

Received Date : 13-May-2016  
Revised Date : 06-Aug-2016  
Accepted Date : 29-Aug-2016  
Article type : Original Research Article

## **Sensitivity of first trimester ultrasound in the detection of congenital anomalies in twin pregnancies: population study and systematic review**

**Running headline:** First trimester detection of congenital anomalies in twins

Francesco D'Antonio<sup>1</sup>, Alessandra Familiari<sup>1</sup>, Basky Thilaganathan<sup>1</sup>, Aris T. Papageorghiou<sup>1</sup>, Lamberto Manzoli<sup>2</sup>, Asma Khalil<sup>1</sup> & Amar Bhide<sup>1</sup>

<sup>1</sup>Fetal Medicine Unit, Division of Developmental Sciences, St George's University of London, London, UK, <sup>2</sup>Department of Medicine and Aging Sciences, University of Chieti-Pescara and EMISAC, CeSI Biotech, Chieti, Italy.

### **Correspondence:**

Amar Bhide

Fetal Maternal Medicine Unit, St George's University of London, London SW17 0RE, UK

E-mail: abhide@sgul.ac.uk

### **Conflict of interest statement**

All authors have no conflict of interest to declare.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/aogs.13017

This article is protected by copyright. All rights reserved.

## Abstract

*Introduction:* The first aim of this study was to ascertain the diagnostic performance of first trimester ultrasound in detecting congenital anomalies in twins. The secondary aim was to explore the strength of association between different pregnancy characteristics and early detection of structural anomalies in a large unselected population of twin pregnancies. A systematic review of the published literature was also carried out. *Material and methods:* Retrospective analysis of prospectively collected data from consecutive twin pregnancies booked for antenatal care from 1996 till 2014. Predictive accuracy of those covariates independently associated with the occurrence of fetal anomalies were assessed with logistic regression analysis and ROC curves. *Results:* 1064 twin pregnancies (820 dichorionic and 264 monochorionic) were included in the analysis. 42 pregnancies had one or more fetuses with structural abnormalities. Detection of structural abnormalities using ultrasound was possible in the first trimester in 27.3% (95% CI: 15.0-42.8) of twin pregnancies. Monochorionicity (OR 2.3, 95% CI: 1.1-4.7) and discordance in crown-rump length and nuchal translucency were associated with an increased risk of fetal anomalies. However, their predictive accuracy was only moderate (AUC: 0.67, 95% CI: 0.6-0.8 and 0.68, 95% CI: 0.6 – 0.8, for crown-rump length and nuchal translucency discrepancy respectively). *Conclusions:* First trimester detection of structural abnormalities in twin pregnancies is possible in 27.3% (95% CI: 15.0-42.8) of cases. The likelihood for first trimester detection of structural anomalies in twins was maximum for cranial vault, midline brain and abdominal wall defects. Monochorionicity and increasing discrepancy in crown-rump length and nuchal translucency were associated with fetal structural abnormalities, although their predictive performance was only moderately good.

## Keywords

Prenatal diagnosis, multiple gestations, ultrasound, first trimester, congenital anomalies.

## Abbreviations:

CRL: crown-rump length

NT: nuchal translucency

OR, odds ratio

CI, confidence interval

DOR diagnostic odd ratio

### **Key message**

First trimester detection of structural abnormalities in twin pregnancies using ultrasound is possible in 27.3 % (95% CI: 15.0-42.8) of twin pregnancies. Monochorionicity and increasing discrepancy in crown-rump length and nuchal translucency are associated with fetal structural abnormality in twins, although their predictive performance is only moderately good.

### **Introduction**

Twin pregnancies are at increased risk of structural anomalies and aneuploidies compared to singletons (1-4). Several risk factors such as maternal age, assisted reproductive techniques, monochorionicity and discordance in either fetal size or nuchal translucency (NT) have been advocated to carry an increased risk of these anomalies and multiple pregnancies are usually scanned more frequently than singletons in order to promptly diagnose these conditions (1-5).

Advances in prenatal imaging techniques have led to an increased detection rate of structural anomalies in singletons (6-8). Examination of fetal anatomy in the first trimester scan depends upon several factors, such as fetal size, imaging protocol adopted, availability of transvaginal scan, maternal habitus and training of the sonographer (6). This has led in turn to an increase in the diagnosis of fetal anomalies during the first trimester of pregnancy, especially in those cases with increased NT (7,8).

A recent systematic review assessing the diagnostic performance of ultrasound in detecting congenital anomalies in singleton pregnancies showed that approximately half of fetal malformations can be detected at the 11-14 weeks' scan (9). Diagnostic accuracy was higher for cranial vault, midline and abdominal wall defects and lower for renal anomalies and heart defects (9).

There are no large studies assessing the accuracy of first trimester ultrasound in twins and it has not been completely elucidated yet whether the findings from singletons can be applicable to multiple gestations.

The first aim of this study was to ascertain the diagnostic performance of first trimester ultrasound in detecting congenital anomalies in twins. The secondary aim was to explore the strength of association between different pregnancy and fetal characteristics and early detection of structural anomalies in a large unselected population of twin pregnancies. A systematic review of the published literature was also carried out.

## **Material and methods**

### *Population study*

In this historical cohort study, all consecutive twin pregnancies booked for antenatal care at St. George's Hospital from 1996 till 2014 were included in the analysis. All women registering for routine antenatal care by 11 weeks of gestation were considered suitable for the analysis. Cases referred to the hospital for specialist advice were excluded in order to avoid referral bias. Scan details were obtained by a computerized search of the obstetric ultrasound computer database of the hospital. Details of pregnancy outcome were obtained from the computerized records of the maternity and neonatal unit. All data included in the analysis were collected prospectively but was analysed in retrospect. Written confirmation was obtained from the ethics committee that a formal approval for this retrospective study was not necessary.

All scans were performed trans-abdominally in the first instance. Transvaginal scanning was performed if visualisation was sub-optimal on trans-abdominal route. Gestational age was determined by the crown-rump length (CRL) of the larger twin at the 11-14 weeks' scan (10). At this time, NT was measured and fetal anatomy was assessed in order to exclude major structural abnormalities. Departmental protocol for anatomical evaluation was congruent with the International Society of Ultrasound in Obstetrics and Gynaecology (ISUOG) first trimester scan guideline (11). This consisted of demonstrating presence of skull and two symmetrical hemispheres, neck (evaluation of NT), presence of heart in the left side of the chest, presence of stomach and urinary bladder, integrity of the anterior abdominal wall and presence of four limbs (proximal and distal long bones). Detailed examination of the heart

and spine was not required. Examination of number of umbilical arteries in the cord was also not necessary. The examination protocol was the same from the beginning of the study and there have not been any change made during the study period.

Chorionicity was determined according to the presence of the lambda or T-sign and number of placentas and confirmed at birth (12) A detailed fetal structural survey was carried out at 20-22 weeks. In addition, all monochorionic twins had scans at 17 and 19 weeks in order to identify early features of twin-to-twin transfusion syndrome (TTTS). CRL discordance was calculated as  $100 \times (\text{larger CRL} - \text{smaller CRL}) / \text{larger CRL}$ . NT discordance was reported in mm.

For the purpose of the analysis, the structural anomalies were classified according to that reported by in the study by Whitlow *et al.* as follows (13):

- Central nervous system
- Face
- Neck
- Cardiovascular
- Lung
- Gastro-intestinal
- Renal
- Skeletal

In addition, those cases which did not fit into any of the above or those with multiple abnormalities were all classified into a miscellaneous class. For the purpose of the analysis, increased NT was not considered a structural malformation.

Distribution of the data was checked by Shapiro-Wilk test. Differences in the prevalence of fetal anomalies at birth were initially evaluated using chi-squared test for categorical variables and t-test or Kruskal-Wallis test for normally and non-normally distributed continuous variables respectively. The potential independent predictors of fetal anomalies at birth were then evaluated using stepwise forward logistic regression. All covariates were tested for inclusion in the final model, in which only those significant at either univariate or adjusted analysis were retained. To reduce potential over-fitting, the

overall number of covariates was limited to 1/10 of the anomalies in all phases of model fitting. The goodness-of-fit was checked using Hosmer-Lemeshow test, and the predictive power assessed through C-statistics (area under the Receiving Operator Curve). Standard post-estimation tests were used to check the final model validity, performing multicollinearity and influential observation analyses (using standardized residuals, change in Pearson and deviance chi-square). Missing values were less than 1% for all variables, thus no missing imputation technique was required.

We also analysed a sample restricted to the 42 fetuses with an anomaly (Supporting Information Table S1) confirmed at birth, in order to evaluate the potential predictors of a missed diagnosis at the first trimester scan. The differences in the rate of missed diagnosis were evaluated using the same approach above described for the anomalies at birth, but the logistic regression model was fitted with less than three covariates in all steps of model building (n. of missed diagnoses = 30).

The actual diagnostic performance of first trimester ultrasound is strictly dependent upon the imaging protocol adopted to screen for fetal anomalies. In order to evaluate the actual diagnostic performance of first trimester ultrasound we divided fetal anomalies in three categories as previously reported by Syngelaki *et al.* (7) and calculated the ratio between the anomalies detected at first trimester scan compared to those diagnosed later on in pregnancy or at birth. According to this proposed classification, fetal anomalies are divided into three different categories: those always detectable (body stalk anomaly, anencephaly, lobar holoprosencephaly, exomphalos, gastroschisis, megacystis), undetectable (microcephaly, agenesis of the corpus callosum, lobar holoprosencephaly, ventriculomegaly, fetal tumours, ovarian cysts, echogenic lung lesions, duodenal atresia, gastro intestinal obstruction, severe hydronephrosis not associated with megacystis) and potentially detectable (facial cleft, renal agenesis, multicystic kidneys, clubfeet, cardiac defects, spina bifida, diaphragmatic hernia, posterior fossa anomalies).

Predictive accuracy of those covariates independently associated with the occurrence of fetal anomalies were assessed by using receiver operating characteristics curve (ROC) and the values for sensitivity, specificity, positive (PPV) and negative (NPV) predictive values and positive (PLR) and negative (NLR) likelihood ratios and diagnostic odd ratio (DOR) were calculating accordingly. STARD guidelines for studies of diagnostic accuracy were followed (11). Statistical significance was defined as a two-sided p-value<0.05, and all analyses were carried out using Stata, version 13.1 (Stata Corp., College Station, Texas, USA, 2013).

### *Systematic review*

#### ***Eligibility criteria, information sources and search.***

This review was performed according to a prospectively designed protocol recommended for systematic reviews and meta-analysis (114-17). MEDLINE, EmBASE, CINAHL and Cochrane database were searched electronically from the 1<sup>st</sup> of January 2000 to the 2<sup>nd</sup> June 2015 utilizing combinations of the relevant medical subject heading (MeSH) terms, key words, and word variants for “first trimester”, “ultrasound,” and “twins” (Supporting Information Appendix S1). The search and selection criteria were restricted to English language. Reference lists of relevant articles and reviews were hand searched for additional reports. PRISMA guidelines were followed (18).

#### ***Study selection, data collection and data items***

Studies were assessed according to the following criteria: population, study design, type of imaging protocol. Only studies reporting a prospective diagnosis of fetal anomalies in the first trimester of pregnancy in an unselected population of twin pregnancies were considered suitable for the inclusion. Prospective and retrospective cohorts, case-control studies, case reports and case series were acceptable. Studies assessing the predictive accuracy of first trimester ultrasound exclusively in pregnancies at risk, such those with high NT or CRL discrepancy were not considered suitable for inclusion in this systematic review. Conference abstracts, case reports, case series with fewer than three cases and larger case series with a lack of information on false negatives were also excluded in order to avoid publication bias.

Two reviewers (FD, AF) independently extracted data. Inconsistencies were discussed by the reviewers and consensus reached. Quality of studies was assessed using the revised tool for the quality assessment of diagnostic accuracy studies (QUADAS-2). Each item was scored a “yes”, “no”, or “unclear” if there was insufficient information to make an accurate judgment (19).

The overall predictive accuracy of first trimester ultrasound in detecting congenital anomalies was aimed to be reported in terms of detection rate. Detection rate was calculated as the number of structural anomalies identified at the scan divided by the total number of those diagnosed after birth according to the classification proposed by Syngelaki *et al.* (7).

### ***Risk of bias, summary measures and synthesis of the results***

We aimed to use meta-analyses of proportions to combine data on the detection rate and obtain summary estimates of the sensitivity for each outcome (20). Limited number of studies did not permit meaningful stratified meta-analyses to explore the test performance in subgroups of patients that may be less or more susceptible to bias. The assessment of the potential publication bias was also problematic, both because of the outcome nature (rates with the left side limited to the value zero) which limits the reliability of funnel plots, and because of the small number of individual studies, which strongly limits the reliability of formal tests (20-22). In any case, funnel plots displaying the outcome rate from individual studies versus their precision (1/standard error) were carried out with an exploratory aim.

Between-study heterogeneity was explored using the  $I^2$  statistic, which represents the percentage of between-study variation that is due to heterogeneity rather than chance. A value of 0% indicates no observed heterogeneity, whereas  $I^2$  values of  $\geq 50\%$  indicate a substantial level of heterogeneity. A fixed effects model was used if substantial statistical heterogeneity was not present. On the contrary, if there was evidence of significant heterogeneity between studies included, a random effect model was used. All proportion meta-analyses were carried out using StatsDirect 2.7.9 (StatsDirect Ltd, Altrincham, UK).



## Results

### *Population study*

1064 twin pregnancies (820 dichorionic and 264 monochorionic) were included in the analysis. 44 fetuses in 42 pregnancies had abnormalities. General characteristics of the study population is shown in Table 1 and Supporting Information Table S1. The prevalence of fetal anomalies in the population analyzed was 4.0% (95% CI 2.9-5.3; 42/1064) per pregnancy (2.1% per fetus, 95% CI 1.5-2.8) (Table 2). Anomalies were present in both twins in two monochorionic pregnancies (four fetuses). Miscarriage and perinatal loss occurred in 8.93% (95% CI 7.3-10.8) and 1.60% (95% CI 0.9-2.5) of the cases respectively.

At univariate analysis, variables that were significantly associated with the presence of a fetal anomaly were monochorionicity, CRL discrepancy and NT discrepancy between twins (Table 3). There was no association with the use of *in-vitro* fertilization.

Multivariate analyses confirmed the results of the univariate analysis. The probability of fetal anomalies at birth significantly increased with increasing NT discrepancy between twins (odds ratio (OR): 1.05; 95% confidence interval (CI): 1.0-1.1 for each mm increase), and when a twin pregnancy was monochorionic rather than dichorionic (OR: 2.3; 95% CI: 1.1-4.7).

The rate of fetal anomalies detected in the first trimester was 27.3% per pregnancy (95% CI: 15.0%-42.8%) (29.6%, 95% CI: 16.8-45.2 per fetus). The diagnostic accuracy of ultrasound was excellent for brain anomalies involving cranial vault or midline structures, such as anencephaly, abdominal wall defects and complex anomalies, while it was poor for heart, kidney, face, limb anomalies and subtle brain anomalies such as ventriculomegaly (Table 3). When analyzing the diagnostic performance of first trimester ultrasound according to the classification system proposed by Syngelaki *et al.* (7), 100% (95% CI 66.4-100; 9/9) of the anomalies which should be always detected were identified during the first trimester scan, compared to only 16.7% (95% CI 3.6-41.4; 3/18) of the potentially detectable and none of those never undetectable (0%, 95% CI 0-21.8; 0/15) (Table 4).

When the analyses were restricted to the 42 pregnancies with anomalies at birth, CRL and NT discordance were significantly higher in fetuses with a first trimester ultrasound diagnosis of fetal anomaly ( $p < 0.001$  and  $p = 0.003$  respectively); the likelihood of having a

first trimester diagnosis significantly decreased with decreasing CRL and NT discrepancy between the twins. (Table 5).

The overall diagnostic performance of CRL and NT discrepancy in detecting fetal anomalies was only moderately good (CRL discrepancy: AUC: 0.7, 95% CI: 0.6-0.8; sensitivity: 76.2%, 95% CI: 60.6-88.0; specificity: 54.6%, 95% CI: 51.5-57.7; PPV: 6.5%, 95% CI: 4.5-9.0; NPV: 98.2%, 95% CI: 96.8-99.2; PLR: 1.7, 95% CI: 1.3-1.9; NLR: 0.4, 95% CI: 0.2-0.7; DOR: 3.9, 95% CI: 1.8-8.9 (Figure 1a). NT discrepancy: AUC: 0.7, 95% CI: 0.6 – 0.8, sensitivity: 71.4%, 95% CI: 55.4-84.3; specificity: 60.2%, 95% CI: 57.1-63.2; PPV: 6.9%, 95% CI: 4.7-9.7; NPV: 98.1%, 95% CI: 96.7-99.0; PLR: 1.8, 95% CI: 1.4-2.1; NLR: 0.5, 95% CI: 0.3-0.7; DOR: 3.8, 95% CI: 1.9-8.2) (Figure 1b).

#### *Systematic review*

The search yielded 1249 possible citations; of these, 1216 were excluded by reviewing the title or the abstract, as they did not meet the selection criteria. The remaining 33 full-text manuscripts were retrieved and only one study could be included (23) (Supporting Information Figure S1 and Table S2). In this prospective multicentre observational study, 495 pregnancies were examined from 1999 to 2003. Twenty-four cases (1.92% of the entire population) with malformations were diagnosed. Seven (29%) out of these were diagnosed antenatally, whose only three (12.5%, 95% CI 2.7-32.4) at the time of the 11-14 weeks scan (23) (Table 4).

#### **Discussion**

The findings from this study showed that the overall sensitivity of first trimester ultrasound in detecting structural fetal anomalies in twin pregnancies is 27.3 % (95% CI: 15.0-42.8),. Monochorionicity and increased discrepancy in either CRL and NT were associated with the occurrence of fetal structural abnormalities CRL and NT discordance were also independently associated with an early diagnosis of fetal anomalies; however, their predictive accuracy was only moderately good, thus limiting their adoption as a screening test for fetal structural anomalies in clinical practice. In the only other study identified through the systematic review, the diagnostic performance of first trimester ultrasound was high for cranial vault anomalies and low for facial, spinal and cardiac defects, thus reflecting the findings of this study (23).

Advances in prenatal imaging techniques have led to an increased detection rate of fetal structural anomalies (6-8).

A recent systematic review assessing the diagnostic performance of ultrasound in detecting structural anomalies in singleton pregnancies showed that approximately half of such anomalies can be detected during the first trimester of pregnancy; the diagnostic accuracy was higher for cranial vault, midline and abdominal wall defects and lower for conditions like kidney anomalies and heart defects (9). However, the diagnostic performance of first trimester ultrasound was extremely variable among the included studies. Several factors, such as operator's experience, ultrasound equipment, type of scan (trans-abdominal vs trans-vaginal), gestational age at scan, imaging protocol adopted and time of follow-up play a major role in this scenario, thus partially explaining the conflicting results reported in previous studies.

The findings from this study are comparable to those reported in singletons and showed an overall high detection rate for cranial vault and midline anomalies, while diagnostic performance of first trimester ultrasound in detecting cardiac and renal anomalies was low. Renal anomalies, such as hydronephrosis, are usually detected during the second and third trimesters of pregnancy and may not be apparent on first trimester ultrasound. In this scenario, inclusion of visualisation of the kidneys in the anatomical first trimester survey is unlikely to improve the detection rate for these anomalies.

Increasing discordance in CRL or NT has been associated with a higher risk of adverse perinatal outcomes such as fetal loss, weight discordance, fetal anomalies and preterm delivery in previous series (26-29). Impaired fetal growth in early pregnancy and the presence of underlying fetal chromosomal or structural anomalies have been hypothesized to explain this phenomenon. A previous systematic review showed that, although associated, neither CRL nor NT discordance were predictive of adverse pregnancy outcome. In that study, the diagnostic performance of CRL discordance in predicting the occurrence of fetal structural anomalies could not be assessed due to the paucity of data (28). In the present study, discordance in either NT and CRL was associated with a higher chance of first trimester detection of structural anomalies in twins. Increasing discordance in NT or CRL is likely to be a trigger for a more thorough search for structural abnormalities, thus partially explaining the results reported in this study. In a recent prospective study Sperling *et al.*

reported that combining NT with an early scan at 19 weeks of gestation gave a detection rate of 83% for the diagnosis of congenital anomalies (23).

Based on this data, it appears clear that ultrasound assessment of twin pregnancies in centre with high level of expertise in fetal medicine is warranted.

The retrospective nature of the study and short period of follow-up are the major limitations of this study. Furthermore, the study period covers a wide time window (20 years), thus potentially affecting the figures for the diagnostic accuracy reported. However, we did not find a significant improvement of first trimester ultrasound across time (Table 3). Finally, a sub-analysis according to the gestational age at scan could not be performed in view of the short time window considered (11-14 weeks).

Large sample size, exclusion of referrals from the population analysed and the adoption of a specific imaging protocol for early detection of structural anomalies during the entire study period represent the main strengths of the present study.

### *Conclusions*

First trimester detection of structural abnormalities in twin pregnancies is possible in 27.27 % (95% CI: 15.0-42.8) of cases. The likelihood for first trimester detection of congenital anomalies was maximum for cranial vault, midline brain and abdominal wall defects. Monochorionicity and increasing discrepancy in CRL and NT were associated with fetal structural abnormalities, although their diagnostic performance was only moderately good. First trimester assessment in centres with high level of expertise in fetal medicine is warranted in order early diagnose these anomalies and to correctly stratify the antenatal risk of these pregnancies.

### **Funding**

No funding was obtained for this study

## REFERENCES

1. Allen SR, Gray LJ, Frentzen BH, Cruz AC. Ultrasonographic diagnosis of congenital anomalies in twins. *Am J Obstet Gynecol* 1991;165:1056-60.
2. Hardin J, Carmichael SL, Selvin S, Lammer EJ, Shaw GM. Increased prevalence of cardiovascular defects among 56,709 California twin pairs. *Am J Med Genet A*. 2009;149A: 877-86.
3. Rider RA, Stevenson DA, Rinsky JE, Feldkamp ML. Association of twinning and maternal age with major structural birth defects in Utah, 1999 to 2008. *Birth Defects Res A Clin Mol Teratol*. 2013;97:554-63.
4. Schinzel AA, Smith DW, Miller JR. Monozygotic twinning and structural defects. *J Pediatr* 1979;95:921-30.
5. Glinianaia SV, Rankin J, Wright C. Congenital anomalies in twins: a register-based study. *Hum Reprod*. 2008;23:1306-11.
6. Souka AP, Pilalis A, Kavalakis Y, Kosmas Y, Antsaklis P, Antsaklis A. Assessment of fetal anatomy at the 11-14-week ultrasound examination. *Ultrasound Obstet Gynecol*. 2004;24:730-4.
7. Syngelaki A, Chelemen T, Dagklis T, Allan L, Nicolaides KH. Challenges in the diagnosis of fetal non-chromosomal abnormalities at 11-13 weeks. *Prenat Diagn*. 2011;31:90-102.
8. Carvalho MH, Brizot ML, Lopes LM et al. Detection of fetal structural abnormalities at the 11-14 week ultrasound scan. *Prenat Diagn*. 2002;22:1-4.
9. Rossi AC, Prefumo F. Accuracy of ultrasonography at 11-14 weeks of gestation for detection of fetal structural anomalies: a systematic review. *Obstet Gynecol*. 2013;122:1160-67
10. Dias T, Mahsud-Dornan S, Thilaganathan B, Papageorgiou A, Bhide A. First-trimester ultrasound dating of twin pregnancy: are singleton charts reliable? *BJOG*. 2010;117:979-984.
11. Salomon LJ, Alfirevic Z, Bilardo CM, Chalouhi GE, Ghi T, Kagan KO, et al. ISUOG practice guidelines: performance of first-trimester fetal ultrasound scan. *Ultrasound Obstet Gynecol*. 2013;41:102-13.
12. Dias T, Arcangeli T, Bhide A, Napolitano R, Mahsud-Dornan S, Thilaganathan B. First-trimester ultrasound determination of chorionicity in twin pregnancy. *Ultrasound Obstet Gynecol*. 2011;8:530-32.

- Accepted Article
13. Whitlow BJ, Chatzipapas IK, Lazanakis ML, Kadir RA, Economides DL. The value of sonography in early pregnancy for the detection of fetal abnormalities in an unselected population. *Br J Obstet Gynaecol.* 1999;106:929-36.
  14. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015;351:h5527.
  15. Henderson LK, Craig JC, Willis NS, Tovey D, Webster AC. How to write a Cochrane systematic review? *Nephrology (Carlton).* 2010;15:617-24.
  16. NHS Centre for Reviews and Dissemination. *Systematic reviews: CRD's guidance for undertaking reviews in health care.* University of York: York (UK), 2009.
  17. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Ann Intern Med.* 2008;149:889-897.
  18. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6:e1000097.
  19. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155:529-536.
  20. Manzoli L, De Vito C, Salanti G, D'Addario M, Villari P, Ioannidis JP. Meta-analysis of the immunogenicity and tolerability of pandemic influenza A 2009 (H1N1) vaccines. *PLoS One.* 2011;6: e24384.
  21. Hunter JP, Saratzis A, Sutton AJ, Boucher RH, Sayers RD, Bown MJ. In meta-analyses of proportion studies, funnel plots were found to be an inaccurate method of assessing publication bias. *J Clin Epidemiol.* 2014;67:897-903.
  22. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ.* 1997;315:629-34.
  23. Sperling L, Kiil C, Larsen LU, Brocks V, Wojdemann KR, Qvist I, et al. Detection of chromosomal abnormalities, congenital abnormalities and transfusion syndrome in twins. *Ultrasound Obstet Gynecol.* 2007;29:517-26.
  24. Pilalis A, Basagiannis C, Eleftheriades M, Faros E, Troukis E, Arnelidou E, Papastefanou I, Souka AP. Evaluation of a two-step ultrasound examination protocol for the detection of major fetal structural defects. *J Matern Fetal Neonatal Med.* 2012;25:1814-7.

25. Windham GC, Bjerkedal T. Malformations in twins and their siblings, Norway, 1967-79. *Acta Genet Med Gemellol (Roma)*. 1984;33:87-95.
26. D'Antonio F, Khalil A, Mantovani E, Thilaganathan B; on Behalf of the Southwest Thames Obstetric Research Collaborative (STORK). Embryonic growth discordance and early fetal loss: the STORK multiple pregnancy cohort and systematic review. *Hum Reprod*. 2013;28: 2621-7.
27. D'Antonio F, Khalil A, Dias T, Thilaganathan B; Southwest Thames Obstetric Research Collaborative. Crown-rump length discordance and adverse perinatal outcome in twins: analysis of the Southwest Thames Obstetric Research Collaborative (STORK) multiple pregnancy cohort. *Ultrasound Obstet Gynecol*. 2013;41:621-6.
28. D'Antonio F, Khalil A, Pagani G, Papageorgiou AT, Bhide A, Thilaganathan B. Crown-rump length discordance and adverse perinatal outcome in twin pregnancies: systematic review and meta-analysis. *Ultrasound Obstet Gynecol*. 2014;44:138-46
29. Edwards MS, Ellings JM, Newman RB, Menard MK. Predictive value of antepartum ultrasound examination for anomalies in twin gestations. *Ultrasound Obstet Gynecol*. 1995;6:43-49.
30. Buskens E, Grobbee DE, Hess J, Wladimiroff JW. Prenatal diagnosis of congenital heart disease; prospects and problems. *Eur J Obstet Gynecol Reprod Biol*. 1995;60:5-11.

**Supporting Information legends:**

**Table S1:** Individual patient information in dichorionic-diamniotic (DCDA) monochorionic-diamniotic (MCDA) twin pregnancies with structural anomalies.

**Appendix S1.** Search strategy.

**Figure S1.** Systematic review flowchart.

**Table S2:** Excluded studies and reason for the exclusion.

**Legends figure and tables:**

**Figure 1.** Receiver operating characteristics curves (AUC) showing A) the diagnostic accuracy of crown-rump length (CRL) and B) nuchal translucency (NT) discordance in detecting structural anomalies in the first trimester in twin pregnancies.

**Table 1:** General characteristics of the sample (n=1064).

**Table 2:** Rate of structural fetal anomalies detected in twin pregnancy at first trimester ultrasound.

**Table 3.** Selected maternal and fetal characteristics, stratified by fetal anomaly at birth.

**Table 4:** Expected and observed detection of anomalies in the first trimester in twins.

**Table 5.** Analysis restricted to the 42 pregnancies with confirmed fetal anomalies at birth: selected maternal and fetus characteristics, stratified by first trimester diagnosis.



**Table 1:** General characteristics of the sample (n=1064).

<i>Variables</i>	<b>Overall sample</b>
Mean maternal age (SD), years	32.4 (5.4)
Mean maternal BMI (SD), kg/m <sup>2</sup>	24.5 (4.7)
<i>Year of scan, %</i>	
- 1996-2002	21.4
- 2003-2008	39.2
- 2009-2014	39.4
<i>Mode of conception, %</i>	
- Spontaneous	72.8
- IVF	27.2
<i>Parity, %</i>	
- Nulliparous	87.5
- Parous	12.5
<i>Chorionicity, %</i>	
- MCDA	22.9
- DCDA	77.1
Mean gestational age in weeks at the first trimester scan (SD)	12.2 (0.8)
Mean CRL discrepancy (SD), %	4.7 (6.6)
Mean NT discrepancy (SD), %	15.3 (14.3)
Mean NT discrepancy (SD), mm	0.30 (0.6)
Fetal anomalies at birth, %	3.9
Type of fetus anomalies at birth, %	
- None	96.1
- SNC	0.6
- Face	0.2
- Neck	0
- Cardiovascular	0.9
- Lung	0
- Gastrointestinal	0.5
- Renal	1.0
- Skeletal	0.5
- Miscellaneous	0.6

IVF = In vitro fertilization. MCDA = Monochorionic diamniotic. DCDA = Dichorionic diamniotic. CRL = Crown rump length. NT = Nuchal translucency.

**Table 2:** Rate of structural fetal anomalies detected in twin pregnancy at first trimester ultrasound

Type of anomaly at birth	1 <sup>st</sup> trimester diagnosis n/N (%)
SNC	2/6 (33.3)
Face	0/2 (0.0)
Cardiovascular	2/10 (20.0)
Gastro-intestinal	3/4 (75.0)
Renal	1/11 (9.1)
Skeletal	0/5 (0.0)
Miscellaneous	5/6 (83.3)

**Table 3.** Selected maternal and fetal characteristics, stratified by fetal anomaly at birth.

Variables	Fetus anomalies at birth		Unadjusted OR (95% CI)	Adjusted OR (95% CI)
	No (n=1022)	Yes (n=42)		
Mean maternal age (SD), y	32.4 (5.4)	33.0 (5.6)	1.0 (0.96-1.1)	--
Mean maternal BMI (SD), Kg/ m <sup>2</sup>	24.5 (4.7)	24.4 (3.8)	1.0 (0.9-1.1)	--
<i>Year of scan, %</i>				
- 1996-2002	21.2	26.2	1 (Ref. Cat.)	
- 2003-2008	39.4	33.3	0.7 (0.4-1.5)	--
- 2009-2014	39.3	40.5	0.8 (0.4-1.8)	--
<i>Mode of conception, %</i>				
- Spontaneous	72.9	71.4	1 (Ref. Cat.)	
- IVF	27.1	28.6	1.1 (0.5-2.1)	--
<i>Parity, %</i>				
- Nulliparous	89.6	35.7	1 (Ref. Cat.)	1 (Ref. Cat.)
- Parous	10.4	64.3	15.6 (8.0-30.2)	17.1 (8.5-34.7)
<i>Chorionicity, %</i>				
- DCDA	77.6	64.3	1 (Ref. Cat.)	1 (Ref. Cat.)
- MCDA	22.4	35.7	1.9 (1.0-3.7)	2.3 (1.1-4.7)
Mean gestational age in weeks at the first trimester scan (SD)	12.2 (0.8)	12.0 (0.8)	0.6 (0.3-1.5)	--
Mean CRL discrepancy (SD), %	4.6 (6.6)	7.4 (5.9)	1.0 (1.0-1.1)	1.0 (0.98-1.1)
Mean NT discrepancy (SD), mm	0.3 (0.4)	1.0 (1.7)	2.1 (1.6-2.9)	1.9 (1.3-2.7)

OR = Odds ratio. CI = Confidence Interval. Ref. cat. = Reference category. IVF = In vitro fertilization. MCDA = Monochorionic diamniotic. DCDA = Dichorionic diamniotic. CRL = Crown-rump length. NT = Nuchal translucency.

**Table 4.** Analysis restricted to the 42 pregnancies with confirmed fetal anomalies at birth: selected maternal and fetus characteristics, stratified by first trimester diagnosis.

	<b>Non diagnosed At 1<sup>st</sup> trimester</b> (n=30; 71.4%)	<b>Diagnosed At 1<sup>st</sup> trimester</b> (n=12; 28.6%)	<b>OR (95% CI)</b>
Mean maternal age (SD), y	32.9 (5.2)	33.2 (6.8)	--
Mean maternal BMI (SD), Kg/ m <sup>2</sup>	23.8 (3.2)	25.8 (5.0)	--
<i>Year of scan, %</i>			--
- 1996-2002	72.7	27.3	
- 2003-2008	78.6	21.4	
- 2009-2014	64.7	35.3	
<i>Mode of conception, %</i>			--
- Spontaneous	73.3	26.7	
- IVF	66.7	23.3	
<i>Parity, %</i>			--
- Nulliparous	86.7	13.3	
- Multiparous	63.0	37.0	
<i>Chorionicity, %</i>			--
- DCDA	66.7	33.3	
- MCDA	80.0	20.0	
Mean gestational age in weeks at the first trimester scan (SD)	12.3 (0.9)	11.9 (0.8)	--
Mean CRL discrepancy (SD), %	5.3 (4.6)	12.5 (5.7)	0.8 (0.7-0.96)
Mean NT discrepancy (SD), mm	0.5 (0.9)	2.2 (2.5)	0.6 (0.3-1.1)

IVF = In vitro fertilization. MCDA = Monochorionic diamniotic. DCDA = Dichorionic diamniotic. CRL = Crown-rump length. NT = Nuchal translucency. NS = Not significant in neither univariate nor multivariate analysis and not included in the final model. OR = Odds ratio. CI = Confidence Interval.

**Table 5:** Expected and observed detection of anomalies in the first trimester in twins.

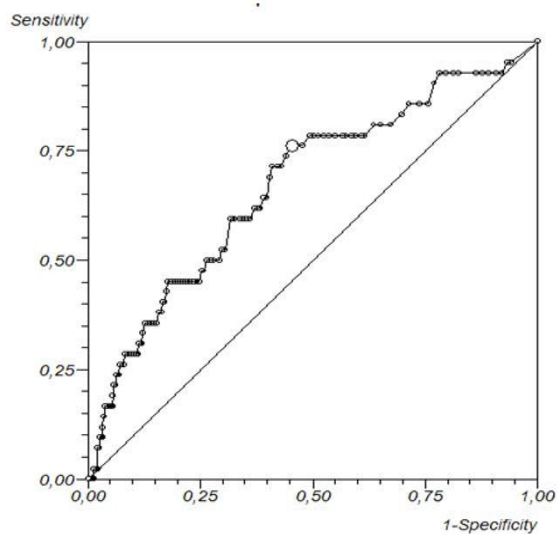
Study	Anomalies	Expected detection rate of anomaly		
		<i>Always detectable</i>	<i>Potentially detectable</i>	<i>Undetectable</i>
<b>Present study</b> (n = 1064)	Total number (n = 42)	9	18	15
	First trimester diagnosis (n = 12)	9	3	0
<b>Sperling et al</b> (n = 495)	Total number (n = 24)	1	20	3
	First trimester diagnosis (n = 3)	1	2	0

**Always detectable:** Body stalk anomaly, anencephaly, alobar holoprosencephaly, exomphalos, gastroschisis, megacystis

**Undetectable:** Microcephaly, agenesis of the corpus callosum, lobar holoprosencephaly, ventriculomegaly, fetal tumours, ovarian cysts, echogenic lung lesions, duodenal atresia, gastro intestinal obstruction, severe hydronephrosis not associated with megacystis.

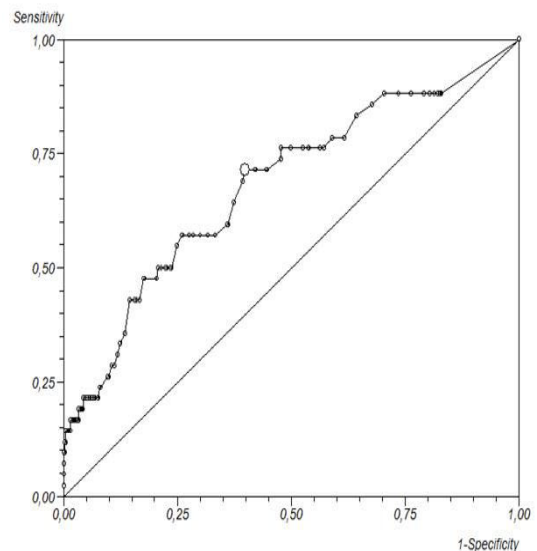
**Potentially detectable:** Facial cleft, renal agenesis, multicystic kidneys, clubfeet, cardiac defects, spina bifida, diaphragmatic hernia, posterior fossa anomalies

**CRL discrepancy**



a)

**NT discrepancy**



b)