Systematic review of first trimester ultrasound screening in detecting fetal structural anomalies and factors affecting screening performance

Affiliations

Jehan N. Karim[1], Nia W. Roberts[2], Laurent J Salomon[3], Aris T Papageorghiou[1]

1. Nuffield Department of Obstetrics & Gynaecology, University of Oxford, Oxford, UK
2. Bodleian Health Care Libraries, University of Oxford, Oxford, UK
3. Maternité Necker-Enfants Malades, AP-HP, Université Paris Descartes, Paris, France

Keywords

First trimester, congenital abnormalities, fetal anomalies, prenatal screening, prenatal diagnosis, ultrasound.

Abstract

Objectives To determine the sensitivity and specificity of first trimester ultrasound for the detection of fetal abnormalities; and to establish which factors might impact this screening.

Methods

Systematic review and meta-analysis of all relevant publications assessing the diagnostic accuracy of first trimester 2D (transabdominal and transvaginal) ultrasound in the detection of congenital fetal anomalies prior to 14 weeks gestation was performed. The reference standard used was the detection of abnormalities at birth or postmortem. Factors that may impact detection rates were evaluated including population characteristics, gestation, healthcare setting, ultrasound modality, use of an anatomical checklist for first trimester anomaly detection and what types of malformations were included in the study. In an effort to

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: 10.1002/uog.17246

reduce the impact of study heterogeneity on results of the meta-analysis, data from the studies were analyzed within subgroups of major anomalies versus all types of anomalies; and low risk / unselected populations versus high risk populations.

Results:

An initial electronic search identified 2,225 citations, from which a total of 30 relevant studies, published between 1991 and 2015, were selected for inclusion. For low risk or unselected populations (19 studies, 115,731 fetuses) the pooled estimate for detection of major abnormalities was 46.10% (95% C.I. 36.88-55.46). The detection rate for all abnormalities in low risk or unselected populations was 32.35% (95% C.I. 22.45-43.12), in 14 studies (97,976 fetuses); while the detection rate in high risk populations for the presence of all types of anomalies (six studies, 2,841 fetuses) was 61.18% (95% C.I. 37.71-82.19). Of the factors examined impacting detection rates there was a statistically significant relationship between the use of an anatomical protocol during first trimester anomaly screening and sensitivity for the detection of fetal anomalies in all subgroups ($P<0.0001$).

Conclusion

Detection rates for first trimester anomalies range from 32% in low risk, to over 60% in high risk groups. This demonstrates that first trimester ultrasound has the potential to identify a large proportion of fetuses affected with structural anomalies. The use of a standardized anatomical protocol improves the sensitivity of first trimester ultrasound screening for all anomalies and major anomalies in populations of varying risk. The development and introduction of international protocols with standard anatomical views should be undertaken, in order to optimize rates of first trimester anomaly detection.

Introduction

The main objectives of prenatal ultrasound at 11-13+6 weeks of gestation are confirming fetal viability; establishing an accurate gestational age from the measurement of fetal crown-rump length (CRL); identifying multiple pregnancies and determining their chorionicity and amnionicity; and screening for major fetal anomalies - both structural abnormalities and aneuploidies [(1)]. In many settings, screening for chromosomal anomalies is undertaken by measurement of fetal nuchal translucency (NT), in combination with maternal age; other ultrasound markers (e.g. the fetal nasal bone, ductus venosus flow, fetal heart rate and assessment of tricuspid valve flow); and measurement of maternal serum free beta hCG and PAPP-A, in the form of a combined test. This screening method is associated with high sensitivity and a relatively low false positive rate [(2-4)].

The recent development of cell-free DNA screening using maternal blood is transforming our expectations of first trimester aneuploidy detection[(5)]. As this innovative technology becomes increasingly accessible and cost-effective, it will complement (and may ultimately supersede) current combined screening. Thus, the objectives of the first trimester ultrasound scan will need to evolve once again. The many advantages of first trimester ultrasound mean it is likely to continue; this should include measurement of fetal NT, as increased NT is linked to structural congenital anomalies, notably major cardiac defects[(6)]; and may also be indicative of chromosomal aberrations that are not detectable with cell-free DNA screening. In addition, first trimester screening for

pregnancy complications , such as early pre-eclampsia and very early preterm birth, are commonly used as a tool for pregnancy risk stratification[7].

Nevertheless, it is our opinion that the resulting shift in aneuploidy screening, in combination with improvements in ultrasound technology, will mean that the role of first trimester ultrasound will increasingly be in the visualization of fetal anatomy[8-11]. Thus, while fetal anomaly screening has traditionally been performed in the second trimester, many structural abnormalities can reliably be diagnosed between 11 and 14 weeks[12-14] with obvious advantages. However, the varying sensitivity of the test means there is little consensus internationally as to whether first trimester anomaly screening is valuable for use in daily clinical practice. In addition there is currently a limited understanding of which factors impact the success of first trimester anomaly detection, and how this screening should be optimally performed.

Within this context, our aim was to perform a systematic review and meta-analysis of current literature in order to assess the sensitivity and specificity of first trimester anomaly detection; and crucially, to determine which factors might impact this screening test.

Methods

Search Strategy

A systematic electronic search was conducted in order to identify all relevant publications assessing the diagnostic accuracy of first trimester two-dimensional ultrasound for the

detection of congenital fetal anomalies. The search was conducted using Medline, Embase, Web of Science and The Cochrane Library with no publication-year restrictions. Free-text terms and medical subject headings related to prenatal screening, early pregnancy and congenital anomalies were used (please see Supplement A for full search strategy). The electronic search was completed on July 29th, 2015.

Study selection was performed in multiple stages. Initially, the database of studies collected from the electronic search was screened using article titles and abstracts, where available. On this basis, a list of potentially suitable articles for inclusion in the systematic review was formulated. The full texts of these articles were then assessed in order to determine which studies met the inclusion criteria. Reference lists of all eligible studies were screened for additional citations, which may not have been identified by the initial electronic search.

Study Selection

All studies, which reported the detection of fetal structural abnormalities using two-dimensional ultrasound prior to 14 weeks gestation, were included. Prospective observational studies, retrospective observational studies and randomized control trials were all eligible for inclusion. Literature reviews, abstracts, case reports, editorial letters, personal communications and non-English language publications were excluded. Every attempt was made to identify incidences where multiple publications from the same group shared population subjects. In these cases, only the most recent study from the group was included in the review.

Studies reporting sensitivity of first trimester anomaly screening in either singleton or multiple pregnancies in any healthcare setting and pregnancies of all levels of risk were eligible for inclusion. Prospective studies were included based on their intention to perform ultrasound screening prior to 14 weeks, with the understanding that the reality of clinical practice means that all scans would not necessarily be performed within this time period.

This review included studies collecting data on all types of structural abnormalities, which included lethal, major, moderate and minor abnormalities as defined by the Royal College of Obstetrics & Gynaecology[15]. Only those studies which gave an individual breakdown of the fetal structural anomalies detected within their population cohort were eligible for inclusion. Publications focused on a specific malformation, or specific groups of anatomical malformations (e.g. cardiac malformations only) were excluded. Studies where the aims were solely to investigate the use of first-trimester ultrasound for the detection of fetal chromosomal abnormalities, or soft markers, were also excluded. Studies utilizing various modes of two-dimensional ultrasound, including transvaginal ultrasound, transabdominal ultrasound and a combination of both, were eligible for inclusion. However, studies evaluating fetuses using three-dimensional ultrasound modalities were excluded.

The reference standard used to determine the accuracy of first trimester anomaly screening was the detection of fetal structural abnormalities at birth or later. As such, studies which did not perform a post-natal examination or which did not obtain data

regarding neonatal outcome for the purposes of confirming true positive, false positive, true negative and false negative results, were excluded. Regarding post-mortem examination of all fetuses we took a pragmatic approach: this did not form a requirement for inclusion in the review, as it must be accepted that this often not possible following first trimester termination of pregnancy.

Data Extraction

Review of all articles included within this meta-analysis and the reporting of all results was conducted based on the Meta-Analysis of Observational Studies in Epidemiology (MOOSE) guidelines[16] and the Synthesizing Evidence from Diagnostic Accuracy Tests (SEDATE) guidance[17].

For each study, the following data were recorded: name of the author, year of publication, sample size; as well as a list of modifiable and non—modifiable factors that may impact the detection rates. These were: population characteristics, gestation at which ultrasound was performed, type of healthcare setting, the index test used (i.e. ultrasound modality – either transvaginal ultrasound, transabdominal ultrasound or both), whether an anatomical checklist was used for first trimester anomaly detection, whether a heart examination was specifically performed during screening and what types of malformations were included in the study. With respect to the use of an anatomical checklist for performing the first trimester anomaly scan, studies were graded as either having a basic checklist, a detailed checklist or none; studies which did not declare the use of an anatomical checklist were declared as having none. A basic checklist was

defined as one which requires observation of certain anatomical regions or the presence of specific organs. A detailed checklist was defined as one which directs the sonographer to visualize multiple specific regions of at least one of the organs listed; and which may also include planes of insonation or taking of measurements; or uses advanced ultrasound markers such as intracranial translucency or retronasal triangle. For example, a checklist which lists 'head' would be considered a basic checklist; while one which lists 'cranial ossification, interhemispheric falx, butterfly-shaped choroid plexus' would be considered detailed.

An initial attempt was made to collect data with respect to the level of experience and training of sonographers in each study. However, the majority of studies reported no such data, and those that did provided data which was not amenable to comparison, so this was abandoned.

Finally, a search of all studies was undertaken to determine what modifiable and non-modifiable factors were specifically cited by study authors as having a significant impact on the accuracy and efficiency of first trimester anomaly screening.

When assessing the sensitivity of first trimester anomaly screening, it is possible to define this as the number of abnormalities detected; and also as the number of fetuses identified with one or more anomaly. Both outcomes are important: the first determines the accuracy of first trimester ultrasound in detecting individual anomalies of varying severity, whereas the second provides an understanding of how many fetuses are

impacted by first trimester anomaly screening. Therefore, the number of abnormalities present in each study population and their proportion detected at first trimester screening were documented. In addition, the number of fetuses in each study cohort affected with one or more structural malformations was noted along with the number of these fetuses identified in the first trimester.

Due to the heterogeneity of the studies included in this review, considerable efforts were made to ensure that the results of the studies were comparable. Our review was therefore required to develop strict definitions for what constitutes 'one detected structural abnormality'. Firstly, it was decided that all bilateral defects such as bilateral renal agenesis would be counted as two individual structural anomalies. This was done so that our review would be able to distinguish between a fetus with exclusive unilateral renal agenesis diagnosed in the first trimester ('one correctly diagnosed anomaly') and a fetus, diagnosed with unilateral renal agenesis, with a missed bilateral defect ('one correctly identified anomaly, one missed anomaly diagnosis'). This is particularly important because it provides the most accurate data for analysis of first trimester screening sensitivity. Second, fetuses which were diagnosed with a syndrome (eg. Dandy-Walker syndrome) were considered to have one abnormality. The diagnosis of a syndrome is often made based on the findings of multiple malformations on ultrasound. However, in the majority of studies, there was no specification as to how many constituent anomalies were detected in order to make the diagnosis of a specific syndrome and moreover whether these anomalies were all detected in the first trimester. As such, the only way to unify the analysis of all the studies was to treat the ultrasound diagnosis of one syndrome

as one anomaly. Thirdly, in several studies, a single fetus would be attributed to having multiple anomalies within one organ system (eg. 'multiple skeletal abnormalities'). In this case, the fetus was considered to have one structural ("a skeletal") abnormality. Finally, single umbilical artery was considered a structural variant of normal anatomy and therefore excluded from data collection, as were soft markers for fetal aneuploidy (including increased nuchal translucency and absent nasal bone). Of note, cystic hygromas were excluded from analysis. The diagnosis of a cystic hygroma is often defined as the presence of a bilateral, cystic structure within the occipitocervical region, distinguished by the presence of septations within the cystic fluid. However, evidence suggests that cystic hygromas should not be considered as a distinct entity from increased nuchal translucency[18] and do not confer any 'risk status', which is independent of that related to increased nuchal translucency[18]. As such, within our review, cystic hygromas were considered soft markers for aneuploidy, much like increased nuchal translucency, and therefore were excluded from our analysis of structural anomalies.

Data regarding the number of false positive diagnoses made during the first trimester screening process was also collected. In many studies, women were offered the option of anomaly screening in the second and/ or third trimesters of pregnancy in addition to their first trimester anomaly scan. In these cases, the number of antenatal diagnoses made outside of the first trimester was recorded. All data were collected and extracted from tables or text for each study on two independent occasions.

Quality Assessment of Studies

Assessment of the quality of studies included within this review was performed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2)[19]. This is a tool designed to evaluate the risk of bias within each study and assess its applicability to the systematic review. It provides a framework for evaluating studies within four key domains: patient selection, the index test, the reference standard and the flow of patients through the study along with the timing of the index test. Each domain is assessed with respect to bias, and the first three domains with respect to applicability. A judgment of 'low, high or unclear' risk of bias and lack of applicability was made for each study based on a series of signaling questions developed specific to our review (please see Supplement B).

Data Analysis

In an effort to reduce the impact of study heterogeneity on the results of the meta-analysis, data from the studies was analyzed within subgroups, which reflected the types of abnormalities included and the type of population assessed by each study. Extracted data was assessed within one of three subgroups:

(1) Studies reporting on major anomalies in a low risk or unselected population,

(2) Studies assessing all types of anomalies in a low risk or unselected population and

(3) Studies examining all types of anomalies in high risk populations.

When studies published adequate data on two distinct cohorts, then the relevant data from each study was included within more than one subgroup. In studies performed in unselected populations, where anomaly type was defined using RCOG criteria, anomalies labeled as lethal or severe were considered 'major anomalies' and therefore analyzed as

part of subgroup 1 whereas anomalies labeled as 'lethal, severe, moderate or minor' were included as part of subgroup 2. In studies not based on RCOG criteria, the definition provided by the study itself was used to determine what constituted a major anomaly. Meta-analysis of data extracted from eligible studies was performed in two steps. First, summary statistics with 95% C.I. were derived for each study with respect to both the sensitivity of first trimester anomaly screening for detecting fetal anomalies and for detecting fetuses affected by one or more abnormality. Second, individual study statistics within each subgroup were combined in order to obtain a pooled summary estimate. In those studies providing adequate data for the construction of 2x2 tables, pooled summary estimates of sensitivity and specificity were calculated. The pooled summary statistics were estimated using random effects models[17]. Heterogeneity between studies was estimated using I^2. Further analysis was undertaken within each subgroup in order to determine whether factors such as year of study publication, mode of ultrasound and the use of an anatomical protocol impacted outcomes.

All statistical analysis was performed using StatsDirect statistical software (England, 2013).

Results

Study Selection and Description of Included Studies

The initial electronic search yielded 2,225 citations, from which a total of 30 relevant studies were selected for inclusion in the systematic review (Figure 1). The studies evaluated were published between 1991 and 2015. The gestational age at which scans

were performed ranged from 9 to 15+6 weeks gestation, with the vast majority completed prior to fourteen weeks. Studies were performed in a variety of healthcare settings with the majority taking place, at least in part, in either a University Hospital, Tertiary Care Centre or Research Facility (n=20 studies). Three studies involved multi-centre data collection.

Several studies in the review included adequate data on several distinct cohorts, allowing for data from these studies to be analyzed as part of multiple subgroups (see Supplement C).

In total, there were 26 studies evaluating unselected and low risk populations (Table 1) of which

- 19 study cohorts focused on the detection of major anomalies (n =115,731 fetuses)
- 14 study cohorts assessed all types of anomalies (n=97,976 fetuses).

There were 6 studies which focused on detecting all types of anomalies in high-risk populations representing a total of 2,841 fetuses (Table 2).

Subgroup 1 – Sensitivity of first trimester anomaly screening for the detection of major anomalies in a low risk/unselected population

There were nineteen study cohorts (n =115,731 fetuses) which evaluated a low risk or unselected population for the presence of major anomalies (Table 3). There were 1165 anomalies (mean number of major structural abnormalities present per 100 fetuses = 1.01,

95% C.I. 0.95-1.07). Of these, 529 were detected during the first trimester giving a sensitivity of first trimester ultrasound for the detection of major fetal abnormalities (pooled estimate) of 46.10% (95% C.I. 36.88-55.46, Figure 2(i). Heterogeneity as estimated by I^2 was 90.1% (95% C.I. 86.5-92.4%)

In fifteen of these study cohorts (n= 77,664 fetuses), an additional ultrasound after the first trimester of pregnancy was performed. In these studies, the abnormalities detected in the first trimester represented 53.47% (95% C.I. 43.42-63.37) of all antenatal ultrasound anomaly diagnoses.

Twelve of these study cohorts (n= 61,930 fetuses) provided data on the number of fetuses affected with one or more structural abnormalities: 573 were affected with structural anomalies, of which 264 were detected during the first trimester. The mean prevalence of affected fetuses was 0.93% (95% C.I. 0.85-1.00). The mean sensitivity of first trimester ultrasound for the detection of fetuses affected by one or more anomalies (pooled estimate – Figure 2(ii) was 45.25% (95% C. I. 38.44-52.14). It was possible to create complete 2x2 tables for the three studies providing false positive rates (pooled sensitivity – 41.98% [95% C.I. 23.83-61.33], pooled specificity – 99.96% [95% C.I. 99.90-100.00]).

Subgroup 2 – Sensitivity of first trimester anomaly screening for the detection of all types of anomalies in a low risk/unselected population

There were fourteen study cohorts (n=97,976 fetuses), which evaluated a low risk or unselected population for the presence of all types anomalies (Table 4). Thirteen of these study cohorts (n= 84,253 fetuses) reported 1,521 anomalies (mean number per 100 fetuses = 1.81, 95% C.I. 1.72-1.90). Of these, 526 were detected during the first trimester giving a pooled sensitivity of first trimester ultrasound for the detection of fetal abnormalities of 32.35% (95% C.I. 22.45-43.12, Figure 3(i)). Heterogeneity as estimated by I^2 was 93.5% (95% C.I. 91.1-95.0%).

In twelve study cohorts (n=77,561 fetuses) reporting on the detection of individual anomalies, antenatal ultrasound was performed on an additional occasion after the first trimester of pregnancy. In these studies, the abnormalities detected in the first trimester represented 41.10% (95% C.I. 32.13-50.38) of all antenatal ultrasound anomaly diagnoses.

Nine studies (n= 77,186 fetuses) provided data on the number of fetuses affected with one or more structural abnormalities (Table 4): 1256 were affected with structural anomalies (prevalence 1.63%, 1.54-1.72). Of these, 435 were detected during the first trimester (pooled estimated sensitivity 35.56%, 26.27-45.44, Figure 3(ii)). It was possible to create complete 2x2 tables for the three studies providing false positive rates (pooled sensitivity – 44.44% [95% C.I. 32.76-56.44], pooled specificity – 99.86% [95% C.I. 99.82-99.89])

Subgroup 3 – Studies assessing the sensitivity of first trimester anomaly screening for the detection of all types of anomalies in a high risk population

There were six studies, which evaluated a high risk population for the presence of all types anomalies (Table 5). Within these studies (n=2,841 fetuses), there were 186 anomalies present, of which 116 were detected during the first trimester. The mean number of major structural abnormalities present per 100 fetuses was 6.55 (95% C.I. 5.66-7.52), confirming the high risk status of this population. The mean sensitivity of first trimester ultrasound for the detection of all types of fetal abnormalities (pooled estimate - Figure 4(i) was 61.18% (95% C.I. 37.71-82.19). Heterogeneity as estimated by I^2 was 90.5% (95% C.I. 82.1-94.0%)

Abnormalities detected in the first trimester represented 66.29% (95% C.I. 43.47-85.69) of all antenatal ultrasound anomaly diagnoses.

Five studies (2,345 fetuses) provided data on the number of fetuses affected with one or more structural abnormalities within their population cohort. Based on the 2,345 fetuses reported in these five studies, 88 were affected with structural anomalies, of which 48 were detected during the first trimester. The mean prevalence of affected fetuses was 3.75% (95% C.I.: 3.02-4.60). The mean sensitivity of first trimester ultrasound for the detection of fetuses affected by one or more anomalies (pooled estimate – Figure 4(ii) was 62.42% (95% C.I: 33.40-87.24). Complete 2x2 tables for the three studies providing false positive rates were compiled (pooled sensitivity – 79.85% [95% C.I. 43.87-99.29], pooled specificity – 97.78% [95% C.I. 90.96-100.00])

Factors Impacting Detection Rates in First Trimester Anomaly Screening

A chi-squared test demonstrated a statistically significant relationship between the use of an anatomical protocol during first trimester anomaly screening and sensitivity for the detection of fetal anomalies in all subgroups (subgroup 1: chi-squared 60.95, $P<0.0001$, subgroup 2: chi-squared 112.46, $P<0.0001$, subgroup 3: chi-squared 24.71, $P<0.0001$).

In subgroups 1 and 2 there was a statistically significant chi-squared for linear trend ($P<0.0001$) suggesting that use of an increasingly detailed protocol resulted in higher sensitivity for the detection of fetal anomalies.

Simple linear regression analysis showed no statistically significant relationship between the year of publication study and sensitivity for anomaly detection in either of subgroup 1 ($R^2 = 0.066$), subgroup 2 ($R^2 = 0.030$) or subgroup 3 ($R^2 = 0.44$).

The impact of the mode of ultrasound used for screening on detection rates was also explored. The vast majority of studies used a combination of TA and TV ultrasound, often beginning with the former and then complementing with the latter when organ visualization was suboptimal. In all three subgroups, there were insufficient studies using TA or TV ultrasound exclusively in order to make useful comparisons between detection rates using one of the three methods (TA, TV or combination of both).

Finally, with respect to the collection of qualitative data, studies evaluated as part of this systematic review cited numerous factors as having an impact on first trimester detection rates for fetal anomalies. Non-modifiable factors cited included: the small size of fetal anatomy at this gestational age[20-26], the progressive patho-physiology of certain fetal anomalies and the fact that some anomalies are not yet present in the first trimester[14, 21, 22, 24-32], raised maternal BMI[24, 33] and the presence of uterine fibroids[27, 33]. A number of modifiable factors were considered to impact first trimester anomaly screening including: gestational age at time of scan[24], mode of ultrasound use[20, 24, 29, 34, 35], time allocated for screening[14, 20, 24, 25, 30, 33], use of an anatomical protocol with standard sonographic views[14, 20, 27, 35-37], sonographer experience and training[14, 20, 23-25, 27, 29, 30, 33, 35-42], system in place for regular audit[30, 35, 36], knowledge of fetal embryology including normal developmental milestones in the first trimester[21, 25, 26, 31] and knowledge of easily recognizable markers for the diagnosis of anomalies such as spina bifida and facial clefts[14].

Methodological Quality Assessment of Studies

Results of the QUADAS-2 assessment are summarized in Figure 5. With respect to bias in patient selection, eighteen of thirty studies were scored low risk for bias. Two studies were deemed at high risk of bias due to inappropriate exclusions from their patient cohorts. Nine studies were labeled 'unclear' with respect to bias due to lack of information provided on the methods used to enroll patients. In four studies, there was no information pertaining to patient exclusions and therefore they were also labeled unclear. With respect to the index test, none of the studies declared whether their sonographers

were blinded to the history of patients prior to performing anomaly screening, and therefore all studies were labeled as unclear in this regard. Five studies excluded cardiac examinations from their first trimester anomaly screening and were therefore considered at high risk of bias. Four additional studies provided no specification as to the types of anomalies included in their assessment and as such were labeled as high risk. All thirty studies were found to be low risk for bias relating to the reference standard; this was after all a criterion for inclusion in this study. In terms of flow and timing, fifteen out of thirty studies were labeled as high risk for bias because they included no data pertaining to false positive diagnoses. One further study was labeled as high risk because a reference standard was performed in less than 90% of the patients enrolled in the study. With respect to applicability, there were no concerns raised regarding the thirty studies included in this review.

Discussion

In this study we show that first trimester ultrasound can identify about half of all major anomalies diagnosed antenatally (Table 6): in unselected and low risk pregnancies about 40% of all antenatally diagnosed anomalies can be identified at this stage. The detection rates for major anomalies are higher (46%) with 45% of fetuses affected with one or more major malformations identified. Finally, in high risk populations, sensitivity for anomaly detection was even higher (61%), with first trimester screening detecting 66% of all antenatal ultrasound diagnoses. This higher rate of detection in high risk populations is likely due to targeted screening [13]: sonologists are aware of the increased risk; and women at high risk may be scanned by more expert examiners. Nevertheless,

theoretically at least, the detection rate of an anomaly should not be influenced by its prevalence in the population cohort.

This suggests that first trimester ultrasound has the potential to identify a high percentage of fetuses affected with structural anomalies in all risk groups: if we have the technology and skill available to achieve sensitivities of over 60% in high risk populations, there is no reason that under optimal conditions these detection rates could not be achievable for all patients. This idea of "if you look you will find" is further supported by an important result from our analysis of all three subgroups: a significant association ($P<0.001$) between the sensitivity of first trimester ultrasound and the use of an anatomical protocol for screening; and a trend suggesting that the more detailed the protocol used, the greater the detection rates. We therefore suggest that international protocols with standard anatomical views should be used in practice, in order to optimize rates of first trimester anomaly detection.

Most studies used a combination of TV and TA ultrasound, which meant that a statistically useful comparison between sensitivities from studies using one of the three methods (TA, TV or combination of both) could not be performed. Findings from studies evaluating fetal organ detection suggest that optimal visualization rates are obtained with a combination of TA and TV ultrasound [8]. For TA, raised BMI, fibroids or retroversion of the uterus will decrease image quality. In contrast, TV ultrasound provides much higher resolution, but has the disadvantage of limited probe maneuverability[43]; some women may also find TV ultrasound to be a less acceptable test. [34, 44]. We believe that

adopting a flexible, patient tailored approach, which may require use of TA and TV modalities, should be encouraged as this will be the only way of determining the true potential of first trimester anomaly detection.

A number of other factors affecting first trimester ultrasound were suggested by authors. These provide insights into how the process might be optimized: several studies demonstrate a significant "learning curve" associated with first trimester anomaly screening; [38, 45] [30] knowledge of fetal embryological development is also important [46]; as is the time allocated for screening. Apart from using an anatomical protocol, Syngelaki et al.[14] highlights the positive impact of having an easily recognizable marker for the diagnosis of anomalies, such as spina bifida or facial clefts, on overall detection rates.

Non-modifiable factors impacting detection rates include the small size of anomalies and fetal crown rump length; and the presence of anomalies with progressive pathophysiology. Studies assessing the types of anomalies detected [14] [13] suggest that there are some conditions which are nearly always detectable in the first trimester; others that are never identifiable; and some that may potentially be diagnosed, depending on maternal, fetal, sonographer and equipment factors. Undetectable anomalies mostly relate to structures not yet fully developed prior to 14 weeks, e.g. cerebellar anomalies and echogenic lung cysts; or those diagnosed secondary to changes in amniotic fluid, e.g. duodenal atresia, bowel obstruction or renal agenesis. Thus, first trimester anomaly screening will never replace that at later gestations completely. What is clear from our

assessment is that expectations and future objectives for first trimester anomaly screening must be tailored to the types of anomalies amenable to detection at this gestation.

Limitations of this review include the fact that, despite subgroup analysis, there remained considerable heterogeneity between studies: extensive variation between the studies existed in inclusion and exclusion criteria; the age at postnatal follow; use of anatomical protocols used; and outcome reporting. The types of anomalies examined was different, even amongst those aiming to assess solely 'major anomalies'. We would therefore recommend the use of international definitions, such as the RCOG, EUROCAT or March of Dimes criteria in future studies.

A number of studies excluded cardiac anomalies from their analysis, presumably because these anomalies are often difficult to diagnose in the first trimester and require significant sonographer skill/. The latest BINOCAR data[47] suggest overall antenatal detection rates are low (53.1%, 95% C.I. 49.1-57.1) and this must be seen in the context of them being common anomalies (82.2 per 10,000 fetuses, 95% C.I. 81.30-83.03)[48]. It is impossible to develop a proper understanding of the overall impact of first trimester screening on antenatal care, or to compare first trimester and second trimester anomaly screening, if studies do not include anomalies belonging to all organ systems within their analysis.

Some studies only reported the number of *anomalies* within their cohort, providing no data with respect to the number of affected fetuses. This limits the ability to understand

fully the number of fetuses impacted on by first trimester screening initiatives; such information must form a minimum standard to be reported.

Finally, it must be highlighted that the majority of studies did not include false positive rates. One of the inevitable consequences of first trimester screening is the offer of early termination of pregnancy when major anomalies are seen. There is a self-evident concern regarding termination after first trimester anomaly screening without a full understanding of a false positive diagnosis; however, the rate of false positives is thought to be much lower than second trimester screening[38, 45]. It is also not that easy to work out what a false positive is, as anomalies evolve: for example, a significant proportion of megacystis (in particular those ≤15mm) spontaneously resolve later in pregnancy[49]. In the study by Syngelaki et al. [14], a large proportion of false positive diagnoses involved either megacystis or bowel-only exomphalos. As such, it is critical to not only understand the rate of false positive diagnoses in first trimester screening, but also what types of anomalies are most likely to spontaneously resolve.

Based on recommendations from QUADAS-2[19], the design of a perfect evaluation of first trimester anomaly screening would involve blinding of sonographers to patient histories; prevention of referral bias in tertiary centre trials; post-mortem analysis of every terminated case; standardized neonatal assessment of internal anomalies with neonatal echocardiography in all fetuses and blinding of neonatal assessors to the prenatal sonography results. Such a rigorous examination of first trimester anomaly screening is unlikely to ever be performed on a large scale or in fact to be considered ethical. A

historical assessment of second trimester anomaly screening reveals that at the time, there were also concerns that it was not adequately evaluated prior to widespread adoption in prenatal care, particularly with respect to the burden of false positive diagnoses, optimal timing for screening, cost-benefit analysis and the potential for increasing parental anxiety[50, 51]; and little consensus or uniformity as to how screening should be performed[52]. However, the process of standardized anatomy protocols, a systematic approach to training of sonographers and an emphasis on quality assessment has allowed the 20 week scan to become a fundamental component of prenatal care; this is now also happening with first trimester ultrasound.

Conclusions

This systematic review has used subgroup analysis, strict 'anomaly' criteria and manual extraction of data to obtain the most accurate picture possible of the true sensitivity of first trimester anomaly screening. Our findings demonstrate that first trimester ultrasound should be considered as a valuable clinical addition to prenatal anomaly screening in low risk, high risk and unselected populations. We have demonstrated that the currently presented literature represents only the beginning of what is achievable for first trimester anomaly screening. It is clear that greater sensitivities can be achieved with the use of a detailed anatomical protocol, increased attention to sonographer training and an appreciation of the learning curve involved in acquiring these skills.

Legends of figures included in the paper:

Figure 1 – Flowchart demonstrating the search strategy and selection of studies to be included in the systematic review and meta-analysis.

Figure 2 – Forest plots demonstrating sensitivity of first trimester ultrasound in low risk and unselected pregnancies (i). For the detection of major fetal anomalies. (ii). For the detection of fetuses affected with major anomalies.

Figure 3 – Forest plots demonstrating sensitivity of first trimester ultrasound in low risk and unselected pregnancies (i). For the detection of all types of fetal anomalies. (ii). For the detection of fetuses affected with all types of anomalies.

Figure 4 – Forest plots demonstrating sensitivity of first trimester ultrasound in high risk pregnancies (i). For the detection of all fetal anomalies. (ii). For the detection of fetuses affected with all types of anomalies.

Figure 5 – Results from QUADAS-2 Quality Assessment of Studies. (i). Proportion of studies with low, high, or unclear risk of bias. (ii). Proportion of studies with low, high or unclear concerns regarding applicability.

References

1. Salomon LJ, Alfirevic Z, Bilardo CM, Chalouhi GE, Ghi T, Kagan KO, Lau TK, Papageorghiou AT, Raine-Fenning NJ, Stirnemann J, Suresh S, Tabor A, Timor-Tritsch I, Toi A, Yeo G. ISUOG practice guidelines: performance of first-trimester fetal ultrasound scan. Ultrasound in Obstetrics & Gynecology. 2013;41(1):102.

2. Nicolaides KH. Screening for fetal aneuploidies at 11 to 13 weeks. Prenatal diagnosis. 2011;31(1):7-15.

3. Maiz N, Valencia C, Kagan KO, Wright D, Nicolaides KH. Ductus venosus Doppler in screening for trisomies 21, 18 and 13 and Turner syndrome at 11-13 weeks of gestation. Ultrasound in Obstetrics & Gynecology. 2009;33(5):512-7.

4. Kagan KO, Cicero S, Staboulidou I, Wright D, Nicolaides KH. Fetal nasal bone in screening for trisomies 21, 18 and 13 and Turner syndrome at 11-13 weeks of gestation. Ultrasound in Obstetrics and Gynecology. 2009;33(3):259-64.

5. Gil MM, Quezada MS, Revello R, Akolekar R, Nicolaides KH. Analysis of cell•free DNA in maternal blood in screening for fetal aneuploidies: updated meta•analysis. Ultrasound in Obstetrics & Gynecology. 2015;45(3):249-66.

6. Souka AP, Snijders RJ, Novakov A, Soares W, Nicolaides KH. Defects and syndromes in chromosomally normal fetuses with increased nuchal translucency thickness at 10-14 weeks of gestation. Ultrasound in Obstetrics & Gynecology. 1998;11(6):391-400.

7. Nicolaides KH. A model for a new pyramid of prenatal care based on the 11 to 13 weeks' assessment. Prenatal diagnosis. 2011;31(1):3-6.

8. Braithwaite JM, Armstrong MA, Economides DL. Assessment of fetal anatomy at 12 to 13 weeks of gestation by transabdominal and transvaginal sonography. British Journal of Obstetrics & Gynaecology. 1996;103(1):82-5.

9. Guariglia L, Rosati P. Transvaginal sonographic detection of embryonic-fetal abnormalities in early pregnancy. Obstetrics & Gynecology. 2000;96(3):328-32.

10. Whitlow BJ, Economides DL. The optimal gestational age to examine fetal anatomy and measure nuchal translucency in the first trimester. Ultrasound in Obstetrics & Gynecology. 1998;11(4):258-61.

11. Luchi C, Schifano M, Sacchini C, Nanini C, Sceusa F, Capriello P, Genazzani AR. Detailed fetal anatomy assessment in the first trimester at 11, 12 and 13 weeks of gestation. Journal of Maternal-Fetal & Neonatal Medicine. 2012;25(6):675-8.

12. Farraposo S, Montenegro N, Matias A. Evaluation of the role of first-trimester obstetric ultrasound in the detection of major anomalies: a systematic review. Journal of Perinatal Medicine. 2014;42(2):141-9.

13. Rossi AC, Prefumo F. Accuracy of ultrasonography at 11–14 weeks of gestation for detection of fetal structural anomalies: a systematic review. Obstetrics & Gynecology. 2013;122(6):1160-7.

14. Syngelaki A, Chelemen T, Dagklis T, Allan L, Nicolaides KH. Challenges in the diagnosis of fetal non•chromosomal abnormalities at 11–13 weeks. Prenatal diagnosis. 2011;31(1):90-102.

15. Whittle M, Chitty LS, Neilson JP, Shirley IM, Smith NC, Ville YG, Wyldes MP. Ultrasound screening for fetal abnormalities: report of the RCOG working party. London, The Royal College of Obstetricians and Gynaecologists. 1997.

16. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Journal of the American Medical Association. 2000;283(15):2008-12.

17. Sotiriadis A, Papatheodorou SI, Martins WP. Synthesizing Evidence from Diagnostic Accuracy TEsts: the SEDATE guideline. Ultrasound in Obstetrics & Gynecology. 2016
(47):386-95.

18. Molina FS, Avgidou K, Kagan KO, Poggi S, Nicolaides KH. Cystic hygromas, nuchal edema, and nuchal translucency at 11–14 weeks of gestation. Obstetrics & Gynecology. 2006;107(3):678-83.

19. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011;155(8):529-36.

20. Pilalis A, Basagiannis C, Eleftheriades M, Faros E, Troukis E, Armelidou E, Papastefanou I, Souka AP. Evaluation of a two-step ultrasound examination protocol for the detection of major fetal structural defects. Journal of Maternal-Fetal & Neonatal Medicine. 2012;25(9):1814-7.

21. Chen M, Lee CP, Lam YH, Tang RYK, Chan BCP, Wong SF, Tse LHY, Tang MHY. Comparison of nuchal and detailed morphology ultrasound examinations in early pregnancy for fetal structural abnormality screening: a randomized controlled trial. Ultrasound in Obstetrics & Gynecology. 2008;31(2):136-46; discussion 46.

22. Li Q, Guan B, Li D. Detection of fetal structural abnormalities by early pregnancy ultrasound in China. International Journal of Gynaecology & Obstetrics. 2008;100(3):277-8.

23. Dane B, Dane C, Sivri D, Kiray M, Cetin A, Yayla M. Ultrasound screening for fetal major abnormalities at 11-14 weeks. Acta Obstetricia Et Gynecologica Scandinavica. 2007;86(6):666-70.

24. Souka AP, Pilalis A, Kavalakis I, Antsaklis P, Papantoniou N, Mesogitis S, Antsaklis A. Screening for major structural abnormalities at the 11-to 14-week ultrasound scan. American journal of obstetrics and gynecology. 2006;194(2):393-6.

25. Chen M, Lam YH, Lee CP, Tang MHY. Ultrasound screening of fetal structural abnormalities at 12 to 14 weeks in Hong Kong. Prenatal diagnosis. 2004;24(2):92-7.

26. 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. British Journal of Obstetrics & Gynaecology. 1999;106(9):929-36.

27. Wang L, Wu Q-q, Chen Y, Ma Y-q, Yao L, Li M. Ultrasound screening of fetal structural abnormalities by standard ultrasound views during the first trimester. Chinese Medical Journal. 2013;126(5):986-7.

28. Abu-Rustum RS, Daou L, Abu-Rustum SE. Role of ultrasonography in early gestation in the diagnosis of congenital heart defects. Journal of Ultrasound in Medicine. 2010;29(5):817-21.

29. Saltvedt S, Almstrom H, Kublickas M, Valentin L, Grunewald C. Detection of malformations in chromosomally normal fetuses by routine ultrasound at 12 or 18 weeks of gestation-a randomised controlled trial in 39,572 pregnancies. BJOG: An International Journal of Obstetrics & Gynaecology. 2006;113(6):664-74.

30. Taipale P, Ammala M, Salonen R, Hiilesmaa V. Learning curve in ultrasonographic screening for selected fetal structural anomalies in early pregnancy. Obstetrics & Gynecology. 2003;101(2):273-8.

31. den Hollander NS, Wessels MW, Niermeijer MF, Los FJ, Wladimiroff JW. Early fetal anomaly scanning in a population at increased risk of abnormalities. Ultrasound in Obstetrics & Gynecology. 2002;19(6):570-4.

32. Chatzipapas IK, Whitlow BJ, Economides DL. The 'Mickey Mouse' sign and the diagnosis of anencephaly in early pregnancy. Ultrasound in Obstetrics & Gynecology. 1999;13(3):196-9.

33. Hernadi L, Torocsik M. Screening for fetal anomalies in the 12th week of pregnancy by transvaginal sonography in an unselected population. Prenatal diagnosis. 1997;17(8):753-9.

34. Abu-Rustum RS, Daou L, Abu-Rustum SE. Role of first-trimester sonography in the diagnosis of aneuploidy and structural fetal anomalies. Journal of Ultrasound in Medicine. 2010;29(10):1445-52.

35. Cedergren M, Selbing A. Detection of fetal structural abnormalities by an 11-14-week ultrasound dating scan in an unselected Swedish population. Acta Obstetricia Et Gynecologica Scandinavica. 2006;85(8):912-5.

36. Vavilala S, Geeta K. The 11-13+6 Weeks Scan: Where do We Stand? A 5-year Review at Fernandez Hospital. International Journal of Infertility and Fetal Medicine. 2011;2(2):65-9.

37. Hildebrand E, Selbing A, Blomberg M. Comparison of first and second trimester ultrasound screening for fetal anomalies in the southeast region of Sweden. Acta Obstetricia Et Gynecologica Scandinavica. 2010;89(11):1412-9.

38. Iliescu D, Tudorache S, Comanescu A, Antsaklis P, Cotarcea S, Novac L, Cernea N, Antsaklis A. Improved detection rate of structural abnormalities in the first trimester using an extended examination protocol. Ultrasound in Obstetrics & Gynecology. 2013;42(3):300-9.

39. Grande M, Arigita M, Borobio V, Jimenez JM, Fernandez S, Borrell A. First-trimester detection of structural abnormalities and the role of aneuploidy markers. Ultrasound in Obstetrics & Gynecology. 2012;39(2):157-63.

40. Oztekin O, Oztekin D, Tinar S, Adibelli Z. Ultrasonographic diagnosis of fetal structural abnormalities in prenatal screening at 11-14 weeks. Diagnostic & Interventional Radiology. 2009;15(3):221-5.

41. Drysdale K, Ridley D, Walker K, Higgins B, Dean T. First-trimester pregnancy scanning as a screening tool for high-risk and abnormal pregnancies in a district general hospital setting. Journal of Obstetrics & Gynaecology. 2002;22(2):159-65.

42. Carvalho MHB, Brizot ML, Lopes LM, Chiba CH, Miyadahira S, Zugaib M. Detection of fetal structural abnormalities at the 11-14 week ultrasound scan. Prenatal diagnosis. 2002;22(1):1-4.

43. McAuliffe FM, Fong KW, Toi A, Chitayat D, Keating S, Johnson J-A. Ultrasound detection of fetal anomalies in conjunction with first-trimester nuchal translucency screening: a feasibility study. American Journal of Obstetrics & Gynecology. 2005;193(3 Pt 2):1260-5.

44. Ebrashy A, El Kateb A, Momtaz M, El Sheikhah A, Aboulghar MM, Ibrahim M, Saad M. 13-14-week fetal anatomy scan: a 5-year prospective study. Ultrasound in Obstetrics & Gynecology. 2010;35(3):292-6.

45. D'Ottavio G, Mandruzzato G, Meir YJ, Rustico MA, Fischer-Tamaro L, Conoscenti G, Natale R. Comparison of first and second trimester screening for fetal anomalies. In: Levi S, Chervenak FA, editors. Ultrasound Screening for Fetal Anomalies: Is It Worth It?: Screening Revisited after the Eurofetus Data. Annals of the New York Academy of Sciences. 8471998. p. 200-9.

46. Johnson SP, Sebire NJ, Snijders RJ, Tunkel S, Nicolaides KH. Ultrasound screening for anencephaly at 10-14 weeks of gestation. Ultrasound in Obstetrics & Gynecology. 1997;9(1):14-6.

47. Boyd PA, Tonks AM, Rankin J, Rounding C, Wellesley D, Draper ES. Monitoring the prenatal detection of structural fetal congenital anomalies in England and Wales: register-based study. Journal of Medical Screening. 2011;18(1):2-7.

48. EUROCAT. Prevalence Tables http://www.eurocat-network.eu/accessprevalencedata/prevalencetables2008-2012 [July 2014].

49. Fong KW, Toi A, Salem S, Hornberger LK, Chitayat D, Keating SJ, McAuliffe F, Johnson J-A. Detection of Fetal Structural Abnormalities with US during Early Pregnancy1. Radiographics. 2004;24(1):157-74.

50. Chitty LS. Ultrasound screening for fetal abnormalities. Prenatal diagnosis. 1995;15(13):1241-57.

51. Roberts T, Henderson J, Mugford M, Bricker L, Neilson J, Garcia J. Antenatal ultrasound screening for fetal abnormalities: a systematic review of studies of cost and cost effectiveness. BJOG: An International Journal of Obstetrics & Gynaecology. 2002;109(1):44-56.

52. Pathak S, Lees C. Ultrasound structural fetal anomaly screening: An update. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2009;94(5):F384-F90.

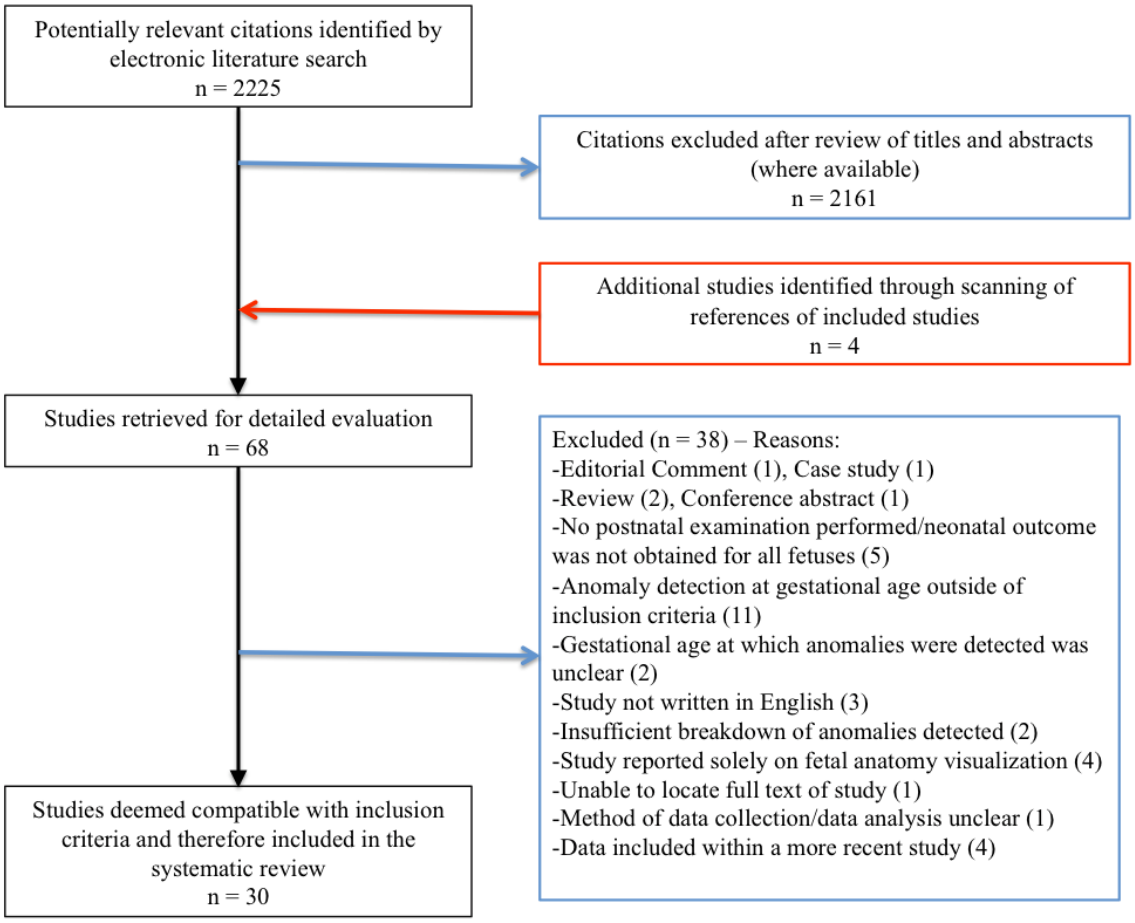

Figure 1

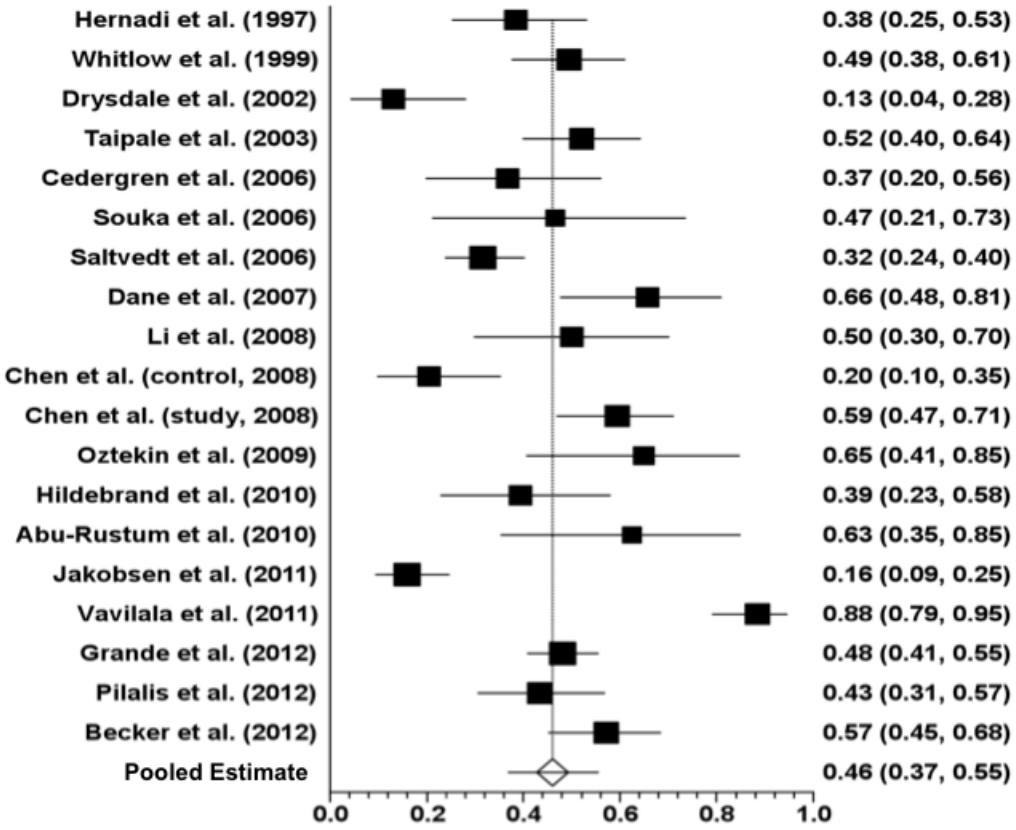

Figure 2(i).

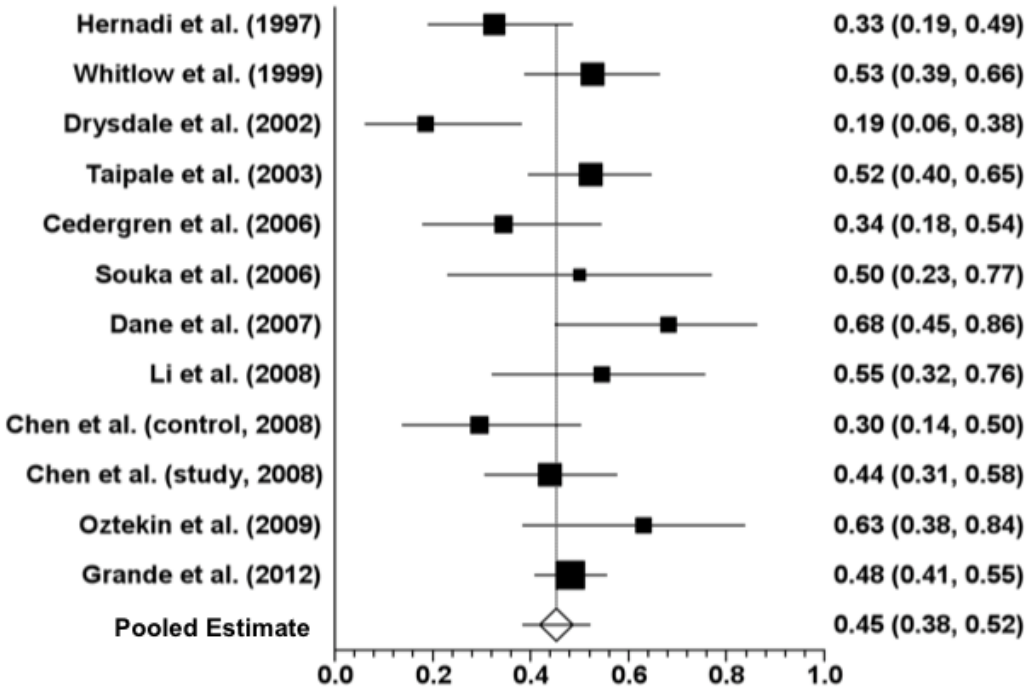

Figure 2(ii).

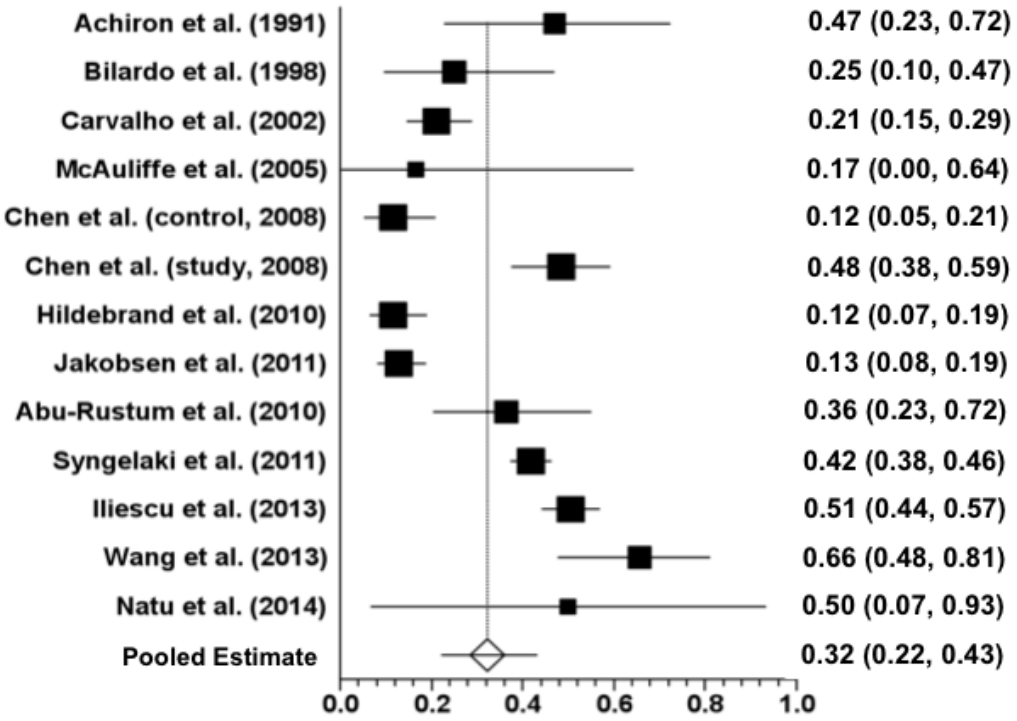

Figure 3(i).

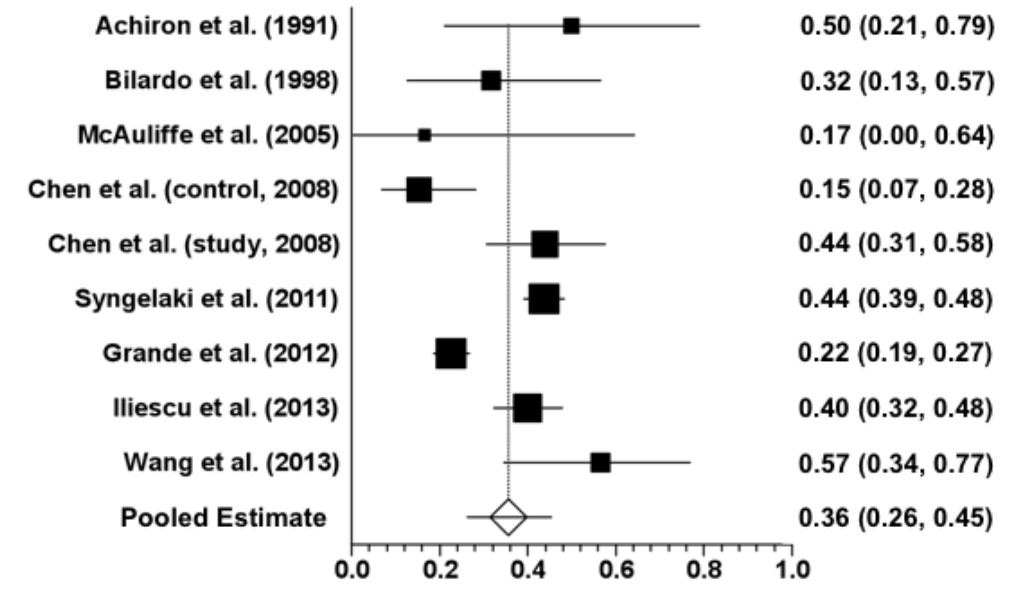

Figure 3(ii).

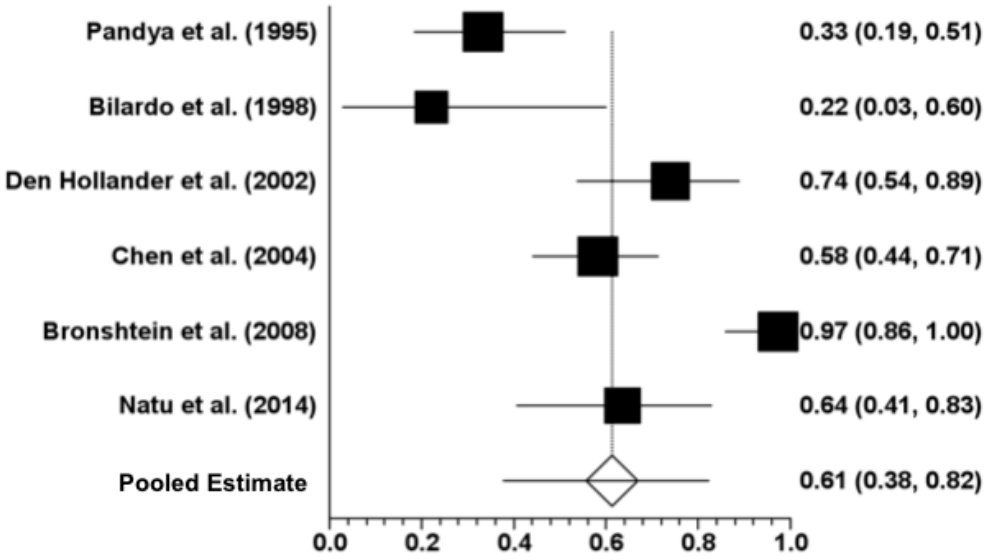

Figure 4(i).

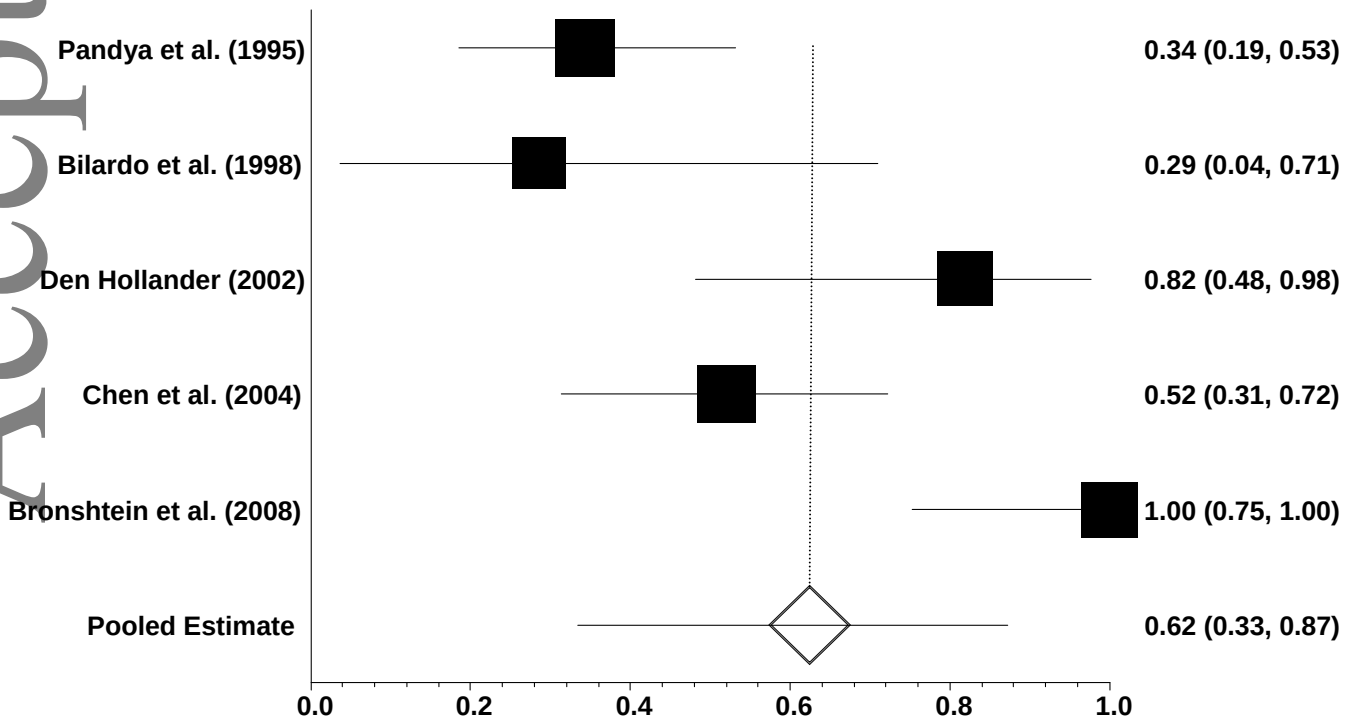

Figure 4(ii).

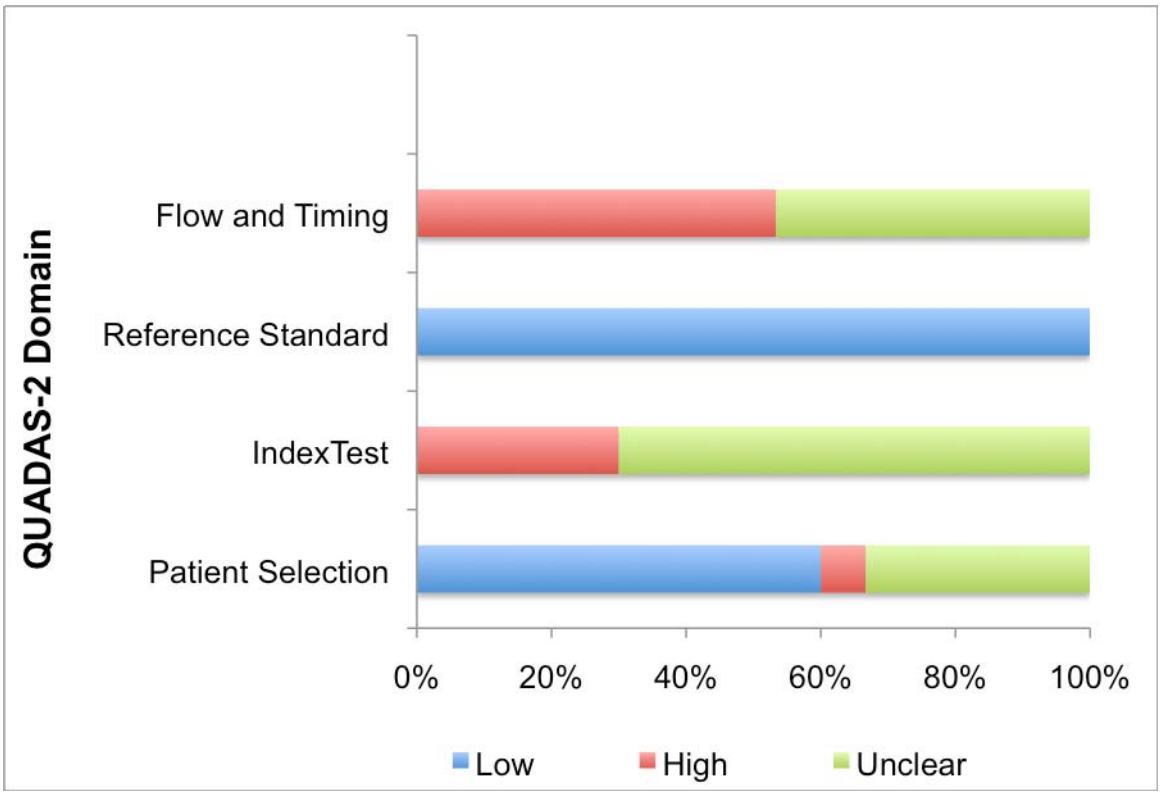

Figure 5(i).

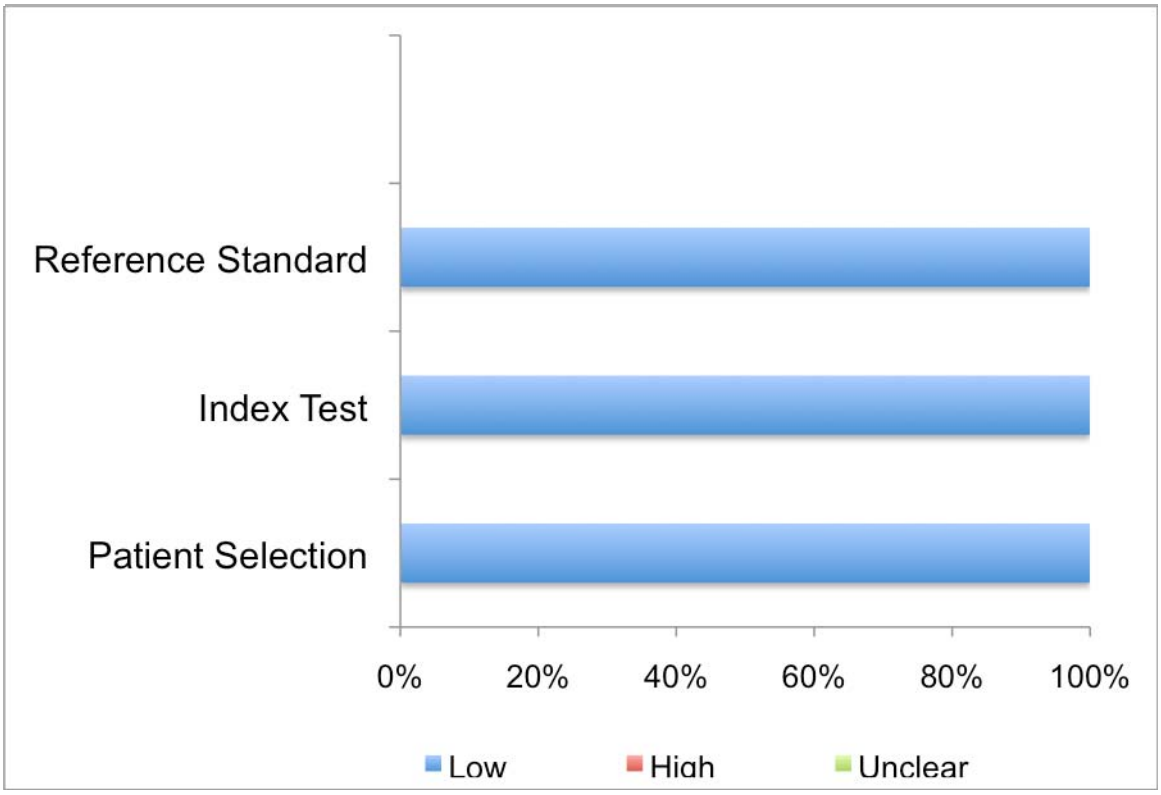

Figure 5(ii).

Tables

Table 1 – Characteristics of studies reporting on the detection of structural anomalies in low risk and unselected populations using first trimester ultrasound.

Notes:
(i). In studies where both TA and TV ultrasound were used, the number in parentheses adjacent to the ultrasound modality refers to the percentage of the study population which received this screening test (when available).
(ii). In studies where aneuploid fetuses were included, the percentage of the study population confirmed by karyotyping as aneuploid was indicated in parentheses (where available).
(iii). The subgroup analysis column identifies the group(s) in which the respective study's data was analyzed.
* Highlights studies where TV ultrasound was only performed in situations where visualization with TA ultrasound was deemed suboptimal.
**Cardiac exam performed at time of first trimester scan, but cardiac malformations excluded from study analysis.
***For the purposes of this review, only the cohort of known euploid fetuses was included in analysis (as insufficient data was provided on the entire cohort reported in the study).

Group	Year	N	Gestation (weeks)	Population	Health-Care Setting	Aneuploid Included?	Index Test	Anatomy Checklist	Cardiac Exam Done?	Subgroup for data Analysis
Achiron[62]	1991	800	9-13	Mixed indications: vaginal bleeding, dating and early anomaly screening	Unclear	Yes	TV/TA	Basic	Yes	2
Hernadi[37]	1997	3991	11-14	Unselected	Unclear	Yes (0.2%)	TV	Basic	No	1

Bilardo[25] (Low risk)	1998	1543	10-14	Consecutive, Singleton pregnancies, Normal NT (<3.0mm),	University Hospital	No	TA/TV	None	Unclear	2
Whitlow[32]	1999	6443	$11\text{-}14^{+6}$	Unselected, consecutive recruitment	University Hospital	Yes (0.7%)	TA/ TV(20.1%)	Detailed	Yes	1
Carvalho[44]	2002	2853	11-14	Unselected	University Hospital, tertiary care	Yes (0.9%)	TA/TV*	Basic	No**	2
Drysdale[43]	2002	917	12-14	Unselected	District General Hospital	Yes	TA/TV	None	No	1
Taipale[34]	2003	20,751	$11\text{-}15^{+6}$	Unselected, consecutive recruitment	Local hospital	Yes (0.3%)	TV/ TA(<1%)	Detailed	Yes**	1
McAuliffe (46)	2005	325	$11\text{-}13^{+6}$	Unselected	University Hospital, tertiary care	No	TA/TV (24.6%)*	Detailed	Yes	2
Cedergren (39)	2006	2708	11-14	Unselected, consecutive recruitment	University Hospital	Yes (0.3%)	TA	None	Unclear	1
Souka (30)	2006	1148	11-14	Unselected	Unclear	Yes	TA/TV	Detailed	Yes	1
Saltvedt[24]	2006	18053	$11^{+5}\text{-}13^{+5}$	Unselected	Multi-centre (8)	No	TA/TV*	Detailed	Yes	1

Dane[29]	2007	1290	11-14	Unselected	Research hospital	Yes	TA/TV	Basic	No	1
Li[28]	2008	2232	11-14	Unselected, consecutive recruitment	Unclear	Yes	TA/ TV (2.0%)*	None	Unclear	1
Chen[19] (Control group)	2008	3693	10-14^{+6}	Unselected, consecutively randomized,	Multi-centre (one university & one regional hospital)	Yes	TA/TV*	None	No	1,2
Chen[19] (Study group)	2008	3949	12-14^{+6}	Unselected, consecutively randomized	Multi-centre (One university & one regional hospital)	Yes	TA/TV*	Detailed	Yes	1,2
Oztekin[42]	2009	1085	11-14	Unselected	Research hospital	Yes	TA/TV*	Detailed	Yes	1
Hildebrand[21]	2010	6692	11-15	Unselected, consecutive recruitment	University Hospital	Yes (0.2%)	TA	None	No	1,2
Abu-Rustum[38]	2010	1370	11-13^{+6}	Unselected, retrospective	Private Practice	Yes (4.4%)	TA/TV*	Detailed	Yes	1,2
Syngelaki[14]	2011	44,859	11-13^{+6}	Unselected, Retrospective	University Hospital, tertiary care	N	TA/ TV(1%)	Detailed	Y	2

Jakobsen[22]	2011	9324	11-15	Unselected, Retrospective	University Hospital	Yes	TA/TV*	None	No	1,2
Vavilala[63]	2011	7916	11-13+6	Unselected	Tertiary Care	Yes	TA/TV*	Detailed	Yes	1
Grande[23]	2012	13723	11-14	Unselected retrospective	Tertiary Care	No	TA/TV	Detailed	Yes	1,2
Pilalis[27]	2012	3902	11-14	Unselected, retrospective	Private maternity hospital	Yes	TA/TV	Detailed	No	1
Becker[64]	2012	6544	11-13^{+6}	Normal NT (≤ 95th centile)	University Hospital	Yes (0.6%)***	TA/ TV* (23.4%)	Detailed	Yes	1
Iliescu[41]	2013	5472	12-13^{+6}	Unselected	Multi-centre (2)	Yes (0.4%)	TA/ TV(7.8%)	Detailed	Yes	2
Wang[33]	2013	2822	11-14	Not stated	University Hospital	Yes	TA	Detailed	Yes	2
Natu[26] (Low Risk)	2014	551	11-14wks	Low risk: age<30, no FH, no co-morbidity	Unclear	Yes	Unclear	Detailed	Yes	2

n – Number of fetuses included in the study population. TA – Transabdominal Ultrasound. TV – Transvaginal Ultrasound. FH – Family history. NT – Nuchal Translucency.

Table 2 – Characteristics of studies reporting on the detection of all types of fetal structural abnormalities using first trimester ultrasound in high risk pregnancies.

Group	Year	N	Gestation (weeks)	Population	Health-Care Setting	Aneuploid Included?	Index Test	Anatomy Checklist	Cardiac Exam Done?	Types of Anomalies Included
Pandya[65]	1995	565	10-14	Euploid fetuses with raised NT (≥3.0mm)	University Hospital, tertiary care	No	TA	None	No	Not specified
Bilardo[25] (High risk)	1998	47	10-14	Euploid fetuses with raised NT (≥3.0mm)	University Hospital	No	TA/TV	None	Unclear	Not specified
Den Hollander[35]	2002	101	11-14	Women with previously affected infants (92%), fetuses with parental consanguinity	Tertiary care	Yes	TA/TV	Detailed	Yes	Not specified
Chen[31]	2004	1609	12-14wks	Women aged ≥35 years	University Hospital	Yes	TA/TV	Detailed	Yes	Not specified
Bronshtein[45]	2008	23	11-14wks	Fetuses with raised NT (≥3.5mm)	Unclear	Yes	TV	Used, but not provided	Yes	All anomalies

Natu[26] (High Risk)	2014	496	11-14wks	Mixed indications including age >30, prev. affected child, FHx of anomalies, multiple pregnancy, hx of smoking/ETOH, maternal RF.	Unclear	Yes	Unclear	Detailed	Yes	All anomalies

n - Number of fetuses included in the study population. NT – Nuchal Translucency. TA – Transabdominal Ultrasound. TV – Transvaginal Ultrasound.

Table 3 – Summary of results from studies assessing the sensitivity of first trimester ultrasound for the detection of major fetal structural abnormalities in low risk/unselected pregnancies (Subgroup 1)

Group	Number of Anomalies Present Within Study per 100 fetuses [95% C.I.]	Prevalence of Affected Fetuses within Study Population* (%) [95% C.I.]	Anomalies Detected in 1st Trimester (TP)	Total Anomalies Present in Study (TP+FN)	False Positives Detected during 1st Trimester USS (FP)	Sensitivity of 1st Trimester USS for Anomaly Detection (%) [95% C.I.]	Sensitivity of 1st Trimester USS for Detection of Affected Fetuses* (%) [95% C.I.]	Antenatal Diagnoses Made Within 1st Trimester** (%)
Hernadi	1.30 [0.97-1.71]	1.08 [0.78-1.45]	20	52	NA	38.46 [25.30-52.98]	32.56 [19.08-48.54]	48.78 [32.88-64.87]
Whitlow	1.20 [0.94-1.49]	0.85 [0.64-1.11]	38	77	3	49.35 [37.76-61.00]	52.73 [38.80-66.35]	57.58 [44.79-69.66]
Drysdale	4.14 [2.95-5.64]	2.94 [1.95-4.26]	5	38	1	13.16 [4.41-28.09]	18.52 [6.30-38.08]	11.76 [3.30-27.45]
Taipale	0.34 [0.27-0.43]	0.32 [0.25-0.41]	37	71	2	52.11 [39.92-64.12]	52.24 [39.67-64.60]	NA
Cedergren	1.11 [0.75-1.58]	1.07 [0.72-1.53]	11	30	NA	36.67 [19.93-56.14]	34.48 [17.94-54.33]	NA
Souka	1.30]1 [0.73-2.15]	1.22 [0.67-2.04]	7	15	3	46.67 [21.27-73.41]	50.00 [23.04-76.96]	50.00 [23.04-76.96]
Saltvedt	0.74 [0.62-0.87]	NA	42	133	2	31.58 [23.80-40.20]	NA	34.43 [26.06-43.57]
Dane	2.71 [1.90-3.75]	1.71 [1.07-2.57]	23	35	NA	65.71 [47.79-80.87]	68.18 [45.13-86.14]	71.88 [53.25-86.25]
Li	1.16 [0.76-1.70]	0.99 [0.62-1.49]	13	26	NA	50.00 [29.93-70.07]	54.55 [32.21-75.61]	65.00 [40.78-84.61]
Chen	1.19	0.73	9	44	NA	20.45	29.63	20.45

(Control group)	[0.87-1.60]	[0.48-1.06]				[9.80-35.30]	[13.75-50.18]	[9.80-35.3]
Chen (Study group)	1.75 [1.36-2.21]	1.44 [1.10-1.87]	41	69	NA	59.42 [46.92-71.09]	43.86 [30.74-57.64]	61.19 [48.50-72.86]
Oztekin	1.81 [1.11-2.78]	1.75 [1.06-2.72]	13	20	NA	65.00 [40.78-84.61]	63.16 [38.36-83.71]	72.22 [46.52-90.31]
Hildebrand	0.49 [0.34-0.69]	NA	13	33	NA	39.39 [22.91-57.86]	NA	NA
Abu-Rustum	1.17 [0.67-1.89]	NA	10	16	1	62.50 [35.43-84.80]	NA	66.67 [38.38-88.18]
Jakobsen	1.07 [0.87-1.30]	NA	16	100	NA	16.00 [9.43-24.68]	NA	33.33 [20.40-48.41]
Vavilala	0.99 [0.78-1.23]	NA	69	78	NA	88.46 [79.22-94.59]	NA	NA
Grande	1.39 [1.20-1.60]	1.39 [1.20-1.60]	92	191	NA	48.17 [40.90-55.50]	48.17 [40.90-55.50]	51.11 [43.57-58.62]
Pilalis	1.54 [1.18-1.97]	NA	26	60	NA	43.33 [30.59-56.76]	NA	44.07 [31.16-57.60]
Becker	1.18 [0.93-1.47]	NA	44	77	NA	57.14 [45.35-68.37]	NA	64.71 [52.17-75.92]
Total Included	1.01 [0.95-1.07]	0.93 [0.85-1.00]	529	1165	12	46.10 [36.88-55.46]	45.25 [38.44-52.14]	53.47 [43.42-63.37]

TP – True positives. FP – False Positives, FN – False Negatives. USS – Ultrasound. NA – Not available.
*Those studies which did not provide data on the number of affected fetuses within their cohorts were not included in the pooled estimates of 'Prevalence of affected fetuses' and of 'Sensitivity of 1st trimester USS for detection of affected fetuses'.
**Those studies which only performed one antenatal USS on fetuses during pregnancy were not included in the pooled estimate of 'Antenatal anomalies diagnosed during the 1st trimester'.

Note: The specificity of first trimester ultrasound for major anomaly detection was not calculated due to the small numbers of studies, which provided data on false positive diagnoses.

Table 4 – Summary of results from studies assessing the sensitivity of first trimester ultrasound for the detection of all types of fetal structural abnormalities in a low risk/unselected pregnancies (Subgroup 2).

Group	Number of Anomalies Present Within Study per 100 fetuses [95% C.I.]	Prevalence of Affected Fetuses within Study Population* (%) [95% C.I.]	Anomalies Detected in 1st Trimester (TP)	Total Anomalies Present in Study (TP+FN)	False Positives Detected during 1st Trimester USS (FP)	Sensitivity of 1st Trimester USS for Anomaly Detection (%) [95% C.I.]	Sensitivity of 1st Trimester USS for Detection of Affected Fetuses* (%) [95% C.I.]	Antenatal Diagnoses Made with 1st Trimester USS** (%)
Achiron	2.12 [1.24-3.38]	1.50 [0.78-2.61]	8	17	NA	47.06 [22.98-72.19]	50.00 [21.09-78.91]	53.33 [26.59-78.73]
Bilardo (Low risk)	1.56 [1.00-2.31]	1.23 [0.74-1.92]	6	24	NA	25.00 [9.77-46.71]	31.58 [12.58-56.55]	37.50 [15.20-64.57]
Carvalho	4.98 [4.21-5.84]	NA	30	142	NA	21.13 [14.73-28.77]	NA	29.13 [20.59-38.90]
McAuliffe	1.85 [0.68-3.97]	1.85 [0.68-3.97]	1	6	1	16.67 [0.42-64.12]	16.67 [0.42-64.12]	20.00 [0.51-71.64]
Chen (Control)	2.11 [1.67-2.63]	1.41 [1.05-1.84]	9	78	NA	11.54 [5.41-20.78]	15.38 [6.88-28.08]	16.36 [7.77-28.80]
Chen (Study)	2.30 [1.86-2.82]	1.44 [1.10-1.87]	44	91	NA	48.35 [37.74-59.07]	43.86 [30.74-57.64]	61.97 [49.67-73.24]
Abu-Rustum	2.41 [1.66-3.37]	NA	12	33	1	36.36 [20.40-54.87]	NA	37.50 [21.10-56.31]
Hildebrand	1.79 [1.49-2.14]	NA	14	120	NA	11.67 [6.53-18.80]	NA	NA
Syngelaki	1.18 [1.09-1.29]	1.09 [0.99-1.19]	222	531	62	41.81 [37.57-46.13]	43.65 [39.19-48.18]	42.86 [38.55-47.25]

Jakobsen (All)	1.92 [1.65-2.22]	NA	23	179	NA	12.85 [8.32-18.65]	NA	21.90 [14.42-31.03]
Grande (All)	NA	3.18 [2.89-3.48]	NA	NA	NA	NA	22.48 [18.64-26.69]	NA
Iliescu	4.77 [4.22-5.37]	2.98 [2.54-3.46]	132	261	187	50.57 [44.34-56.80]	39.88 [32.30-47.83]	53.44 [47.01-59.79]
Wang	1.24 [0.87-1.72]	0.82 [0.52-1.22]	23	35	3	65.71 [47.79-80.87]	56.52 [34.49-76.81]	69.70 [51.29-84.41]
Natu (Low risk)	0.73 [0.20-1.85]	NA	2	4	0	50.00 [6.76-93.24]	NA	50.00 [6.76-93.24]
Pooled Results	1.81 [1.72-1.90]	1.63 [1.54-1.72]	526	1521	254	32.35 [22.45-43.12]	35.56 [26.27-45.44]	41.10 [32.13-50.38]

TP – True positives. FP – False Positives, FN – False Negatives. USS – Ultrasound. NA – Not available.
Note: The specificity of first trimester ultrasound for major anomaly detection was not calculated due to the small numbers of studies, which provided data on false positive diagnoses.
*Those studies which did not provide data on the number of affected fetuses within their cohorts were not included in the pooled estimates of 'Prevalence of affected fetuses' and of 'Sensitivity of 1st trimester USS for detection of affected fetuses'.
**Those studies which only performed one antenatal USS on fetuses during pregnancy were not included in the pooled estimate of 'Antenatal anomalies diagnosed during the 1st trimester'.

Table 5 – Summary of results from studies assessing the sensitivity of first trimester ultrasound for the detection of all types of fetal structural abnormalities in high risk pregnancies (Subgroup 3).

Group	Number of Anomalies Present Within Study per 100 fetuses [95% C.I.]	Prevalence of Affected Fetuses within Study Population* (%) [95% C.I.]	Anomalies Detected in 1st Trimester (TP)	Total Anomalies Present in Study (TP+FN)	False Positives Detected during 1st Trimester USS (FP)	Sensitivity of 1st Trimester USS for Anomaly Detection (%) [95% C.I.]	Sensitivity of 1st Trimester USS for Detection of Affected Fetuses* (%) [95% C.I.]	Antenatal Diagnoses Made with 1st Trimester USS (%)
Pandya	6.37 [4.50-8.71]	5.66 [3.91-7.90]	12	36	NA	33.33 [18.56-50.97]	34.38 [18.57-53.19]	35.29 [19.75-53.51]
Bilardo (High risk)	19.15 [9.15-33.26]	14.89 [6.2-28.31]	2	9	NA	22.22 [2.81-60.01]	28.57 [3.67-70.96]	33.33 [4.33-77.72]
Den Hollander	26.73 [18.41-36.46]	10.89 [5.56-18.65]	20	27	0	74.07 [53.72-88.89]	81.82 [48.22-97.72]	83.33 [62.62-95.26]
Chen	3.42 [2.59-4.43]	1.55 [1.01-2.29]	32	55	5	58.18 [44.11-71.35]	52.00 [31.31-72.20]	64.00 [49.19-7.08]
Bronshtein	94.87 [82.68-99.37]	33.33 [19.09-50.22]	36	37	2	97.30 [85.84-99.93]	100.00 [75.29-100.00]	97.30 [85.84-99.93]
Natu (High Risk)	4.44 [2.80-6.64]	NA	14	22	0	63.64 [40.66-82.80]	NA	63.64 [40.66-82.80]
Pooled Results	6.55 [5.66-7.52]	3.75 [3.02-4.60]	116	186	7	61.18 [37.71-82.19]	62.42 [33.40-87.24]	66.29 [43.47-85.69]

Abbreviations: TP – True positives. FP – False Positives, FN – False Negatives. USS – Ultrasound. NA – Not available.
*Those studies which did not provide data on the number of affected fetuses within their cohorts were not included in the pooled estimates of 'Prevalence of affected fetuses' and of 'Sensitivity of 1st trimester USS for detection of affected fetuses'.

Table 6 – Summary of Results from Meta-Analysis

Subgroup	Population/Anomaly Type	Outcomes		
		Number of Anomalies Present Within Study per 100 fetuses [95% C.I.]	**Sensitivity of 1st Trimester USS for Anomaly Detection (%) [95% C.I.]**	**Antenatal Diagnoses Made with 1st Trimester USS (%) [95% C.I.]**
1	Major Anomalies in a Low Risk/Unselected Population	1.01 [0.95-1.07]	46.10 [36.88-55.46]	53.47 [43.42-63.37]
2	All Types of Anomalies in a Low Risk/Unselected Population	1.81 [1.72-1.90]	32.35 [22.45-43.12]	41.10 [32.13-50.38]
3	All Types of Anomalies in a High Risk Population	6.55 [5.66-7.52]	61.18 [37.71-82.19]	66.29 [43.47-85.69]

Supplement A: Search strategy for systematic review of diagnostic accuracy of first trimester two-dimensional ultrasound for fetal structural abnormalities.

The search was conducted using Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R)[OvidSP](1946-present), Embase [OvidSP](1974-2014 July 17), Cochrane Central Register of Controlled Trials [Cochrane Library, Wiley](Issue 6 of 12, 2014), Database of Abstracts of Reviews of Effects [Cochrane Library, Wiley](Issue 2 of 4, 2014), Science Citation Index-Expanded [Web of Science Core Collection, Thomson Reuters](1945-present).

Search #	Searches Conducted	Results
1	Ultrasonography, Prenatal/	24571
2	Prenatal diagnosis/ and exp ultrasonography/	6667
3	(ultrasound* or ultra-sound or ultrasonogra* or ultra-sonogra* or sonogra* or echocardiogra*).ti,ab.	350062
4	((fetal or foetal or fetus or foetus or prenat* or pre-nat* or prepart* or pre-part*) adj3 (screen* or scan* or structural assessment* or structural survey*)).ti,ab.	4946
5	1 or 2 or 3 or 4	363227
6	Pregnancy Trimester, First/	13090
7	(1st trimester or first trimester).ti,ab.	17031
8	(early pregnan* or early gestation*).ti,ab.	14840
9	(10 week? or 11 week? or 12 week? or 13 week? or 14 week?).ti,ab.	95426
10	(10week? or 11week? or 12week? or 13week? or 14week?).ti,ab.	535
11	((ten*2 or eleven*2 or twel*3 or thirteen*2 or fourteen*2) adj week?).ti,ab.	5950
12	6 or 7 or 8 or 9 or 10 or 11	131973

13	exp *Congenital Abnormalities/	368002
14	(congenital* adj2 (defect? or malformation? or abnormalit* or anomal*)).ti,ab.	46663
15	((fetal or foetal or fetus or foetus) adj2 (defect? or malformation? or abnormalit* or anomal*)).ti,ab.	7081
16	(structural adj2 (defect? or malformation? or abnormalit* or anomal*)).ti,ab.	12112
17	((non-chromosomal or nonchromosomal) adj2 (defect? or malformation? or abnormalit* or anomal*)).ti,ab.	80
18	13 or 14 or 15 or 16 or 17	405021
19	5 and 12 and 18	2386
20	((fetal or foetal or fetus or foetus) adj (anatomy or defect? or malformation? or abnormalit* or anomal*) adj5 (ultrasound* or ultra-sound or ultrasonogra* or ultra-sonogra* or sonogra* or echocardiogra* or scan* or screen* or survey* or assessment?)).ti,ab.	891
21	exp Congenital Abnormalities/us [Ultrasonography]	16703
22	20 or 21	17312
23	12 and 22	1383
24	((early pregnan* or early gestation* or 1st trimester or first trimester) adj3 (ultrasound* or ultra-sound or ultrasonogra* or ultra-sonogra* or sonogra* or echocardiogra* or scan* or screen* or survey* or assessment?)).ti,ab.	2264
25	((10 week? or 11 week? or 12 week? or 13 week? or 14 week?) adj3 (ultrasound* or ultra-sound or ultrasonogra* or ultra-sonogra* or sonogra* or echocardiogra* or scan* or screen* or survey* or assessment?)).ti,ab.	1150
26	((10week? or 11week? or 12week? or 13week? or 14week?) adj3 (ultrasound* or ultra-sound or ultrasonogra* or ultra-sonogra* or sonogra* or echocardiogra* or scan* or screen* or survey* or	10

	assessment?)).ti,ab.	
27	((ten*2 or eleven*2 or twel*3 or thirteen*2 or fourteen*2) adj week? adj3 (ultrasound* or ultra-sound or ultrasonogra* or ultra-sonogra* or sonogra* or echocardiogra* or scan* or screen* or survey* or assessment?)).ti,ab.	34
28	24 or 25 or 26 or 27	3332
29	18 and 28	1113
30	19 or 23 or 29	2744
31	limit 30 to "reviews (maximizes specificity)"	44

Supplement B - QUADAS -2 Assessment Tool:

Defining the review question:

1. What is the sensitivity of first trimester ultrasound for structural fetal malformations? Is it a sensitive enough tool for use in daily clinical practice?

2. What factors might impact detection rates?

-Patient selection: pregnant women with gestational age prior to 14 weeks, mothers with all levels of risk and with either singleton or multiple pregnancies were included

-Index Test: Transvaginal and/or Transabdominal 2D Ultrasound prior to 14 weeks gestational age.

-Reference Standard: Postnatal examination of fetus for structural abnormalities, or postmortem of fetus for structural abnormalities.

-Target condition: all types of congenital fetal structural anomalies (lethal, severe, moderate, and minor as defined by the RCOG).

Domain 1: Patient Selection

A. Risk of Bias : Could the selection of patients have introduced bias?

LOW/HIGH/UNCLEAR

i. Was a consecutive or random sample of patients enrolled? YES/NO/UNCLEAR

ii. Did the study avoid inappropriate exclusions? YES/NO/UNCLEAR

B. Applicability

i. Are there concerns that the included patients and setting do not match the review question (i.e. severity of the target condition, demographic features, presence of co-morbidity, setting)? LOW/HIGH/UNCLEAR

Domain 2: Index Test

A. Risk of Bias - Could the conduct or interpretation of the index test have introduced bias? LOW/HIGH/UNCLEAR

i. Were sonographers blinded to the history (risk profile) of the patients? YES/NO/UNCLEAR

ii. Were all major anatomical organs included in the index test examination? YES/NO/UNCLEAR

iii. Did the study adequately and clearly specify what types of abnormalities were to be assessed by 1st term USS? YES/NO/UNCLEAR

B. Applicability

i. Are there concerns that the index test, its conduct, or interpretation differ from the review question?

LOW/HIGH/UNCLEAR

Domain 3: Reference Standard

A. Risk of Bias – Could the reference standard, its conduct, or its interpretation have introduced bias?

LOW/HIGH/UNCLEAR

i. Was an appropriate reference standard used to correctly classify the target condition?

YES/NO/UNCLEAR

B. Applicability

i. Are there concerns that the target condition as defined by the reference standard does not match the question?

LOW/HIGH/UNCLEAR

Domain 4: Flow and Timing

A. Risk of Bias – Could the patient flow have introduced bias?

LOW/HIGH/UNCLEAR

i. Was a reference standard performed for all appropriate patients enrolled in the study? (including post-mortems for still-births/TOPs in those with diagnosed malformations)

YES/NO/UNCLEAR

ii. Were all patients enrolled in the study (and who had complete follow-up data) included in the analysis?

YES/NO/UNCLEAR

iii. Were all measures of 1st trimester ultrasound detection accuracy (eg. TP, FP, TN, FN) reported?

YES/NO/UNCLEAR

Supplement C: Table of studies included and details regarding which subgroup analysis they contributed to:

Chen et al. (2008)[19], Abu-Rustum et al. (2010)[20], Hildebrand et al. (2010)[21], Jakobsen et al. (2011)[22] and Grande et al. (2012)[23]	Data on fetuses with major abnormalities and on those with a wider range of anomalies were presented, allowing these studies to be analyzed as part of both subgroups 1 and 2.
Saltvedt et al. (2006)[24]	Reported on the detection of all types of abnormalities. However, they only provided a breakdown of major abnormalities and as such was evaluated as part of subgroup 1.
Chen et al. (2008)[19]	In this prospective randomized control trial, which compared the first term detection rates between a control group (randomized to receive only a nuchal

	scan at 10-14+6 weeks) and a study cohort (randomized to receive a nuchal scan at 10-14+6 weeks in addition to a detailed anomaly scan at 12-14+6 weeks). We collated data on the two population cohorts of this study separately, as one group had a detailed first trimester anomaly scan with the use of an anatomical checklist, whereas the other did not.
Bilardo et al. (1998)[25]	Provided data on both a cohort of euploid fetuses with normal nuchal scans (low risk) and a cohort of euploid fetuses with raised nuchal translucency (high risk), which were analyzed

	separately in subgroups 2 and 3 respectively.
Natu et al.(2014)[26]	Assessed unspecified types of anomalies in both low risk and high risk populations and therefore data from this study was analyzed in both subgroups 2 and 3.