Cochrane Library

Cochrane Database of Systematic Reviews

Ultrasound for diagnosis of birth weight discordance in twin pregnancies (Review)

Jahanfar S, Ho JJ, Jaafar SH, Abraha I, Noura M, Ross CR, Pammi M

Jahanfar S, Ho JJ, Jaafar SH, Abraha I, Noura M, Ross CR, Pammi M.
Ultrasound for diagnosis of birth weight discordance in twin pregnancies.
Cochrane Database of Systematic Reviews 2021, Issue 3. Art. No.: CD012553.
DOI: 10.1002/14651858.CD012553.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

ABSTRACT 1
PLAIN LANGUAGE SUMMARY 2
SUMMARY OF FINDINGS 3
BACKGROUND 5
OBJECTIVES 7
METHODS 7
RESULTS 9
Figure 1. 10
Figure 2. 11
Figure 3. 12
Figure 4. 14
Figure 5. 15
Figure 6. 16
Figure 7. 16
Figure 8. 17
Figure 9. 18
Figure 10. 19
Figure 11. 19
DISCUSSION 19
AUTHORS' CONCLUSIONS 20
ACKNOWLEDGEMENTS 20
REFERENCES 22
CHARACTERISTICS OF STUDIES 28
DATA 83
Test 1. Cut-off 20% 84
Test 2. 20% abdominal circumference 84
Test 3. 20% femur length 84
Test 4. Cut-off 25% 85
Test 5. 25% abdominal circumference 85
Test 6. 25% femur length 85
ADDITIONAL TABLES 85
APPENDICES 86
HISTORY 92
CONTRIBUTIONS OF AUTHORS 92
DECLARATIONS OF INTEREST 92
SOURCES OF SUPPORT 93
DIFFERENCES BETWEEN PROTOCOL AND REVIEW 93
INDEX TERMS 93

[Diagnostic Test Accuracy Review]

Ultrasound for diagnosis of birth weight discordance in twin pregnancies

Shayesteh Jahanfar[1], Jacqueline J Ho[2], Sharifah Halimah Jaafar[3], Iosief Abraha[4], Mohaddesseh Noura[5], Cassandra R Ross[6], Mohan Pammi[7]

[1]MPH Program, Department of Public Health and Community Medicine, Tufts University School of Medicine, Michigan, USA. [2]Department of Paediatrics, RCSI & UCD Malaysia Campus (formerly Penang Medical College), George Town, Malaysia. [3]Department of Obstetrics and Gynaecology, Regency Specialist Hospital, Johor Bahru, Malaysia. [4]Servizio Immunotrasfusionale, Azienda Unita' Sanitaria Locale Umbria 2, Foligno (PG), Italy. [5]Department of Midwifery/Nursing, Golestan University of Medical Sciences, Gorgan, Iran. [6]School of Health Sciences, Central Michigan University, Mt. Pleasant, Michigan, USA. [7]Section of Neonatology, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA

Contact: Shayesteh Jahanfar, jahanfar2000@yahoo.com, jahan2s@cmich.edu.

Editorial group: Cochrane Pregnancy and Childbirth Group.
Publication status and date: New, published in Issue 3, 2021.

Citation: Jahanfar S, Ho JJ, Jaafar SH, Abraha I, Noura M, Ross CR, Pammi M. Ultrasound for diagnosis of birth weight discordance in twin pregnancies. *Cochrane Database of Systematic Reviews* 2021, Issue 3. Art. No.: CD012553. DOI: 10.1002/14651858.CD012553.pub2.

ABSTRACT

Background

There is a need to standardize monitoring in obstetric research of twin pregnancies. Identification of birth weight discordance (BWD), defined as a difference in the birth weights of twins, is a well-documented phenomenon in twin pregnancies. Ultrasound for the diagnosis of BWD informs complex decision making including whether to intervene medically (via laser photo coagulation) or deliver the twins to avoid fetal morbidities or even death. The question is, how accurate is this measurement?

Objectives

To determine the diagnostic accuracy (sensitivity and specificity) of ultrasound estimated fetal weight discordance (EFWD) of 20% and 25% using different estimated biometric ultrasound measurements compared with the actual BWD as the reference standard in twin pregnancies.

Search methods

The search for this review was performed on 15 March 2019. We searched CENTRAL, MEDLINE (Ovid), Embase (Ovid), seven other databases, conference proceedings, reference lists and contacted experts. There were no language or date restrictions applied to the electronic searches, and no methodological filters to maximize sensitivity.

Selection criteria

We selected cohort-type studies with delayed verification that evaluated the accuracy of biometric measurements at ultrasound scanning of twin pregnancies that had been proposed for the diagnosis of estimated BWD, compared to BWD measurements after birth as a reference standard. In addition, we only selected studies that considered twin pregnancies and applied a reference standard for EFWD for the target condition of BWD.

Data collection and analysis

We screened all titles generated by electronic database searches. Two review authors independently assessed the abstracts of all potentially relevant studies. We assessed the identified full papers for eligibility, and extracted data to create 2 × 2 tables. Two review authors independently performed quality assessment using the QUADAS-2 tool. We excluded studies that did not report data in sufficient

detail to construct 2 × 2 tables, and where this information was not available from the primary investigators. We assessed the certainty of the evidence using the GRADE approach.

Main results

We included 39 eligible studies with a median study sample size of 140. In terms of risk of bias, there were many unclear statements regarding patient selection, index test and use of proper reference standard. Twenty-one studies (53%) were of methodological concern due to flow and timing. In terms of applicability, most studies were of low concern.

Ultrasound for diagnosis of BWD in twin pregnancies at 20% cut-off

Twenty-two studies provided data for a BWD of 20% and the summary estimate of sensitivity was 0.51 (95% CI 0.42 to 0.60), and the summary estimate of specificity was 0.91 (95% CI 0.89 to 0.93) (8005 twin pregnancies; very low-certainty evidence).

Ultrasound for diagnosis of BWD in twin pregnancies at 25% cut-off

Eighteen studies provided data using a BWD discordance of 25%. The summary estimate of sensitivity was 0.46 (95% CI 0.26 to 0.66), and the summary estimate of specificity was 0.93 (95% CI 0.89 to 0.96) (6471 twin pregnancies; very low-certainty evidence).

Subgroup analyses were possible for both BWD of 20% and 25%. The diagnostic accuracy did not differ substantially between estimation by abdominal circumference and femur length but femur length had a trend towards higher sensitivity and specificity. Subgroup analyses were not possible by sex of twins, chorionicity or gestational age due to insufficient data.

Authors' conclusions

Very low-certainty evidence suggests that EFWD identified by ultrasound has low sensitivity but good specificity in detecting BWD in twin pregnancies. There is uncertain diagnostic value of EFWD; this review suggests there is insufficient evidence to support this index as the sole measure for clinical decision making to evaluate the prognosis of twins with growth discordance. The diagnostic accuracy of other measures including amniotic fluid index and umbilical artery Doppler resistive indices in combination with ultrasound for clinical intervention requires evaluation. Future well-designed studies could also evaluate the impact of chorionicity, sex and gestational age in the diagnostic accuracy of ultrasound for EFWD.

PLAIN LANGUAGE SUMMARY

Ultrasound during pregnancy for predicting differences in birth weight between twins

Background

Birth weight differences of more than 20% in twins is associated with poor outcomes for the mother and baby. Clinicians measure the estimated fetal weight differences by ultrasound before birth and compare it to differences in birth weight after the babies are born. In this review, we summarized data on whether the ultrasound measurements are accurate enough to predict birth weight differences in twins.

Study characteristics

We searched medical databases to March 2019 for studies comparing ultrasound measurements to birth weight differences and we identified 39 studies. Twenty-two studies provided data on birth weight differences of 20% and 18 studies provided data on birth weight differences of 25%.

Quality of the evidence

We assessed the quality of individual studies using a tool called "Quality Assessment of Diagnostic Accuracy Studies" (QUADAS-2) and the overall quality by a recommended method called GRADE to find out the reliability of the evidence.

Key results

We found that ultrasound estimation of fetal weight differences compared to birth weight differences was not reliable. On average, ultrasound detected birth weight differences of 20% and 25% only half the time. The quality of evidence was very low.

There is insufficient evidence to support the use of ultrasound as the sole measure for detecting birth weight differences in twins, or poor outcomes. The diagnostic accuracy of other measures including amniotic fluid volume (the fluid surrounding the babies in the womb) or Doppler studies (which use sound waves to detect the movement of blood in the babies' blood vessels and the umbilical cord) in combination with ultrasound to inform clinical decisions needs to be evaluated. Future well-designed studies could also research the impact of whether the babies share a placenta (or not), the sex of the babies, and gestational age (time from woman's last menstrual period), in the diagnostic accuracy of ultrasound for estimated birth weight differences.

Summary of findings 1. Ultrasound for diagnosis of birth weight discordance in twin pregnancies at 20% cut-off

Sensitivity	0.51 (95% CI 0.42 to 0.60)		**Prevalences[a]**	18%			15%			28%	
Specificity	0.91 (95% CI 0.89 to 0.93)										
Outcome	**No. of studies and participants**	**Study design**	**Factors that may decrease certainty**					**Effect per 1000 women tested**			
			Risk of bias	**Indirectness**	**Inconsistency**	**Imprecision**	**Publication bias**	**Pretest probability of 18%**	**Pretest probability of 15%**	**Pretest probability of 28%**	**Test accuracy (certainty of the evidence)**
True positives (women with diagnosis of birth weight discordance)	22 studies, 1462 participants	Cohort-type studies with delayed verification (cohort type accuracy study)	Very serious[b]	Not serious	Very serious[c]	Not serious	None	92 (76 to 108)	77 (63 to 90)	143 (118 to 168)	⊕⊖⊖⊖ **Very low** [d]
False negatives (women incorrectly classified as not having diagnosis of birth weight discordance)								88 (72 to 104)	73 (60 to 87)	137 (112 to 162)	
True negatives (women without diagnosis of birth weight discordance)	22 studies, 6453 participants	Cohort-type studies with delayed verification (cohort type accuracy study)	Very serious[b]	Not serious	Very serious[c]	Not serious	None	746 (730 to 763)	774 (757 to 791)	655 (641 to 670)	⊕⊖⊖⊖ **Very low** [d]
False positives (women incorrectly classified as having diagnosis of birth weight discordance)								74 (57 to 90)	76 (59 to 93)	65 (50 to 79)	

CI: confidence interval.

[a]The prevalence used to represent the pretest probability are the median, first quartile and third quartile of the prevalences of included studies.

[b]In more than 50% of the studies there were unclear statements regarding index test, use of proper reference standard and flow and timing elements; in 1/3 of the studies, it was unclear how the participants were selected.

[c]Very high unexplained heterogeneity in terms of sensitivity ranging from 0.16 to 1.00.

[d]GRADE certainty of evidence downgraded one level for risk of bias and two levels for inconsistency.

Summary of findings 2. Ultrasound for diagnosis of birth weight discordance in twin pregnancies at 25% cut-off

Sensitivity	0.46 (95% CI 0.26 to 0.66)		**Prevalences[a]**	19%			9%			27%	
Specificity	0.93 (95% CI 0.89 to 0.96)										
Outcome	**No. of studies and participants**	**Study design**	**Factors that may decrease certainty**					**Effect per 1000 women tested**			
			Risk of bias	**Indirectness**	**Inconsistency**	**Imprecision**	**Publication bias**	**Pretest probability of 19%**	**Pretest probability of 9%**	**Pretest probability of 27%**	**Test accuracy (certainty of the evidence)**
True positives (women with diagnosis of birth weight discordance)	18 studies, 1679 participants	Cohort-type studies with delayed verification (cohort type accuracy study)	Very serious[b]	Not serious	Very serious[c]	Not serious	None	87 (49 to 125)	41 (23 to 59)	124 (70 to 178)	⊕⊖⊖⊖ **Very low** [d]
False negatives (women incorrectly classified as not having diagnosis of birth weight discordance)								103 (65 to 141)	49 (31 to 67)	146 (92 to 200)	
True negatives (women without diagnosis of birth weight discordance)	18 studies 4792 participants	Cohort-type studies with delayed verification (cohort type accuracy study)	Very serious[b]	Not serious	Very serious[c]	Not serious	None	753 (721 to 778)	846 (810 to 874)	679 (650 to 701)	⊕⊖⊖⊖ **Very low** [d]
False positives (women incorrectly classified as having diagnosis of birth weight discordance)								57 (32 to 89)	64 (36 to 100)	51 (29 to 80)	

CI: confidence interval.

[a]The prevalence used to represent the pretest probability are the median, first quartile and third quartile of the prevalences of included studies.

[b]At least 50% of the studies had unclear statements regarding index test, use of proper reference standard and flow and timing elements.

[c]Very high unexplained heterogeneity in terms of sensitivity ranging from 0.1 to 1.00.

[d]GRADE certainty of evidence downgraded one level for risk of bias and two levels for inconsistency.

BACKGROUND

Target condition being diagnosed

Birth weight discordance (BWD), defined as a difference in the birth weights of twins, is a well-documented phenomenon in twin pregnancies (Bagchi 2006; Mahony 2006). Estimated fetal weight discordance (EFWD) estimated by ultrasound is measured using the formula: ((larger twin estimated weight – smaller twin estimated weight)/larger twin estimated weight) × 100. EFWD of 20% or greater has been shown to have significant impact on fetal and perinatal outcomes. Growth discordance may occur as a physiological BWD, when both twins are appropriate-for-gestational-age, or as pathological BWD when at least one twin is small-for-gestational age (Appleton 2007).

BWD is one of the major risk factors for adverse fetal, neonatal and maternal outcomes (Demissie 2002; Kilic 2006; Wen 2006). Demissie and colleagues found that the odds ratio of fetal death varied from 1.26 to 12.75 and the odds ratio of neonatal death varied from 1.02 to 3.43, depending on the degree of BWD. Kilic and colleagues reported higher frequency of mortality, sepsis, polycythaemia, hypoglycaemia, anaemia and respiratory distress syndrome among discordant twins. Wen and coworkers reported higher odds of maternal hypertension, eclampsia and other medical complications associated with BWD. Depending on the chorionicity, degree of discordance and the threshold used, BWD occurs in about 15% to 30% of twin gestations (Lewi 2008; Lopriore 2012; Miller 2012). Fetal growth abnormalities are associated with multiple conditions that affect fetal or neonatal well-being, such as twin-to-twin transfusion syndrome (TTTS), chromosomal aberrations or structural defects. Excluding these conditions, discordant growth by itself is recognized as an independent risk factor for adverse perinatal outcomes (Mazhar 2010; Morin 2011; Suzuki 2009). BWD of 20% and 25% are the most common thresholds used in the literature to identify adverse perinatal outcomes (Jahanfar 2016a). One study suggested that BWD of 30% and greater is significantly associated with adverse perinatal outcomes irrespective of chorionicity (Jahanfar 2016b).

Several mechanisms have been proposed for growth discordance of fetuses exposed to the same intrauterine environment (Victoria 2001). For monochorionic (MC) twin pregnancies, the mechanism is explained through conditions such as TTTS and intrauterine growth retardation (IUGR). These two conditions result in greater risk of fetal and neonatal mortality compared to dichorionic (DC) twins. Technology is available to manage TTTS including amnio-reduction and septotomy (Behrendt 2016), hence monitoring is usually started as early as 20 weeks.

In the case of DC twin pregnancies, intrauterine growth restriction and placenta pathology can cause growth discordance. Abnormal cord insertion and insufficient placental implantation in the uterine wall are other proposed mechanisms for BWD. These conditions are non-treatable; hence, the main management is to determine fetal well-being and decide the optimal time for delivery to save the growth discordant twins.

If the accuracy of ultrasound in assessing the 20% to 25% cut-off is known, it helps the complex decision making that the clinician goes through in deciding whether to intervene medically (via laser photo coagulation) or deliver the twins to avoid fetal morbidities or even death.

Index test(s)

Ultrasound has been an important tool in the diagnosis of BWD since 1972 (Hoopmann 2011; Woo 1939). Diagnostic ultrasound is a sophisticated electronic technology, which utilizes pulses of high-frequency sound to produce an image. In the past, before ultrasound became available, abdominal palpation was used, which is extremely poor for detecting growth discordance. Radiological examination is not recommended since it is not safe for the fetus.

Diagnosis of EFWD can be made via two modalities of ultrasound.

1. A series of diagnostic two-dimensional (2D) ultrasound examinations is used to assess fetal growth of both twins; identify chorionicity; and diagnose problems with cord, amnion layers, congenital abnormalities and TTTS. Doppler ultrasound is used to detect abnormal blood flow patterns in fetal/placenta circulation which may indicate poor fetal prognosis (Alfirevic 2003).
2. Three-dimensional (3D) ultrasound to facilitate the assessment of the placenta, such as surface-rendered imaging and volume measurement, quantitative and qualitative assessments of the vascularization, and blood flow of the placenta (Hata 2011).

The standard diagnostic test for fetal growth discordance is ultrasound. Growth discordance can be estimated by measuring crown rump length (CRL: the distance measured from the top of the head to the bottom of the buttocks) at or before 12 weeks of gestation (Tai 2007). Later, during the second and third trimesters, other ultrasound measures such as biparietal diameter (BPD), abdominal circumference (AC) and femur length (FL) are used to estimate fetal weight and calculate the degree of BWD (Simoes 2011).

It is still unclear whether first or second trimester findings can accurately predict BWD, what sonographic parameters assessed at each trimester are reliable to assess for discordance or whether any particular ultrasound modality (e.g. 2D or 3D) is superior to the other.

Apart from EFWD, information on chorionicity is essential for the management of a twin pregnancy for the following reasons.

1. MC twins are at greater risk of fetal morbidity and mortality due to shared vascularization.
2. If monochorionicity is suspected at early ultrasound screening, subsequent screening should be serially scheduled during the second and third trimester.

Early ultrasound detection narrows down the differential diagnosis for the underlying causes such as placenta sharing and subsequent vascular anastomosis as an underlying cause of EFWD. Better prognosis is expected for DC twin pregnancies as they do not share placentas (Miller 2012). In simpler terms, MC twins have higher risk of complications and should be assessed separately.

Second trimester ultrasound screening, from 16 to 24 weeks' gestation is useful in measuring AC, amniotic fluid pockets, identification of dividing membrane between amniotic sacs and umbilical artery Doppler studies (Giles 1998; Rizzo 1993).

Third trimester ultrasound, after 24 weeks' gestation, aims at identifying discordance or insufficient fetal growth. Identification

of a discordant twin is crucial if one of the twin pairs is small-for-gestational age rather than both being appropriate-for-gestational-age.

It should be noted that there is overlap in the use of these diagnostic tests from one trimester to another.

The standard of practice is to identify the growth discordance in twin pregnancies using ultrasound (2D or 3D, or both) using measurements (CRL, BPD, AC or FL) in either early (for CRL) or late pregnancy (for BPD, AC and FL) and compare with the actual birth weight. Since EFWD of 20% and 25% are associated with adverse perinatal outcomes, we studied the diagnostic accuracy of ultrasound at those threshold values.

Clinical pathway

In the management of a twin pregnancy, evaluation of fetal growth is particularly important because growth restriction and prematurity are major causes of morbidity and mortality reported in twins compared with a singleton pregnancy (Adegbite 2004). Most clinicians start monitoring MC twins for growth delay and discordant growth from 18 weeks' gestation every two to three weeks. Each twin's growth, amniotic volume and fetal bladder volume are monitored for signs of oligohydramnios or polyhydramnios. in contrast, in DC twins, fetal growth monitoring by ultrasound usually begins after 20 weeks of gestation every four to six weeks as fetal growth deceleration leading to discordance is optimally detected between 20 and 28 weeks of gestation (Gonzalez-Quintero 2003).

When discordant growth is identified in the absence of TTTS or congenital abnormalities, intensive fetal monitoring of fetal well-being is employed until delivery becomes indicated. In mildly discordant twins, expectant management by frequent fetal assessment is preferable to preterm delivery, given the limitation of ultrasound diagnosis of growth discordance. In one large non-randomized study of 2399 twin pregnancies, the sonographic diagnosis of discordance greater than 25% was a poor predictor of BWD, fetal loss after 22 weeks' and 28 weeks' gestation, perinatal death and preterm birth before 34 weeks of gestation (D'Antonio 2014). In complicated twin pregnancies, in particular those with fetal growth restriction and discordant growth, fetal well-being assessment by a non-stress test (NST), biophysical profile (BPP), amniotic fluid index (AFI) measurement and Doppler velocimetry may be useful to identify which fetuses would benefit from early delivery (Devoe 1995).

Ideally, a diagnostic test is expected to correctly identify all patients with the assessed condition and to exclude all patients without it; that is, to have a sensitivity and specificity of 100%. In practice, however, it is extremely rare to find a test with equally high sensitivity and specificity, when compared to the current gold standard (BWD). For most tests, there is usually a trade-off between the measures. An approach to the test might vary depending on the clinical context, which relates to the consequences of missed diagnosis and magnitude of subsequent interventions if the test is positive. If clinical priority would be to avoid missed diagnosis, an adequate test in that case would be expected to have a high sensitivity (low false-negative results), with lower specificity (higher number of false-positive results). Thus, no woman with a growth discordant twin pregnancy is missed by a test, but some still get unnecessary interventions (intensive surveillance, early delivery, consequential issues pertaining to prematurity such as admission to the neonatal intensive care unit, costs associated with care for the premature baby and increased distress for parents). An alternative approach gives priority to a test that avoids unnecessary invasive interventions. In this scenario, the emphasis should be on a high specificity (close to 100%) with lower sensitivity (preferably above 50%), which will rule out growth discordance of the twins, but will not detect some women with this condition. Ruling out growth discordance eliminates the need for further clinical intervention. Ideally, a test with high sensitivity and specificity is most useful in a clinical setting, which will help in efficient patient counselling and ongoing pregnancy management. Improving the selection of patients who might benefit from interventions such as laser coagulation or amnioreduction or early delivery would be of considerable benefit to neonates, mothers, families and the community.

Rationale

Guidelines provide conflicting advice about the time, frequency and type of ultrasound measurements that reliably diagnose growth discordance. According to the American College of Obstetricians and Gynecologists (ACOG), growth discordance is defined as a difference of AC of 20 mm or estimated fetal weight (EFW) difference of 20% (ACOG Committee 2004). The Society of Obstetricians and Gynecologists of Canada recommends that the EFW be derived from bi-parietal diameter with AC or a combination of AC and FL (Okun 2000). ACOG does not provide recommendations on the frequency of ultrasound examinations for twins while the green guideline from the Royal College of Obstetricians and Gynecologists advocates screening ultrasound for MC twins every two to three weeks from 16 weeks' gestation onwards (NICE 2011).

For DC twins, the issues related to feto-placental perfusion leading to growth discordance can be clarified by screening ultrasound. However, there are no guidelines on the optimal gestation at which to start screening, or the frequency and gestations for subsequent screenings. Moreover, gender is an important factor that has interaction with BWD and perinatal outcomes (Melamed 2009; Miller 2012). Identifying the gender mix by ultrasound improves predictability of antenatal outcomes (Di 2007). Thus, the accuracy of ultrasound to identify gender is also crucial. As a part of this review, we investigated the impact of gender on BWD as a confounding factor.

The variation in recommendations made in the guidelines arises from the fact that the ability to estimate abnormal growth is challenging. It is not clear if the antenatal prediction of a critical BWD by fetal weight estimation formulae ((birth weight of heavier twin – birth weight of lighter twin)/birth weight of heavier twin) two to three weeks prior to delivery (Caravello 1997), provides an accurate estimation of BWD (Kalish 2003; Klam 2003; van Mieghem 2009). The pooled sensitivity for ultrasound prediction of BWD of 25% or greater, in three studies, was 63% for a false-positive rate of 2% (Caravello 1997; Diaz-Garcia 2010; Gernt 2001).

The existing uncertainties are further complicated by an inconclusive predictive value of early (first-second trimester) versus late detection (two to three weeks prior to delivery) of growth discordance, use of different sonographic estimates (AC versus FL), and reliance on the retrospective nature of study designs and small sample sizes of existing literature. Literature suggests

that biometric measurements of EFWD at early gestation or during the second trimester have significantly different precisions (Banks 2008; Tai 2007). Moreover, the efficacy of a single biometric measurement, such as CRL or AC with or without other measurement(s), in predicting BWD, is controversial (Banks 2008; Bhide 2009; Chamberlain 1991; Chitkara 1985). The most popular current methods for predicting discordant growth in twin gestations have limited accuracy when held to a standard for discordance that requires a birth weight difference of at least 20% (Caravello 1997), as sonography tends to underestimate the degree of discordance (Chang 2006).

We reviewed and summarized the evidence for the diagnostic accuracy of antenatal ultrasound for EFWD compared to the BWD to inform clinical practice.

OBJECTIVES

To determine the diagnostic accuracy (sensitivity and specificity) of ultrasound estimated fetal weight discordance (EFWD) of 20% and 25% using different estimated biometric ultrasound measurements compared with the actual BWD as the reference standard in twin pregnancies.

Secondary objectives

To explore heterogeneity related to gestational age, sex and chorionicity. We planned to perform subgroup analyses for:

1. assessing the sensitivity and specificity of available diagnostic ultrasound tests in subgroups of twin pregnancies at various gestational ages by week (less than 28 weeks, 28 to 32 weeks, 32 to 36 weeks and more than 36 weeks);
2. assessing the sensitivity and specificity of each diagnostic ultrasound test in twin pregnancies with same-sex versus opposite-sex twins;
3. assessing the sensitivity and specificity of each diagnostic ultrasound test in dichorionic diamniotic versus monochorionic diamniotic twins.

We anticipate the following potential sources of heterogeneity.

1. Clinical factors: characteristics of study population (gestational age, chorionicity, inclusion of high-risk pregnancy complicated by underlying maternal–fetal conditions, breastfeeding).
2. Methodological factors: study design (patient selection, prospective versus retrospective studies, time of test performance (time between index test and reference standard), clinical settings (tertiary centre versus community health care), multiple testing versus single testing for diagnosis of high-risk pregnancy complicated by maternal–fetal conditions).
3. Other factors: geographic area (high-, middle- and low-income countries), year of publication.

METHODS

Criteria for considering studies for this review

Types of studies

We included cohort-type studies with delayed verification design that assessed EFWD of 20% compared to BWD of 20% and EFWD of 25% compared to BWD of 25%. We included only studies that reported both the index test and reference standard.

Participants

Twin pregnancies with ultrasound measurements to determine EFWD at any stage of pregnancy. The twin pregnancies included any type of chorionicity, any type of conception, and any maternal age and body mass index. We excluded pregnancies with inappropriate comparisons that were likely to distort an assessment of the diagnostic value of antenatal ultrasound (e.g. pregnancies that includes twins with one stillbirth, or other multiple pregnancies such as triplets or quadruplets).

Index tests

Any type of biometric measurements assessing EFWD, when either used as a single measurement, combinations or defined formulas, inclusive of CRL, BPD, AC and FL. The measurements were eligible if performed using transabdominal or transvaginal ultrasound, machines of any brand based on 2D or 3D ultrasound methods.

Target conditions

The target condition was BWD in twin pregnancies, which occurs when there is a disparity in birth weight between the larger and smaller infants of a twin set. For this review, we included studies that had investigated growth discordance of 20% or greater and growth discordance of 25% or greater.

Reference standards

Ultrasound EFWD of 20% was used to diagnose a prespecified BWD of 20% and an ultrasound EFWD of 25% was used to diagnose a prespecified BWD of 25%. BWD was calculated using the formula: ((larger estimated weight – smaller estimated weight)/ larger estimated weight) × 100 and was categorized as those below or above a 20% threshold. Birth weight was accepted as estimated by using either electronic or mechanical scales of any type (bench-top, portable, hanging, compact or not identified) from any manufacturer. Measurements were considered if performed in the hospital (labour ward, nursery or newborn intensive care unit) by trained medical personnel (doctor, nurse, midwife, paramedic). We included only the measurements performed within seven days of birth. Where available, the data on scale calibration, and whether baby was wet or dried before weighing were recorded and incorporated in the QUADAS-2 quality assessment tool.

Search methods for identification of studies

We adopted a comprehensive search of multiple sources for eligible studies. We searched for unpublished and published studies using databases and conference proceedings as outlined below.

A librarian developed the search strategy following the recommendations in the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (de Vet 2013). The searches were not limited to particular types of study design or have language or publication date restrictions. The search strategy incorporated words in the title, abstract, text words across the record and the subject headings. The search strategies for major databases are presented in Appendix 1.

Electronic searches

We searched the following electronic databases from inception to March 2020.

- Cochrane Central Register of Controlled Trials (CENTRAL; 2020, Issue 3) via the Cochrane Library.
- MEDLINE via Ovid (from 1946).
- Embase via Ovid (from 1980).
- CINAHL (from 1982).
- ISI Web of Science Core Collection (from 1900).
- Trip Database (from 1997).
- PubMed Systematic Reviews subset (from 1946).
- DARE and NHS EED via the University of York (1994 to 2015).
- HTA (2003 to 2018).
- Prospero via the University of York (2011).

Searching other resources

Additional searches included:

1. a hand search of Australasian Journal of Ultrasound in Medicine (2009); Canadian Journal of Medical Sonography (2013), the reference lists of all the included studies and the seminal reviews from the field; and
2. communication with at least five experts in the field asking them to review our reference list and identify any studies that may be missing.

Data collection and analysis

We performed data extraction and handling, assessment of methodological quality and statistical analyses based on the recommendations of the Diagnostic Test Accuracy (DTA) group and their Internet-based tutorials (methods.cochrane.org/sdt/dta-author-training-online-learning).

Selection of studies

One review author scanned the titles of studies identified by our search to remove any clearly irrelevant articles and scanned the titles and abstracts of the remaining studies to select potentially relevant articles. Two review authors independently reviewed full-text versions of the articles selected by title and abstract and assessed their eligibility for inclusion. We resolved any disagreements by discussion and, if necessary, with a third review author, who was an expert in the field and in methodological aspects of Cochrane systematic reviews.

When we identified the reports that updated previous publications of the same study population at different recruitment points, the earlier records were classified as excluded. The most complete data set that superseded previous publications were used to avoid double counting participants or studies.

We retrieved missing data by contacting the authors of identified studies directly to clarify the study eligibility. Potentially relevant studies in languages other than English were translated where possible.

For excluded studies, we documented the reasons for exclusion with details of which criteria were not met. We produced Characteristics of included studies, Characteristics of excluded studies and Characteristics of studies awaiting classification tables.

A single failed eligibility criterion was sufficient for a study to be excluded from the review.

Data extraction and management

We used a structured, piloted form to extract data from included studies. Two review authors independently extracted study characteristics and resolved disagreements by discussion. If disagreements persisted, a third review author resolved the issues. We extracted information on: author, year of publication, journal; study design; timing of data collection (prospective, retrospective); setting (inpatients, outpatients); study population (age, parity, obesity); type of index test and reference standard and data on index and reference test operators. We used the reported number of true positives (TP), false negatives (FN), true negatives (TN) and false positives (FP) to construct a 2 × 2 table for each index test. If these values were not reported, we attempted to reconstruct the 2 × 2 tables from the summary estimates presented in the article. We entered data into Review Manager 5, and used this to graphically display the quality assessment, study level data as forest plots and summary estimates as summary receiver operating characteristics (SROC) plots with studies and summary points (Review Manager 2014).

Assessment of methodological quality

We used the Quality Assessment of Diagnostic Test Accuracy Studies-2 (QUADAS-2) to assess the methodological quality of included studies. The QUADAS-2 tool was applied in four phases: summarize the review question, tailor the tool and produce review-specific guidance, construct a flow diagram for the primary study, and judge bias and applicability (Whiting 2011). Each paper was judged at 'low', 'high' or 'unclear' risk of bias for each of four domains, and concerns about applicability were assessed in three domains. The review-specific QUADAS-2 tool and explanatory document are presented in Appendix 2. Two review authors independently piloted the review-specific tool to rate three of the included studies. The tool was utilized if there was a high level of agreement at the pilot stage. Similarly, two review authors independently applied the QUADAS-2 tool to the full text of each study. We resolved disagreements by discussion, or if needed, by a third review author. We used Review Manager 5 to construct methodological quality summary graphs (Review Manager 2014).

We assessed the certainty of evidence and reported it according to GRADE for diagnostic test studies (Schünemann 2019; Schünemann 2020a; Schünemann 2020b). We used GRADEpro software to generate 'Summary of findings' tables (GRADEpro GDT).

The prevalences for assessing pretest probability were the median, lower and higher quartiles of the prevalences of the target condition (BWD) in the included studies (Oerbekke 2020).

Statistical analysis and data synthesis

We performed statistical analyses following recommendations presented in Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Macaskill 2013).

The unit of analysis in the included studies was twin pregnancy, rather than twins themselves, as BWD is a single calculated measure estimated using biometric measurements of two fetuses. We extracted the absolute counts of TP, FP, FN and TN from each study. TP was defined as ultrasound positive for EFWD greater than or equal to 20% or 25% by biometric ultrasound measurements (CRL, AC, FL) confirmed by BWD of greater than or equal to 20% or 25% after birth. FP was defined as ultrasound positive for EFWD,

as defined above, without BWD after birth. FN was a negative ultrasound for EFWD of greater than or equal to 20% or 25% with the diagnosis of BWD of greater than or equal to 20 or 25%. Finally, TN was a negative ultrasound for EFWD of 20% or 25% and without BWD of 20% or 25% after birth.

Two review authors extracted the counts separately and discussed disagreements before engaging a third review author to resolve any disagreements. We then transferred the data into Review Manager 5 to produce plots and estimates (Review Manager 2014).

We used the counts (TP, FP, FN and TN) to construct 2 × 2 tables to estimate sensitivity and specificity with 95% confidence intervals (CIs) for each study. We plotted estimates of the sensitivities and specificities, with their 95% CIs, in forest plots (Chu 2006). The results of the index tests were dichotomous (positive or negative). The bivariate random-effects approach enabled us to calculate the summary estimates of sensitivity and specificity with 95% CI) (Deeks 2019; Macaskill 2013). The bivariate model also deals with variation beyond chance in sensitivity and specificity between studies and any correlation that may exist between sensitivity and specificity. We calculated summary estimates of sensitivity and specificity using 'xtmelogit' using 'metandi' and MIDAS packages (Dwamena 2009; Harbord 2007; Harbord 2009) for Stata (Stata Corp, College Station, TX). We generated forest plots with 95% CIs for sensitivity and specificity and SROC plots in Review Manager 5 (Review Manager 2014).

When a meta-analysis was possible, we planned to use test-level covariates in the bivariate logit-normal model to identify statistically significant differences but there were insufficient data for covariates of sex and chorionicity. We planned to stratify our analysis using categories of gestational age (below or equal to 24 weeks of gestation, below or equal to 32 weeks of gestation, below or equal to 37 weeks of gestation). However, the data were too variable, and the recorded gestational age was presented as ranges rather than one cut-off point for gestational age. Hence, we decided not to pool the data in respect of gestational age. We calculated positive (LR+) and negative (LR–) likelihood ratios by using summary sensitivity and specificity.

Investigations of heterogeneity

We investigated heterogeneity for the diagnostic tests where there were sufficient data (more than 10 included studies). Steps taken were as follows.

We planned to start the investigation of heterogeneity through visual examination of the forest plots by grouping according to covariates all the items listed as potential sources of heterogeneity (e.g. type of ultrasound, year of publication, geographic areas (high-income versus low- and middle-income countries), consecutive enrolment, blinding of the operators to clinical data, modifications applied to the widely accepted method of imaging techniques (such as 2D and 3D acquisition), number of index test operators and missing data. There were insufficient data to allow grouping according to the above covariates.

We planned to investigate heterogeneity by meta-regression if data were available for the covariates; type of ultrasound (2D, 3D), gender determination (yes/no), gestational age (continuous variable), type of study (prospective versus retrospective), chorionicity determination (yes/no), estimated BWD threshold (10% or greater versus 20% or greater) and time frame between index test and delivery. There were insufficient data to allow grouping according to the above covariates.

We planned to investigate methodological sources of heterogeneity using meta-regression analysis. These were to be listed as verification bias, incorporation bias, diagnostic review bias, and clinical review bias. Prespecified variables were inclusive of clinical settings (tertiary centre versus community health care) and high-risk pregnancy complicated by underlying maternal–fetal conditions (yes/no). Maternal complications included hypertensive disorders, gestational diabetes and antepartum haemorrhage. Fetal complications includes TTTS, major structural congenital anomalies, stillbirth of one twin and IUGR. We were unable to perform meta-regression analyses for the above-mentioned covariates due to lack of data.

Sensitivity analyses

We planned to conduct sensitivity analyses based on 'type of study design' (prospective versus retrospective study designs) and the individual quality items of QUADAS-2 tool, but sufficient data were not available.

Assessment of reporting bias

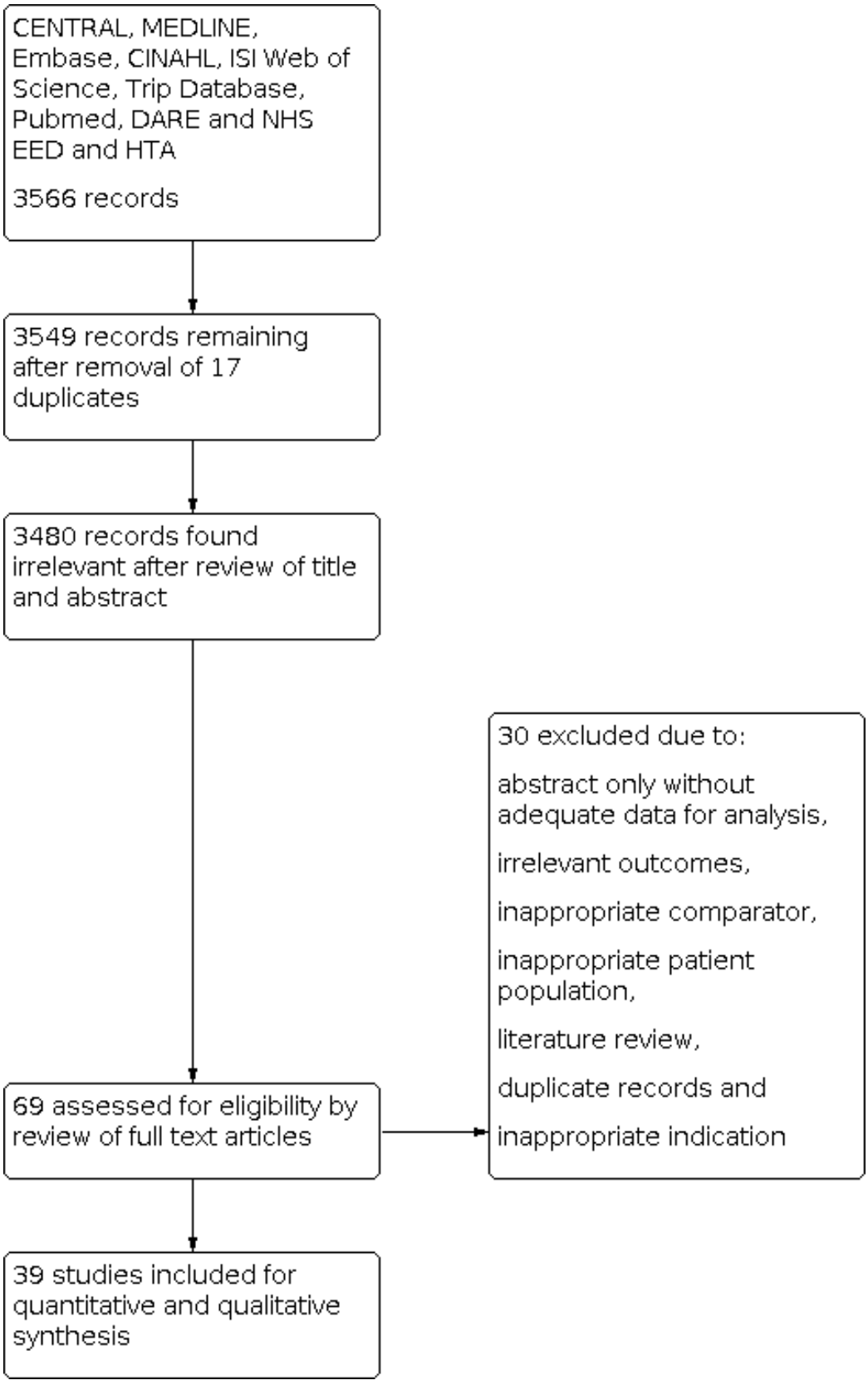
We did not plan to use funnel plots to evaluate the impact of publication bias or other biases associated with small studies, because, according to Leeflang and colleagues, the tests commonly used in interventional systematic reviews for publication bias are not useful for diagnostic testing reviews (Leeflang 2008). Deeks and colleagues verified that the use of an asymmetric effective sample size plot to detect publication bias lacks power in situations where sample variability is present (Deeks 2005). We attempted to use unpublished data to minimise reporting bias.

RESULTS

Results of the search

Results of database searches to March 2019 yielded 3566 records. Of these, we removed 17 duplicates and 3480 records at the title and abstract level (Figure 1). Of the remaining 69 articles, we excluded 30 at full paper screening due to a variety of reasons (abstract only, irrelevant outcomes, inappropriate comparator, inappropriate patient population, literature review, duplicate records and inappropriate indication; see Characteristics of excluded studies table). We finally included 39 studies, all of which evaluated the index test of EFWD by ultrasound measurements compared to the reference standard of actual birth weight measured at birth (Bimson 2012; Bimson 2014a; Bimson 2014b; Blickstein 1989; Blickstein 1996; Caravello 1997; Chamberlain 1991; Chang 2006; Chauhan 1995; Crane 1980; D'Antonio 2013a; D'Antonio 2014; Danon 2008; Diaz-Garcia 2010; Fox 2011; Hehir 2017; Hill 1994; Hoopmann 2011; Jensen 1995; Johansen 2014; Kalish 2003; Kim 2015; Klam 2005; Machado 2007; MacLean 1992; Murray 2014; Nakayama 2014; O'Connor 2013; Reberdao 2010; Sayegh 1993; Secher 1985; Shahshahan 2011; Simoes 2011; Storlazzi 1987; van de Waarsenburg 2015; van Mieghem 2009; Watson 1991; Zipori 2016; Zuckerwise 2015).

Figure 1. PRISMA flow diagram outlining the study selection process.



The list and details of the included studies are presented in the Characteristics of included studies table. The 39 eligible studies had a median sample size of 140 and range of 38 to 2161 women per study. Of these studies, 16 were conducted in Europe, seven in Asia, 14 in North America and two in other geographical areas. Nine studies were conducted at university hospitals, eight studies were at tertiary centres, 11 studies were conducted in hospitals, six studies were conducted at other settings and five studies did not indicate their settings. Three studies were published prior to 1990, eight were published between 1991 and 2000, eight between 2001 and 2010, and the remaining 14 after 2010. All studies used ultrasound by 2D, and none evaluated EFWD using 3D ultrasound.

Methodological quality of included studies

We assessed all 39 studies using the QUADAS-2 framework (Figure 2; Figure 3). All studies were observational cohort-type studies with delayed verification. From the studies that reported BWD, four received financial support. There were no studies whose authors declared a conflict of interest. Thirty-one studies gave no information on either of these two potential biases.

Figure 2. QUADAS-2 risk of bias and applicability concerns graph including review authors' judgements about each domain presented as percentages across included studies.

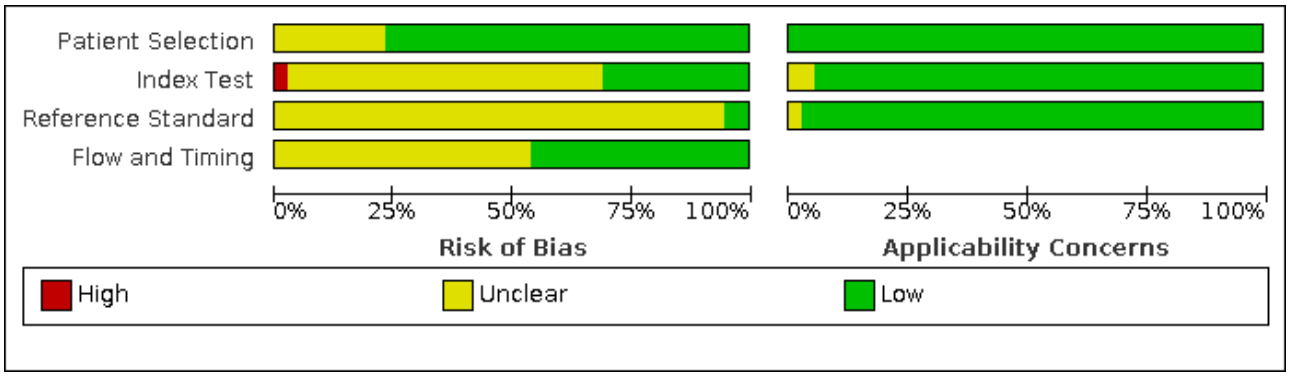
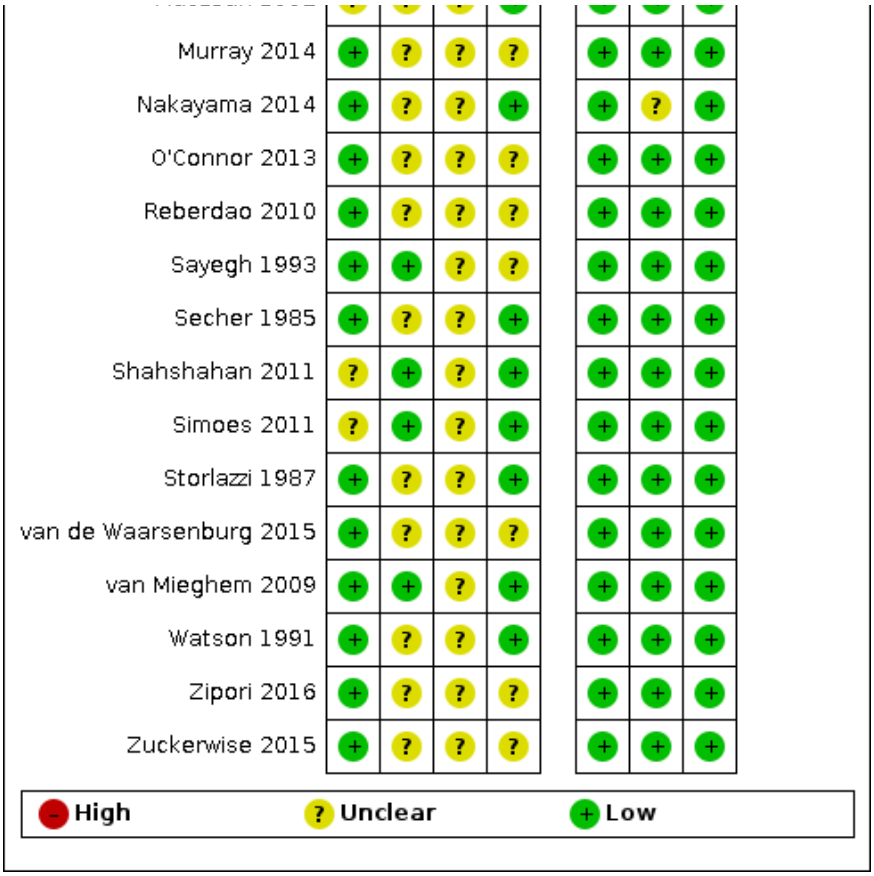


Figure 3. QUADAS-2 risk of bias and applicability concerns summary including review authors' judgements about each domain for each included study.

Figure 3. (Continued)



In terms of risk of bias, there were many unclear statements regarding patient selection, index test, use of proper reference standard, and flow and timing elements of the studies. In terms of applicability, most studies were high quality, and there was low concern. There was one study at high risk of bias based on the judgments made about the index test. There were no studies at high risk of bias based on the judgements made about the reference standard, and, in 37 (90%), the risk was unclear. Twenty-one studies (53%) were of methodological concern due to flow and timing.

We judged 36 studies (92%) at low concern of applicability concerns in all three domains (Figure 3). Two studies had an unclear risk of bias for the index test (Danon 2008; Nakayama 2014), and one study with an unclear risk of bias for a reference standard (Hoopmann 2011).

We used the GRADE approach to evaluate the certainty of the evidence. For both diagnostic performance of ultrasound at 20% and 25% BWD, the certainty of evidence was downgraded one level for risk of bias and two levels for inconsistency to give a final estimation of very low certainty of evidence.

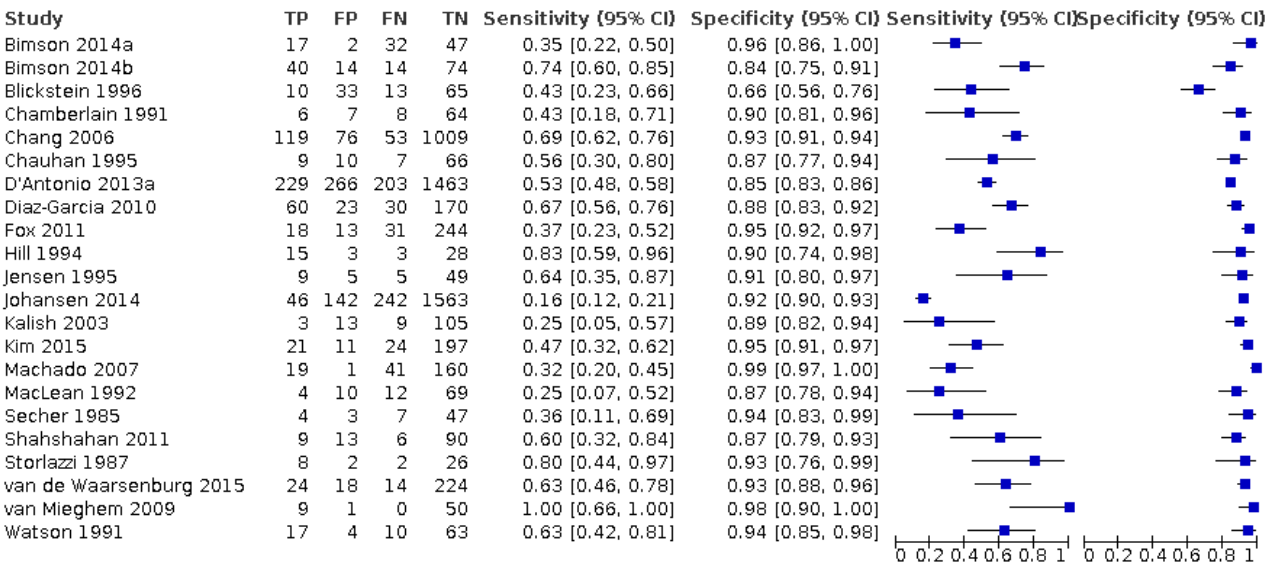
Findings

1. Diagnostic performance of ultrasound (20% difference in birth weight as reference)

Twenty-two studies that estimated EFWD by ultrasound measurements provided data for a discordance of equal to or greater than 20% (Bimson 2014a; Bimson 2014b; Blickstein 1996; Chamberlain 1991; Chang 2006; Chauhan 1995; D'Antonio 2013a; Diaz-Garcia 2010; Fox 2011; Hill 1994; Jensen 1995; Johansen 2014; Kalish 2003; Kim 2015; Machado 2007; MacLean 1992; Secher 1985; Shahshahan 2011; Storlazzi 1987; van de Waarsenburg 2015; van Mieghem 2009; Watson 1991). Median BWD of 20% prevalence across the studies was 18% (range 9% to 50%). The findings and certainty of evidence are detailed in Summary of findings 1.

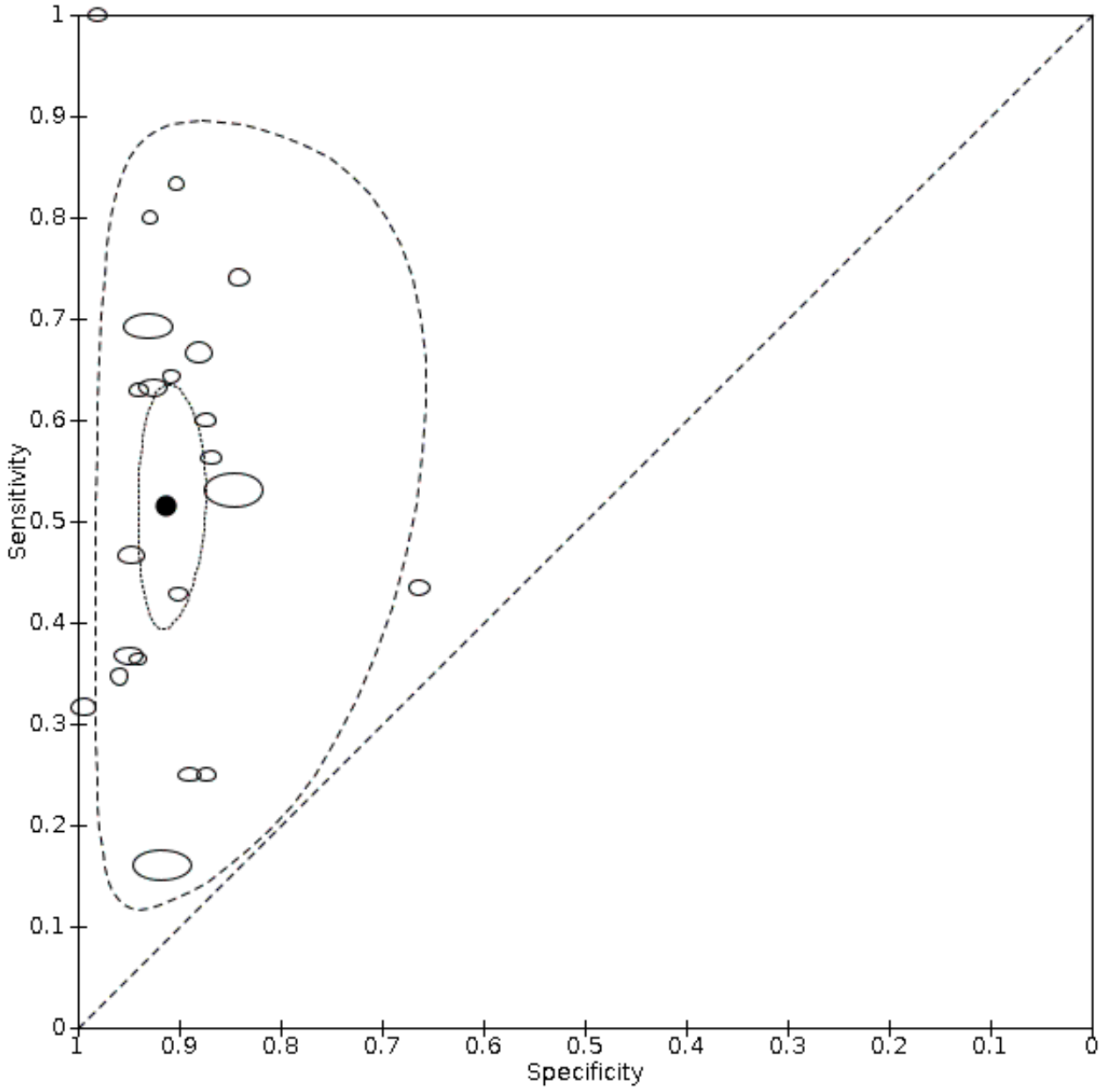
Study level estimates of sensitivities and specificities are shown in Figure 4.

Figure 4. A forest plot representing study level sensitivities and specificities that used an estimated fetal weight discordance (EFWD) of 20%.



The sensitivities ranged between 16% and 100%, whereas the specificities ranged between 66% and 99%. The summary estimate of sensitivity was 0.51 (95% CI 0.42 to 0.60), and the summary estimate of specificity was 0.91 (95% CI 0.89 to 0.93) (Table 1). The summary receiver operating plot with the summary point is illustrated in Figure 5. Most of the studies on the SROC plot were close to the Y axis indicating high specificity but variable sensitivity. The confidence regions were narrow for specificity but not sensitivity. The large prediction region indicated the variability of the diagnostic accuracy of future studies. The LR+ for EFWD by ultrasound was 5.9 (95% CI 4.3 to 8.1), and for LR– was 0.53 (95% CI 0.44 to 0.64).

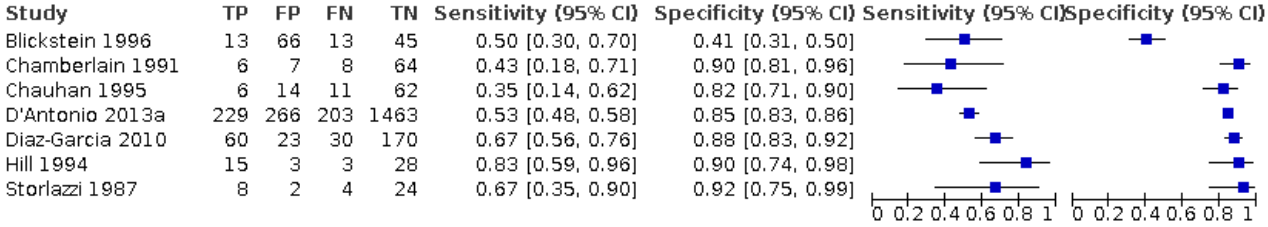
Figure 5. Summary receiver operating characteristic plot of studies assessing the accuracy of ultrasound based on an estimated fetal weight discordance (EFWD) of 20%. Each study is represented as an ellipse with size of the ellipse adjusted to the sample size of the study. The filled circle represents the summary point indicating summary sensitivity and specificity of the meta-analytic estimate. Dotted closed line represents 95% confidence region and the dashed line represents the 95% prediction region around the summary point.



Subgroup analyses of the 20% discordance based on measurements by abdominal circumference versus femur length

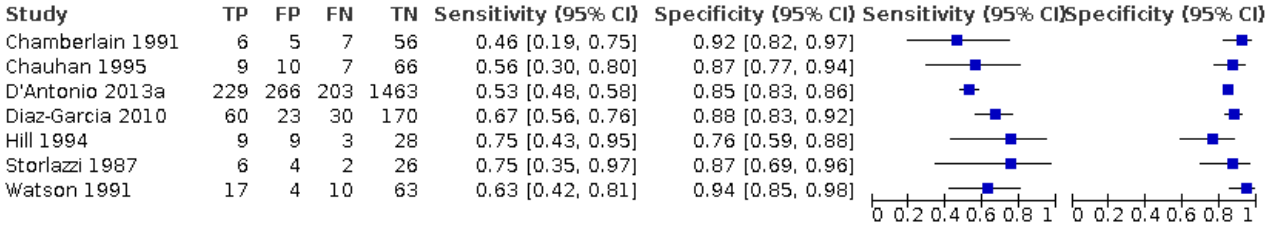
Of the 22 studies that reported estimates with 20% cut-off, seven used AC (Blickstein 1996; Chamberlain 1991; Chauhan 1995; D'Antonio 2013a; Diaz-Garcia 2010; Hill 1994; Storlazzi 1987) (Figure 6). Median disease prevalence across the studies was 26.3 (range 15.3 to 36.7). For AC as a measurement for EFWD, the summary estimates of sensitivity was 0.57 (95% CI 0.48 to 0.66), and the summary estimate of specificity was 0.84 (95% CI 0.72 to 0.92) (Table 1). The LR+ for EFWD by ultrasound was 3.6 (95% CI 1.8 to 7.4), and for LR– was 0.51 (95% CI 0.38 to 0.68).

Figure 6. A forest plot representing sensitivities and specificities of the studies that used ultrasound abdominal circumference to detect estimated fetal weight discordance (EFWD) of 20%. FN: false negative; FP: false positive; TN: true negative; TP: true positive.



Among the 22 studies that reported estimates with 20% difference in birth weight as reference, seven studies used FL for EFWD (Chamberlain 1991; Chauhan 1995; D'Antonio 2013a; Diaz-Garcia 2010; Hill 1994; Storlazzi 1987; Watson 1991). The study level sensitivity and specificity for ultrasound using FL are depicted in Figure 7. For FL as a measurement for EFWD, the summary estimate of sensitivity was 0.60 (95% CI 0.53 to 0.67), and the summary estimate of specificity was 0.87 (95% CI 0.84 to 0.90) (Table 1). LR+ was 4.6 (3.4 to 6.1), and LR– was 0.46 (0.38 to 0.56).

Figure 7. A forest plot representing sensitivities and specificities of the studies that used femoral length by ultrasound to detect estimated fetal weight discordance (EFWD) of 20%. FN: false negative; FP: false positive; TN: true negative; TP: true positive.



No subgroup analyses were possible based on gestational age, sex discordance or chorionicity due to lack of sufficient data for analysis.

2. Diagnostic performance of ultrasound (25% difference in birth weight as reference)

Eighteen studies that EFWD by ultrasound measurements provided data using a 25% difference in birth weight as reference (Bimson 2014a; Bimson 2014b; Blickstein 1996; Caravello 1997; Chamberlain 1991; Chang 2006; Crane 1980; D'Antonio 2013a; Danon 2008; Diaz-Garcia 2010; Hoopmann 2011; Klam 2005; Nakayama 2014; Reberdao 2010; Sayegh 1993; Simoes 2011; van Mieghem 2009; Zipori 2016). Study level estimates of sensitivities and specificities are shown in Figure 8. The sensitivities ranged between 1% and 100% whereas the specificities ranged between 79% and 100%. The summary estimate of sensitivity was 0.46 (95% CI 0.26 to 0.66), and the summary estimate of specificity was 0.93 (95% CI 0.89 to 0.96) (Table 1). The summary receiver operating plot along with the summary point is illustrated in Figure 9. Most of the studies on the SROC plot were close to the Y axis indicating high specificity but variable sensitivity. The confidence regions were narrow for specificity but not sensitivity. The large prediction region indicated the variability of the diagnostic accuracy of future studies. The LR+ for EFWD by ultrasound was 6.7 (95% CI 3.0 to 14.9) and for LR– was 0.58 (95% CI 0.39 to 0.88). Findings and the certainty of evidence (very low) are detailed in the Summary of findings 2.

Figure 8. A forest plot representing sensitivities and specificities of all the studies that used an estimated fetal weight discordance (EFWD) of 25%. FN: false negative; FP: false positive; TN: true negative; TP: true positive.

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bimson 2014a	10	1	15	50	0.40 [0.21, 0.61]	0.98 [0.90, 1.00]
Bimson 2014b	24	3	3	15	0.89 [0.71, 0.98]	0.83 [0.59, 0.96]
Blickstein 1996	3	50	12	74	0.20 [0.04, 0.48]	0.60 [0.50, 0.68]
Caravello 1997	9	12	71	150	0.11 [0.05, 0.20]	0.93 [0.87, 0.96]
Chamberlain 1991	3	7	5	70	0.38 [0.09, 0.76]	0.91 [0.82, 0.96]
Chang 2006	68	46	23	1120	0.75 [0.65, 0.83]	0.96 [0.95, 0.97]
Crane 1980	9	0	0	126	1.00 [0.66, 1.00]	1.00 [0.97, 1.00]
D'Antonio 2013a	253	157	287	1464	0.47 [0.43, 0.51]	0.90 [0.89, 0.92]
Danon 2008	28	26	26	198	0.52 [0.38, 0.66]	0.88 [0.83, 0.92]
Diaz-Garcia 2010	41	19	20	203	0.67 [0.54, 0.79]	0.91 [0.87, 0.95]
Hoopmann 2011	16	21	11	148	0.59 [0.39, 0.78]	0.88 [0.82, 0.92]
Klam 2005	39	25	70	369	0.36 [0.27, 0.46]	0.94 [0.91, 0.96]
Nakayama 2014	8	5	17	96	0.32 [0.15, 0.54]	0.95 [0.89, 0.98]
Reberdao 2010	1	8	6	105	0.14 [0.00, 0.58]	0.93 [0.87, 0.97]
Sayegh 1993	12	5	3	58	0.80 [0.52, 0.96]	0.92 [0.82, 0.97]
Simoes 2011	4	6	566	22	0.01 [0.00, 0.02]	0.79 [0.59, 0.92]
van Mieghem 2009	7	1	2	50	0.78 [0.40, 0.97]	0.98 [0.90, 1.00]
Zipori 2016	2	0	5	82	0.29 [0.04, 0.71]	1.00 [0.96, 1.00]

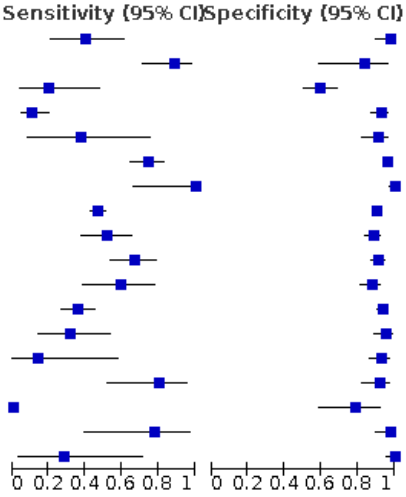
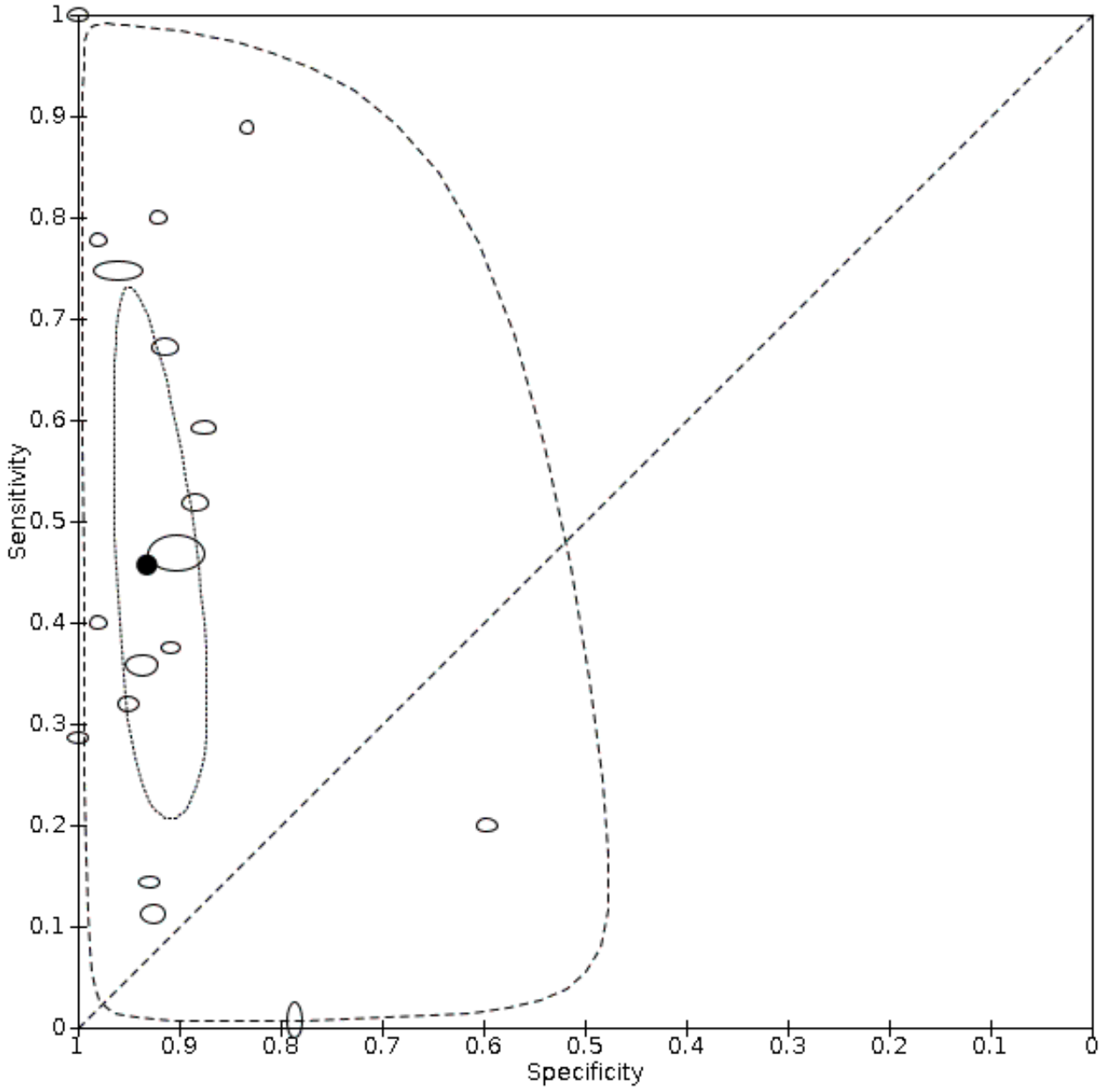


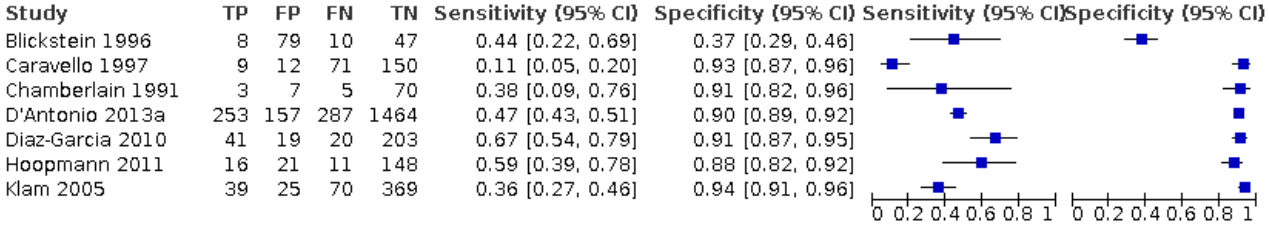
Figure 9. Summary receiver operating characteristic plot of studies assessing the accuracy of ultrasound based on an estimated fetal weight discordance (EFWD) of 25%. Each study is represented as an ellipse with size of the ellipse adjusted to the sample size of the study. The filled circle represents the summary point indicating summary sensitivity and specificity of the meta-analytic estimate. Dotted closed line represents 95% confidence region and the dashed line represents the 95% prediction region around the summary point.



Subgroup analyses of the 25% discordance based on measurements of abdominal circumference measurement versus femur length

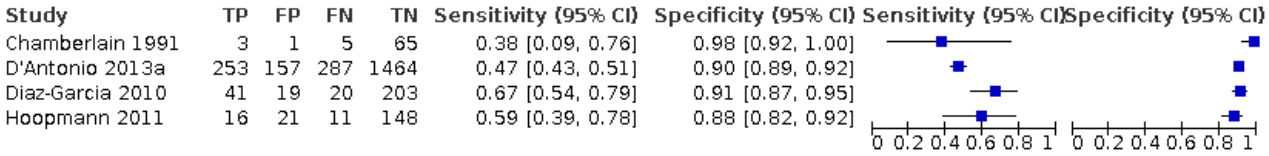
Among the studies that reported estimates with 25% difference in birth weight as reference, seven studies used AC (Blickstein 1996; Caravello 1997; Chamberlain 1991; D'Antonio 2013a; Diaz-Garcia 2010; Hoopmann 2011; Klam 2005). Study level sensitivity and specificity for ultrasound using AC are depicted in Figure 10. The pooled estimate of sensitivity was 0.42 (95% CI 0.27 to 0.58) whereas that of the specificity was 0.88 (95% CI 0.76 to 0.94) (Table 1). LR+ was 3.3 (95% CI 1.6 to 6.9) and LR– was 0.67 (95% CI 0.51 to 0.87).

Figure 10. A forest plot representing sensitivities and specificities of the studies that used ultrasound abdominal circumference to detect estimated fetal weight discordance (EFWD) of 25%. FN: false negative; FP: false positive; TN: true negative; TP: true positive.



Among the studies that reported estimates with 25% difference in birth weight as reference, we only found four studies used FL (Chamberlain 1991; D'Antonio 2013a; Diaz-Garcia 2010; Hoopmann 2011). Study level sensitivity and specificity for ultrasound using FL are depicted in Figure 11. The pooled estimate of sensitivity was 0.55 (95% CI 0.44 to 0.65) whereas that of the specificity was 0.91 (95% CI 0.89 to 0.92) (Table 1). LR+ was 5.8 (95% CI 4.3 to 7.9) and LR– was 0.50 (95% CI 0.40 to 0.64).

Figure 11. A forest plot representing sensitivities and specificities of the studies that used ultrasound femoral length to detect estimated fetal weight discordance (EFWD) of 25%. FN: false negative; FP: false positive; TN: true negative; TP: true positive.



No subgroup analyses were possible based on gestational age, sex discordance or chorionicity due to lack of sufficient data for analysis.

DISCUSSION

Summary of main results

We evaluated the diagnostic accuracy of EFWD of 20% and 25% using biometric ultrasound measurements compared with BWD as the reference standard. We identified 39 studies that matched our inclusion criteria. The summary estimate of sensitivity was 0.51 (95% CI 0.42 to 0.60), and the summary estimate of specificity was 0.91 (95% CI 0.89 to 0.93) at 20% discordance in BWD (Table 1). The summary estimate of sensitivity was 0.46 (95% CI 0.26 to 0.66), and the summary estimate of specificity was 0.93 (95% CI 0.89 to 0.96) at 25% discordance in BWD. Sensitivity and specificity did not differ substantively between the two cut-off points (20% and 25%). We found that there was high variability in sensitivity ranging from 0.1 to 1.0 but specificity was high with tight ranges for both 20% and 25% discordances in BWD. The certainty of the evidence was very low due to high risk of bias in the included studies and inconsistency.

Subgroup analyses were possible for ultrasound measurements using AC and FL for both 20% and 25% discordance in BWD. The sensitivities and specificities were not substantially different between these subgroups for both discordance cut-offs, but sensitivity and specificity were marginally better in studies that used FL for EFWD estimation. No subgroup analyses were possible based on gestational age, sex discordance or chorionicity due to lack of sufficient data for analysis.

Overall, this systematic review found very low-certainty evidence that ultrasound EFWD discordance between twins has poor sensitivity but good specificity in predicting actual BWD and suggested that FL performs slightly better than AC.

Leombroni 2017 reviewed limited data sources (PubMed, Embase and CINAHL) and fewer studies (20). The review authors could not perform comprehensive data synthesis for each AC discordance cut-off and were unable to report data on FL measurement. They concluded that the optimal diagnostic performance of AC discordance was for prediction of BW discordance of 25% or greater, with a sensitivity of 70.8% and specificity of 86.4%. Our findings suggest that there was no difference in ultrasound diagnosis of EFWD between 20% and 25% discordance.

Strengths and weaknesses of the review

The strength of this review is implementing a comprehensive search in a variety of databases and a large number of publications reviewed. We followed the standard recommendations of the Cochrane DTA (methods.cochrane.org/sdt/). We provide the most up-to-date overall assessment of ultrasound in predicting BWD in twins.

There are several limitations to this review.

1. We found considerable heterogeneity in sensitivity of the EFWD measurement by ultrasound. The low sensitivity of EFWD would miss a significant number of twins with BWD, which would be clinically consequential. However, twins often have regular screening so a missed diagnosis in one scan could potentially be detected in a subsequent scan.

2. We were able to analyze only two cut-off points (20% and 25%).
3. Lack of stratification by gestational age; gestational age at ultrasound examination is of great importance when assessing the predictive accuracy of ultrasound for growth discordance. We were unable to pool the data based on different gestational age due to the large variation in the extracted data. According to the literature, growth discrepancy in twins usually occurs in the third trimester (Puccio 2014). Hence, an ultrasound assessment performed early in pregnancy cannot reliably predict discordance.
4. Unable to conduct a subgroup analysis based on chorionicity due to the failure of included studies in reporting chorionicity. Since the pathophysiology of growth discordance differs in MC and DC twins, lack of chorionicity information in our analysis may be reflected in the predictive accuracy of ultrasound in identifying twins affected by abnormal growth.
5. Studies included in the analysis used different EFWD formulae (Hadlock, Shepard, Ferrero) to predict fetal weight, and some studies did not specify which formula they used to measure AC or FL. The absence of stratification according to each specific ultrasound weight formula might have hypothetically changed the approximation of EFW for each twin, consequently affecting the estimated EFWD.
6. It is also worth noting that some of the included studies are old; hence the technology used at the time may not have created as sharp an image as the new ultrasound machine.

Applicability of findings to the review question

Performing ultrasound on 1000 twin pregnancies to detect BWD of 20% or more with a median prevalence of BWD of 18% and using a summary sensitivity of 0.51 and specificity of 0.91, 88 women will miss a correct diagnosis (FN) and miss a chance to have beneficial interventions while 74 women will be diagnosed FP and may be subject to unnecessary medical interventions. The FN rate increases as the prevalence increases from 18% and more women will have a missed diagnosis but the FP rate decreases (Summary of findings 1).

Similarly, performing ultrasound on 1000 twin pregnancies to detect BWD of 25% or more with a median prevalence of BWD of 19% and using a summary sensitivity of 0.46 and specificity of 0.93, 103 women will miss a correct diagnosis (FN) and miss a chance to have beneficial interventions while 57 women will be diagnosed FP and may be subject to unnecessary interventions (Summary of findings 2).

BWD is associated with adverse perinatal outcomes in twin pregnancy (Jahanfar 2017a). Therefore, it is crucial to measure EFWD during routine ultrasound examinations. Our findings suggest that ultrasound EFWD has low sensitivity but good specificity in detecting discordance in actual BWD. There are a few points to consider when interpreting the result of this review. First, the reference standard that we used in this review (BWD) is a reliable reference standard, and so the findings are applicable to the review question. However, it should be noted that the tests were performed in a health clinic or tertiary hospitals, and our findings are, therefore, applicable only in these settings. Publications selected for this review included twin pregnancies with viable fetuses. Hence, the results of our review are not applicable to pregnancies with stillbirth, triplets or quadruplets. We were careful not to include the studies that did not define BWD or did not report the EFWD formula as identified by our review ((larger estimated weight – smaller estimated weight)/ larger estimate weight) × 100). In addition, we included only the measurements performed within seven days of birth and measured by a healthcare professional using either electronic or mechanical scales. Examinations included in the analysis were 2D ultrasounds performed during any stage of pregnancy. Since the test works best in the third trimester, it is unlikely that it would be useful for identifying and treating TTTS.

AUTHORS' CONCLUSIONS

Implications for practice

Very low-certainty evidence suggests that estimated fetal weight discordance (EFWD) identified by ultrasound has a low sensitivity but good specificity in detecting birth weight discordance (BWD) in twin pregnancies. There is uncertain diagnostic value of EFWD; this review suggests there is insufficient evidence to support this index as the sole measure for clinical decision making to evaluate the prognosis of twins with growth discordance. The diagnostic accuracy of other measures including amniotic fluid Index and umbilical artery Doppler resistive indices in combination with ultrasound for clinical intervention needs evaluation. Other clinical factors that should be considered are gestational age, chorionicity and real-time information about twin's heart rates and are important in the management of twin pregnancies with growth discordance. Currently, ultrasound is the only technology available to identify BWD. Until a better technology or test with better diagnostic accuracy is available, serial ultrasound to investigate the growth of fetuses is necessary to evaluate the degree of intertwin growth discordance.

Implications for research

Well-designed prospective cohort studies with delayed verification that assess the diagnostic performance of EFWD is needed. Stratification by gestational age, with different fetal weight discordance values (20% and 25%), chorionicity information and sex discordance information are needed and may be useful to reliably determine the diagnostic accuracy of ultrasound in predicting growth discordance in twin pregnancies.

ACKNOWLEDGEMENTS

As part of the prepublication editorial process, the review has been commented on by three peers (an editor and two referees who are external to the editorial team) and a member of the Pregnancy and Childbirth Group's international panel of consumers. The authors are grateful to the following peer reviewers for their time and comments: Gabriele Saccone, Department of Neuroscience, Reproductive Sciences and Dentistry, School of Medicine, University of Naples Federico II, Naples, Italy; Alexis Shub, Mercy Hospital for Women, Melbourne, Australia.

We thank Ursula Ellis for preparing the search strategies for this review. We also thank Vicki Nisenblat for her contribution to the protocol for this review.

This project was supported by the National Institute for Health Research (NIHR), via Cochrane Infrastructure funding to Cochrane Pregnancy and Childbirth. The views and opinions expressed herein are those of the authors and do not necessarily reflect those

of the Evidence Synthesis Programme, the NIHR, National Health Service (NHS) or the Department of Health and Social Care.

REFERENCES

References to studies included in this review

Bimson 2012 *{unpublished data only}*

Bimson B, Schwartz RA, Morrissey M, Lust B, Brustman L. How accurate is ultrasound in predicting birth weight discordance in twin gestations? *Reproductive Sciences. 59th Annual Scientific Meeting of the Society for Gynecologic Investigation, San Diego, CA, United States* 2012;**19**(3 Suppl 1):146A.

Bimson 2014a *{published data only}*

Bimson BE, Brustman L, Murphy E, Foroutan J, Herrera K, Rosenn B. Does ultrasound accurately predict growth discordance in dichorionic diamniotic twins? *Reproductive Sciences. 61st Annual Scientific Meeting of the Society for Gynecologic Investigation, Florence, Italy* 2014;**21**(3 Suppl):169A-70A.

Bimson 2014b *{published data only}*

Bimson BE, Brustman L, Murphy E, Herrera K, Foroutan J, Rosenn B. Using ultrasound, what best predicts pathologic discordance in diamniotic dichorionic twins? *Reproductive Sciences. 61st Annual Scientific Meeting of the Society for Gynecologic Investigation, Florence, Italy* 2014;**21**(3):169A.

Blickstein 1989 *{published data only}*

Blickstein I, Friedman A, Caspi B, Lancet M. Ultrasonic prediction of growth discordancy by intertwin difference in abdominal circumference. *International Journal of Gynaecology and Obstetrics* 1989;**29**(2):121-4.

Blickstein 1996 *{published data only}*

Blickstein I, Manor M, Levi R, Goldchmit R. Is intertwin birth weight discordance predictable? *Gynecologic and Obstetric Investigation* 1996;**42**(2):105-8.

Caravello 1997 *{published data only}*

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

Chamberlain 1991 *{published data only}*

Chamberlain P, Murphy M, Comerford FR. How accurate is antenatal sonographic identification of discordant birthweight in twins? *European Journal of Obstetrics, Gynecology, and Reproductive Biology* 1991;**40**(2):91-6.

Chang 2006 *{published data only}*

Chang YL, Chang TC, Chang SD, Cheng PJ, Chao AS, Hsieh PC, et al. Sonographic prediction of significant intertwin birth weight discordance. *European Journal of Obstetrics, Gynecology, and Reproductive Biology* 2006;**127**(1):35-40.

Chauhan 1995 *{published data only}*

Chauhan SP, Washburne JF, Roberts WE, Magann EF, Morrison JC, Martin JN. Estimates of birth-weight among twins by the use of femur length alone. *Journal of Maternal-fetal Investigation* 1995;**5**(2):113-6.

Crane 1980 *{published data only}*

Crane JP, Tomich PG, Kopta M. Ultrasonic growth patterns in normal and discordant twins. *Obstetrics and Gynecology* 1980;**55**(6):678-83.

D'Antonio 2013a *{published data only}*

D'Antonio F, Khalil A, Dias T, Thilaganathan B. Weight discordance and perinatal mortality in twins: analysis of the Southwest Thames Obstetric Research Collaborative (STORK) multiple pregnancy cohort. *Ultrasound in Obstetrics & Gynecology* 2013;**41**(6):643-8.

D'Antonio 2014 *{published data only}*

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

Danon 2008 *{published data only}*

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

Diaz-Garcia 2010 *{published data only}*

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

Fox 2011 *{published data only}*

Fox NS, Saltzman DH, Schwartz R, Roman AS, Klauser CK, Rebarber A. Second-trimester estimated fetal weight and discordance in twin pregnancies. *Journal of Ultrasound in Medicine* 2011;**30**(8):1095-101.

Hehir 2017 *{published data only}*

Hehir MP, Breathnach FM, Hogan JL, McAuliffe FM, Geary MP, Daly S, et al. Prenatal prediction of significant intertwin birthweight discordance using standard second and third trimester sonographic parameters. *Acta Obstetricia et Gynecologica Scandinavica* 2017;**96**(4):472-8.

Hill 1994 *{published data only}*

Hill LM, Guzick D, Chenevey P, Boyles D, Nedzesky P. The sonographic assessment of twin growth discordancy. *ACOG Current Journal Review* 1994;**2**(8):20.

Hoopmann 2011 *{published data only}*

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 Diagnosis and Therapy* 2011;**30**(1):29-34.

Jensen 1995 *{published data only}*

Jensen OH, Jenssen H. Prediction of fetal weights in twins. *Acta Obstetricia et Gynecologica Scandinavica* 1995;**74**(3):177-80.

Johansen 2014 *{published data only}*

Johansen ML, Oldenburg A, Rosthøj 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* 2014;**43**(3):277-83.

Kalish 2003 *{published data only}*

Kalish RB, Chasen ST, Gupta M, Sharma G, Perni SC, Chervenak FA. First trimester prediction of growth discordance in twin gestations. *American Journal of Obstetrics & Gynecology* 2003;**189**(3):706-9.

Kim 2015 *{published data only}*

Kim JH, Sung JH, Kim JY, Chang KH, Choi SJ, Oh SY, et al. Pregnancy outcome of isolated cases of inter-twin fetal weight discordance estimated by antenatal ultrasound. *International Journal of Gynecology & Obstetrics* 2015;**131**(S5):E461.

Klam 2005 *{published data only}*

Klam SL, Rinfret D, Leduc L. Prediction of growth discordance in twins with the use of abdominal circumference ratios. *American Journal of Obstetrics & Gynecology* 2005;**192**(1):247-51.

Machado 2007 *{published data only}*

Machado RC, Brizot ML, Liao AW, Cabar FR, Zugaib M. Prenatal sonographic prediction of twin growth discordance. *Twin Research and Human Genetics* 2007;**10**(1):198-201.

MacLean 1992 *{published data only}*

MacLean M, Mathers A, Walker JJ, Cameron AD, Howat R. The ultrasonic assessment of discordant growth in twin pregnancies. *Ultrasound in Obstetrics & Gynecology* 1992;**2**(1):30-4.

Murray 2014 *{published data only}*

Murray A, Breathnach F, McAuliffe F, Geary M, Daly S, Higgins J, et al. 187: Prediction of significant birth weight discordance in twin pregnancies with second and third trimester ultrasound. *American Journal of Obstetrics & Gynecology* 2014;**210**(1):S104.

Nakayama 2014 *{published data only}*

Nakayama S, Ishii K, Kawaguchi H, Yamamoto R, Murata M, Hayashi S, et al. Perinatal complications of monochorionic diamniotic twin gestations with discordant crown-rump length determined at mid-first trimester. *Journal of Obstetrics and Gynaecology Research* 2014;**40**(2):418-23.

O'Connor 2013 *{published data only}*

O'Connor C, McAuliffe FM, Breathnach FM, Geary M, Daly S, Higgins JR, et al. Prediction of outcome in twin pregnancy with first and early second trimester ultrasound. *Journal of Maternal-fetal & Neonatal Medicine* 2013;**26**(10):1030-5.

Reberdao 2010 *{published data only}*

Reberdao MA, Martins L, Torgal M, Viana R, Seminova T, Casal E, et al. The source of error in the estimation of intertwin birth weight discordance. *Journal of Perinatal Medicine* 2010;**38**(6):671-4.

Sayegh 1993 *{published data only}*

Sayegh SK, Warsof SL. Ultrasonic prediction of discordant growth in twin pregnancies. *Fetal Diagnosis and Therapy* 1993;**8**(4):241-6.

Secher 1985 *{published data only}*

Secher NJ, Kaern J, Hansen PK. Intrauterine growth in twin pregnancies: prediction of fetal growth retardation. *Obstetrics and Gynecology* 1985;**66**(1):63-8.

Shahshahan 2011 *{published data only}*

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* 2011;**16**(9):1224.

Simoes 2011 *{published data only}*

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* 2011;**39**(1):43-6.

Storlazzi 1987 *{published data only}*

Storlazzi E, Vintzileos AM, Campbell WA, Nochimson DJ, Weinbaum PJ. Ultrasonic diagnosis of discordant fetal growth in twin gestations. *Obstetrics and Gynecology* 1987;**69**(3 Pt 1):363-7.

van de Waarsenburg 2015 *{published data only}*

van de Waarsenburg MK, Hack KE, Rijpma RJ, Mulder EJ, Pistorius L, Derks JB. Ultrasonographic prediction of birth weight discordance in twin pregnancies. *Prenatal Diagnosis* 2015;**35**(9):906-12.

van Mieghem 2009 *{published data only}*

van Mieghem T, Deprest J, Klaritsch P, Gucciardo L, Done E, Verhaeghe J, et al. Ultrasound prediction of intertwin birth weight discordance in monochorionic diamniotic twin pregnancies. *Prenatal Diagnosis* 2009;**29**(3):240-4.

Watson 1991 *{published data only}*

Watson WJ, Valea FA, Seeds JW. Sonographic evaluation of growth discordance and chorionicity in twin gestation. *American Journal of Perinatology* 1991;**8**(05):342-4.

Zipori 2016 *{published data only}*

Zipori Y, Reidy K, Gilchrist T, Doyle LW, Umstad MP. The outcome of monochorionic diamniotic twins discordant at 11 to 13+ 6 weeks' gestation. *Twin Research and Human Genetics* 2016;**19**(6):692-6.

Zuckerwise 2015 *{published data only}*

Zuckerwise L, Nayeri U, Abdel-Razeq S, Copel J, Bahtiyar MO. Doppler abnormalities in monochorionic diamniotic twin pregnancies with discordant growth. *Journal of Perinatology* 2015;**35**(6):387.

References to studies excluded from this review

Allaf 2014 *{published data only}*

Allaf MB, Campbell WA, Vintzileos AM, Haeri S, Javadian P, Shamshirsaz AA, et al. Does early second-trimester sonography predict adverse perinatal outcomes in monochorionic diamniotic twin pregnancies? *Journal of Ultrasound in Medicine* 2014;**33**(9):1573-8.

Banks 2008 *{published data only}*

Banks CL, Nelson SM, Owen P. First and third trimester ultrasound in the prediction of birthweight discordance in dichorionic twins. *European Journal of Obstetrics, Gynecology, and Reproductive Biology* 2008;**138**(1):34-8.

Ben-Ami 2015 *{published data only}*

Ben-Ami I, Daniel-Spiegel E, Battino S, Melcer Y, Floeck A, Geipel A, et al. The association of crown rump length discrepancy with birthweight discordance in spontaneous versus IVF monochorionic twins: a multicenter study. *Prenatal Diagnosis* 2015;**35**(9):864-9.

Bora 2009 *{published data only}*

Bora SA, Bourne T, Bottomley C, Kirk E, Papageorghiou AT. Twin growth discrepancy in early pregnancy. *Ultrasound in Obstetrics & Gynecology* 2009;**34**(1):38-42.

Brown 1987 *{published data only}*

Brown CE, Guzick DS, Leveno KJ, Santos-Ramos R, Whalley PJ. Prediction of discordant twins using ultrasound measurement of biparietal diameter and abdominal perimeter. *Obstetrics and Gynecology* 1987;**70**(5):677-81.

Chauhan 1996 *{published data only}*

Chauhan SP, Cowan BD, Brost BD, Washburne JF, Magann EF, Morrison JC. Estimating birth weight in twins. Comparison of eight sonographic models. *Journal of Reproductive Medicine* 1996;**41**(6):403-8.

Chauhan 2004 *{published data only}*

Chauhan SP, Shields D, Parker D, Sanderson M, Scardo JA, Magann EF. Detecting fetal growth restriction or discordant growth in twin gestations stratified by placental chorionicity. *Journal of Reproductive Medicine* 2004;**49**(4):279-84.

D'Antonio 2013b *{published data only}*

D'Antonio F, Khalil A, Mantovani E, Thilaganathan B. Embryonic growth discordance and early fetal loss: the STORK multiple pregnancy cohort and systematic review. *Human Reproduction* 2013;**28**(10):2621-7.

Esinler 2017 *{published data only}*

Esinler D, Aldemir OB, Alici Davutoglu E, Karahanoglu E, Salihoglu KN, Kuzu E, et al. A new mathematical formula to predict the foetal weight in twin pregnancies: a comparison of it with 19 different formulas. *Journal of Obstetrics and Gynaecology* 2017;**37**(1):53-7.

Fajardo-Expósito 2011 *{published data only}*

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

Gaziano 1998 *{published data only}*

Gaziano E, Gaziano C, Brandt D. Doppler velocimetry determined redistribution of fetal blood flow: correlation with growth restriction in diamniotic monochorionic and dizygotic twins. *American Journal of Obstetrics & Gynecology* 1998;**178**(6):1359-67.

Harper 2012a *{published data only}*

Harper L, Roehl K, Odibo A, Cahill A. 359: Crown-rump length discordance and adverse pregnancy outcomes in dichorionic twins. *American Journal of Obstetrics and Gynecology* 2012;**206**(1):S168-9.

Harper 2012b *{published data only}*

Harper L, Weis M, Odibo A, Roehl K, Macones G, Cahill A. 137: Are normally grown twin pregnancies with birth weight discordance at risk for adverse outcomes? *American Journal of Obstetrics and Gynecology* 2012;**206**(1):S73.

Harper 2012c *{published data only}*

Harper LM, Roehl KA, Odibo AO, Cahill AG. First-trimester growth discordance and adverse pregnancy outcome in dichorionic twins. *Ultrasound in Obstetrics & Gynecology* 2012;**41**(6):627-31.

Harper 2013 *{published data only}*

Harper LM, Weis MA, Odibo AO, Roehl KA, Macones GA, Cahill AG. Significance of growth discordance in appropriately grown twins. *American Journal of Obstetrics & Gynecology* 2013;**208**(5):393. e1-5.

Jensen 1992 *{published data only}*

Jensen OH. Doppler velocimetry in twin pregnancy. *European Journal of Obstetrics & Gynecology* 1992;**45**(1):9-12.

Kremkau 1978 *{published data only}*

Kremkau FW, Nelson LH. Diagnostic ultrasound and its obstetric applications. *American Family Physician* 1978;**17**(5):148-57.

Leombroni 2017 *{published data only}*

Leombroni M, Liberati M, Fanfani F, Pagani G, Familiari A, Buca D, et al. Diagnostic accuracy of ultrasound in predicting birth weight discordance in twin pregnancy: systematic review and meta-analysis. *Ultrasound in Obstetrics & Gynecology* 2017;**50**(4):442-50.

Nakano 2015 *{published data only}*

Nakano JC, Liao AW, Brizot MdeL, Miyadahira M, Francisco RP, Zugaib M. Fetal growth according to different reference ranges in twin pregnancies with placental insufficiency. *Clinics* 2015;**70**(12):816-9.

O'Brien 1986 *{published data only}*

O'Brien WF, Knuppel RA, Scerbo JC, Rattan PK. Birth weight in twins: an analysis of discordancy and growth retardation. *Obstetrics and Gynecology* 1986;**67**(4):483-6.

Ocer 2011 *{published data only}*

Ocer F, Aydin Y, Atis A, Kaleli S. Factors affecting the accuracy of ultrasonographical fetal weight estimation in twin pregnancies. *Journal of Maternal-fetal & Neonatal Medicine* 2011;**24**(9):1168-72.

Papaioannou 2011 *{published data only}*

Papaioannou GI, Syngelaki A, Maiz N, Ross JA, Nicolaides KH. Prediction of outcome in dichorionic twin pregnancies at 6-10 weeks' gestation. *American Journal of Obstetrics & Gynecology* 2011;**205**(4):348. e1-5.

Ropacka-Lesiak 2012 *{published data only}*

Ropacka-Lesiak M, Breborowicz G, Dera A. Blood flow changes in dichorionic twins with growth discordance. *Twin Research and Human Genetics* 2012;**15**(6):781-7.

Senoo 2000 *{published data only}*

Senoo M, Okamura K, Murotsuki J, Yaegashi N, Uehara S, Yajima A. Growth pattern of twins of different chorionicity evaluated by sonographic biometry. *Obstetrics and Gynecology* 2000;**95**(5):656-61.

Shivkumar 2015 *{published data only}*

Shivkumar S, Himes KP, Hutcheon JA, Platt RW. An ultrasound-based fetal weight reference for twins. *American Journal of Obstetrics & Gynecology* 2015;**213**(2):224. e1-9.

Snijder 1998 *{published data only}*

Snijder MJ, Wladimiroff JW. Fetal biometry and outcome in monochorionic vs. dichorionic twin pregnancies; a retrospective cross-sectional matched-control study. *Ultrasound in Medicine and Biology* 1998;**24**(2):197-201.

Stirrup 2016 *{published data only}*

Stirrup OT, Khalil A, D'Antonio F, Thilaganathan B. 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 and Therapy* 2016;**41**(2):100-7.

Suzuki 2009 *{published data only}*

Suzuki S, Shimizu E, Kinoshita M, Araki S. Accuracy of ultrasonographic fetal weight estimation in Japanese twin pregnancies. *Journal of Medical Ultrasonics* 2009;**36**(3):157-8.

Weissmann-Brenner 2012 *{published data only}*

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* 2012;**40**(5):489-93.

Xu 1995 *{published data only}*

Xu B, Deter RL, Milner LL, Hill RM. Evaluation of twin growth status at birth using individualized growth assessment: comparison with conventional methods. *Journal of Clinical Ultrasound* 1995;**23**(5):277-86.

Additional references

ACOG Committee 2004

American College of Obstetricians and Gynecologists Committee on Practice Bulletins-Obstetrics, Society for Maternal-Fetal Medicine, ACOG Joint Editorial Committee. ACOG Practice Bulletin #56: Multiple gestation: complicated twin, triplet, and high-order multifetal pregnancy. *Obstetrics and Gynecology* 2004;**104**(4):869-83.

Adegbite 2004

Adegbite AL, Castille S, Ward S, Bajoria R. Neuromorbidity in preterm twins in relation to chorionicity and discordant birth weight. *American Journal of Obstetrics and Gynecology* 2004;**190**(1):156-8.

Alfirevic 2003

Alfirevic Z, Sundberg K, Brigham S. Amniocentesis and chorionic villus sampling for prenatal diagnosis. *Cochrane Database of Systematic Reviews* 2003, Issue 3. Art. No: CD003252. [DOI: 10.1002/14651858.CD003252] [PMID: 12917956]

Appleton 2007

Appleton C, Pinto L, Centeno M, Clode N, Cardoso C, Graca LM, et al. Near term twin pregnancy: clinical relevance of weight discordance at birth. *Journal of Perinatal Medicine* 2007;**35**(1):62-6.

Bagchi 2006

Bagchi S, Salihu HM. Birth weight discordance in multiple gestations: occurrence and outcomes. *Journal of Obstetrics and Gynaecology* 2006;**26**(4):291-6.

Behrendt 2016

Behrendt N, Galan HL. Twin-twin transfusion and laser therapy. *Current Opinion in Obstetrics & Gynecology* 2016;**28**(2):79-85. [PMID: 26871537]

Bhide 2009

Bhide A, Sankaran S, Sairam S, Papageorghiou AT, Thilaganathan B. Relationship of intertwin crown-rump length discrepancy to chorionicity, fetal demise and birth-weight discordance. *Ultrasound in Obstetrics and Gynecology* 2009;**34**(2):131-5.

Chitkara 1985

Chitkara U, Berkowitz GS, Levine R, Riden DJ, Fagerstrom RM Jr, Chervenak FA, et al. Twin pregnancy: routine use of ultrasound examinations in the prenatal diagnosis of intrauterine growth retardation and discordant growth. *American Journal of Perinatology* 1985;**2**(1):49-54.

Chu 2006

Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. Journal of Clinical Epidemiology 2006;**59**(12):1331-2. [PMID: 17098577]

de Vet 2013

de Vet HC, Eisinga A, Riphagen II, Aertgeerts B, Pewsner D. Chapter 7: Searching for studies. In: Deeks JJ, Bossuyt PM, Gatsonis C, editor(s). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration, 2013. Available from methods.cochrane.org/sdt/handbook-dta-reviews.

Deeks 2005

Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology* 2005;**58**(9):882-93.

Deeks 2019

Deeks JJ, Higgins JP, Altman DG. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor(s). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated August 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

Demissie 2002

Demissie K, Ananth CV, Martin J, Hanley ML, MacDorman MF, Rhoads GG, et al. Fetal and neonatal mortality among twin gestations in the United States: the role of intrapair birth weight discordance. *Obstetrics and Gynecology* 2002;**100**(3):474-80.

Devoe 1995

Devoe LD, Wre DJ. Antenatal assessment of twin gestation. *Seminars in Perinatology* 1995;**19**(5):413.

Di 2007

Di Renzo GC, Rosati A, Sarti RD, Cruciani L, Cutuli AM. Does fetal sex affect pregnancy outcome? *Gender Medicine* 2007;**4**(1):19-30. [PMID: 17584623]

Dwamena 2009

Dwamena B. MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies, Statistical Software Components S456880, Boston College Department of Economics. ideas.repec.org/c/boc/bocode/s456880.html (accessed prior to 26 January 2021).

Gernt 2001

Gernt PR, Mauldin JG, Newman RB, Durkalski VL. Sonographic prediction of twin birth weight discordance. *Obstetrics and Gynecology* 2001;**97**(1):53-6.

Giles 1998

Giles WB. Doppler ultrasound in multiple pregnancies. *Bailliere's Clinical Obstetrics and Gynaecology* 1998;**12**(1):77-89.

Gonzalez-Quintero 2003

Gonzalez-Quintero VH, Luke B, O'Sullivan MJ, Misiunas R, Anderson E. Antenatal factors associated with significant birthweight discordancy in twin gestations. *American Journal of Obstetrics and Gynecology* 2003;**189**(3):83.

GRADEpro GDT [Computer program]

McMaster University (developed by Evidence Prime) GRADEpro GDT. Hamilton (ON): McMaster University (developed by Evidence Prime), accessed 1 July 2019. Available at gradepro.org.

Harbord 2007

Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;**8**:239-51.

Harbord 2009

Harbord RM, Whiting P. Metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata Journal* 2009;**9**:211-29.

Hata 2011

Hata T, Tanaka H, Noguchi J, Hata K. Three-dimensional ultrasound evaluation of the placenta. *Placenta* 2011;**32**(2):105-15.

Jahanfar 2016a

Jahanfar S, Lim K. A systematic review of birth weight discordance and perinatal outcomes. *Fetal and Maternal Medicine Review* 2016;**1**(1):1-31.

Jahanfar 2016b

Jahanfar S, Lim K. Optimal threshold for birth weight discordance. *Journal of Perinatology* 2016;**2**:1-9.

Jahanfar 2017a

Jahanfar S, Lim K, Ovideo-Joekes E. Birth weight discordance and adverse perinatal outcomes. *Journal of Perinatal Medicine* 2017;**45**(5):603-11. [PMID: 28622142]

Kilic 2006

Kilic M, Aygun C, Kaynar-Tuncel E, Küçüködük S. Does birth weight discordance in preterm twins affect neonatal outcome? *Journal of Perinatology* 2006;**26**(5):268-72.

Klam 2003

Klam S, Rinfret DE, Leduc L. Abdominal circumference ratio: a useful predictor of twin birth weight discordance. *American Journal of Obstetrics and Gynecology* 2003;**189**(6 Suppl):231.

Leeflang 2008

Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Annals of Internal Medicine* 2008;**149**(12):889-97.

Lewi 2008

Lewi L, Gucciardo L, Huber A, Jani J, Van Mieghem T, Done E, et al. Clinical outcome and placental characteristics of monochorionic diamniotic twin pairs with early- and late-onset discordant growth. *American Journal of Obstetrics and Gynecology* 2008;**199**(5):511.e1-7.

Lopriore 2012

Lopriore E, Pasman SA, Klumper FJ, Middeldorp JM, Walther FJ, Oepkes D, et al. Placental characteristics in growth-discordant monochorionic twins: a matched case-control study. *Placenta* 2012;**33**(3):171-4.

Macaskill 2013

Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editor(s). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration, 2013. Available from srdta.cochrane.org.

Mahony 2006

Mahony R, Foley M, Mcauliffe F, Mcparland P, Carroll S. Birth weight discordance, obstetric and perinatal outcome in twin pregnancies at term (37 weeks gestation). *American Journal of Obstetrics and Gynecology* 2006;**195**(6):S114.

Mazhar 2010

Mazhar SB, Kanwal S. Twin birth weight discordance: associated factors and outcome. *Journal of the College of Physicians and Surgeons--Pakistan : JCPSP* 2010;**20**(6):391-4.

Melamed 2009

Melamed N, Yogev Y, Glezerman M. Effect of fetal sex on pregnancy outcome in twin pregnancies. *Obstetrics and Gynecology* 2009;**114**(5):1085-92.

Miller 2012

Miller J, Chauhan SP, Abuhamad AZ. Discordant twins: diagnosis, evaluation and management. *American Journal of Obstetrics and Gynecology* 2012;**206**(1):10-20.

Morin 2011

Morin L, Lim K. Ultrasound in twin pregnancies. *Journal of Obstetrics and Gynaecology Canada : JOGC* 2011;**33**(6):643-56.

NICE 2011

National Institute for Health and Care Excellence. Multiple pregnancy: the management of twin and triplet pregnancies in the antenatal period multiple pregnancy: the management of twin and triplet pregnancies in the antenatal period (CG129). www.nice.org.uk/guidance/cg129 (accessed October 2014).

Oerbekke 2020

Oerbekke MS, Jenniskens K, Scholten RJ, Hooft L. Data sources and methods used to determine pretest probabilities in a cohort of Cochrane diagnostic test accuracy reviews. *BMC Medical Research Methodology* 2020;**20**(1):85.

Okun 2000

Okun N, Pastuzak A, Seaward G, Wilkinson J, Wilson D, Amankwah K, et al. The SOCG consensus statement: management of twin pregnancies (part 1). sogc.org/wp-content/uploads/2013/01/91E-CONS1-July2000.pdf (accessed October 2014).

Puccio 2014

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 Pediatrics* 2014;**40**:43. [PMID: 24887062]

Review Manager 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration Review Manager (RevMan). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Rizzo 1993

Rizzo G, Arduini D, Romanini C. Uterine artery Doppler velocity waveforms in twin pregnancies. *Obstetrics and Gynecology* 1993;**82**(6):978-83.

Schünemann 2019

Schünemann HJ, Mustafa RA, Brozek J, Santesso N, Bossuyt PM, Steingart KR, et al. GRADE guidelines: 22. The GRADE approach for tests and strategies – from test accuracy to patient-important outcomes and recommendations. Clinical Epidemiology 2019;**111**:69-82.

Schünemann 2020a

Schünemann HJ, Mustafa RA, Brozek J, Steingart KR, Leeflang M, Murad MH, et al. GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. Journal of Clinical Epidemiology 2020;**122**:142-52.

Schünemann 2020b

Schünemann HJ, Mustafa RA, Brozek J, Steingart KR, Leeflang M, Murad MH, et al. GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. *Journal of Clinical Epidemiology* 2020;**122**:129-41.

Tai 2007

Tai J, Grobman WA. The association of crown-rump length discordance in twin gestations with adverse perinatal outcomes. *American Journal of Obstetrics and Gynecology* 2007;**197**(4):369.e1-4.

Victoria 2001

Victoria A, Mora G, Arias F. Perinatal outcome, placental pathology, and severity of discordance in monochorionic and dichorionic twins. *Obstetrics and Gynecology* 2001;**97**(2):310-5. [PMID: 11165601]

Wen 2006

Wen SW, Tan H, Walker M. The association between intratwin birthweight discordance and preterm birth in twin pregnancy. *Australian & New Zealand Journal of Obstetrics & Gynaecology* 2006;**46**(5):402-6.

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529-36.

Woo 1939

Woo J. A short history of the development of ultrasound in obstetrics and gynecology. www.ob-ultrasound.net/history3.html (accessed 28 January 2017).

References to other published versions of this review

Jahanfar 2017b

Jahanfar S, Ho JJ, Jaafar SH, Abraha I, Nisenblat V, Ellis UM, et al +. Ultrasound for diagnosis of birth weight discordance in twin pregnancies. *Cochrane Database of Systematic Reviews* 2017, Issue 3. Art. No: CD012553. [DOI: 10.1002/14651858.CD012553]

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Bimson 2012

Study characteristics			
Patient Sampling	Twin pregnancies with documented births; GA ≥ 24 weeks at delivery; US within 7 days of delivery		
Patient characteristics and setting	241 twins, GA: mean 35.8 (SD 2.4) weeks		
Index tests	EFW		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	No excluded women indicated. US within 7 days of delivery		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		

Bimson 2012 *(Continued)*

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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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	

Bimson 2014a

Study characteristics			
Patient Sampling	DADC twin pregnancies with EFW within 2 weeks of delivery in the 3rd trimester		
Patient characteristics and setting	231 twin sets (January 2006 to March 2013); GA: mean 36 weeks 2 days (SD 2 weeks), DADC twins		
Index tests	EFW		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	EFW within 2 weeks of delivery		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	

Bimson 2014a *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Bimson 2014b

Study characteristics	
Patient Sampling	DADC twins
Patient characteristics and setting	336 twin sets (January 2006 March 2013); GA: N/A; DADC twins
Index tests	EFW
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	1st trimester CRL, 2nd and 3rd trimester EFW

Bimson 2014b *(Continued)*

Comparative
Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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		

Bimson 2014b *(Continued)*

Could the patient flow have introduced bias?	Low risk

Blickstein 1989

Study characteristics			
Patient Sampling	Twin pairs delivered after 32 weeks' gestation and 1 US within 2 weeks of delivery		
Patient characteristics and setting	178 twin pregnancies (1 September 1983 to 31 August 1986); GA: discordants: mean 36.4 (SD 2.3) weeks, concordants: mean 37.1 (SD 2.3) weeks; hospital		
Index tests	AC and FL		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women without AC, FL in both twins, and congenital malformations in either twin. US was within 2 weeks of delivery		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Blickstein 1989 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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 risk	

Blickstein 1996

Study characteristics			
Patient Sampling	Twin pregnancies with complete sets of US measurements within 2 weeks of delivery delivered at Kaplan Hospital, Rehovot, Israel		
Patient characteristics and setting	90 twin pregnancies with US within 2 weeks of delivery; GA: mean 36.1 (SD 2.9) weeks; hospital		
Index tests	EFW and AC		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women who had measurements performed > 2 weeks before delivery and had incomplete measurements		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		

Blickstein 1996 *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Caravello 1997

Study characteristics	
Patient Sampling	Live-born twin pairs born at tertiary centre with GA > 23 weeks, no anomalies and US within 3 weeks of delivery

Caravello 1997 *(Continued)*

Patient characteristics and setting	242 twin pregnancies; GA: mean 32.8 (SD 4.1) weeks; tertiary centre
Index tests	EFW or AC
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	US within 3 weeks of delivery
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Caravello 1997 *(Continued)*

DOMAIN 4: Flow and Timing			
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	

Chamberlain 1991

Study characteristics			
Patient Sampling	Twin pregnancies with last US performed within 7–14 days of delivery		
Patient characteristics and setting	85 twin pregnancies (enrolled from January 1985 to December 1988); GA at delivery: 36.9 weeks; hospital		
Index tests	EFW using either AC alone or AC and FL		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women had an interval between US and delivery > 14 days, intrauterine death, BW was not within 6 hours of delivery or EFW was unable to be determined		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			

Chamberlain 1991 *(Continued)*

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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
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?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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	

Chang 2006

Study characteristics	
Patient Sampling	Twin births with GA ≥ 24 weeks, and US within 28 days of delivery
Patient characteristics and setting	1257 twin pregnancies (January 1991 to August 2002); GA: mean 35.2 (SD 2.6) weeks; hospital
Index tests	EFWD using a combination of BPD, AC, HC and FL
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	Excluded women had incomplete data. US was within 28 days of delivery
Comparative	

Chang 2006 *(Continued)*

Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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	

Chauhan 1995

Study characteristics			
Patient Sampling	Twin pregnancies with FL, BPD and AC within 72 hours of delivery at medical centre		
Patient characteristics and setting	92 twin gestations; GA: mean 32.3 (SD 4.5) weeks; hospital		
Index tests	EFW using FL, BPD or AC		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	EFW within 72 hours of delivery. Excluded women were those with fetal death or known structural or chromosomal abnormality		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		

Chauhan 1995 *(Continued)*

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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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	

Crane 1980

Study characteristics			
Patient Sampling	Twin pregnancies		
Patient characteristics and setting	18 twin pairs; GA: N/A; university		
Index tests	BPD		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women not indicated		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	

Crane 1980 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

D'Antonio 2013a

Study characteristics	
Patient Sampling	All twin pregnancies registering for antenatal care by 11 weeks
Patient characteristics and setting	2161 twin pregnancies (enrolled 2000–2010); large regional cohort of 9 hospitals over ten years
Index tests	EFW using HC, AC and FL
Target condition and reference standard(s)	Birth weight discordance

D'Antonio 2013a *(Continued)*

Flow and timing	Excluded women not described		
Comparative			
Notes			

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		

D'Antonio 2013a *(Continued)*

Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

D'Antonio 2014

Study characteristics			
Patient Sampling	Twin pregnancies registering for routine antenatal care by 11 weeks at antenatal care		
Patient characteristics and setting	2399 twin pregnancies; GA: N/A; antenatal care in hospital		
Index tests	EFW based on HC, AC and FL		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women with termination of pregnancy, structural or chromosomal abnormalities, pregnancies of unknown chorionicity, monochorionic, monoamniotic, and higher-order multiple gestations		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	

D'Antonio 2014 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Danon 2008

Study characteristics			
Patient Sampling	Twin pregnancies > 500 g; GA > 24 weeks: absence of fetal malformations or hydrops		
Patient characteristics and setting	278 twin pregnancies and 834 pregnancies in control; GA: mean 34.2 (SD 2.6); tertiary centre		
Index tests	EFW using Hadlock equation		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	EFW within 3 days of delivery. Excluded women had pregnancies complicated by gestational or pregestational diabetes		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Danon 2008 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Diaz-Garcia 2010

Study characteristics			

Diaz-Garcia 2010 *(Continued)*

Patient Sampling	Twin pregnancies with delivery ≥ 22 weeks and 1 US within 15 weeks of delivery
Patient characteristics and setting	283 participants (3-year period); GA mean 33.29 (SD 4.12) weeks; tertiary referral centre
Index tests	EFWD using a combination of formulas including BPD, HC, AC and FL
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	Excluded women did not have a 1st trimester US, had chromosomal abnormalities or congenital malformations
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
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?	Yes		

Diaz-Garcia 2010 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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	

Fox 2011

Study characteristics			
Patient Sampling	Twin pregnancies > 24 weeks cared for and delivered in maternal-fetal medicine practice centre		
Patient characteristics and setting	343 twin deliveries (August 2005 to June 2010). Delivered by Maternal-Fetal Medicine practice		
Index tests	EFW		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women with monoamniotic twins, GA < 24 weeks, pregnancies with didelphic or unicornuate uterus, and those with major fetal anomalies		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	

Fox 2011 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Hehir 2017

Study characteristics	
Patient Sampling	Twin pregnancies with 2 weekly US surveillance from 24 weeks' gestation with surveillance of monochorionic twins 2-weekly from 16 weeks at an academic perinatal centre
Patient characteristics and setting	1001 twin pairs (May 2007 to October 2009); GA: discordant: mean 35.3 (SD 3.0) weeks, concordant: mean 36.3 (SD 2.4) weeks; perinatal centres
Index tests	EFW based on BPD, HC, AC and FL

Hehir 2017 *(Continued)*

Target condition and reference standard(s)	Birth weight discordance
Flow and timing	Excluded women with monochorionic pregnancies with TTTS, a major structural abnormality, fetal aneuploidy or monoamnionicity
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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?	No		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Hehir 2017 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Hill 1994

Study characteristics			
Patient Sampling	Twin pregnancies with US ≥ 15 weeks		
Patient characteristics and setting	203 twin gestations; GA: 24.4–39.7 weeks, hospital		
Index tests	EFW from HC and AC		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women not reported		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	

Hill 1994 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Hoopmann 2011

Study characteristics			
Patient Sampling	Twin pregnancies with US at 18–25 weeks' gestation and 1 US within 14 days prior to delivery		
Patient characteristics and setting	196 twin pregnancies; GA: median 35.6 (interquartile range 32.4–37.0); university obstetrics/gynaecology department		
Index tests	EFW using BPD, HC, AC and FL		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Fetuses with structural or chromosomal defects were excluded. EFW within 14 days of delivery		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Hoopmann 2011 *(Continued)*

DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
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	

Jensen 1995

Study characteristics			

Jensen 1995 *(Continued)*

Patient Sampling	Twin pregnancies delivered at university hospital with final US within 7 days of delivery
Patient characteristics and setting	73 twin pregnancies (1 January 1990 to 31 March 1993); GA: mean 37 (range 19–40) weeks; university hospital
Index tests	EFW using BPD and AC
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	US was within 7 days of delivery
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	

Jensen 1995 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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	

Johansen 2014

Study characteristics			
Patient Sampling	Twin pregnancies with 2 live fetuses and chorionicity determined at the time of NT scan		
Patient characteristics and setting	2038 diamniotic twin pregnancies (1 January 2004 to 31 December 2006); GA: N/A; Department of Obstetrics/Gynaecology		
Index tests	CRL discordance		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women with pregnancies of unknown chorionicity, monochorionic monoamniotic pregnancies and pregnancies with known reduction from a higher number of multiples		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern

Johansen 2014 *(Continued)*

DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Kalish 2003

Study characteristics	
Patient Sampling	Twin pregnancies with US at 11–14 weeks' gestation
Patient characteristics and setting	157 dichorionic twin pregnancies (April 2000 to April 2002); GA: mean 35.5 (SD 2.7) weeks; medical centre
Index tests	CRL
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	Excluded women with monochorionic pregnancies, chromosomal anomalies and the loss of 1 or both fetuses

Kalish 2003 *(Continued)*

Comparative
Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		

Kalish 2003 *(Continued)*

Could the patient flow have introduced bias?	Unclear risk

Kim 2015

Study characteristics			
Patient Sampling	Twin pregnancies with delivery at ≥ 35 weeks' gestations in medical centre		
Patient characteristics and setting	253 twin pregnancies (January 2007 to December 2013); GA: N/A; medical centre		
Index tests	EFW		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women had complicated pregnancies such as preterm labour, premature rupture of membranes, placenta previa, pre-eclampsia, diabetes, TTTS, monoamniotic twin and congenital fetal anomaly		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	

Kim 2015 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Klam 2005

Study characteristics	
Patient Sampling	Diamniotic twin pregnancies with 2 live births from tertiary care centre
Patient characteristics and setting	503 diamniotic twin pregnancies (1 April 1994 to 11 January 2002); GA: growth discordant twins: mean 34.4 (SD 2.9) weeks, growth concordant twins: mean 35.6 (SD 2.9) weeks; tertiary care centre
Index tests	EFW and AC
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	Excluded women with chromosomal and structural anomalies, pregnancies complicated by TTTS and intrauterine death of 1 or both fetuses
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Klam 2005 *(Continued)*

DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Machado 2007

Study characteristics			

Machado 2007 *(Continued)*

Patient Sampling	Twin pregnancies with US at university medical school
Patient characteristics and setting	221 twin pregnancies; GA: mean 35.6 (SD 2.7) weeks; medical school
Index tests	EFW with HC, AC, BPD and FL
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	Excluded women with fetal malformation, TTTS, fetal death or unknown outcome
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	

Machado 2007 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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	

MacLean 1992

Study characteristics			
Patient Sampling	Twin pregnancies beyond 28 weeks' gestation, with on 2 US and 1 within 3 weeks of delivery recruited from antenatal clinic		
Patient characteristics and setting	107 twin pregnancies (1984–1987); GA: mean 36.1 (SD 2.4) weeks; antenatal clinic		
Index tests	EFW using AC formula and AC and BPD formula		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	US within 3 weeks of delivery		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			

MacLean 1992 *(Continued)*

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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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	

Murray 2014

Study characteristics	
Patient Sampling	Twin pregnancies from academic centres with 2-weekly US from 24 weeks' gestation with surveillance of monochorionic twins 2-weekly from 16 weeks
Patient characteristics and setting	960 twin pairs without TTTS; GA: N/A; academic centres
Index tests	EFW, BPD, AC and FL
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	Excluded women with twin pairs with TTTS
Comparative	

Murray 2014 *(Continued)*

Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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?	No		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Nakayama 2014

Study characteristics			
Patient Sampling	Monochorionic twin pregnancies delivered at single referral centre, CRL at 8–10 weeks' gestation		
Patient characteristics and setting	126 MCDA twin pregnancies; GA: mean 37.4 weeks; single referral centre		
Index tests	CRL		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women with terminated pregnancies, 1 or both fetal death before 10 weeks, twin-reversed arterial perfusion sequence, triplets or reduction of high-order multiples, and anomalies diagnosed before 10 weeks' gestation		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			

Nakayama 2014 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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	

O'Connor 2013

Study characteristics			
Patient Sampling	Twin pregnancies with CRL in 1st trimester, with BPD, HC, AC and FL measured. Included women had 2-weekly growth scans from 24 weeks' gestation		
Patient characteristics and setting	1001 twin pregnancies (May 2007 to October 2009); GA: N/A; perinatal centres		
Index tests	EFW, BPD, AC, FL and CRL		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women had monoamnionicity, a major structural abnormality or fetal aneuploidy		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		

O'Connor 2013 *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Reberdao 2010

Study characteristics	
Patient Sampling	Twin pregnancies with US within 2 weeks of delivery, and both twins were alive at antenatal diagnosis centre

Reberdao 2010 *(Continued)*

Patient characteristics and setting	124 twin pairs; GA: N/A; antenatal diagnosis centre
Index tests	EFW
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	US within 2 weeks of delivery. Excluded women not reported
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Reberdao 2010 *(Continued)*

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Sayegh 1993

Study characteristics			
Patient Sampling	Twin pregnancies with US at GA > 23 weeks		
Patient characteristics and setting	78 twin pairs; GA: N/A; general hospital		
Index tests	EFW		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women with congenital anomalies		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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		

Sayegh 1993 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Secher 1985

Study characteristics			
Patient Sampling	Twin pregnancies delivered at university hospital		
Patient characteristics and setting	80 twin pregnancies; GA: mean 251 (SD 22) days; hospital		
Index tests	EFW using BPD and abdominal diameter		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women not reported		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Secher 1985 *(Continued)*

DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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	

Shahshahan 2011

Study characteristics

Shahshahan 2011 *(Continued)*

Patient Sampling	Twin pregnancies with US at 7–14 week of pregnancy at hospital
Patient characteristics and setting	118 twin pregnancies; GA: mean 33.86 (SD 2.36) weeks; hospital
Index tests	CRL
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	Excluded women with monochorionic twins, and with 1st or 2nd trimester pregnancy terminations
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	

Shahshahan 2011 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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	

Simoes 2011

Study characteristics			
Patient Sampling	Twin gestations with US within 2 weeks before birth and both twins were born alive at university hospital		
Patient characteristics and setting	661 twin pregnancies (1 January 1994 to 30 June 2008); GA: concordant: mean 35.6 (SD 2.2) weeks, appropriate-for-gestational-age: mean 35.7 (SD 1.0) weeks, small-for-gestational-age: mean 33.5 (SD 2.7) weeks; hospital		
Index tests	EFW and AC		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	US measured within 2 weeks before birth		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern

Simoes 2011 *(Continued)*

DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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	

Storlazzi 1987

Study characteristics	
Patient Sampling	Twin pregnancies at university health centre, with US every 2 weeks, and US within 2 weeks of delivery
Patient characteristics and setting	43 twin pregnancies; GA: 33.5 weeks; university health centre
Index tests	EFW based on BPD and AC or FL and AC
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	Excluded women with congenital anomalies. US within 2 weeks of delivery

Storlazzi 1987 *(Continued)*

Comparative
Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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		

Storlazzi 1987 *(Continued)*

Could the patient flow have introduced bias?	Low risk

van de Waarsenburg 2015

Study characteristics	
Patient Sampling	Twin pregnancies with antenatal surveillance at university medical centre
Patient characteristics and setting	281 twin pregnancies (2008–2011); GA: median 35 + 0 weeks; medical centre
Index tests	EFW based on HC, AC and FL
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	Excluded women with monoamniotic twin pregnancies, selective feticide, congenital disorders, intrauterine fetal death, GA at birth < 22 weeks, or BW < 500 g
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

van de Waarsenburg 2015 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

van Mieghem 2009

Study characteristics			
Patient Sampling	MCDA twin pregnancies with EFW at 16, 20 and 26 weeks' gestation, and within 2 weeks of birth at university hospital		
Patient characteristics and setting	68 MCDA twin pregnancies; GA 35.7 weeks; hospital		
Index tests	EFW		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women with intrauterine fetal death or twin-reverse arterial perfusion sequence. EFW within 2 weeks of delivery		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		

van Mieghem 2009 *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
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	

Watson 1991

Study characteristics	
Patient Sampling	Twin pregnancies with fetal weight within 14 days of delivery
Patient characteristics and setting	94 twin pregnancies; GA: mean 33 (SD 2.5) weeks; hospital

Watson 1991 *(Continued)*

Index tests	BPD, mean abdominal diameters and FL
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	Excluded women with congenital anomalies. EFW within 14 days of delivery
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Watson 1991 *(Continued)*

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	

Zipori 2016

Study characteristics			
Patient Sampling	MCDA twin gestations with NT and CRL measurements on US at 11–13 weeks' gestation		
Patient characteristics and setting	89 twin pregnancies (August 2003 to August 2012); GA: mean 34.1 (SD) 3.3 weeks; hospital		
Index tests	EFW, CRL and NT		
Target condition and reference standard(s)	Birth weight discordance		
Flow and timing	Excluded women with lethal anomalies at the time of scan, loss of 1 or both twins prior to scan or altered chorionicity findings		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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		

Zipori 2016 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

Zuckerwise 2015

Study characteristics	
Patient Sampling	Twin pregnancies with MCDA placentation with antenatal care at single academic centre
Patient characteristics and setting	73 twin pairs; GA: non-discordant: mean 34.6 (SD 4.0) weeks, discordant: mean 32.4 (SD 4.6) weeks; academic centre
Index tests	EFW
Target condition and reference standard(s)	Birth weight discordance
Flow and timing	Excluded women with pregnancies complicated by TTTS
Comparative	
Notes	
Methodological quality	

Zuckerwise 2015 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
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	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
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	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
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	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
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	

AC: abdominal circumference; BPD: biparietal diameter; BW: birth weight; CRL: crown rump length; DADC: diamniotic dichorionic; EFW: estimated fetal weight; FL: femur length; GA: gestational age; HC: head circumference; MCDA: monochorionic diamniotic; N/A: not available; NT: nuchal translucency; SD: standard deviation; TTTS: twin-to-twin syndrome; US: ultrasound.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Allaf 2014	Research question did not include the reference standard.
Banks 2008	Index test difference in birth weight as reference did not match research question.
Ben-Ami 2015	Birthweight discordance cut-off did not match research question.
Bora 2009	Research question did not include the reference standard.
Brown 1987	Birthweight discordance cut-off did not match research question.
Chauhan 1996	Research question did not include the reference standard.
Chauhan 2004	Unable to construct 2 × 2 table from data.
D'Antonio 2013b	Research question did not include the reference standard.
Esinler 2017	Research question did not include the reference standard.
Fajardo-Expósito 2011	Index test cut-off did not match research question.
Gaziano 1998	Study did not address the index test.
Harper 2012a	Index test cut-off did not match research question.
Harper 2012b	Study did not address the index test.
Harper 2012c	Research question did not include the reference standard.
Harper 2013	Study did not address the index test.
Jensen 1992	Study did not address the index test.
Kremkau 1978	Article was not a published study.
Leombroni 2017	Article was a systematic review and meta-analysis.
Nakano 2015	Research question did not include the reference standard.
O'Brien 1986	Study did not address the index test.
Ocer 2011	Unable to construct 2 × 2 table from data.
Papaioannou 2011	Research question did not include the reference standard.
Ropacka-Lesiak 2012	Study did not address the index test.
Senoo 2000	Research question did not include the reference standard.
Shivkumar 2015	Research question did not include the reference standard.
Snijder 1998	Research question did not include the reference standard.

Study	Reason for exclusion
Stirrup 2016	Research question did not include the reference standard.
Suzuki 2009	Birthweight discordance cut-off did not match research question.
Weissmann-Brenner 2012	Birthweight discordance cut-off did not match research question.
Xu 1995	Research question did not include the reference standard.

DATA

Presented below are all the data for all of the tests entered into the review.

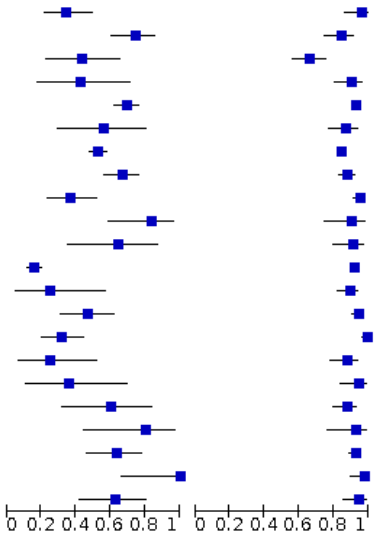
Table Tests. Data tables by test

Test	No. of studies	No. of participants
1 Cut-off 20%	22	8005
2 20% abdominal circumference	7	2846
3 20% femur length	7	2791
4 Cut-off 25%	18	6471
5 25% abdominal circumference	7	3614
6 25% femur length	4	2714

Test 1. Cut-off 20%

Cut-off 20%

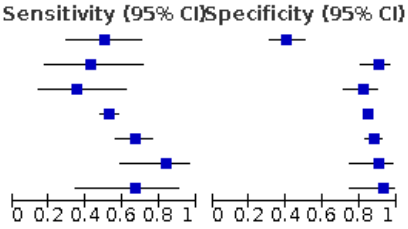
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bimson 2014a	17	2	32	47	0.35 [0.22, 0.50]	0.96 [0.86, 1.00]
Bimson 2014b	40	14	14	74	0.74 [0.60, 0.85]	0.84 [0.75, 0.91]
Blickstein 1996	10	33	13	65	0.43 [0.23, 0.66]	0.66 [0.56, 0.76]
Chamberlain 1991	6	7	8	64	0.43 [0.18, 0.71]	0.90 [0.81, 0.96]
Chang 2006	119	76	53	1009	0.69 [0.62, 0.76]	0.93 [0.91, 0.94]
Chauhan 1995	9	10	7	66	0.56 [0.30, 0.80]	0.87 [0.77, 0.94]
D'Antonio 2013a	229	266	203	1463	0.53 [0.48, 0.58]	0.85 [0.83, 0.86]
Diaz-Garcia 2010	60	23	30	170	0.67 [0.56, 0.76]	0.88 [0.83, 0.92]
Fox 2011	18	13	31	244	0.37 [0.23, 0.52]	0.95 [0.92, 0.97]
Hill 1994	15	3	3	28	0.83 [0.59, 0.96]	0.90 [0.74, 0.98]
Jensen 1995	9	5	5	49	0.64 [0.35, 0.87]	0.91 [0.80, 0.97]
Johansen 2014	46	142	242	1563	0.16 [0.12, 0.21]	0.92 [0.90, 0.93]
Kalish 2003	3	13	9	105	0.25 [0.05, 0.57]	0.89 [0.82, 0.94]
Kim 2015	21	11	24	197	0.47 [0.32, 0.62]	0.95 [0.91, 0.97]
Machado 2007	19	1	41	160	0.32 [0.20, 0.45]	0.99 [0.97, 1.00]
MacLean 1992	4	10	12	69	0.25 [0.07, 0.52]	0.87 [0.78, 0.94]
Secher 1985	4	3	7	47	0.36 [0.11, 0.69]	0.94 [0.83, 0.99]
Shahshahan 2011	9	13	6	90	0.60 [0.32, 0.84]	0.87 [0.79, 0.93]
Storlazzi 1987	8	2	2	26	0.80 [0.44, 0.97]	0.93 [0.76, 0.99]
van de Waarsenburg 2015	24	18	14	224	0.63 [0.46, 0.78]	0.93 [0.88, 0.96]
van Mieghem 2009	9	1	0	50	1.00 [0.66, 1.00]	0.98 [0.90, 1.00]
Watson 1991	17	4	10	63	0.63 [0.42, 0.81]	0.94 [0.85, 0.98]



Test 2. 20% abdominal circumference

20% abdominal circumference

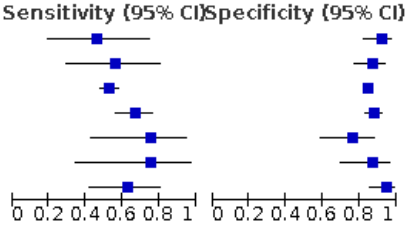
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Blickstein 1996	13	66	13	45	0.50 [0.30, 0.70]	0.41 [0.31, 0.50]
Chamberlain 1991	6	7	8	64	0.43 [0.18, 0.71]	0.90 [0.81, 0.96]
Chauhan 1995	6	14	11	62	0.35 [0.14, 0.62]	0.82 [0.71, 0.90]
D'Antonio 2013a	229	266	203	1463	0.53 [0.48, 0.58]	0.85 [0.83, 0.86]
Diaz-Garcia 2010	60	23	30	170	0.67 [0.56, 0.76]	0.88 [0.83, 0.92]
Hill 1994	15	3	3	28	0.83 [0.59, 0.96]	0.90 [0.74, 0.98]
Storlazzi 1987	8	2	4	24	0.67 [0.35, 0.90]	0.92 [0.75, 0.99]



Test 3. 20% femur length

20% femur length

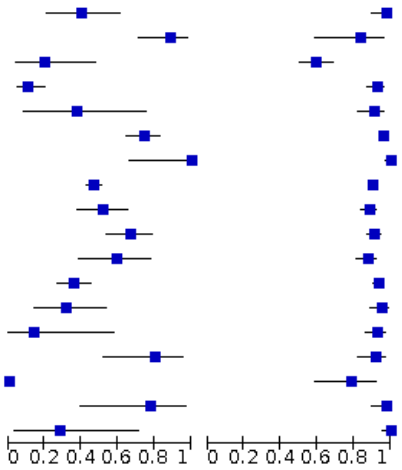
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chamberlain 1991	6	5	7	56	0.46 [0.19, 0.75]	0.92 [0.82, 0.97]
Chauhan 1995	9	10	7	66	0.56 [0.30, 0.80]	0.87 [0.77, 0.94]
D'Antonio 2013a	229	266	203	1463	0.53 [0.48, 0.58]	0.85 [0.83, 0.86]
Diaz-Garcia 2010	60	23	30	170	0.67 [0.56, 0.76]	0.88 [0.83, 0.92]
Hill 1994	9	9	3	28	0.75 [0.43, 0.95]	0.76 [0.59, 0.88]
Storlazzi 1987	6	4	2	26	0.75 [0.35, 0.97]	0.87 [0.69, 0.96]
Watson 1991	17	4	10	63	0.63 [0.42, 0.81]	0.94 [0.85, 0.98]



Test 4. Cut-off 25%

Cut-off 25%

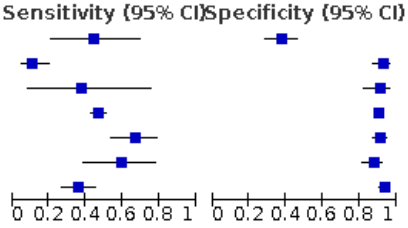
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bimson 2014a	10	1	15	50	0.40 [0.21, 0.61]	0.98 [0.90, 1.00]
Bimson 2014b	24	3	3	15	0.89 [0.71, 0.98]	0.83 [0.59, 0.96]
Blickstein 1996	3	50	12	74	0.20 [0.04, 0.48]	0.60 [0.50, 0.68]
Caravello 1997	9	12	71	150	0.11 [0.05, 0.20]	0.93 [0.87, 0.96]
Chamberlain 1991	3	7	5	70	0.38 [0.09, 0.76]	0.91 [0.82, 0.96]
Chang 2006	68	46	23	1120	0.75 [0.65, 0.83]	0.96 [0.95, 0.97]
Crane 1980	9	0	0	126	1.00 [0.66, 1.00]	1.00 [0.97, 1.00]
D'Antonio 2013a	253	157	287	1464	0.47 [0.43, 0.51]	0.90 [0.89, 0.92]
Danon 2008	28	26	26	198	0.52 [0.38, 0.66]	0.88 [0.83, 0.92]
Diaz-Garcia 2010	41	19	20	203	0.67 [0.54, 0.79]	0.91 [0.87, 0.95]
Hoopmann 2011	16	21	11	148	0.59 [0.39, 0.78]	0.88 [0.82, 0.92]
Klam 2005	39	25	70	369	0.36 [0.27, 0.46]	0.94 [0.91, 0.96]
Nakayama 2014	8	5	17	96	0.32 [0.15, 0.54]	0.95 [0.89, 0.98]
Reberdao 2010	1	8	6	105	0.14 [0.00, 0.58]	0.93 [0.87, 0.97]
Sayegh 1993	12	5	3	58	0.80 [0.52, 0.96]	0.92 [0.82, 0.97]
Simoes 2011	4	6	566	22	0.01 [0.00, 0.02]	0.79 [0.59, 0.92]
van Mieghem 2009	7	1	2	50	0.78 [0.40, 0.97]	0.98 [0.90, 1.00]
Zipori 2016	2	0	5	82	0.29 [0.04, 0.71]	1.00 [0.96, 1.00]



Test 5. 25% abdominal circumference

25% abdominal circumference

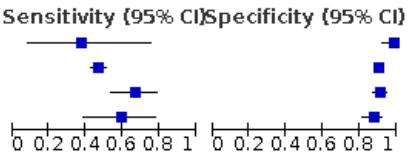
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Blickstein 1996	8	79	10	47	0.44 [0.22, 0.69]	0.37 [0.29, 0.46]
Caravello 1997	9	12	71	150	0.11 [0.05, 0.20]	0.93 [0.87, 0.96]
Chamberlain 1991	3	7	5	70	0.38 [0.09, 0.76]	0.91 [0.82, 0.96]
D'Antonio 2013a	253	157	287	1464	0.47 [0.43, 0.51]	0.90 [0.89, 0.92]
Diaz-Garcia 2010	41	19	20	203	0.67 [0.54, 0.79]	0.91 [0.87, 0.95]
Hoopmann 2011	16	21	11	148	0.59 [0.39, 0.78]	0.88 [0.82, 0.92]
Klam 2005	39	25	70	369	0.36 [0.27, 0.46]	0.94 [0.91, 0.96]



Test 6. 25% femur length

25% femur length

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chamberlain 1991	3	1	5	65	0.38 [0.09, 0.76]	0.98 [0.92, 1.00]
D'Antonio 2013a	253	157	287	1464	0.47 [0.43, 0.51]	0.90 [0.89, 0.92]
Diaz-Garcia 2010	41	19	20	203	0.67 [0.54, 0.79]	0.91 [0.87, 0.95]
Hoopmann 2011	16	21	11	148	0.59 [0.39, 0.78]	0.88 [0.82, 0.92]



ADDITIONAL TABLES

Table 1. Meta-analysis summary

Groups	No. of studies	Sensitivity (95% CI)	Specificity (95% CI)	LR+	LR–
BWD of 20%	**22**	**0.51 (0.42 to 0.60)**	**0.91 (0.89 to 0.93)**	**5.9 (4.3 to 8.1)**	**0.53 (0.44 to 0.64)**
EFWD by AC	7	0.57 (0.48 to 0.66)	0.84 (0.72 to 0.92)	3.6 (1.8 to 7.4)	0.51 (0.38 to 0.68)

Table 1. Meta-analysis summary *(Continued)*

EFWD by FL	7	0.60 (0.53 to 0.67)	0.87 (0.84 to 0.90)	4.6 (3.4 to 6.1)	0.46 (0.38 to 0.56)
BWD of 25%	**18**	**0.46 (0.26 to 0.66)**	**0.93 (0.89 to 0.96)**	**6.7 (3.0 to 14.9)**	**0.58 (0.39 to 0.88)**
EFWD by AC	7	0.42 (0.27 to 0.58)	0.88 (0.76 to 0.94)	3.3 (1.6 to 6.9)	0.67 (0.51 to 0.87)
EFWD by FL	4	0.55 (0.44 to 0.65)	0.91 (0.89 to 0.92)	5.8 (4.3 to 7.9)	0.50 (0.40 to 0.64)

AC: abdominal circumference; BWD: birth weight discordance; CI: confidence interval; EFWD: estimated fetal weight discordance; FL: femur length; LR+: likelihood ratio of a positive test; LR–: likelihood ratio of a negative test.

APPENDICES

Appendix 1. Search strategy

CENTRAL

(ultraso* or sonogra* or doppler) and (twin* or (multiple and (birth* or pregnanc*)) and (weigh* or size or grow* or discordan*))

Ovid MEDLINE

1. ultrasonography/ or exp ultrasonography, doppler/ or exp ultrasonography, prenatal/

2. Laser-Doppler Flowmetry/

3. ultraso*.mp.

4. sonogra*.mp.

5. doppler.mp.

6. acoustic microscop*.mp

7. us.fs.

8. or/1-7

9. exp Twins/ or Twin discordance/ or exp Pregnancy, Multiple/

10. twin*.mp.

11. (multiple adj3 (pregnanc* or birth*)).mp.

12. or/9-11

13. exp Birth Weight/ or Fetal Weight/ or exp Fetal Growth Retardation/ or exp Infant, Low Birth Weight/

14. (weigh* adj3 (discordan* or differ*)).mp.

15. bwd.mp.

16. (growth adj3 (discordan* or differ* or retard*)).mp.

17. ((birth or newborn or neonat* or fetal or foetal or fetus or foetus) adj3 (weigh* or grow* or size)).mp.

18. or/13-17

19. 8 and 12 and 18

Ovid Embase

1. echography/ or doppler echography/ or doppler flowmetry/ or exp fetus echography/

2. ultrasound/

3. exp fetal ultrasound monitor/

4. exp ultrasound transducer/

5. echograph/ or exp doppler device/ or exp ultrasound scanner/

6. (ultraso* or sonogra* or doppler or acoustic microscop*).mp.

7. or/1-6

8. exp twins/ or exp multiple pregnancy/

9. twin*.mp.

10. (multiple adj3 (pregnanc* or birth*)).mp.

11. or/8-10

12. exp birth weight/

13. exp intrauterine growth retardation/

14. (weigh* adj3 (discordan* or differ*)).mp.

15. bwd.mp.

16. (growth adj3 (discordan* or differ* or retardat*)).mp.

17. exp prenatal growth/

18. ((birth or newborn or neonat* or fetal or foetal or fetus or foetus) adj3 (weigh* or grow* or size)).mp.

19. or/12-18

20. 7 and 11 and 19

CINAHL

1. (MH "Ultrasonography") OR (MH "Ultrasonography, Doppler+") OR (MH "Ultrasonography, Prenatal+")

2. ultraso* or sonogra* or doppler or acoustic microscop*

3. twins

4. (MH "Twins") OR (MH "Pregnancy, Multiple+")

5. twin* or (multiple N3 (pregnanc* OR birth*)

6. (MH "Birth Weight") OR (MH "Fetal Weight")

7. (MH "Fetal Growth Retardat*")

8. weigh* N3 (discordan* or differ*)

9. bwd

10. growth N3 (discordan* OR differ* or retard*)

11. (birth or newborn or neonat* or fetal or foetal or fetus or foetus) N3 (weigh* or grow* or size)

12. s1 or s2

13. s3 or s4 or s5

14. s6 or s7 or s8 or s9 or s10 or s11

15. s12 and s13 and s14

Web of Science Core Collection

TOPIC: (ultraso* or sonogra* or doppler) AND TOPIC: (twin* or (multiple NEAR/2 birth*) or (multiple NEAR/2 pregnanc*)) AND TOPIC: (weigh* or size or grow* or discordan*)

DARE, NHS EED, HTA, Prospero, Trip database, PubMed systematic review subset

(ultraso* or sonogra* or doppler) and (twin* or (multiple and (birth* or pregnanc*)) and (weigh* or size or grow* or discordan*))

Appendix 2. QUADAS-2 risk of bias assessment tool

Domain 1 –participant selection	
Description	Describe methods of participant selection; describe included participants (previous testing, presentation, intended use of index test and setting)
Type of bias assessed	Selection bias, spectrum bias
Review question	Twin pregnancies with ultrasound measurements at any stage of pregnancy. The twin pregnancies included any type of chorionicity, any type of conception, any body mass index, any maternal age and any maternal condition
Information collected	Study objectives, study population, selection (inclusion/exclusion criteria), study design, clinical presentation, age, number enrolled and number available for analysis, setting, place and period of the study
Signalling question 1	***Did the study avoid inappropriate exclusions?***
Yes	If all the study participants with alive twin pregnancies were included
No	If the study selected participants based on particular clinical features. This section should include something indicated that the population was preselected and hence may not reflect the accurate findings applied to general population. Example: only participants with or without specific medical conditions; participants selected based on body-mass index; only participants attending certain number of previous ultrasound examinations; only participants selected on the basis of other antenatal tests
Unclear	If the study did not provide a clear definition of the selection (inclusion/exclusion) criteria and 'no' judgement was not applicable
Signalling question 2	***Was a consecutive or random sample of participants enrolled?***
Yes	If a consecutive sample or a random sample of the eligible participants was included in the study
No	If a non-consecutive sample or non-random sample of the eligible participants was included in the study
Unclear	If this information was unclear
Signalling question 3	***Was a case-control design avoided?***
Yes	If the study had a single set of inclusion criteria, defined by the clinical presentation (i.e. only participants with alive twin pregnancy)
No	If the study had > 1 set of inclusion criteria in respect to clinical presentation (i.e. participants with alive twin pregnancy and participants with singleton pregnancies or additional group of participants used as historical cohort)

(Continued)

Unclear	If it was unclear whether a case-control deign was avoided or not
Risk of bias	**Could the selection of participants have introduced bias?**
High	If 'no' classification for any of the above 3 questions
Low	If 'yes' classification for all the above 3 questions
Unclear	If 'unclear' classification for any of the above questions and 'high' risk judgement was not applicable
Concerns about applicability	**Are there concerns that the included participants do not match the review question?**
High	If the study population differed from the population defined in the review question in terms of demographic features and comorbidity
Low	If the study included only clinically relevant population that would have undergone the index test in real practice
Unclear	If this information was unclear; only studies with sufficient information on study population were included, therefore, none of the included studies were anticipated to be classified as 'unclear'
Domain 2 – index test	
Description	Describe the index test, how it was conducted and interpreted
Type of bias assessed	Test review bias, clinical review bias, interobserver variation bias
Review question	Any type of ultrasound method and any type of biometric measurements for assessment of estimated fetal weight
Information collected	Index test name, description of positive case definition by index test as reported, examiners (number, level of expertise, blinding), interobserver variability, conflict of interests
Signalling question 1	***Were the index test results interpreted without knowledge of the results of the reference standard?***
Yes	If the operators interpreting the index test were unaware of the results of reference standard. All the index tests are expected to be performed and interpreted before execution of the reference standard; therefore, none of the included studies are anticipated be classified as 'Yes' in response to this question
No	If the operators interpreting the index test were not blinded to the results of reference standard
Unclear	If this information was unclear
Signalling question 2	***Did the study provide a clear prespecified definition of what was considered to be a 'positive' result of index test?***
Yes	If study provided clear definition of positive findings, which was defined before execution/interpretation of index test
No	If definition of the positive result was not provided or if study described the findings derived from the index test and not defined prior to its execution
Unclear	If it was unclear whether the criteria were prespecified or not

(Continued)

Signalling question 3	***Was the level of expertise of the index test operator provided?***
Yes	If test was performed/interpreted either by single operator or interpreted after collegial discussion of the case
No	If test was performed/interpreted independently by several operators
Unclear	If this information was unclear
Signalling question 4	***Were the same clinical data available when the index test results were interpreted as would be available when the test is used in practice?***
Yes	If operators performing/interpreting the test were aware of suspected birth weight discordance or of the clinical history (or both), but were not aware of the results of other imaging tests or of previous diagnosis of birth weight discordance, including the results of previous ultrasounds
No	If operators performing/interpreting the test were not blinded to the results of other imaging tests or tests raising suspicion for birth weight discordance
Unclear	If this information was unclear
Risk of bias	**Could the conduct or interpretation of the index test have introduced bias?**
High	If 'no' classification for any of the above 4 questions
Low	If 'yes' classification for all the above 4 questions
Unclear	If 'unclear' classification at least for 1 question and 'high' risk judgement was not applicable
Concerns about applicability	**Are there concerns that the index test, its conduct or interpretation differ from the review question?**
High	We plan not to consider the studies where index tests other than ultrasound estimated fetal weight measurements were included or where index test looked at other target conditions not specified in the review (e.g. studies aimed at identifying maternal diseases); therefore, none of the included studies is expected to be classified as 'high concern'
Low	We considered all types of ultrasound methods and biometric parameters as eligible; therefore, all the included studies are expected to be classified as 'low concern'
Unclear	Only studies with sufficient information on index was included; therefore, none of the included studies were classified as 'unclear'
Domain 3 –reference standard	
Description	Describe the reference standard, how it was conducted and interpreted
Type of bias assessed	Verification bias, bias in estimation of diagnostic accuracy due to inadequate reference standard
Review question	Target condition – birth weight discordance
Information collected	Target condition, prevalence of target condition in the sample, reference standard, description of positive case definition by reference test as reported, examiners (number, level of expertise, blinding)
Signalling question 1	***Is the reference standard likely to correctly classify the target condition?***

(Continued)

Yes	If the study reported ≥ 1 of the following: calibration of scales, electronic scales, timing of weighing in relation to delivery or drying of neonate before weighing
No	If the reference standard did not classify the target condition correctly; considering the inclusion criteria, none of the studies are expected to be classified as 'no' for this item
Unclear	If information on execution of the reference standard, its interpretation or operators was unclear
Signalling question 2	***Were the reference standard results interpreted without knowledge of the results of the index tests?***
Yes	If operators performing the reference test were unaware of the results of index test
No	If operators performing the reference test were aware of the results of index test
Unclear	If this information was unclear
Risk of bias	**Could the reference standard, its conduct or its interpretation have introduced bias?**
High	If 'no' classification for any of the above 2 questions
Low	If 'yes' classification for all the above 2 questions
Unclear	If 'unclear' classification for any of the above 2 questions and 'high risk' judgement was not applicable
Concerns about applicability	**Are there concerns that the target condition as defined by the reference standard does not match the question?**
High	We excluded studies where participants did not have ultrasound measurement; therefore, none of the included studies are expected to be classified as 'high concern'
Low	Considering the inclusion criteria, all the studies are expected to be classified as 'low concern'
Unclear	Only studies where scales to assess birth weight and universal formula to calculate birth weight discordance served as a reference test was included; therefore, none of the included studies is expected to be classified as 'unclear concern'
Domain 4 –flow and timing	
Description	Describe any participants who did not receive the index tests or reference standard or who were excluded from the 2 × 2 table, describe the interval and any interventions between index tests and the reference standard
Type of bias assessed	Disease progression bias, bias of diagnostic performance due to missing data
Review question	Birth weight discordance assessed within 7 days of delivery
Information collected	Time interval between index test and reference standard, withdrawals (overall number of reported and if were explained)
Signalling question 1	***Was there an appropriate interval between index test and reference standard?***
Yes	If time interval was reported and was < 7 days
No	We excluded all studies where the time interval was > 7 days; therefore, none of the included studies is anticipated to be classified as 'no' for this item

(Continued)

Unclear	If time interval was not stated clearly, but authors description allowed the reader to assume that the interval was reasonably short
Signalling question 2	***Did all participants receive the same reference standard?***
Yes	If all participants underwent birth weight assessment and there are sufficient data to calculate birth weight discordance; considering the inclusion criteria, all the studies are expected to be classified as 'yes' for this item
No	If all participants did not undergo birth weight assessment or if only a subset of participants had this assessment, but the information on this population was not available in isolation
Unclear	If this information was unclear
Signalling question 3	***Were all participants included in the analysis?***
Yes	If all the participants were included in the analysis or if the participants were excluded prior to execution of index test or if the withdrawals were < 2% of the enrolled population (arbitrary selected cut-off)
No	If any participants were excluded from the analysis because of uninterpretable results, inability to undergo either index test or reference standard or unclear reasons
Unclear	None of the studies expected to be classified as 'unclear' for this item
Risk of bias	**Could the patient flow have introduced bias?**
High	If 'no' classification for any of the above 3 questions
Low	If 'yes' classification for all the above 3 questions
Unclear	If 'unclear' classification for any of the above 3 questions and 'high risk' judgement was not applicable

HISTORY

Protocol first published: Issue 3, 2017
Review first published: Issue 3, 2021

CONTRIBUTIONS OF AUTHORS

SJ: conceptualized the idea, wrote the review, incorporated reviewers' comments, and edited the text.

JH: made clinical recommendations.

SH: made clinical recommendations.

IA: made methodological and analytical recommendations.

MN: helped with the screening process.

CR: helped with quality assessment and data extraction.

MP: reviewed and revised the final draft, reanalyzed data including subgroup analyses, assisted in making revisions according to the DTA editorial team.

DECLARATIONS OF INTEREST

SJ: none.

JH: none.

SH: none.

IA: none.

MN: none.

CR: none.

MP: none.

SOURCES OF SUPPORT

Internal sources

- DTA group, Other

  DTA group provided technical support

External sources

- No sources of support supplied

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We were unable to identify an adequate number of papers looking into 18% or 30% BWD. Therefore, we analyzed the cut-off points of 20% and 25%.

We reanalyzed data using MIDAS package in STATA.

INDEX TERMS

Medical Subject Headings (MeSH)

Bias; *Birth Weight; Cohort Studies; Femur [anatomy & histology] [diagnostic imaging]; *Pregnancy, Twin; Reference Standards; Reproducibility of Results; Sensitivity and Specificity; Ultrasonography, Prenatal [*methods]; Waist Circumference

MeSH check words

Female; Humans; Pregnancy