the Division of MFM at Eastern Virginia Medical School. **Exclusion Criteria** Accurate EFW NC.	**Tests** **Screening tests** Intertwin EFW difference of ≥15%, ≥20% and ≥25%. Calculation of EFW was based on BPD and AC, according to Shepard's formula (1982). **Reference test** Intertwin BWD of ≥25%	**Methods** Only data from scans performed at more than 23 weeks of pregnancy, when EFW could be calculated, were used in the analysis. Scans were reviewed by the authors without knowledge of birthweight outcomes. Details of equipment and methods reported. Information regarding the frequency and duration of screening was not reported.	**Results** Prediction of BWD ≥25% using EFWD ≥15%: TP=NR, FP=NR, FN=NR, TN=NR sensitivity: 71% specificity: 88% Prediction of BWD ≥25% using EFWD ≥20%: TP=NR, FP=NR, FN=NR, TN=NR sensitivity: 74% specificity: 90% Prediction of BWD ≥25% using EFWD ≥25%: TP=10, FP=5, FN=3, TN=60	**Limitations** The study included one twin pregnancy with FFTS. RoB was assessed using QUADAS-II **Patient Sampling** **A. RoB** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Low risk **B. Concerns regarding applicability** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Unclear concern: one

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
established neonatal outcome. **Study dates** July 1984 to June 1987 **Source of funding** Not reported.					pregnancy with FFTS as included in the analyses **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? Low risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low **Other information** Study information transcribed from CG129 Multiple Pregnancy appendix H: Evidence tables.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					The RoB assessment was conducted by the NGA technical team
Full citation Shah, Y. G., Sherer, D. M., Gragg, L. A., Casaceli, C. J., Woods, J. R., Jr., Diagnostic accuracy of different ultrasonographic growth parameters in predicting discordancy in twin gestation: a different approach, American Journal of Perinatology, 11, 199-204, 1994 **Ref Id** 758456 **Country/ies where the study was carried out** USA **Study type** Retrospective cohort study **Aim of the study** To examine the predictability of intrapair percentage differences of ultrasonic fetal biometric	**Sample size** N=90 twin pregnancies, included in the analysis max=85 and min=54. **Characteristics** Details of ethnicity and chorionicity not reported. **Inclusion Criteria** All women with twin pregnancies that underwent USS of both fetuses within 7 days of a live twin birth in the perinatal US unit, Strong Memorial Hospital, New York between 1 January 1983 and 31 May 1988, and in whom measurements of BPD, HC, AC, FL, and EFW were obtained. **Exclusion Criteria** Maternal gestational or type 1 diabetes; fetal anomalies and congenital toxoplasmosis, rubella, cytomegalovirus, herpes complex (TORCH) infection.	**Tests** **Screening tests** Intrapair differences in: 1) BPD 2) HC 3) AC 4) FL 5) EFW ≥20% EFW was computed by the method of Warsof et al. (1977) using FL and AC **Reference test** Intertwin BWD ≥20%	**Methods** Intrapair difference of 5% and 10% for all biometric measurements (BPD, HC, AC, and FL) were considered to be critical values for predicting discordancy and were compared with BW. Details of techniques and equipment reported. Information regarding the frequency and duration of screening was not reported.	**Results** <u>Prediction of BWD ≥20% using US measurements with intrapair difference >5%:</u> BPD: TP=8, FP=19, FN=6, TN=31 HC: TP=7, FP=11, FN=4, TN=32 AC: TP=16, FP=27, FN=2, TN=40 FL: TP=8, FP=13, FN=9, TN=49 Prediction of BWD ≥20% using US measurements with intrapair difference >10%: BPD: TP=5, FP=3, FN=9, TN=47 HC: TP=2, FP=3, FN=9, TN=40 AC:	**Limitations** RoB was assessed using QUADAS-II **Patient Sampling** **A. RoB** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Low risk **B. Concerns regarding applicability** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
parameters in detecting twin discordancy. **Study dates** January 1983 – May 1988 **Source of funding** Not reported.				TP=11, FP=7, FN=7, TN=60 FL: TP=3, FP=4, FN=14, TN=58 <u>Prediction of birthweight discordance ≥20% using EFW difference ≥20%:</u> TP=10, FP=5, FN=4, TN=43	of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? No Could the patient flow have introduced bias? Unclear risk **Other information** Study information transcribed from CG129 Multiple Pregnancy appendix H: Evidence tables. The RoB assessment was conducted by the NGA technical team.

Full citation Shahshahan,Z, Hashemi,M., Crown-rump length discordance in twins in the first trimester and its correlation with perinatal complications, Journal of Research in Medical Sciences, 16, 1224-1227, 2011 **Ref Id** 795528 **Country/ies where the study was carried out** Iran **Study type** To be decided **Aim of the study** To evaluate discordance in CRL in the 1st trimester and its correlation with perinatal complications. **Study dates** Not reported **Source of funding** Not reported	**Sample size** N=118 women with twin pregnancy **Characteristics** Maternal age (mean, SD): 28.4 years (4.6); gestational age at birth (mean): 33.9 weeks (range 28 - 38); CRL discrepancy was normal (<11%) in n=96 (81%), birth weight discordance was normal (<20%) in n=103 (87%); Mean CRL discrepancy (SD): 6.5% (5.8), mean birth weight difference (SD): 7.5 (7.7). **Inclusion Criteria** Women with twin pregnancy in the 1st trimester **Exclusion Criteria** MC twins and women who underwent 1st or 2nd trimester pregnancy termination.	**Tests** Ultrasound **Index test** Discrepancy in CRL (threshold >11%) measured at 7 to 14 weeks' gestation. **Reference standard** Intertwin weight discordance (threshold >20%)	**Methods** The value of CRL discrepancy was calculated as the difference between CRL in twins in the 1st trimester divided by the larger CRL. The difference >11% was considered as abnormal CRL discordance. Weight difference was calculated as the difference in birth weight between the twins divided by the birth weight of the larger twin. A difference >20% was considered as abnormal birth weight discordance. Information regarding the frequency and duration of screening was not reported.	**Results** Diagnostic accuracy of CRL discordance (>11%) to predict SGA (defined as intertwin weight discordance >20%): TP=9, FP=13, FN=6, TN=90	**Limitations** RoB was assessed using QUADAS-II **A. RoB** **Patient Sampling** Was a consecutive or random sample of patients enrolled? Unclear Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes (n=2 women who underwent 1st or 2nd trimester pregnancy termination were excluded from the analysis) Could the selection of patients have introduced bias? Unclear risk **B. Concerns regarding applicability:** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Unclear concern (CRL was measured earlier than in other studies; i.e. at 7 to 14 weeks' gestation.) **Index Test** **A. RoB**

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standards likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					B. Concerns regarding applicability Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low concern **Other information** None

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Full citation Sklar, C, Yaskina, M, Ross, S, Naud, K., Accuracy of Prenatal Ultrasound in Detecting Growth Abnormalities in Triplets: A Retrospective Cohort Study, Twin Research & Human Genetics: the Official Journal of the International Society for Twin Studies, 20, 84-89, 2017 **Ref Id** 794336 **Country/ies where the study was carried out** Canada **Study type** Retrospective cohort study **Aim of the study** To ascertain the sensitivity, specificity, positive, and negative predictive values of modern tertiary level prenatal ultrasounds to predict growth abnormalities (FGR, SGA, growth discordance) in triplet pregnancies.	**Sample size** N=78 triplet pregnancies **Characteristics** Chorionicity: MC n=6 (7.7%), DC n=27 (34.6%), TC n=45 (57.7%). maternal age (median): 31 years; FGR detected n=15, SGA present n=15, SGA absent n=0, inconclusive n=0; FGR not detected n=63, SGA present n=12, SGA absent n=51, inconclusive n=0; pregnancy with an EFW discordance >=25% n=12 (15.4%) **Inclusion Criteria** Triplet pregnancies>18 weeks were included when documented on ultrasound; prenatal ultrasounds were performed at the Royal Alexandra Perinatal Clinic, and all triplets were born at the Royal Alexandra Hospital between the study dates. **Exclusion Criteria** Those cases where birth occurred <23^{+0} weeks' gestation, if spontaneous reduction or	**Tests** Ultrasound **Index test** 1) FGR defined as EFW <10th percentile for GA using for reference the Canadian Perinatal Surveillance System singleton growth curves (Kramer et al., 2001) 2) EFW discordance (>25%) which was calculated using Hadlock's formula, based on HC, AC, femur length (Hadlock et al., 1985) Data for the analyses used were measured closest to date of birth (median interval between last US and birth was 8 days (IQR 0 - 21); median 30.9 weeks of gestational age). **Reference standard**	**Methods** All triplet pregnancies were identified using medical coding in the Alberta Perinatal Health Program Database. The final US before birth was performed at median 30.9 weeks' gestation, with a median interval between last US and birth of 8 days, range between 0 and 21 days. EFW discordance (%) was defined as (Largest triplet EFW - Smallest triplet EFW)/ (Largest triplet EFW)*100. For each set of newborn triplet, ABW discordance (%) was defined as (Largest triplet ABW – Smallest triplet ABW)/ (Largest triplet ABW) × 100. Information regarding the frequency and duration of screening was not reported.	**Results** Diagnostic accuracy of EFW <10th percentile for gestational age to predict SGA (defined as actual birth weight <10th percentile for gestational age using for reference the Canadian Perinatal Surveillance System singleton growth curves (Kramer et al., 2001)): TP=15, FP=0, FN=12, TN=51 sensitivity: 55.6% (35.3 - 74.5) specificity: 100% (93 - 100) Diagnostic accuracy of EFW discordance (>25%) to predict birth-weight discordance of >25%: TP=8, FP=4, FN=2, TN=64 sensitivity: 80% (44.4 - 97.5) specificity: 94.1% (85.6 - 98.4)	**Limitations** RoB was assessed using QUADAS-II **A. RoB** **Patient Sampling** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Low risk **B. Concerns regarding applicability:** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Study dates From January 2004 to May 2015. **Source of funding** This research has been funded by generous supporters of the Lois Hole Hospital for Women through the Women and Children's Health Research Institute.	multi-fetal reduction of a higher order multiple pregnancy into a triplet pregnancy occurred, if fetal reduction (spontaneous or not) of a triplet pregnancy into a twin/singleton pregnancy occurred, if the most recent US was performed more than 21 days before the birth or if the triplet pregnancy had no prenatal care or prenatal US.	1) SGA defined as actual birth weight <10th percentile for gestational age using for reference the Canadian Perinatal Surveillance System singleton growth curves (Kramer et al., 2001) 2) Inter-triplet weight discordance >25%			Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standards likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Flow and Timing **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Unclear concern **Other information** None
Full citation Storlazzi, E., Vintzileos, A. M., Campbell, W. A., Nochimson, D. J., Weinbaum, P. J., Ultrasonic diagnosis of discordant fetal growth in twin gestations, Obstetrics & Gynecology, 69, 363-7, 1987 **Ref Id** 758477 **Country/ies where the study was carried out** USA	**Sample size** N=43 consecutive twin pregnancies with last USS within 2 weeks of birth. **Characteristics** An attempt was made to measure BPD, AC and FL in both fetuses. Babies were weighed within 24 hours of birth. Details of chorionicity and ethnicity not reported. **Inclusion Criteria** Consecutive women with twin pregnancy who gave	**Tests** **Screening tests** Intertwin EFW difference ≥20%. EFW calculation was based on BPD and AC, using the formula of Shepard et al. (1982) or on AC and FL using the formula of Hadlock (1984), when BPD was unobtainable. **Reference test** Intertwin BWD ≥20%	**Methods** All patients had an US examination upon admission to confirm the presence of twin gestation. The US evaluations were repeated every two weeks until birth. Only the results of the last scan were considered for analysis. Cut-offs used for discordancy were as follows: BPD (6mm), AC (20mm), FL (5mm). Details of methods and equipment reported.	**Results** Prediction of BWD ≥20% by EFWD ≥20%: TP=8, FP=2, FN=2, TN=26	**Limitations** RoB was assessed using QUADAS-II **Patient Sampling** **A. RoB** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Low risk **B. Concerns regarding applicability**

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Study type Retrospective cohort study **Aim of the study** To investigate the value of intrapair difference in BPD, AC, FL and EFW in predicting discordant fetal growth. **Study dates** Not reported. **Source of funding** Not reported.	birth at the Connecticut Health Centre, USA. **Exclusion Criteria** Congenital anomalies.				Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB** Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Yes Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low risk **Other information** Study information transcribed from CG129

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Multiple Pregnancy appendix H: Evidence tables. The RoB assessment was conducted by the NGA technical team.
Full citation van de Waarsenburg, M. K, Hack, K. E, Rijpma, R. J, Mulder, E. J, Pistorius, L, Derks, J. B., Ultrasonographic prediction of birth weight discordance in twin pregnancies, Prenatal Diagnosis, 35, 906-12, 2015 **Ref Id** 794340 **Country/ies where the study was carried out** The Netherlands **Study type** Retrospective cohort study **Aim of the study** To assess the accuracy of the various sonographical estimations of size or weight discordance at	**Sample size** N=281 twin pregnancies **Characteristics** n=206 DC, n=75 MC twins maternal age (mean): 32.9 years; interval between the last US and birth (median): 8 days (IQR 0 - 59); gestational age at birth (median): 35 weeks (IQR 23+3 - 41+0); n=42 twin pairs (15%) showed a BWD of ≥20%. **Inclusion Criteria** Twin pregnancy **Exclusion Criteria** Monoamniotic twin pregnancies, pregnancies with a selective feticide, complicated by congenital disorders or intrauterine fetal death, a GA at birth of less than 22 weeks or fetuses with a birth weight less than 500 g and cases in	**Tests** US **Index test** 1) CRL discordance (thresholds ≥11% and ≥20%) measured in the 1st trimester; 2) IUGR (at least 1 twin) defined as an EFW <10th percentile based on the last ultrasound before birth (median interval between the last US and birth was 8 days (IQR 0 - 59); 3) Amniotic fluid amount (oligohydramnios defined as the deepest vertical pocket of amniotic fluid of less than 2 cm), not	**Methods** Twin pregnancies were monitored according to a protocol based on chorionicity which included a 1st trimester determination of chorionicity, detailed anomaly scan at 20 weeks' gestation and US assessment of growth and amniotic fluid volume at 20, 26, 30, 32, 34 and 36 weeks for DC twin gestations and fortnightly from 14 weeks onwards. Oligohydramnios was defined as the deepest vertical pocket of amniotic fluid of less than 2 cm. The EFW was calculated by a formula of Hadlock (1991) based on the HC, AC and femur length measurements. CRL difference was calculated as the difference between the	**Results** Diagnostic accuracy of CRL discordance (≥11%) to predict birth weight discordance (≥20%): TP=4, FP=11, FN=38, TN=228 Diagnostic accuracy of CRL discordance (≥20%) to predict birth weight discordance (≥20%): TP=1, FP=2, FN=41, TN=237 Diagnostic accuracy of IUGR (at least 1 twin, defined as EFW<10th percentile based on the last US before birth) to predict BWD (≥20%): TP=29, FP=48, FN=13, TN=191 Diagnostic accuracy of amniotic fluid amount (defined as oligohydramnios = the deepest vertical pocket	**Limitations** RoB was assessed using QUADAS-II **A. RoB** **Patient Sampling** Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes Could the selection of patients have introduced bias? Low risk **B. Concerns regarding applicability:** Patient characteristics and setting Are there concerns that the included patients and setting do not match the review question? Low concern **Index Test** **A. RoB**

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
various gestational ages in the prediction of severe weight discordance at birth. **Study dates** Between 2008 and 2011 **Source of funding** Not reported	which it was impossible to calculate a BWD.	reported when it was measured. **Reference standard** inter-twin weight discordance >=20%	size or weight of the larger and smaller twins, divided by the size or weight of the larger twin. Severe size or weight discordance was defined as the difference in CRL ≥20%. A CRL discordance of 11% was also considered. Weight discordance at birth was calculated as the intertwin difference in BW expressed as a % of the heaviest twin; a value of >=20% was defined as severe BW discordance. IUGR was defined as an EFW <10th percentile based on the last US before birth.	of amniotic fluid of less than 2 cm) to predict BWD (≥20%): TP=7, FP=35, FN=35, TN=204	Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes Could the conduct or interpretation of the index test have introduced bias? Low risk **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concern **Reference Standard** **A. RoB** Is the reference standards likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index tests? Unclear Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk

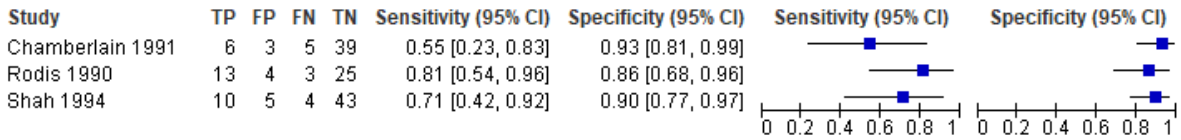
Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					B. Concerns regarding applicability Are there concerns that the target condition as defined by the reference standard does not match the question? Low concern **Flow and Timing** **A. RoB** Was there an appropriate interval between index test and reference standard? Yes. Did all patients receive the same reference standard? Yes Were all patients included in the analysis? Yes Could the patient flow have introduced bias? Low concern for CRL, unclear for IUGR, high for amniotic fluid **Other information** None

ABW: actual birth weight; AC: abdominal circumference; AUC: area under the curve; BPD: biparietal diameter; BW: birth weight; BWD: birth weight discordance; CG: clinical guidelines; CI: confidence interval; CRL: crown rump length; CRL-S: crown rump length smaller twin; CRL-L: crown rump length larger twin; DC: dichorionic; EBWD: estimated birth weight discordance; EDD: estimated due date; EFW: estimated fetal weight; ESPRiT: Evaluation of Sonographic Predictors of Restricted growth in Twins; FFTS: feto-fetal transfusion syndrome; FGR: fetal growth rate; FL: femur length; FN: false negative; FP: false positive; GA: gestational age HC: head circumference; HSROC: hierarchical summary receiver operating characteristic; IQR: interquartile range; IUGR: intrauterine growth restriction; MC: monochorionic; MFM: Maternal Fetal Medicine; NC: not calculable; NT: nuchal translucency; RoB: risk of bias; SD: standard deviation; SFH: symphysis-fundal height; sGA: small gestational age; STORK: Southwest Thames region of London Obstetric Research Collaborative; TAPS: twin anemia-polycythemia sequence; TC: trichorionic; TCD: transverse cerebellar diameter; TN: true negative; TP: true positive; TTTS: twin to twin transfusion syndrome; US: ultrasound; USS: ultrasound scan

Appendix E – Forest plots and receiver operating characteristic curves

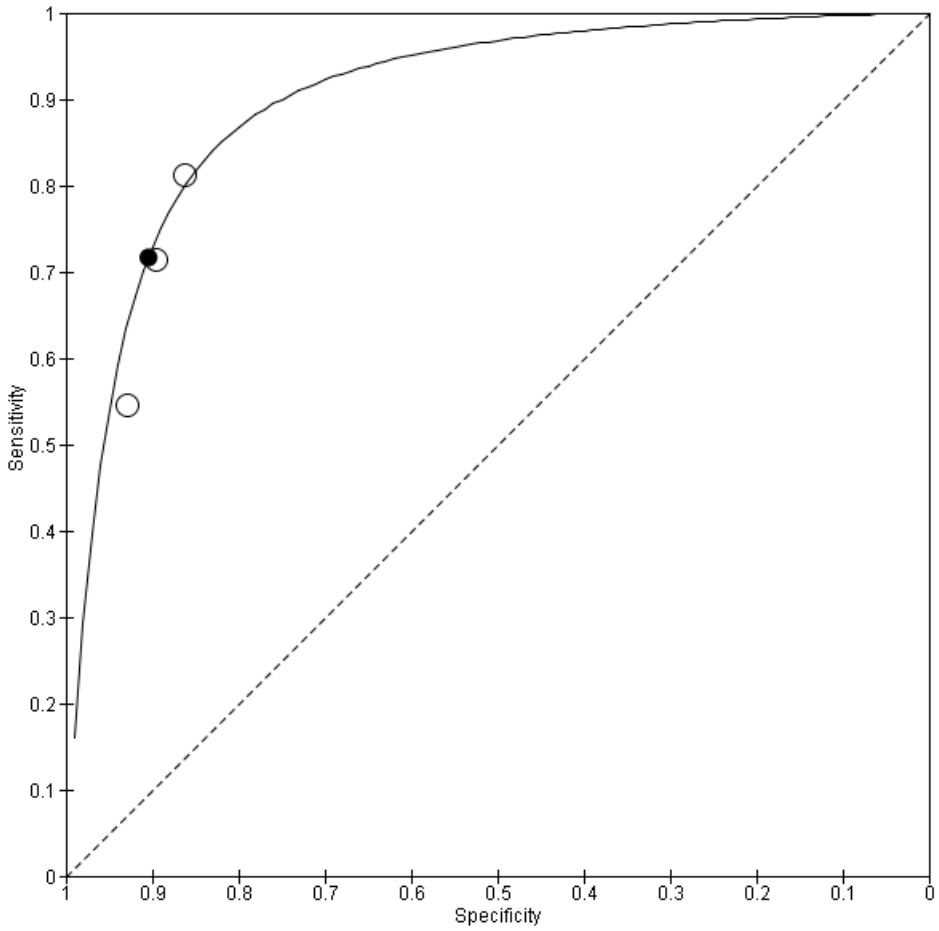
Forest plots and receiver operating characteristic (ROC) curves for review question: What is the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy?

Figure 2: Forest plot for estimated fetal weight discordancy ≥20% in 2nd trimester (estimated fetal weight based on abdominal circumference and femur length)



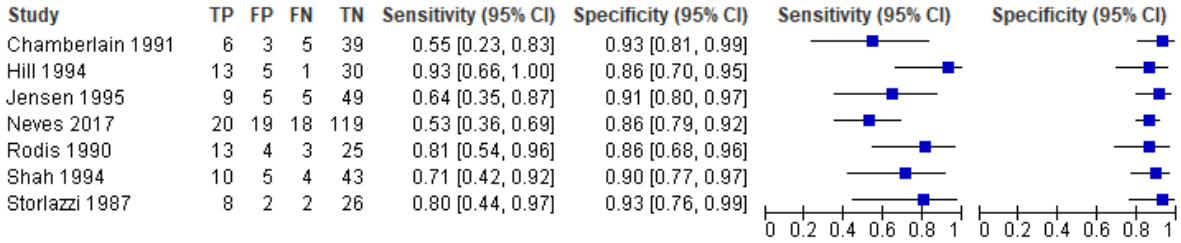
Sensitivity (95%CI): 0.70 (0.34 to 0.93); specificity (95%CI): 0.89 (0.69 to 0.98)

Figure 3: ROC curve for estimated fetal weight discordancy ≥20% in 2nd trimester (estimated fetal weight based on abdominal circumference and femur length)



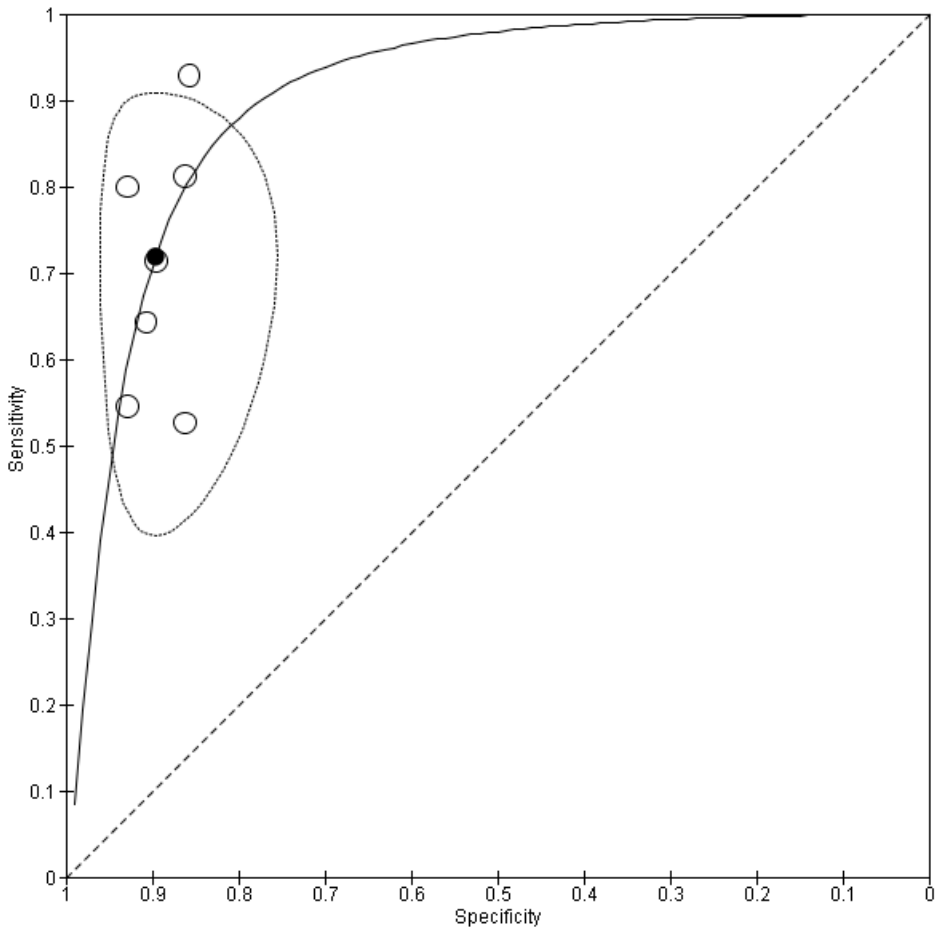
ROC: receiver operating characteristic curve which represents the estimates from different studies and an overall estimate

Figure 4: Forest plot for estimated fetal weight discordancy ≥20% in 2nd trimester (overall)



Sensitivity (95%CI): 0.71 (0.54 to 0.85); specificity (95%CI): 0.89 (0.83 to 0.94)

Figure 5: ROC curve for estimated fetal weight discordancy ≥20% in 2nd trimester (overall)



ROC: receiver operating characteristic curve which represents the estimates from different studies and an overall estimate

Appendix F – GRADE tables

GRADE profile for review question: What is the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy?

Table 4: Clinical evidence profile for screening to identify a small-for-gestational-age baby or intertwin birth weight discordancy in twin pregnancy in first trimester (11^{+0} to 13^{+6} weeks' gestation)

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Quality of the evidence (GRADE)	Importance
SGA defined as birth weight <5th centile											
CRL discordancy (continuous) - overall for DC and MC twins	1	2155	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	-	-	0.56 (0.53 to 0.59)	⊕⊕⊕⊖ MODERATE	IMPORTANT
CRL discordancy (continuous) - for MC twins only	1	420	Serious[1]	No serious inconsistency	No serious indirectness	Very serious[2]	-	-	0.57 (0.49 to 0.66)	⊕⊖⊖⊖ VERY LOW	IMPORTANT
SGA defined as birth weight <10th percentile											
CRL discordancy >15% - overall for DC and MC twins	1	46	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.10 (0 to 0.45)	0.94 (0.81 to 0.99)	-	⊕⊕⊕⊖ MODERATE	CRITICAL
CRL discordancy >15% - for DC twins only	1	35	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.13 (0 to 0.53)	0.96 (0.81 to 1)	-	⊕⊕⊕⊖ MODERATE	CRITICAL

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Quality of the evidence (GRADE)	Importance
CRL discordancy >15% - for MC twins only	1	11	Serious[1]	No serious inconsistency	No serious indirectness	Serious[3]	0 (0 to 0.84)	0.89 (0.52 to 1)	-	⨁⨁⨀⨀ LOW	CRITICAL
Intertwin birth weight discordancy >15%											
CRL discordancy >15% - overall for DC and MC twins	1	46	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.13 (0.02 to 0.40)	0.97 (0.83 to 1)	-	⨁⨁⨁⨀ MODERATE	CRITICAL
CRL discordancy >15% - for DC twins only	1	35	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.08 (0 to 0.38)	0.96 (0.78 to 1)	-	⨁⨁⨁⨀ MODERATE	CRITICAL
CRL discordancy >15% - for MC twins only	1	11	Serious[1]	No serious inconsistency	No serious indirectness	Very serious[3]	0.33 (0.01 to 0.91)	1 (0.63 to 1)	-	⨁⨀⨀⨀ VERY LOW	CRITICAL
Intertwin birth weight discordancy ≥15%											
CRL discordancy (continuous)	1	660	Very serious[4]	No serious inconsistency	No serious indirectness	Serious[5]	-	-	0.59 (0.54 to 0.65)	⨁⨀⨀⨀ VERY LOW	IMPPORTANT
Intertwin birth weight discordancy ≥18%											
CRL discordancy >20%	1	260	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.02 (0 to 0.11)	1 (0.97 to 1)	-	⨁⨁⨁⨀ MODERATE	CRITICAL

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Quality of the evidence (GRADE)	Importance
Intertwin birth weight discordancy ≥20%											
CRL discordancy ≥5% - DC twins only	1	108	Very serious[6]	No serious inconsistency	No serious indirectness	Serious[2]	0.59 (0.36 to 0.79)	0.60 (0.48 to 0.72)	-	⨁⊖⊖⊖ VERY LOW	CRITICAL
CRL discordancy (continuous) - DC twins only	1	108	Very serious[6]	No serious inconsistency	No serious indirectness	Very serious[7]	-	-	0.55 (0.44 to 0.66)	⨁⊖⊖⊖ VERY LOW	CRITICAL
CRL discordancy ≥10% - overall for DC and MC twins	1	1993	Very serious[8]	No serious inconsistency	No serious indirectness	No serious imprecision	0.24 (0.19 to 0.31)	0.87 (0.85 to 0.88)	-	⨁⨁⊖⊖ LOW	CRITICAL
CRL discordancy ≥10% - for DC twins only	1	1733	Very serious[8]	No serious inconsistency	No serious indirectness	No serious imprecision	0.24 (0.17 to 0.31)	0.86 (0.85 to 0.88)	-	⨁⨁⊖⊖ LOW	CRITICAL
CRL discordancy ≥10% - for MC twins only	1	260	Very serious[8]	No serious inconsistency	No serious indirectness	No serious imprecision	0.28 (0.14 to 0.47)	0.89 (0.84 to 0.92)	-	⨁⨁⊖⊖ LOW	CRITICAL
CRL discordancy ≥11%	1	281	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.10 (0.03 to 0.23)	0.95 (0.92 to 0.98)	-	⨁⨁⨁⊖ MODERATE	CRITICAL
CRL discordancy ≥20%	1	281	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.02 (0 to 0.13)	0.99 (0.97 to 1)	-	⨁⨁⨁⊖ MODERATE	CRITICAL

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Quality of the evidence (GRADE)	Importance
Intertwin birth weight discordancy >20%											
CRL discordancy ≥11%	1	118	Very serious[9]	No serious inconsistency	No serious indirectness	Serious[3]	0.60 (0.32 to 0.84)	0.87 (0.79 to 0.93)	-	⨁⊖⊖⊖ VERY LOW	CRITICAL
Intertwin birth weight discordancy ≥25%											
CRL discordancy (continuous)	1	660	Very serious[4]	No serious inconsistency	No serious indirectness	Serious[5]	-	-	0.63 (0.55 to 0.70)	⨁⊖⊖⊖ VERY LOW	IMPORTANT

AUC: area under the curve (the curve represents different cut-off points); CI: confidence interval; CRL: crown-rump length
1 Unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test
2 The quality of the evidence was downgraded by 2 levels because the 95% CI crosses 2 default cut-offs (0.50 and 0.61)
3 The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results are judged to be very seriously imprecise.
4 No description of the population; unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test
5 The quality of the evidence was downgraded by 1 level because the 95% CI crosses default 1 cut-off (0.61)
6 Unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test; unclear if a consecutive or random sample of participants was enrolled; unclear if the included participants match the review question as CRL was measured at 10 to 14 weeks' gestation; index test threshold was not pre-specified
7 The quality of the evidence was downgraded by 2 levels because the 95% CI crosses 2 default cut-offs (0.50 and 0.61)
8 Unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test; unclear if a consecutive or random sample of participants was enrolled
9 Unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test; unclear if a consecutive or random sample of participants was enrolled; unclear if the included participants match the review question as CRL was measured at 7 to 14 weeks' gestation

Table 5: Clinical evidence profile for diagnostic monitoring to identify intertwin birth weight discordancy ≥15% or more using fetal biometry discordancy in twin pregnancy in second trimester

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Quality of the evidence (GRADE)	Importance
Intertwin birth weight discordancy ≥15%											
Overall AC discordancy (US-to-birth interval within 2 weeks in 2 studies, not reported in 1 study)	3	1090 [1]	Very serious[2]	Not possible to asses as no data reported	No serious indirectness	No serious imprecision	0.27 (0.22 to 0.32)	0.91 (0.89 to 0.92)	-	⨁⨁⊖⊖ LOW	CRITICAL
Intertwin birth weight discordancy ≥20%											
Overall AC discordancy (US-to-birth interval 8 days (range 0-59) or within 3 weeks)	2	371 [1]	Very serious[2]	Not possible to assess as no data reported	No serious indirectness	No serious imprecision	0.32 (0.21 to 0.45)	0.91 (0.88 to 0.94)	-	⨁⨁⊖⊖ LOW	CRITICAL
HC discordancy >5% (US-to-birth interval within 7 days)	1	54	Very serious[3]	No serious inconsistency	No serious indirectness	Serious[4]	0.64 (0.31 to 0.89)	0.74 (0.59 to 0.86)	-	⨁⊖⊖⊖ VERY LOW	CRITICAL
HC discordancy >10% (US-to-birth interval within 7 days)	1	54	Very serious[3]	No serious inconsistency	No serious indirectness	No serious imprecision	0.18 (0.02 to 0.52)	0.93 (0.81 to 0.99)	-	⨁⨁⊖⊖ LOW	CRITICAL

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Quality of the evidence (GRADE)	Importance
AC discordancy >5% (US-to-birth interval - within 7 days)	1	85	Very serious[3]	No serious inconsistency	No serious indirectness	Very serious	0.89 (0.65 to 0.99)	0.60 (0.47 to 0.72)	-	⨁⊖⊖⊖ VERY LOW	CRITICAL
AC discordancy >10% (US-to-birth interval within 7 days)	1	85	Very serious[3]	No serious inconsistency	No serious indirectness	Serious[4]	0.61 (0.36 to 0.83)	0.90 (0.80 to 0.96)	-	⨁⊖⊖⊖ VERY LOW	CRITICAL
FL discordancy >5% (US-to-birth interval within 7 days)	1	79	Very serious[3]	No serious inconsistency	No serious indirectness	No serious imprecision	0.47 (0.23 to 0.72)	0.79 (0.67 to 0.88)	-	⨁⨁⊖⊖ LOW	CRITICAL
FL discordancy >10% (US-to-birth interval within 7 days)	1	79	Very serious[3]	No serious inconsistency	No serious indirectness	No serious imprecision	0.18 (0.04 to 0.43)	0.94 (0.84 to 0.98)	-	⨁⨁⊖⊖ LOW	CRITICAL
BPD discordancy >5% (US-to-birth interval within 7 days)	1	64	Very serious[3]	No serious inconsistency	No serious indirectness	Serious[4]	0.57 (0.29 to 0.82)	0.62 (0.47 to 0.75)	-	⨁⊖⊖⊖ VERY LOW	CRITICAL
BPD discordancy >10% (US-to-birth interval within 7 days)	1	64	Very serious[3]	No serious inconsistency	No serious indirectness	No serious imprecision	0.36 (0.13 to 0.65)	0.94 (0.83 to 0.99)	-	⨁⨁⊖⊖ LOW	CRITICAL
Intertwin birth weight discordancy ≥25%											

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Quality of the evidence (GRADE)	Importance
Overall AC discordancy (US-to-birth interval 3 days, 1.6±0.14 weeks, 2-4 weeks or within 2 or 3 weeks)	5	1609 [1]	Very serious[1]	Not possible to asses as no data reported	No serious indirectness	Serious[4]	0.71 (0.51 to 0.85)	0.86 (0.62 to 0.96)	-	⨁⊖⊖⊖ VERY LOW	CRITICAL
AC discordancy (continuous) - overall for DC and MC twins (US-to-birth interval not reported)	1	2399	Serious [5]	No serious inconsistency	No serious indirectness	Serious[6]	-	-	0.61 (0.58 to 0.63)	⨁⨁⊖⊖ LOW	IMPORTANT
AC discordancy (continuous) – for MC twins only (US-to-birth interval not reported)	1	457	Serious [5]	No serious inconsistency	No serious indirectness	Serious[6]	-	-	0.61 (0.58 to 0.63)	⨁⨁⊖⊖ LOW	IMPORTANT

AC: abdominal circumference; AUC: area under the curve (the curve represents different cut-off points); BPD: biparietal diameter; CI: confidence interval; FL: femur length; HC: head circumference; RoB: risk of bias
1 The number of participants included in meta-analysis was not reported, it was calculated by the NGA 2019 technical team
2 (1 very high RoB) High risk of bias for patient selection and for flow and timing; the review does not report the number of people included in meta-analysis; the heterogeneity between the included studies regarding the US-to-birth interval is high; poor reporting as the included studies do not report the number of live births or stillbirths
3 Unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test; not all participants were included in the analysis

4 The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results are judged to be very seriously imprecise
5 Unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test
6 The quality of the evidence was downgraded by 1 level because the 95%CI crosses default 1 cut-off (0.61)

Table 6: Clinical evidence profile for diagnostic monitoring to identify a small-for-gestational-age baby (defined as recognised reference standard for small for gestational age or intrauterine growth restriction) using growth discordancy in twin and triplet pregnancy in second trimester

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	Quality of the evidence (GRADE)	Importance
Twins										
Growth curves - France										
EFW <3rd percentile – EFW based on Hadlock 1985 curve (includes HC, AC, FL, BPD) (US-to-birth interval less than 30 days before birth)	1	236	Serious[1]	No serious inconsistency	No serious indirectness	Serious[2]	0.64 (0.49 to 0.78)	0.89 (0.86 to 0.92)	⊕⊕⊖⊖ LOW	CRITICAL
EFW <3rd percentile – EFW based on customised curve (includes maternal weight and height, parity, fetal sex, Ego	1	236	Serious[1]	No serious inconsistency	No serious indirectness	Serious[2]	0.66 (0.50 to 0.80)	0.86 (0.82 to 0.89)	⊕⊕⊖⊖ LOW	CRITICAL

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	Quality of the evidence (GRADE)	Importance
2006) (US-to-birth interval less than 30 days before birth)										
EFW <3rd percentile – EFW based on the EPOPé unadjusted curve (Ego 2016) (US-to-birth interval less than 30 days before birth)	1	236	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.57 (0.42 to 0.71)	0.89 (0.86 to 0.92)	⨁⨁⨁⨀ MODERATE	CRITICAL
EFW <3rd percentile – EFW based on the EPOPé adjusted curve (on the fetal sex) curve (Ego 2016) (US-to-birth interval less than 30 days before birth)	1	236	Serious[1]	No serious inconsistency	No serious indirectness	Serious[2]	0.64 (0.49 to 0.78)	0.90 (0.87 to 0.93)	⨁⨁⨀⨀ LOW	CRITICAL
EFW <10th percentile – EFW based on Hadlock 1985 curve (includes	1	236	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.67 (0.60 to 0.74)	0.80 (0.75 to 0.84)	⨁⨁⨁⨀ MODERATE	CRITICAL

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	Quality of the evidence (GRADE)	Importance
HC, AC, FL, BPD) (US-to-birth interval less than 30 days before birth)										
EFW <10th percentile – EFW based on customised curve (includes maternal weight and height, parity, fetal sex, Ego 2006) (US-to-birth interval less than 30 days before birth)	1	236	Serious[1]	No serious inconsistency	No serious indirectness	No serous imprecision	0.63 (0.55 to 0.70)	0.82 (0.76 to 0.86)	⨁⨁⨁⨀ MODERATE	CRITICAL
EFW <10th percentile – EFW based on the EPOPé unadjusted curve (Ego 2016) (US-to-birth interval less than 30 days before birth)	1	236	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.60 (0.52 to 0.68)	0.84 (0.79 to 0.88)	⨁⨁⨁⨀ MODERATE	CRITICAL

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	Quality of the evidence (GRADE)	Importance
EFW <10th percentile – EFW based on the EPOPé adjusted curve (on the fetal sex) curve (Ego 2016) (US-to-birth interval less than 30 days before birth)	1	236	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.57 (0.49 to 0.65)	0.83 (0.79 to 0.87)	⨁⨁⨁⊖ MODERATE	CRITICAL
Triplets										
Growth curves - Canada										
EFW <10th percentile (median US-to-birth interval 8 days (range 0-21))	1	78	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.56 (0.35 to 0.75)	1 (0.93 to 1)	⨁⨁⨁⊖ MODERATE	CRITICAL
EFW discordancy >25% (based on Hadlock et al. 1985, includes HC, AC, FL) (median US-to-birth interval 8 days (range 0-21))	1	78	Serious[1]	No serious inconsistency	No serious indirectness	Very serious[2]	0.80 (0.44 to 0.97)	0.94 (0.86 to 0.98)	⨁⊖⊖⊖ VERY LOW	CRITICAL

AC: abdominal circumference; BPD: biparietal diameter; CI: confidence interval; EFW: estimated fetal weight; EFWD: estimated fetal weight discordancy; FL: femur length; HC: head circumference

1 Unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test
2 The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results are judged to be very seriously imprecise

Table 7: Clinical evidence profile for diagnostic monitoring to identify intrauterine growth restriction or intertwin birth weight discordancy ≥15% or more using growth discordancy in twin pregnancy in second trimester

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Quality of the evidence (GRADE)	Importance
Intertwin birth weight discordancy ≥15%											
EFWD ≥15%– overall accuracy US-to-birth interval 48 h, 3 days (range 1-7), 15 days; within 28 days or 2 weeks)	6	1477 [1]	Very serious[2]	No serious inconsistency	No serious indirectness	No serious imprecision	0.68 (0.62 to 0.73)	0.83 (0.79 to 0.87)	-	⨁⨁⨀⨀ LOW	CRITICAL
Intertwin birth weight discordancy ≥20%											
EFWD ≥20% – overall accuracy (US-to-birth interval 48 h; 3 days (range 1-7) or 8 days (range 0-59), 3, 10 or 15	7	1780 [1]	Very serious[2]	No serious inconsistency	No serious indirectness	No serious imprecision	0.65 (0.58 to 0.72)	0.91 (0.87 to 0.94)	-	⨁⨁⨀⨀ LOW	CRITICAL

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Quality of the evidence (GRADE)	Importance
days; within 28 days)											
EFWD ≥20% - EFFW based on AC and FL (US-to-birth interval within 7 days)	3	160	Very serious[3]	Serious[4]	No serious indirectness	Very serious[5]	0.70 (0.34 to 0.93)	0.89 (0.69 to 0.98)	-	⨁⊖⊖⊖ VERY LOW	CRITICAL
EFWD ≥20% - overall accuracy (US-to-birth interval within 7 or 21 days; within 2 weeks)	7	491	Very serious[6]	Serious[4]	No serious indirectness	Serious[5]	0.71 (0.54 to 0.85)	0.89 (0.83 to 0.94)	-	⨁⊖⊖⊖ VERY LOW	CRITICAL
EFW ≥20% - EFW based on AC and FL (last US to birth interval ≤14 days)	1	74	Serious[7]	No serious inconsistency	No serious indirectness	No serous imprecision	0.46 (0.19 to 0.75)	0.92 (0.82 to 0.97)	-	⨁⨁⨁⊖ MODERATE	CRITICAL
EFW ≥20% - EFW based on Shepard's formula, includes AC and BPD (US-to-birth	1	29	Very serious[8]	No serious inconsistency	No serious indirectness	Very serious[5]	0.86 (0.57 to 0.98)	0.80 (0.52 to 0.96)	-	⨁⊖⊖⊖ VERY LOW	CRITICAL

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Quality of the evidence (GRADE)	Importance
interval within 7 days)											
EFWD ≥20% - overall for DC and MC twins (based on Hadlock's formula 1985) (median US-to-birth interval 2 weeks (range 0-3))	1	176	Very serious[9]	No serious inconsistency	No serious indirectness	Serious[10]	-	-	0.84 (0.76 to 0.92)	⨁⊖⊖⊖ VERY LOW	IMPORTANT
EFWD ≥20% - for DC twins only (based on Hadlock's formula 1985) (median US-to-birth interval 2 weeks (range 0-3))	1	123	Very serious[9]	No serious inconsistency	No serious indirectness	Very serious[11]	-	-	0.85 (0.76 to 0.95)	⨁⊖⊖⊖ VERY LOW	IMPORTANT
EFWD ≥20% - for MC twins only (based on Hadlock's formula 1985) (median US-	1	53	Very serious[9]	No serious inconsistency	No serious indirectness	Very serious[12]	-	-	0.82 (0.68 to 0.96)	⨁⊖⊖⊖ VERY LOW	IMPORTANT

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Quality of the evidence (GRADE)	Importance
to-birth interval 2 weeks (range 0-3))											
EFW <10th percentile (at least 1 twin, based on Hadlock 1991, includes HC, AC, FL) (median US-to-birth interval 8 days (range 0-59))	1	281	Serious[7]	No serious inconsistency	No serious indirectness	Serious[5]	0.69 (0.53 to 0.82)	0.80 (0.74, 0.85)	-	⨁⨁⨀⨀ LOW	CRITICAL
Intertwin birth weight discordancy ≥25%											
EFWD ≥25% - overall accuracy (US-to-birth interval 48 h, 3, 14 or 15 days; within 3, 6 or 28 days; 1.6±0.14 weeks; within 2, 2-4 or 3 weeks)	14	3980 [1]	Very serious[2]	Serious[4]	No serious indirectness	No serious imprecision	0.58 (0.46 to 0.68)	0.95 (0.93 to 0.97)	-	⨁⨀⨀⨀ VERY LOW	CRITICAL

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Quality of the evidence (GRADE)	Importance
EFW ≥25% - EFW based on AC and FL (last US to birth interval ≤7 days)	1	53	Serious[7]	No serious inconsistency	No serious indirectness	Serious[5]	0.50 (0.12 to 0.88)	0.98 (0.89 to 1)	-	⨁⨁⨀⨀ LOW	CRITICAL
EFWD ≥25% - EFW based on AC and FL (last US to birth interval ≤14 days)	1	74	Serious[7]	No serious inconsistency	No serious indirectness	Serious[5]	0.38 (0.09 to 0.76)	0.98 (0.92 to 1)	-	⨁⨁⨀⨀ LOW	CRITICAL
EFWD ≥25% - EFW based on BPD and AC, according to Shepard's formula (1982) (US-to-birth interval 1 to 6 weeks)	1	78	Serious[13]	No serious inconsistency	No serious indirectness	Very serious[5]	0.77 (0.46 to 0.95)	0.92 (0.83 0.97)	-	⨁⨀⨀⨀ VERY LOW	CRITICAL
EFWD discordancy (continuous) – EFW based on Hadlock's formula, includes HC,	1	2399	Serious[14]	No serious inconsistency	No serious indirectness	Serious[15]	-	-	0.63 (0.56 to 0.65)	⨁⨁⨀⨀ LOW	IMPORTNAT

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Quality of the evidence (GRADE)	Importance
AC and FL – overall for dichorionic and monochorionic twins (US-to-birth interval not reported)											
EFWD discordancy (continuous) – EFW based on Hadlock's formula, includes HC, AC and FL – for monochorionic twins only (US-to-birth interval not reported)	1	457	Serious[14]	No serious inconsistency	No serious indirectness	Serious[15]	-	-	0.61 (0.50 to 0.71)	⨁⨁⊖⊖ LOW	IMPORTANT

AC: abdominal circumference; AUC: area under the curve (the curve represents different cut-off points); BPD: biparietal diameter; CI: confidence interval; EFW: estimated fetal weight; EFWD: estimated fetal weight discrepancy; FL: femur length; HC: head circumference; US: ultrasound; RoB: risk of bias
1 The number of participants included in meta-analysis was not reported, it was calculated by the NGA 2019 technical team
2 (1 very high RoB) High risk of bias for patient selection and for flow and timing; the review does not report the number of people included in meta-analysis; the heterogeneity between the included studies regarding the US-to-birth interval is high; poor reporting as the included studies do not report the number of live births or stillbirths
3 (all high RoB) Unclear if selection of participants may have introduced bias in 1 study; unclear if the index test results were interpreted without knowledge of the results of the reference standard in all studies; unclear if the reference standard results were interpreted without knowledge of the results of the index test in all studies; no exclusion criteria were reported in 1 study; not all participants were included in the analysis in 1 study
4 Inconsistency was assessed by inspection of the sensitivity and specificity forest plots across studies, using the point estimates and confidence intervals

5 The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results are judged to be very seriously imprecise
6 (all high RoB) Unclear if selection of participants may have introduced bias in 2 study; unclear if the index test results were interpreted without knowledge of the results of the reference standard in all studies; unclear if the reference standard results were interpreted without knowledge of the results of the index test in all studies; no exclusion criteria were reported in 2 studies; not all participants were included in the analysis in 1 study
7 Unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test
8 Unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test; no exclusion criteria were reported
9 Unclear if a consecutive or random sample of participants was enrolled; unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test
10 The quality of the evidence was downgraded by 1 level because the 95% CI crosses 1 default cut-off (0.81)
11 The evidence was downgraded by 2 because the 95%CI crosses 2 cut-offs (0.81 and 0.91)
12 The quality of the evidence was downgraded by 2 levels because the 95% CI crosses 3 default cut-offs (0.70, 0.80 and 0.92)
13 unclear if the reference standard results were interpreted without knowledge of the results of the index test
14 Unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test
15 The quality of the evidence was downgraded by 1 level because the 95%CI crosses default 1 cut-off (0.61)

Table 8: Clinical evidence profile for diagnostic monitoring to identify intertwin birth weight discordancy ≥20% using amniotic fluid discordancy in twin pregnancy in second trimester

Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	Quality of the evidence (GRADE)	Importance
Oligohydramnios defined as the deepest vertical pocket of amniotic fluid inferior to 2 cm (measured at the last ultrasound) (median US-to-birth interval 2 weeks (range 0-3))	1	176	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.13 (0.04 to 0.28)	0.97 (0.93 to 0.99)	⨁⨁⨁⨀ MODERATE	CRITICAL
Oligohydramnios defined as the deepest vertical pocket of amniotic fluid of less than 2 cm (median US-to-birth interval 8 days (range 0-59))	1	281	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.17 (0.07 to 0.31)	0.85 (0.80 to 0.90)	⨁⨁⨁⨀ MODERATE	CRITICAL

CI: confidence interval
1 Unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test

Table 9: Clinical evidence profile for diagnostic monitoring to identify intertwin birth weight discordancy ≥20% using symphysio-fundal height measurement in twin pregnancy in second trimester

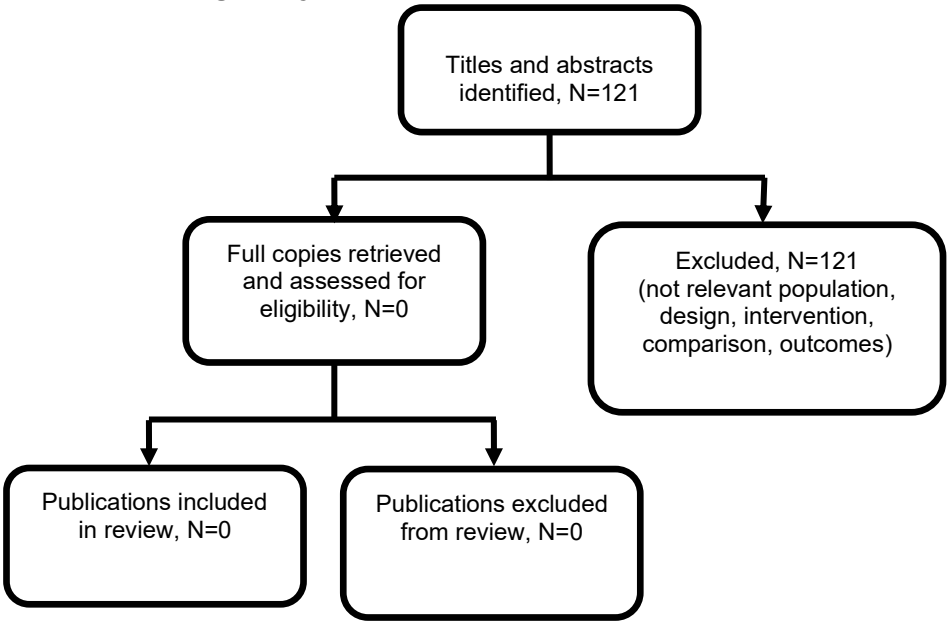
Index test	Number of studies	Number of participants	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (95% CI)	Specificity (95% CI)	Quality of the evidence (GRADE)	Importance
Symphysio-fundal height (no US-to-birth interval reported; US was done between 16 and 36 weeks)	1	160	Very serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	0.24 (0.07 to 0.50)	0.83 (0.75 to 0.88)	⨁⨁⊖⊖ LOW	CRITICAL

CI: confidence interval
1 Unclear if a consecutive or random sample of participants was enrolled; unclear if the index test results were interpreted without knowledge of the results of the reference standard; unclear if the reference standard results were interpreted without knowledge of the results of the index test; unclear if there was an appropriate interval between index test and reference standard

Appendix G – Economic evidence study selection

Economic evidence study selection for review question: What is the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy?

Figure 6: Flow diagram of economic article selection for the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy



Appendix H – Economic evidence tables

Economic evidence tables for review question: What is the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy?

No economic evidence was identified for this review.

Appendix I - Economic evidence profiles

Economic evidence profiles for review question: What is the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy?

No economic evidence was identified for this review.

Appendix J - Economic analysis

Economic analysis for review question: What is the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy?

No economic analysis was conducted for this review.

Appendix K – Excluded studies

Excluded studies for review question: What is the optimal screening programme to detect intrauterine growth restriction (IUGR) in twin and triplet pregnancy?

Clinical studies

Study	Reason for exclusion
Aksam, S., Plesinac, S., Dotlic, J., Tadic, J., Vrzic-Petronijevic, S., Petronijevic, M., Kocijancic-Belovic, D., Buzadzic, S., First trimester ultrasonographic parameters in prediction of the course and outcome of monochorionic twin pregnancies, Turkish Journal of Medical Sciences, 47, 934-941, 2017	No confidence intervals were reported
Alfirevic, Z., Stampalija, T., Gyte, G. M., Fetal and umbilical Doppler ultrasound in high-risk pregnancies, Cochrane Database Syst RevThe Cochrane database of systematic reviews, Cd007529, 2010	Cochrane review regarding fetal assessment in high-risk pregnancies. No diagnostic data on detection of intrauterine growth restriction
Algeri, P., Frigerio, M., Lamanna, M., Petrova, P. V., Cozzolino, S., Incerti, M., Mastrolia, S. A., Roncaglia, N., Vergani, P., Selective IUGR in dichorionic twins: What can Doppler assessment and growth discordancy say about neonatal outcomes?, Journal of Perinatal Medicine., 29, 2017	Non relevant population as all pregnancies were complicated by intrauterine growth restriction at the begging of the study
Ali, M, Miller, J, Chan, C, Fields, J, Houston, L, Bernhard, K, Hawk, A, Sunderji, S, Siddiqui, D, Chang, E, Sandlin, A, Magann, E, Chauhan, S, Chasen, S, Prenatal detection of fetal growth restriction in twins: the TWIG study, Prenatal diagnosis. Conference: 21st international conference on prenatal diagnosis and therapy, ISPD 2017. United states, 37, 101-102, 2017	Conference abstract
Ali, Miami Abd Al Hassan,, Al-Gharny, Hala Abd, Estimation of Fetal Body Weight in Twins: A New Mathematical Model, Iraqi Journal of Community Medicine, 25, 61-65, 2012	Included in Leombroni 2017 review
Allaf, M. B, Campbell, W. A, Vintzileos, A. M, Haeri, S, Javadian, P, Shamshirsaz, A. A, Ogburn, P, Figueroa, R, Wax, J, Markenson, G, Chavez, M. R, Ravangard, S. F, Ruano, R, Sangi-Haghpeykar, H, Salmanian, B, Meyer, M, Johnson, J, Ozhand, A, Davis, S, Borgida, A, Belfort, M. A, Shamshirsaz, A. A., Does early second-trimester sonography predict adverse perinatal outcomes in monochorionic diamniotic twin pregnancies?, Journal of Ultrasound in Medicine, 33, 1573-8, 2014	No diagnostic accuracy data for intrauterine growth restriction were reported (reported only as an adverse composite obstetric outcome)
Allaf, M. B, Vintzileos, A. M, Chavez, M. R, Wax, J. A, Ravangard, S. F, Figueroa, R, Borgida, A, Shamshirsaz, A, Markenson, G, Davis, S, Habenicht, R, Haeri, S, Ozhand, A, Johnson, J, Sangi-Haghpeykar, H, Spiel, M, Ruano, R, Meyer, M, Belfort, M. A, Ogburn, P, Campbell, W. A, Shamshirsaz, A. A., First-trimester sonographic prediction of obstetric and neonatal outcomes in monochorionic diamniotic twin pregnancies, Journal of Ultrasound in Medicine, 33, 135-40, 2014	No data for growth discordance in terms of numbers were reported (reported only in a figure)
Al-Obaidly, S, Parrish, J, Murphy, K. E, Glanc, P, Maxwell, C., The Accuracy of Estimating Fetal Weight and Inter-Twin Weight Discordance by Ultrasound in Twin Pregnancies in Women With Increased Body Mass Index, Journal of	Included in Leombroni 2017 review

Study	Reason for exclusion
obstetrics and gynaecology Canada : JOGC = Journal d'obstetrique et gynecologie du Canada : JOGC, 37, 696-701, 2015	
Athanasiadis, A. P, Michaelidou, A. M, Fotiou, M, Menexes, G, Theodoridis, T. D, Ganidou, M, Tzevelekis, B, Assimakopoulos, E, Tarlatzis, B. C., Correlation of 2nd trimester amniotic fluid amino acid profile with gestational age and estimated fetal weight, Journal of Maternal-Fetal & Neonatal Medicine, 24, 1033-8, 2011	Singleton pregnancies
Barel, O, Maymon, R, Barak, U, Smorgick, N, Tovbin, J, Vaknin, Z., A search for the most accurate formula for sonographic weight estimation by fetal sex - a retrospective cohort study, Prenatal Diagnosis, 34, 1337-44, 2014	Singleton pregnancies
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	Not multiple pregnancy
Barel,O, Vaknin,Z, Tovbin,J, Herman,A, Maymon,R., Assessment of the accuracy of multiple sonographic fetal weight estimation formulas: a 10-year experience from a single center, Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine, 32, 815-823, 2013	Singleton pregnancies
Barnea, E. R., Romero, R., Scott, D., Hobbins, J. C., The value of biparietal diameter and abdominal perimeter in the diagnosis of growth retardation in twin gestation, Am J PerinatolAmerican journal of perinatology, 2, 221-2, 1985	No diagnostic accuracy data were reported
Bartha, J. L., Ling, Y., Kyle, P., Soothill, P. W., Clinical consequences of first-trimester growth discordance in twins, Eur J Obstet Gynecol Reprod BiolEuropean journal of obstetrics, gynecology, and reproductive biology, 119, 56-9, 2005	No relevant diagnostic accuracy data were reported
Baz, E., Hecher, K., Hackeloer, B. J., The clinical relevance of fetal nuchal translucency, Gynakologe, 32, 200-212, 1999	Not in English language
Ben-Ami, I, Daniel-Spiegel, E, Battino, S, Melcer, Y, Floeck, A, Geipel, A, Miron, P, Maymon, R., The association of crown-rump length discrepancy with birthweight discordance in spontaneous versus IVF monochorionic twins: a multicenter study, Prenatal Diagnosis, 35, 864-9, 2015	No relevant comparison as the study compares the correlations between crown-rump length discrepancy and birthweight discordance in spontaneous versus in vitro fertilisation-conceived twin pregnancies
Bennasar, M, Eixarch, E, Martinez, J. M, Gratacos, E., Selective intrauterine growth restriction in monochorionic diamniotic twin pregnancies, Seminars In Fetal & Neonatal Medicine, 22, 376-382, 2017	Narrative article about umbilical artery doppler assessment, classification of selective intrauterine growth restriction and its management in twin pregnancies
Bhide,A., Sankaran,S., Sairam,S., Papageorghiou,A.T., Thilaganathan,B., Relationship of intertwin crown-rump length discrepancy to chorionicity, fetal demise and birth-weight discordance, Ultrasound in Obstetrics and Gynecology, 34, 131-135, 2009	No confidence intervals were reported
Blickstein, I., Friedman, A., Caspi, B., Lancet, M., Ultrasonic prediction of growth discordancy by intertwin difference in abdominal circumference, Int J Gynaecol ObstetInternational	No relevant index test

Study	Reason for exclusion
journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 29, 121-4, 1989	
Blickstein, I., Manor, M., Levi, R., Goldchmit, R., Is intertwin birth weight discordance predictable?, Gynecol Obstet InvestGynecologic and obstetric investigation, 42, 105-8, 1996	Included in Leombroni 2017 review
Breathnach, F. M, McAuliffe, F. M, Geary, M, Daly, S, Higgins, J. R, Dornan, J, Morrison, J. J, Burke, G, Higgins, S, Dicker, P, Manning, F, Mahony, R, Malone, F. D, Perinatal Ireland Research, Consortium, Definition of intertwin birth weight discordance, Obstetrics & Gynecology, 118, 94-103, 2011	The study presents a threshold for birth weight discordance in twin pregnancies; no relevant diagnostic accuracy data were reported/not possible to calculate
Brink Henriksen, T., Villadsen, G. E., Hedegaard, M., Secher, N. J., Prediction of light-for-gestational age at delivery in twin pregnancies: an evaluation of fetal weight deviation and growth discordance measured by ultrasound, European Journal of Obstetrics, Gynecology, & Reproductive Biology, 47, 195-200, 1992	No confidence intervals were reported
Brown, C. E., Guzick, D. S., Leveno, K. J., Santos-Ramos, R., Whalley, P. J., Prediction of discordant twins using ultrasound measurement of biparietal diameter and abdominal perimeter, Obstet GynecolObstetrics and gynecology, 70, 677-81, 1987	No relevant index test
Caravello, J. W., Chauhan, S. P., Morrison, J. C., Magann, E. F., Martin, J. N., Jr., Devoe, L. D., Sonographic examination does not predict twin growth discordance accurately, Obstetrics & Gynecology, 89, 529-33, 1997	Included in Leombroni 2017 review
Casasbuenas,A., Wong,A.E., Sepulveda,W., Nuchal translucency thickness in monochorionic multiple pregnancies: value in predicting pregnancy outcome, Journal of Ultrasound in Medicine, 27, 363-369, 2008	No separate data for twin and triplet pregnancies
Centre for Reviews and Dissemination., Antenatal ultrasound scanning (Structured abstract) , Database of Abstracts of Reviews of Effects, 3, 2010	Structured abstract of a systematic review published in 1994. Original paper reviewed effect of routine ultrasound on perinatal outcome, study population not exclusively twins/triplets
Chang, Y. L., Chang, T. C., Chang, S. D., Cheng, P. J., Chao, A. S., Hsieh, P. C., Soong, Y. K., Sonographic prediction of significant intertwin birth weight discordance, Eur J Obstet Gynecol Reprod BiolEuropean journal of obstetrics, gynecology, and reproductive biology, 127, 35-40, 2006	Included in Leombroni 2017 review
Chang,Y.L., Chang,S.D., Chao,A.S., Hsieh,P.C., Wang,C.N., Wang,T.H., Clinical outcome and placental territory ratio of monochorionic twin pregnancies and selective intrauterine growth restriction with different types of umbilical artery Doppler, Prenatal Diagnosis, 29, 253-256, 2009	No diagnostic accuracy data were reported
Chauhan,S.P., Scardo,J.A., Hayes,E., Abuhamad,A.Z., Berghella,V., Twins: Prevalence, problems, and preterm births, American Journal of Obstetrics and Gynecology, #203, 305-315, 2010	Review article, no new data
Chauhan,S.P., Shields,D., Parker,D., Sanderson,M., Scardo,J.A., Magann,E.F., Detecting fetal growth restriction	Study included twins with feto-fetal transfusion syndrome

Study	Reason for exclusion
or discordant growth in twin gestations stratified by placental chorionicity, Journal of Reproductive Medicine, 49, 279-284, 2004	
Chitkara, U., Berkowitz, G. S., Levine, R., Riden, D. J., Fagerstrom, R. M., Jr., Chervenak, F. A., Berkowitz, R. L., Twin pregnancy: routine use of ultrasound examinations in the prenatal diagnosis of intrauterine growth retardation and discordant growth, Am J PerinatolAmerican journal of perinatology, 2, 49-54, 1985	No confidence intervals were reported
Chittacharoen, A., Leelapattana, P., Phuapradit, W., Umbilical Doppler velocimetry prediction of discordant twins, J Obstet Gynaecol ResThe journal of obstetrics and gynaecology research, 25, 95-8, 1999	No relevant index test
Chittacharoen, A., Leelapattana, P., Rangsiprakarn, R., Prediction of discordant twins by real-time ultrasonography combined with umbilical artery velocimetry, Ultrasound Obstet GynecolUltrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 15, 118-21, 2000	Included in Leombroni 2017 review
Corcoran, S, Breathnach, F, Burke, G, McAuliffe, F, Geary, M, Daly, S, Higgins, J, Hunter, A, Morrison, J. J, Higgins, S, Mahony, R, Dicker, P, Tully, E, Malone, F. D., Dichorionic twin ultrasound surveillance: sonography every 4 weeks significantly underperforms sonography every 2 weeks: results of the Prospective Multicenter ESPRiT Study, American Journal of Obstetrics & GynecologyAm J Obstet Gynecol, 213, 551.e1-5, 2015	The study examines how ultrasound scanning performed at 2- or 4-week intervals impact the prenatal detection of fetal growth restriction, oligohydramnios or abnormal umbilical artery doppler waveforms; no relevant diagnostic accuracy data were reported/not possible to calculate
Daly, S., Higgins, J., Burke, G., Morrison, J., Higgins, S., Mthunzi, A., Gillan, J., Geary, M., O'Malley, A., Kent, E., Breathnach, F., Dicker, P., Manning, F., Malone, F., Dornan, J., Mahony, R., McAuliffe, F., Correlation between histomorphometric placental characteristics and fetal growth restriction in dichorionic twins, American Journal of Obstetrics and Gynecology, 204, S58, 2011	Conference abstract
Danon,D., Melamed,N., Bardin,R., Meizner,I., Accuracy of ultrasonographic fetal weight estimation in twin pregnancies, Obstetrics and Gynecology, 112, 759-764, 2008	Included in Leombroni 2017 review
D'Antonio, F, Khalil, A, Dias, T, Thilaganathan, B, Southwest Thames Obstetric Research, Collaborative, Weight discordance and perinatal mortality in twins: analysis of the Southwest Thames Obstetric Research Collaborative (STORK) multiple pregnancy cohort, Ultrasound in Obstetrics & Gynecology, 41, 643-8, 2013	The study examines the prediction of perinatal loss
D'Antonio, F, Khalil, A, Mantovani, E, Thilaganathan, B, Southwest Thames Obstetric Research, Collaborative, Embryonic growth discordance and early fetal loss: the STORK multiple pregnancy cohort and systematic review, Human Reproduction, 28, 2621-7, 2013	The study examines the prediction of spontaneous single fetal loss
D'Antonio, F, Khalil, A, Morlando, M, Thilaganathan, B., Accuracy of Predicting Fetal Loss in Twin Pregnancies Using Gestational Age-Dependent Weight Discordance Cut-Offs: Analysis of the STORK Multiple Pregnancy Cohort, Fetal Diagnosis & Therapy, 38, 22-Aug, 2015	The study examines whether a single weight discordance cut-off or different cut-offs should be used according to the gestational age at assessment to predict the occurrence of single fetal loss

Study	Reason for exclusion
D'Antonio, F, Khalil, A, Pagani, G, Papageorghiou, A. T, Bhide, A, Thilaganathan, B., Crown-rump length discordance and adverse perinatal outcome in twin pregnancies: systematic review and meta-analysis, Ultrasound in Obstetrics & GynecologyUltrasound Obstet Gynecol, 44, 138-46, 2014	Relevant studies from this review were assessed for a potential inclusion
D'Antonio, F, Odibo, A. O, Prefumo, F, Khalil, A, Buca, D, Flacco, M. E, Liberati, M, Manzoli, L, Acharya, G., Weight discordance and perinatal mortality in twin pregnancies: a systematic review and meta-analysis, Ultrasound in Obstetrics & GynecologyUltrasound Obstet Gynecol, 20, 20, 2017	The systematic review mainly explores the association between birth weight and perinatal mortality
D'Antonio, F, Thilaganathan, B, Laoreti, A, Khalil, A, Southwest Thames Obstetric Research, Collaborative, Birthweight discordance and neonatal morbidity in twin pregnancies: Analysis of the STORK multiple pregnancy cohort, Ultrasound in Obstetrics & Gynecology, 13, 13, 2017	The study evaluates the association between weight discordance and composite neonatal morbidity, and determines the predictive accuracy of different weight discordant cut-offs in predicting neonatal morbidity
D'Antonio, F., Familiari, A., Thilaganathan, B., Papageorghiou, A. T., Manzoli, L., Khalil, A., Bhide, A., Sensitivity of first-trimester ultrasound in the detection of congenital anomalies in twin pregnancies: population study and systematic review, Acta Obstetricia et Gynecologica ScandinavicaActa Obstet Gynecol Scand, 95, 1359-1367, 2016	The article evaluates the diagnostic performance of first-trimester ultrasound in detecting congenital anomalies (central nervous system, face, neck, cardiovascular, lung, gastrointestinal, renal, skeletal) in twins and presents a systematic review on the same topic
DeJesus Allison, S. O, Javitt, M. C, Glanc, P, Andreotti, R. F, Bennett, G. L, Brown, D. L, Dubinsky, T, Harisinghani, M. G, Harris, R. D, Mitchell, D. G, Pandharipande, P. V, Pannu, H. K, Podrasky, A. E, Shipp, T. D, Siegel, C. L, Simpson, L, Wong-You-Cheong, J. J, Zelop, C. M, American College of, Radiology, ACR Appropriateness Criteria Multiple gestations, Ultrasound Quarterly, 28, 149-55, 2012	The papers presents the revised appropriateness criteria to address diagnosing a multiple gestation in the first trimester, and to scan for detailed anatomic evaluation and comparative growth at 18-20 weeks
Deter, R. L., Stefos, T., Harrist, R. B., Hill, R. M., Detection of intrauterine growth retardation in twins using individualized growth assessment. II. Evaluation of third-trimester growth and prediction of growth outcome at birth, Journal of Clinical UltrasoundJ Clin Ultrasound, 20, 579-85, 1992	No confidence intervals were reported
Diaz-Garcia, C, Bernard, J. P, Ville, Y, Salomon, L. J., Validity of sonographic prediction of fetal weight and weight discordance in twin pregnancies, Prenatal Diagnosis, 30, 361-7, 2010	Included in Leombroni 2017 review
Dimassi, K, Karoui, A, Triki, A, Gara, M. F., Performance of ultrasound fetal weight estimation in twins, Tunisie MedicaleTunis Med, 94, 203-9, 2016	Not in English language
Divon, M. Y., Girz, B. A., Sklar, A., Guidetti, D. A., Langer, O., Discordant twins--a prospective study of the diagnostic value of real-time ultrasonography combined with umbilical artery velocimetry, American Journal of Obstetrics & Gynecology, 161, 757-60, 1989	No confidence intervals were reported
Eik-Nes, S. H., Grottum, P., Persson, P. H., Marsal, K., Prediction of fetal growth deviation by ultrasonic biometry. I. Methodology, Acta Obstetricia et Gynecologica Scandinavica, 61, 53-8, 1982	Singleton pregnancies

Study	Reason for exclusion
El Kateb, A., Nasr, B., Nassar, M., Bernard, J. P., Ville, Y., First-trimester ultrasound examination and the outcome of monochorionic twin pregnancies, 27, 922-5, 2007	No relevant diagnostic accuracy data were reported
Erkkola, R., Ala-Mello, S., Piiroinen, O., Kero, P., Sillanpaa, M., Growth discordancy in twin pregnancies: a risk factor not detected by measurements of biparietal diameter, Obstet GynecolObstetrics and gynecology, 66, 203-6, 1985	No relevant index test
Esinler, D, Aldemir, O. B, Alici Davutoglu, E, Karahanoglu, E, Salihoglu, K. N, Kuzu, E, Yerebasmaz, N, Kandemir, O, Yalvac, S., A new mathematical formula to predict the foetal weight in twin pregnancies: A comparison of it with 19 different formulas, Journal of Obstetrics & Gynaecology, 37, 53-57, 2017	The study compares the accuracy of different formulas to predict the fetal weight in twin pregnancies and compares a newly developed formula with these 19 formulas. According to the protocol, the review does not compare different formulas
Evans,M.I., Andriole,S., Screening and testing in multiples, Clinics in Laboratory Medicine, 30, 643-654, 2010	Narrative review on screening in twin pregnancies
Expert Panel on Women's, Imaging, Glanc, P, Nyberg, D. A, Khati, N. J, Deshmukh, S. P, Dudiak, K. M, Henrichsen, T. L, Poder, L, Shipp, T. D, Simpson, L, Weber, T. M, Zelop, C. M., ACR Appropriateness Criteria Multiple Gestations, Journal of the American College of Radiology, 14, S476-S489, 2017	American College of Radiology recommendations regarding ultrasound examinations in multiple pregnancies
Figueras,F, Gardosi,J., Intrauterine growth restriction: new concepts in antenatal surveillance, diagnosis, and management, American Journal of Obstetrics and Gynecology, 204, 288-300, 2011	Narrative review on antenatal surveillance, diagnosis and management of intrauterine growth restriction
Fox, N. S, Saltzman, D. H, Schwartz, R, Roman, A. S, Klauser, C. K, Rebarber, A., Second-trimester estimated fetal weight and discordance in twin pregnancies: association with fetal growth restriction, Journal of Ultrasound in Medicine, 30, 1095-101, 2011	Included in Leombroni 2017 review
Fratelli,N, Prefumo,F, Fichera,A, Valcamonico,A, Marella,D, Frusca,T., Nuchal translucency thickness and crown rump length discordance for the prediction of outcome in monochorionic diamniotic pregnancies, Early Human Development, 87, 27-30, 2011	No relevant reference standard
Gabbay-Benziv, R, Crimmins, S, Contag, S. A., Reference Values for Sonographically Estimated Fetal Weight in Twin Gestations Stratified by Chorionicity: A Single Center Study, Journal of Ultrasound in Medicine, 36, 793-798, 2017	The study describes the development of a set of reference values for sonographic fetal weight in twin gestations
Gandhi, M., Ferrara, L., Belogolovkin, V., Moshier, E., Rebaber, A., Effect of increased body mass index on the accuracy of estimated fetal weight by sonography in twins, Journal of Ultrasound in Medicine, 28, 301-8, 2009	Included in Leombroni 2017 review
Gaziano,E.P., Knox,G.E., Bendel,R.P., Calvin,S., Brandt,D., Is pulsed Doppler velocimetry useful in the management of multiple-gestation pregnancies?, American Journal of Obstetrics and Gynecology, 164, 1426-1431, 1991	Study did not report diagnostic accuracy
Gernt, P. R., Mauldin, J. G., Newman, R. B., Durkalski, V. L., Sonographic prediction of twin birth weight discordance, Obstet GynecolObstetrics and gynecology, 97, 53-6, 2001	Included in Leombroni 2017 review
Gerson, A. G., Wallace, D. M., Bridgens, N. K., Ashmead, G. G., Weiner, S., Bolognese, R. J., Duplex Doppler ultrasound	Includes twins and triplets, and does not report results specifically for twin or triplet pregnancy

Study	Reason for exclusion
in the evaluation of growth in twin pregnancies, Obstet GynecolObstetrics and gynecology, 70, 419-23, 1987	
Giles, W. B., Doppler ultrasound in multiple pregnancies, Baillieres Clinical Obstetrics & GynaecologyBaillieres Clin Obstet Gynaecol, 12, 77-89, 1998	Review article; does not contain original data
Grande, M, Gonce, A, Stergiotou, I, Bennasar, M, Borrell, A., Intertwin crown-rump length discordance in the prediction of fetal anomalies, fetal loss and adverse perinatal outcome, Journal of Maternal-Fetal & Neonatal Medicine, 29, 2883-8, 2016	The study population is mixed as it also includes pregnancies with chromosomal and structural anomalies
Grobman, W. A., Parilla, B. V., Positive predictive value of suspected growth aberration in twin gestations, American Journal of Obstetrics & Gynecology, 181, 1139-41, 1999	No relevant index test
Harper, L. M, Roehl, K. A, Tuuli, M. G, Odibo, A. O, Cahill, A. G., Sonographic accuracy of estimated fetal weight in twins, Journal of Ultrasound in Medicine, 32, 625-30, 2013	No relevant reference standard
Harper,L.M, Roehl,K.A, Odibo,A.O, Cahill,A.G., First-trimester growth discordance and adverse pregnancy outcome in dichorionic twins, Ultrasound in Obstetrics and Gynecology, 41, 627-631, 2013	The study evaluates the association between first-trimester size discordance and dichorionic twin pregnancy outcome (loss of one or both fetuses before 20 weeks' gestation, anomalies in one or both fetuses, preterm birth before 34 weeks' gestation, stillbirth, small-for-gestational age, and admission to the neonatal intensive care unit). No relevant diagnostic accuracy data were reported/not possible to calculate
Hastie, S. J., Danskin, F., Neilson, J. P., Whittle, M. J., Prediction of the small for gestational age twin fetus by Doppler umbilical artery waveform analysis, Obstetrics & Gynecology, 74, 730-3, 1989	No relevant index test
Hata, T., Deter, R. L., Hill, R. M., Individual growth curve standards in triplets: prediction of third-trimester growth and birth characteristics, Obstet GynecolObstetrics and gynecology, 78, 379-84, 1991	No diagnostic accuracy data were reported
Hehir, M. P, Breathnach, F. M, Hogan, J. L, McAuliffe, F. M, Geary, M. P, Daly, S, Higgins, J, Hunter, A, Morrison, J. J, Burke, G, Mahony, R, Dicker, P, Tully, E, Malone, F. D., Prenatal prediction of significant intertwin birthweight discordance using standard second and third trimester sonographic parameters, Acta Obstetricia et Gynecologica Scandinavica, 96, 472-478, 2017	No confidence intervals were reported. No clear from the paper what is the optimal discordance cut-off according to Liu (2011)
Henry, A., Gopikrishna, S., Mahajan, A., Alphonse, J., Meriki, N., Welsh, A. W., Use of the Foetal Myocardial Performance Index in monochorionic, diamniotic twin pregnancy: a prospective cohort and nested case-control study, Journal of Maternal-Fetal and Neonatal Medicine, 1-13, 2018	The study examines whether adding Myocardial Performance Index to routine ultrasonic surveillance provides additional diagnostic or prognostic value in the prediction or monitoring of twins
Hoopmann, M, Kagan, K. O, Yazdi, B, Grischke, E. M, Abele, H., Prediction of birth weight discordance in twin pregnancies by second- and third- trimester ultrasound, Fetal Diagnosis & Therapy, 30, 29-34, 2011	Included in Leombroni 2017 review

Study	Reason for exclusion
Huber, C, Zdanowicz, J. A, Mueller, M, Surbek, D., Factors influencing the accuracy of fetal weight estimation with a focus on preterm birth at the limit of viability: a systematic literature review, Fetal Diagnosis & Therapy, 36, 01-Aug, 2014	Systematic review on possible factors affecting fetal weight estimation
Ishii,K., Murakoshi,T., Takahashi,Y., Shinno,T., Matsushita,M., Naruse,H., Torii,Y., Sumie,M., Nakata,M., Perinatal outcome of monochorionic twins with selective intrauterine growth restriction and different types of umbilical artery Doppler under expectant management, Fetal Diagnosis and Therapy, 26, 157-161, 2009	Study reported on prognosis (not diagnosis)
Jahanfar, S., Lim, K., Oviedo-Joekes, E., Optimal threshold for birth weight discordance: Does knowledge of chorionicity matter?, Journal of Perinatology, 36, 704-12, 2016	The study evaluates the optimal threshold of birth weight discordance for prediction of stillbirth, perinatal mortality and morbidity
Joern, H., Schroeder, W., Sassen, R., Rath, W., Predictive value of a single CTG, ultrasound and Doppler examination to diagnose acute and chronic placental insufficiency in multiple pregnancies, Journal of Perinatal Medicine, 25, 325-32, 1997	The study used a poor methodology; ultrasound measurements were not corrected for gestational age
Kadji, C, Bevilacqua, E, Hurtado, I, Carlin, A, Cannie, M. M, Jani, J. C., Comparison of conventional 2D ultrasound to magnetic resonance imaging for prenatal estimation of birthweight in twin pregnancy, American Journal of Obstetrics & Gynecology, 16, 16, 2017	The study compares ultrasound and magnetic resonance imaging estimated fetal weight with the actual birth weight
Kalish,R.B., Chasen,S.T., Gupta,M., Sharma,G., Perni,S.C., Chervenak,F.A., First trimester prediction of growth discordance in twin gestations, American Journal of Obstetrics and Gynecology, 189, 706-709, 2003	No confidence intervals were reported
Kalish,R.B., Gupta,M., Perni,S.C., Berman,S., Chasen,S.T., Clinical significance of first trimester crown-rump length disparity in dichorionic twin gestations, American Journal of Obstetrics and Gynecology, 191, 1437-1440, 2004	No confidence intervals were reported
Kaponis, A, Thanatsis, N, Papadopoulos, V, Decavalas, G., Intertwin estimated fetal weight or crown rump length discordance and adverse perinatal outcome, Journal of Perinatal Medicine, 44, 863-869, 2016	A narrative review about the importance of inter-twin estimated fetal weight and crown rump length discordance for the prediction of adverse perinatal outcome
Kent, E. M, Breathnach, F. M, Gillan, J. E, McAuliffe, F. M, Geary, M. P, Daly, S, Higgins, J. R, Hunter, A, Morrison, J. J, Burke, G, Higgins, S, Carroll, S, Dicker, P, Manning, F, Tully, E, Malone, F. D., Placental pathology, birthweight discordance, and growth restriction in twin pregnancy: results of the ESPRiT Study, American Journal of Obstetrics & Gynecology, 207, 220.e1-5, 2012	The study evaluates the association between placental pathology and twin growth restriction
Kent,E.M, Breathnach,F.M, Gillan,J.E, McAuliffe,F.M, Geary,M.P, Daly,S, Higgins,J.R, Dornan,J, Morrison,J.J, Burke,G, Higgins,S, Carroll,S, Dicker,P, Manning,F, Malone,F.D., Placental cord insertion and birthweight discordance in twin pregnancies: results of the national prospective ESPRiT Study, American Journal of Obstetrics and Gynecology, 205, 376-377, 2011	The study evaluates the frequency of noncentral cord insertion and birth weight discordance
Khalil, A, D'Antonio, F, Dias, T, Cooper, D, Thilaganathan, B, Southwest Thames Obstetric Research, Collaborative,	Included in Leombroni 2017 review

Study	Reason for exclusion
Ultrasound estimation of birth weight in twin pregnancy: comparison of biometry algorithms in the STORK multiple pregnancy cohort, Ultrasound in Obstetrics & Gynecology, 44, 210-20, 2014	
Khalil, A. A, Khan, N, Bowe, S, Familiari, A, Papageorghiou, A, Bhide, A, Thilaganathan, B., Discordance in fetal biometry and Doppler are independent predictors of the risk of perinatal loss in twin pregnancies, American Journal of Obstetrics & Gynecology, 213, 222.e1-222.e10, 2015	The study examines the role of fetal size, doppler indices and their discordance in the prediction of perinatal loss
Khalil, A., Beune, I., Hecher, K., Wynia, K., Ganzevoort, W., Reed, K., Lewi, L., Oepkes, D., Gratacos, E., Thilaganathan, B., Gordijn, S. J., Consensus definition and essential reporting parameters of selective fetal growth restriction in twin pregnancy: a Delphi procedure, Ultrasound in Obstetrics & Gynecology, 24, 24, 2018	The study describes the expert consensus on a definition of selective fetal growth restriction and essential reporting parameters in twin pregnancies
Klam, S. L., Rinfret, D., Leduc, L., Prediction of growth discordance in twins with the use of abdominal circumference ratios, Am J Obstet GynecolAmerican journal of obstetrics and gynecology, 192, 247-51, 2005	Included in Leombroni 2017 review
Kontopoulos, E., Odibo, A., Wilson, R. D., Current controversies in prenatal diagnosis 2: Are we ready to screen for fetal anomalies with first trimester ultrasound?, Prenatal Diagnosis, 33, 9-12, 2013	The article presents a discussion on the debate question ‘Are we ready to screen for fetal anomalies with first trimester ultrasound?'
Kurmanavicius, J., Hebisch, G., Huch, R., Huch, A., Umbilical artery blood flow velocity waveforms in twin pregnancies, J Perinat MedJournal of perinatal medicine, 20, 307-12, 1992	No relevant index test
Leftwich,H.K., Schmidt,B., Pham,T., Hibbard,J.U., Wilkins,I., Doppler ultrasonography: more than just for intrauterine growth restriction?, Obstetrics and Gynecology, 123 Suppl 1, 193S-194S, 2014	Conference abstract
Lewi, L., Lewi, P., Diemert, A., Jani, J., Gucciardo, L., Van Mieghem, T., Done, E., Gratacos, E., Huber, A., Hecher, K., Deprest, J., The role of ultrasound examination in the first trimester and at 16 weeks' gestation to predict fetal complications in monochorionic diamniotic twin pregnancies, Am J Obstet GynecolAmerican journal of obstetrics and gynecology, 199, 493.e1-7, 2008	No relevant diagnostic accuracy data were reported/not possible to calculate
MacHado Nardozza, L. M., Junior, E. A., Barbosa, M. M., Rabachini Caetano, A. C., Re Lee, D. J., Moron, A. F., Fetal growth restriction: Current knowledge to the general Obs/Gyn, Archives of Gynecology and Obstetrics, 286, 1-13, 2012	Review on the concept, etiology, classification, diagnosis, management, and prognosis of fetal growth restriction
Machado, R. C., Brizot, M. L., Liao, A. W., Cabar, F. R., Zugaib, M., Prenatal sonographic prediction of twin growth discordance, Twin Res Hum GenetTwin research and human genetics : the official journal of the International Society for Twin Studies, 10, 198-201, 2007	No confidence intervals were reported
Maiz,N., Staboulidou,I., Leal,A.M., Minekawa,R., Nicolaides,K.H., Ductus venosus Doppler at 11 to 13 weeks of gestation in the prediction of outcome in twin pregnancies, Obstetrics and Gynecology, 113, 860-865, 2009	Study did not report diagnostic accuracy data
Matias,A, Maiz,N, Montenegro,N, Nicolaides,K., Ductus venosus flow at 11-13 weeks in the prediction of birth weight discordance in monochorionic twins, Journal of Perinatal Medicine, 39, 467-470, 2011	The study examines whether ultrasound at 11-13 weeks’ gestation findings are predictive of discordant fetal growth in the

Study	Reason for exclusion
	second and third trimesters of pregnancy; no relevant diagnostic accuracy data were reported
Memmo,A, Dias,T, Mahsud-Dornan,S, Papageorghiou,A.T, Bhide,A, Thilaganathan,B., Prediction of selective fetal growth restriction and twin-to-twin transfusion syndrome in monochorionic twins, BJOG: An International Journal of Obstetrics and Gynaecology, 119, 417-421, 2012	No relevant reference standard
Miller,J, Chauhan,S.P, Abuhamad,A.Z., Discordant twins: Diagnosis, evaluation and management, American Journal of Obstetrics and Gynecology, 206, Oct-20, 2012	A narrative review about discordant growth among non-anomalous twins (definitions, risk factors, evaluation and management strategies)
Morin, L, Lim, K, Diagnostic Imaging, Committee, Special, Contributor, Genetics, Committee, Maternal Fetal Medicine, Committee, Ultrasound in twin pregnancies, Journal of Obstetrics & Gynaecology Canada: JOGCJ Obstet Gynaecol Can, 33, 643-656, 2011	The article presents the clinical practice guideline on ultrasound use in twin pregnancies
Mundy,D., Heitmann,E., Maulik,D., Umbilical Artery Doppler in the Assessment of Fetal Growth Restriction, Clinics in Perinatology, 38, 65-82, 2011	A narrative review about the doppler velocimetry of the umbilical artery as a fetal monitoring tool
Nabhan,A.F., Abdelmoula,Y.A., Amniotic fluid index versus single deepest vertical pocket as a screening test for preventing adverse pregnancy outcome, Cochrane Database of Systematic Reviews, #2008. Article Number, -, 2008	Population is singleton pregnancies
Neilson, J. P., Detection of the small-for-gestational age twin fetus by a two-stage ultrasound examination schedule, Acta Geneticae Medicae et GemellologiaeActa Genet Med Gemellol (Roma), 31, 235-40, 1982	Study did not report diagnostic accuracy data
Neilson, J. P., Detection of the small-for-dates twin fetus by ultrasound, British Journal of Obstetrics & GynaecologyBr J Obstet Gynaecol, 88, 27-32, 1981	No biparietal diameter discordancy cut-off was reported
O'Brien, W. F., Knuppel, R. A., Scerbo, J. C., Rattan, P. K., Birth weight in twins: an analysis of discordancy and growth retardation, Obstetrics & Gynecology, 67, 483-6, 1986	No relevant index test
Ocer,F, Aydin,Y, Atis,A, Kaleli,S., Factors affecting the accuracy of ultrasonographical fetal weight estimation in twin pregnancies, Journal of Maternal-Fetal and Neonatal Medicine, 24, 1168-1172, 2011	The study measures the factors affecting the accuracy of fetal weight estimation by ultrasonography; no relevant diagnostic accuracy data were reported
Odibo, A. O, Cahill, A. G, Goetzinger, K. R, Harper, L. M, Tuuli, M. G, Macones, G. A., Customized growth charts for twin gestations to optimize identification of small-for-gestational age fetuses at risk of intrauterine fetal death, Ultrasound in Obstetrics & Gynecology, 41, 637-42, 2013	The study compares the association between small for gestational age and intrauterine fetal death in twins using customized growth charts designed for twin gestations vs those designed for singletons; no relevant diagnostic accuracy data were reported
Odibo, A.O., Preface. Prenatal screening and diagnosis, Clinics in Laboratory Medicine, 30, 15-16, 2010	Editorial of special edition of Clinics in Laboratory Medicine Journal. Not a research article
Ong, S., Smith, A. P., Fitzmaurice, A., Campbell, D., Estimation of fetal weight in twins: a new mathematical	Included in Leombroni 2017 review

Study	Reason for exclusion
model, British Journal of Obstetrics & Gynaecology, 106, 924-8, 1999	
Palmer,K., Delpachitra,P., Onwude,J., Rombauts,L., Meagher,S., Tong,S., Association between twin discordance at 6-9 weeks' of gestation and birthweight complications, Twin Research and Human Genetics, 13, 389-392, 2010	Not a prediction study
Puccio, G, Giuffre, M, Piccione, M, Piro, E, Malerba, V, Corsello, G., Intrauterine growth pattern and birthweight discordance in twin pregnancies: a retrospective study, Italian Journal of PediatricsItal, 40, 43, 2014	The study describes the epidemiological characteristics at birth of twins to enhance the understanding of weight discordance
Queiros, A, Blickstein, I, Valdoleiros, S, Felix, N, Cohen, A, Simoes, T., Prediction of birth weight discordance from fetal weight estimations at 21-24 weeks' scans in monochorionic and dichorionic twins, Journal of Maternal-Fetal & Neonatal Medicine, 30, 1944-1947, 2017	Definition of the estimated fetal weigh is not according to the protocol
Reberdao, M. A, Martins, L, Torgal, M, Viana, R, Seminova, T, Casal, E, Hermida, M, Blickstein, I., The source of error in the estimation of intertwin birth weight discordance, Journal of Perinatal Medicine, 38, 671-4, 2010	Relevant diagnostic accuracy measures were reported in a figure and not as row numbers, therefore it was not possible to calculate them
Robert Peter, J., Ho, J. J., Valliapan, J., Sivasangari, S., Symphysial fundal height (SFH) measurement in pregnancy for detecting abnormal fetal growth, Cochrane Database Syst RevThe Cochrane database of systematic reviews, 2009	Protocol of an ongoing Cochrane review
Roberts, W. E., Gnam, E. C., 3rd, Magann, E. F., Martin, J. N., Jr., Morrison, J. C., Labor and membrane rupture in twin gestation. Can they affect the ability to estimate fetal weight?, Journal of Reproductive Medicine, 46, 462-6, 2001	Included in Leombroni 2017 review
Ropacka-Lesiak, M, Breborowicz, G, Dera, A., Blood flow changes in dichorionic twins with growth discordance, Twin Research & Human Genetics: the Official Journal of the International Society for Twin Studies, 15, 781-7, 2012	The study evaluates the usefulness of doppler ultrasonography in the diagnosis of twin pregnancies complicated by discordant fetal growth
Saldana,L.R., Eads,M.C., Schaefer,T.R., Umbilical blood waveforms in fetal surveillance of twins, American Journal of Obstetrics and Gynecology, 157, 712-715, 1987	No diagnostic accuracy data were reported
Salihu, H. M., Aliyu, M. H., Kirby, R. S., In utero nicotine exposure and fetal growth inhibition among twins, Am J PerinatolAmerican journal of perinatology, 22, 421-7, 2005	The study not focused on diagnostic accuracy but on the association between antenatal smoking and fetal growth inhibition
Salomon,L.J., Cavicchioni,O., Bernard,J.P., Duyme,M., Ville,Y., Growth discrepancy in twins in the first trimester of pregnancy, Ultrasound in Obstetrics and Gynecology, 26, 512-516, 2005	No relevant diagnostic accuracy data were reported
Sebire,N.J., D'Ercole,C., Soares,W., Nayar,R., Nicolaides,K.H., Intertwin disparity in fetal size in monochorionic and dichorionic pregnancies, Obstetrics and Gynecology, 91, 82-85, 1998	No diagnostic accuracy data were reported
Secher, N. J., Kaern, J., Hansen, P. K., Intrauterine growth in twin pregnancies: prediction of fetal growth retardation, Obstet GynecolObstetrics and gynecology, 66, 63-8, 1985	Estimated fetal weight was obtained from a formula including abdominal diameter and not circumference measurement

Study	Reason for exclusion
Simoes, T, Julio, C, Cordeiro, A, Cohen, A, Silva, A, Blickstein, I., Abdominal circumference ratio for the diagnosis of intertwin birth weight discordance, Journal of Perinatal Medicine, 39, 43-6, 2011	Included in Leombroni 2017 review
Snijders, R., First-trimester ultrasound, Clinics in Perinatology, 28, 333-52, viii, 2001	Narrative article about the first-trimester ultrasound screening for chromosome defects
Spong, C. Y., Scherer, D. M., Ghidini, A., Pezzullo, J. C., Salafia, C. M., Eglinton, G. S., Midtrimester amniotic fluid tumor necrosis factor-alpha does not predict small-for-gestational-age infants, Am J Reprod ImmunolAmerican journal of reproductive immunology (New York, N.Y. : 1989), 37, 236-9, 1997	No diagnostic accuracy data were reported
Sun, W., Liu, J., Zhang, Y., Cai, A., Yang, Z., Zhao, Y., Wang, Y., Cao, Z., Wei, Q., Quantitative assessment of placental perfusion by three-dimensional power Doppler ultrasound for twins with selective intrauterine growth restriction in one twin, Eur J Obstet Gynecol Reprod Biol, 226, 15-20, 2018	No relevant comparison
Stagnati, V, Pagani, G, Fichera, A, Prefumo, F., Intertwin discrepancy in middle cerebral artery peak systolic velocity and third-trimester fetal growth restriction in monochorionic-diamniotic twin pregnancy, Ultrasound in Obstetrics & Gynecology, 48, 66-71, 2016	No relevant index test, that is middle cerebral artery peak systolic velocity
Stirrup, O. T, Khalil, A, D'Antonio, F, Thilaganathan, B, Southwest Thames Obstetric Research, Collaborative, Fetal growth reference ranges in twin pregnancy: analysis of the Southwest Thames Obstetric Research Collaborative (STORK) multiple pregnancy cohort, Ultrasound in Obstetrics & Gynecology, 45, 301-7, 2015	The study presents reference charts for expected fetal growth in twin pregnancies and compares these with those of singleton pregnancies
Stirrup, O. T, Khalil, A, D'Antonio, F, Thilaganathan, B, Stork,, Patterns of Second- and Third-Trimester Growth and Discordance in Twin Pregnancy: Analysis of the Southwest Thames Obstetric Research Collaborative (STORK) Multiple Pregnancy Cohort, Fetal Diagnosis & Therapy, 41, 100-107, 2017	The study examines patterns of intertwin discordance in abdominal circumference and estimated fetal weight across the second and third trimesters in twin pregnancies; no relevant diagnostic accuracy data were reported
Tchirikov, M.,, Centre for Reviews and Dissemination., Ultrasound screening in pregnancy: a systematic review of the clinical effectiveness, cost-effectiveness and women's views (Structured abstract) , Database of Abstracts of Reviews of Effects, 3, 2010	Structured abstract of a systematic review of use of ultrasound for the detection of fetal abnormalities
Townsend, R., Khalil, A., Fetal growth restriction in twins, Best Practice and Research: Clinical Obstetrics and Gynaecology, 49, 79-88, 2018	Narrative article about diagnosis, classification and management of fetal growth restriction
Valsky, D. V., Eixarch, E., Martinez, J. M., Crispi, F., Gratacos, E., Selective intrauterine growth restriction in monochorionic twins: pathophysiology, diagnostic approach and management dilemmas, Seminars In Fetal & Neonatal Medicine, 15, 342-8, 2010	A full-text copy of the article could not be obtained
Valsky,D.V, Eixarch,E, Martinez,J.M, Gratacos,E., Selective intrauterine growth restriction in monochorionic diamniotic twin pregnancies, Prenatal Diagnosis, 30, 719-726, 2010	Narrative review on some of the aspects of the pathophysiology of selective intrauterine growth restriction in monochorionic twins

Study	Reason for exclusion
	and its implications for the diagnosis and clinical presentation
Van Mieghem, T, Eixarch, E, Gucciardo, L, Done, E, Gonzales, I, Van Schoubroeck, D, Lewi, L, Gratacos, E, Deprest, J., Outcome prediction in monochorionic diamniotic twin pregnancies with moderately discordant amniotic fluid, Ultrasound in Obstetrics & Gynecology, 37, 15-21, 2011	No relevant population as all women had moderately amniotic fluid discordance at the beginning of the study
Van Mieghem, T., Deprest, J., Klaritsch, P., Gucciardo, L., Done, E., Verhaeghe, J., Lewi, L., Ultrasound prediction of intertwin birth weight discordance in monochorionic diamniotic twin pregnancies, Prenatal Diagnosis, 29, 240-4, 2009	Included in Leombroni 2017 review
Vivanti, A. J, Lecarpentier, E, Cordier, A. G, Proulx, F, Tsatsaris, V, Benachi, A., Relevance of routine Doppler sampling at the two umbilical arteries in the follow-up of dichorionic twin pregnancies with intrauterine growth-restricted fetuses, Journal of Gynecology Obstetrics and Human Reproduction, 46, 285-289, 2017	Non relevant population as it includes twin pregnancies complicated by intrauterine growth restriction at the beginning of the study
Watson, W. J., Valea, F. A., Seeds, J. W., Sonographic evaluation of growth discordance and chorionicity in twin gestation, American Journal of Perinatology, 8, 342-4, 1991	No diagnostic accuracy data were reported
Weissman, A, Matanes, E, Drugan, A., Accuracy of ultrasound in estimating fetal weight and growth discordancy in triplet pregnancies, Journal of Perinatal Medicine, 44, 223-7, 2016	The study evaluates the accuracy of fetal weight estimation in triplet pregnancies; no relevant diagnostic accuracy data were reported
Weissmann-Brenner, A, Weisz, B, Achiron, R, Shrim, A., Can discordance in CRL at the first trimester predict birth weight discordance in twin pregnancies?, Journal of Perinatal Medicine, 40, 489-93, 2012	Not possible to calculate 2x2 table
Wilbacher, I, Soares-Weiser, K, Kleijnen, J, Schiller-Fruehwirth, I, Puig, S, Bernardis, D, Endel, G, Systematic review: diagnostic accuracy and outcomes of ultrasound in the first trimenon of pregnancy for detection of complications relevant for Austrian population excluding the screening for Down Syndrom [Down's syndrome] (Structured abstract), European journal of public health, 17, 233, 2007	Conference abstract
Zipori, Y, Reidy, K, Gilchrist, T, Doyle, L. W, Umstad, M. P., The Outcome of Monochorionic Diamniotic Twins Discordant at 11 to 13+6 Weeks' Gestation, Twin Research & Human Genetics: the Official Journal of the International Society for Twin Studies, 19, 692-696, 2016	The study examines the ability of nuchal translucency and crown rump length discordances among twins to predict adverse fetal outcomes (combined)
Aksam, S., Plesinac, S., Dotlic, J., Tadic, J., Vrzic-Petronijevic, S., Petronijevic, M., Kocijancic-Belovic, D., Buzadzic, S., First trimester ultrasonographic parameters in prediction of the course and outcome of monochorionic twin pregnancies, Turkish Journal of Medical Sciences, 47, 934-941, 2017	No confidence intervals were reported

Economic studies

No economic evidence was identified for this review.

Appendix L – Research recommendations

No research recommendations were made for this review.