Diagnostic accuracy and outcomes of ultrasound in the first trimester of pregnancy for detection of complications relevant for Austrian population, exclusive of screening for Down syndrome: a systematic review

Mag. Ingrid Wilbacher, *Federation of Austrian Social Insurance Institutions,* Ingrid.Wilbacher@hvb.sozvers.at – corresponding author

Dr. Karla Soares-Weiser, *Kleijnen Systematic Reviews Ltd.,* ksoares@netvision.net.il

Dr. Jos Kleijnen, *Kleijnen Systematic Reviews Ltd.,* jos@systematic-reviews.com

Dr. Irmgard Schiller-Fruehwirth, *Federation of Austrian Social Insurance Institutions,* Irmgard.Schiller-Fruehwirth@hvb.sozvers.at

Dr. Stefan Puig, *Paracelsus Medical Private University Salzburg,* stefan.puig@pmu.ac.at

Mag. Dagmar Bernardis, *Federation of Austrian Social Insurance Institutions,* Dagmar.Bernardis@hvb.sozvers.at

Dr. Gottfried Endel, *Federation of Austrian Social Insurance Institutions,* Gottfried.Endel@hvb.sozvers.at

Objective:

Review about the diagnostic accuracy of ultrasound in first trimester of pregnancy

Evidence Based Health Care

Federation of Austrian Social Insurance Institutions

In cooperation with

Kleijnen Systematic Reviews Ltd.

Contact:

Ingrid Wilbacher

Evidence Based Health Care

Federation of Austrian Social Insurance Institutions

Kundmanngasse 21

A – 1030 Wien

Tel.: +431/71132-3615

Fax.: +431/71132-3786

Email: Ingrid.Wilbacher@hvb.sozvers.at

Authors:

Mag. Ingrid Wilbacher, *Federation of Austrian Social Insurance Institutions,* Ingrid.Wilbacher@hvb.sozvers.at

Dr. Karla Soares-Weiser, *Kleijnen Systematic Reviews Ltd.,* ksoares@netvision.net.il

Dr. Jos Kleijnen, *Kleijnen Systematic Reviews Ltd.,* jos@systematic-reviews.com

Dr. Irmgard Schiller-Fruehwirth, *Federation of Austrian Social Insurance Institutions,* Irmgard.Schiller-Fruehwirth@hvb.sozvers.at

Dr. Stefan Puig, *Paracelsus Medical Private University Salzburg, stefan.puig@pmu.ac.at*

Mag. Dagmar Bernardis, *Federation of Austrian Social Insurance Institutions,* Dagmar.Bernardis@hvb.sozvers.at

Dr. Gottfried Endel, *Federation of Austrian Social Insurance Institutions,* Gottfried.Endel@hvb.sozvers.at

Contents

Abstract

Diagnostic Accuracy and Outcomes of Ultrasound in the first trimester of pregnancy for detection of complications relevant for Austrian population excluding the screening for Down Syndrome: A Systematic Review

Principal Investigator

Wilbacher, Ingrid

Evidence Based Health Care

Federation of Austrian Social Insurance Institutions

Kundmanngasse 21

1030 Vienna

Tel: +43 1 71132 3618

Fax: +43 1 71132 3786

Email: Ingrid.Wilbacher@hvb.sozvers.at

Background

A systematic review should enlighten details about accuracy and outcomes of ultrasound in the first trimester of pregnancy for the most relevant pregnancy complications.

Methods/Design

We searched in Medline, Embase, Cinahl and Lilacs for accuracy and outcome data of ultrasound examination in the first trimester, regarding the following endpoints: chromosome anomaly excluding Down Syndrome, chorionicity of multiple pregnancy, premature birth, gestational diabetes and gestational age.

References in retrieved articles and systematic reviews were checked, experts were contacted, and a general search for relevant studies was conducted.

Studies were appraised independently by two reviewers using QUADAS and appropriate checklists from CRD Report 4.

Results

415 studies were identified, 25 fulfilled the inclusion criteria, 21 about chromosomal anomalies, 2 about identification of gestational age, one about gestational diabetes and one about chorionicity. No study was found about risk of preterm birth.

We found a pooled sensitivity of 71% (range 41-85%) for NT measurement including T21 with a pooled specificity of 96% (range 87-100%) and also a pooled sensitivity of 71% (range 50-94%) for NT measurement excluding T21 with a pooled specificity of 96% (range 87-100%). Studies, measuring the absence of nasal bone, had sensitivity ranges of 9-77% including T21 with specificity ranges of 97-100%, and sensitivity ranges of 30-88% excluding T21 with specificity ranges of 97-100%. For calculation of risk based on maternal

age and NT the pooled sensitivity was 77% (range 57-89%) including T21 with a pooled specificity of 96% (range 90-98%) and 77% (range 53-86%) excluding T21 with a pooled specificity of 96% (range 90-98%).

For studies describing different ultrasound measurements the sensitivity ranges are 53-100% including T21 with specificity ranges of 94-100%, and sensitivity ranges are 63-100% excluding T21 with specificity ranges of 94-100%: In general we found no major differences when excluding T21 and just focussing on all other chromosomal anomalies.

No evidence was found for detecting the risk for gestational diabetes in an unselected population. Two studies compared the estimation of gestational age either between first and second trimester or between measurement of crown rump length and last menstrual period. Both studies report no significant differences.

The detection of chorionicity in multiple pregnancies by ultrasound in a qualified centre is reported with 100% sensitivity and 98% specificity. Data about the examiner's skills and experience were scarce.

Conclusions

Good evidence for detecting chromosomal anomalies other than Down Syndrome was found, but test results have to be confirmed by karyotyping, and the availability of karyotyping has to be ensured by the health care system. The pregnant women have to be provided with clear information about the consequences of such a test, which should be optional and not obligatory.

Sponsor

Federation of Austrian Social Insurance Institutions

Starting Date: September 2006

Date of Finalisation: July 2007

Reporting Date: January 2008

Background

The Austrian Pregnancy Screening (mother-child-booklet) includes two ultrasound examinations, one in the second and one in the third trimester.

Obstetric experts in Austria recommend a third screening ultrasound examination in the first trimester of pregnancy. Decision making about this issue involves the Austrian Ministry of Health (Bundesministerium für Gesundheit, Familie und Jugend), the highest medical consultant board (Oberster Sanitätsrat) and the Austrian Social Insurance (Österreichische Sozialversicherung).

A systematic review should provide information about accuracy and outcomes of ultrasound in this early period of pregnancy (12+2 weeks) for the most relevant pregnancy complications in Austria.

This review evaluates the medical indication for ultrasound screening in the first trimester, which means

- how accurate is ultrasound screening for endpoints described below
- which added value of the ultrasound screening in the first trimester can be expected versus the ultrasound examination in the second and third trimester for detection of endpoints described below

A protocol was developed and published on the website of the Federation of Austrian Social Insurance Institutions [1].

Methods/Design

Objectives

Based on health insurance data and the international literature the following objectives were considered as the most relevant ones to be addressed in the review by the Steering Group (Dr. Gottfried Endel, Dr. Irmgard Schiller-Fruehwirth, Mag. Ingrid Wilbacher):

Determination of the accuracy of ultrasound examination in the first pregnancy trimester (incl. 12th week) in diagnosing the following disorders:

- Chromosomal anomalies other than Down Syndrome (Chimera 46,XX/46,XY, Chimera 46,XX/46,XY true hermaphrodite, 46,XX with streak-gonads, 46,XY with streak-gonads, pure gonadal dysgenesis, Fragile X-Chromosome, Fragile X-syndrome) (ICD 10 Q99)
- Detection of chorionicity with ultrasound in the first trimester of pregnancy
- Increased risk of preterm birth (ICD 10 P 07)
- Gestational diabetes (ICD 10 O24)
- Determination of gestational age

Determination of the outcomes after ultrasound examination in the first trimester of pregnancy versus ultrasound examination in the second and/or third trimester for the following target disorders:

- Chromosomal anomalies other than Down Syndrome (Chimera 46,XX/46,XY, Chimera 46,XX/46,XY true hermaphrodite, 46,XX with streak-gonads, 46,XY with streak-gonads, pure gonadal dysgenesis, Fragile X-Chromosome, Fragile X-syndrome (ICD 10 Q99)

- Chorionicity with ultrasound in the first trimester of pregnancy
- Increased risk of preterm birth (ICD 10 P 07)
- Gestational diabetes (ICD 10 O24)
- Gestational age

The method for creating the well formulated (PICO) questions is published on the website of the Federation of Austrian Social Insurance Institutions [2].

Criteria for including and excluding studies

These criteria are based on the objectives mentioned above. Diagnostic accuracy studies were included if they allowed extraction 2 by 2 tables of ultrasound findings compared to a reference standard. As reference standard we accepted only assessments according to definitions of the actual outcome, no other tests predicting such an outcome. Any reference standard for studies assessing gestational age was accepted. For studies assessing outcomes of screening in the first trimester compared to later screening, we included randomised trials and controlled observational studies with parallel control groups.

Criteria for including studies

- accuracy studies
- studies that contain early screening vs. later screening
- screening population
- scan in the first trimester, transvaginal + abdominal
- date of publication as of 1.1.1996

- comparison of screening with confirmation of the findings post partum/post abortum/post AC/CVS

Criteria for excluding studies

- doppler and echocardiography
- Down syndrome
- high risk population
- combination with biochemical markers
- studies where animal experiments were involved
- no scan in the first trimester

Literature Search

The following databases were searched: Medline, Embase, DARE, Cinahl, Lilacs and the National Research Register from 1996 to October 2006. Furthermore, references in retrieved articles and systematic reviews were checked, experts were contacted, and the internet was searched via general search engines such as Google for relevant studies. Identified references were downloaded into Reference Manager software for further assessment and handling. The search strategies are listed in the Tables 1-7.

Quality Assessment

Assessing relevance and inclusion: Studies were screened by title and abstract for relevance independently by two reviewers, disagreements were resolved by consensus. Full papers of studies which appeared potentially relevant were ordered and assessed for inclusion by one

reviewer and checked by a second.

Quality assessment of accuracy studies was carried out using the QUADAS [3] instrument, adapted as appropriate. Quality assessment for randomised trials and controlled observational studies was conducted using the appropriate checklists from CRD Report 4 [4]. Quality assessment was used for descriptive purposes of general study quality, and, if possible, as items in meta-regression analysis in order to examine the influence of study quality on the estimates of diagnostic accuracy. Quality assessment was carried out by one reviewer and checked by a second.

Data management

Data extraction

Data extraction forms were developed using Microsoft Access and Excel, these were piloted independently on a small selection of studies and adjusted as necessary. Data were extracted by one reviewer and checked by a second. The following information was extracted for all studies: identifier, aim, study design, location and setting. In addition, information was extracted on test details (provided test, gold standard, details of test performance, at which gestational age, methods, time between tests), participants' details (number of participants, number of imaging tests performed, age, sex, inclusion criteria) and results (data to construct 2 x 2 table). We included all examiners, but recorded the examiners separately as General Practitioner or Gynaecologist or US technician.

Additional items that we extracted from the included studies:

Study characteristics

- Retrospective/prospective study with 2x2 table

- Diagnosis confirmed by pathology/autopsy
- Karyotyping performed
- % of participants with high risk (age >35, maternal diseases, previous malformation) described
- Unclear test result confirmed by expert
- US performed on the 2nd trimester as well

Type of intervention

- Transvaginal ultrasound (details)
- Transabdominal ultrasound (details)
- Diagnostic test performed by doctor, radiologist, nurse, midwife, others
- Years of experience with performing US
- Were examiners trained specifically to perform the 1st trimester screening?

Analysis

For each test, or combination of tests, the range in sensitivity, specificity and likelihood ratios (of both positive and negative tests results) were calculated, together with ranges in positive and negative predictive values which were calculated based on a number of different estimates of disease prevalence. Diagnostic odds ratios were calculated. These have the advantage of being a single indicator of diagnostic accuracy in contrast to most of the other measures, which have to be judged in pairs. The DOR takes values between 0 and infinity, with high values indicating good test performance.

Heterogeneity of the sensitivity, specificity, likelihood ratios and DOR were investigated

using the Q and I-squared statistics and through visual examination of Galbraith plots of study results.

If studies were homogenous in terms of sensitivity and specificity then the pooled sensitivity and specificity were calculated using a random effects model. If either one of these measures showed evidence of heterogeneity further analyses were conducted using DOR. If study homogeneity could not be rejected, DORs were pooled using a random effects model to calculate a sROC curve, separately for each single study type. If there was evidence of heterogeneity, random-effects meta-regression analysis was done, depending on the amount of data.

Results

Literature search

The literature search identified 4239 references. These references were screened by title and abstract for relevance. 415 references were considered to be potentially relevant and full papers ordered. In addition, we obtained a full copy of further 28 references identified through scrutiny of bibliographic references of the potentially included studies. Three references were not obtainable. Among the potentially relevant references, 32 non-English language papers were assessed for this review and only one paper in German (Tercanli) met the inclusion criteria.

Figure 1 shows the flowchart of studies through the review process and the number of studies excluded after applying inclusion/exclusion criteria and assessing the papers.

In summary, 25 diagnostic accuracy studies evaluating the impact of ultrasound performed in the first trimester of pregnancy met the inclusion criteria: 21 studies were performed to detect chromosomal anomalies, one study evaluated chorionicity, one study gestational diabetes, and two studies gestational age. No study evaluated increased risk of preterm birth in the first trimester. Seventeen of the included studies were performed in Europe (Austria, Denmark, Hungary, Italy, Turkey, and UK), two in the USA, two in Brazil, one in Australia, one in Israel, and another one in Taiwan.

Table 8 provides the study characteristics and reasons for excluding potentially relevant studies. Table 9 shows the study characteristics of the included studies and the according database.

Quality assessment

The QUADAS tool for this review involved 14 items described in Table 10. Figure 2 shows the proportion of studies that were rated "yes", "no" and "unclear" for each one of the QUADAS items. In more than 80% of studies the population was likely to be similar to that in Austria, the participants clearly received a reference standard, the reference standard used was likely to classify the target condition correctly, the time between application of reference standard and index test was short, the reference standard was applied independently of the index test, results of the index test were interpreted without knowledge of the results of the reference standard, and the same clinical data available when the test was used in practice were available for the index test interpretation. In contrast, only 32% of studies gave details about the selection of people for inclusion in the study, in only 23% of studies participants received the same reference standard. Details of the index test and the reference standard were provided in 27% and 9% of the studies respectively, 36% reported uninterpretable results. Approximately 50% of studies reported reasons for withdrawals of participants before the end of the study.

Details of the quality assessment of each included study are provided in Table 11.

Assessment of randomised controlled studies

Figure 3 shows the proportion of studies that were rated "yes", "no" and "unclear" to each one of the items to assess the quality of randomised trials. The two identified randomised trials had an adequate procedure to generate randomisation and adequate allocation concealment, but care providers, assessor and patients were not blinded. Eligibility criteria were clearly described and patients' characteristics were similar at baseline. Results were provided indicating variability of the effect measures, there was a description of sample size

calculation, and data were analysed using an intention-to-treat basis.

Diagnostic accuracy of chromosomal anomalies using first trimester foetal ultrasound

A total of 21 studies reported 49 data sets evaluating foetal ultrasound in the first trimester of pregnancy for the detection of chromosomal anomalies in an unselected population of women. Data were collected for all chromosomal anomalies (including T21). For chromosomal anomalies excluding T21 the following softmarkers were considered: nuchal translucency thickness, absence of nasal bone, omphalocele, placental quotient; for risk calculation maternal age and nuchal translucency.

Studies measuring NT to detect all chromosomal anomalies (including T21)

Table 12 details the relevant characteristics of the included studies measuring nuchal translucency thickness to detect all chromosomal anomalies (including T21). It provides sensitivity, specificity, DOR, positive LR, and negative LR, with a 95% confidence interval for each study. Pooled data were presented only if the level of inconsistency was below 75%. All analyses were performed using a fixed effect model on the MetaDisc Software (version 1.1.4) [5].

Ten studies reported 10 data sets on the diagnostic accuracy of nuchal translucency thickness to detect chromosomal anomalies (including T21). All except one retrospective study (Brizot_Brazil 2001) were prospective diagnostic cohort studies. The reference standards used by these studies were karyotyping for all women, or karyotyping for women with a positive finding and pregnancy outcome for those with a negative finding in the ultrasound. All studies included women at 10 to 14 weeks of pregnancy, and cut-off for NT as described as a measurement varying from 2.5 to 4 mm or above the 95% percentile.

Sensitivities (Figure 4) among the included studies ranged from 41 (specificity 100%) to 85% (specificity 98%); pooled sensitivity was 71% (95% CI: 67-76, I^2 = 16.3%). Specificities (Figure 5) ranged from 100% (sensitivity 41%) to 87% (sensitivity 70%). The DOR (Figure 6) ranged from 15.91 to 236.4; pooled DOR was 86.4 (95% CI: 52.1-143.3, I^2 = 72.6%). Positive likelihood ratios ranged from 5.54 (-LR = 0.35) to 106.5 (-LR = 0.59). Negative likelihood ratios ranged from 0.15 (+LR = 35.44) to 0.59 (+LR = 106.5). Significant heterogeneity (I^2 above 75%) was observed for pooled specificity, these results are not presented. Results for each one of the included studies are presented on Table 12.

In addition, Figure 7 shows the dispersion of sensitivity and specificity plotted in a symmetric ROC space (AUC: 0.88, SE: 0.02), but due to significant statistical heterogeneity results should be interpreted with caution.

Multivariate regression analyses

A multivariate regression analysis was carried out to explore sources of heterogeneity among diagnostic accuracy studies of NT measurement for the detection of all chromosomal anomalies (including T21). Using the *Moses-Shapiro-Littenberg* method [6] covariates were added to the model. The antilogarithm transformations of the resulting estimated parameters are interpreted as a RDOR of the corresponding covariable. They indicate the change in diagnostic performance of the NT measurement per unit increase in the covariate [5].

The regression model $D = \alpha + \beta S$ was extended to include variables of the QUADAS items (score above or below 9), mean maternal age (age above or below 35 years-old), and type of cut-off used in the NT measurement (2.5-4mm or above 95% confidence interval). The results of the multivariate regression analysis are shown in Table 13. None of the items were significant for the analysis and no further attempt was made to explain the observed heterogeneity. These results, however, do not discard the possibility of these variables to account for at least part of the observed heterogeneity. In fact, this lack of association may

well be caused by the small number of studies pooled in the meta-regression, or the poor quality of reporting of relevant information in the majority of studies.

Studies measuring NT to detect chromosomal anomalies (excluding T21)

Table 14 details the relevant characteristics of the included studies measuring nuchal translucency thickness to detect chromosomal anomalies (excluding T21). It provides sensitivity, specificity, DOR, positive LR, and negative LR, with 95% confidence intervals for each study. Pooled data where presented only if the level of inconsistency was below 75%. All analyses were performed using fixed effect model in the MetaDisc Software (version 1.1.4) [5].

Nine studies reported 9 data sets on the diagnostic accuracy of nuchal translucency thickness to detect chromosomal abnormalities (excluding T21). All except one retrospective study (Brizot_Brazil 2001) were prospective diagnostic cohort studies. The reference standards used by these studies were karyotyping for all women, or karyotyping for women with a positive finding and pregnancy outcome for those with a negative finding of the ultrasound. All studies included women at 10 to 14 weeks of pregnancy, and cut-off for NT as described as a measurement varying from 2.5 to 4 mm in all studies, except for one that defined the cut-off as the NT measurement above the 95% percentile.

Sensitivities (Figure 8) among the included studies ranged from 50 (specificity 100%) to 94% (specificity 98%); pooled sensitivity was 71% (95% CI: 67, 76, $I^2 = 28.2\%$). Specificities (Figure 9) ranged from 100% (sensitivity 50%) to 87% (sensitivity 63%). The DOR (Figure 10) ranged from 11.18 to 611.5; pooled DOR was 117.3 (95% CI: 54.2, 254.1, $I^2 = 73.8\%$). Positive likelihood ratios ranged from 5.06 (-LR = 0.42) to 129.4 (-LR = 0.50). Negative likelihood ratios ranged from 0.06 (+LR = 39.15) to 0.50 (+LR = 129.4). Significant heterogeneity (I^2 above 75%) was observed for pooled specificity, these results were not

presented. Results for each one of the included studies are presented on Table 14.

Figure 11 shows the dispersion of sensitivity and specificity plotted in a symmetric ROC space (AUC: 0.88, SE: 0.02). As it is customary with meta-analysis of diagnostic tests, due to significant statistical heterogeneity results should be interpreted with caution.

Multivariate regression analyses

A multivariate regression analysis was carried out to explore sources of heterogeneity among diagnostic accuracy studies of NT measurement for the detection of chromosomal anomalies (excluding T21). As reported in session 4.3.1.1, the regression model $D = \alpha + \beta S$ was extended to include variables for the QUADAS items (score above or below 9), mean maternal age (age above or below 35 years-old), and type of cut-off used in the NT measurement (2.5-4mm or above 95% confidence interval).

The results of the multivariate regression analysis are shown in Table 15. None of the items were significant for the analysis and no further attempt was made to explain the observed heterogeneity. These results, however, do not discard the possibility of these variables to account for at least part of the observed heterogeneity. In fact, this lack of association may well be caused by the small number of studies pooled in the meta-regression, or the poor quality of reporting of relevant information in the majority of studies.

Studies measuring absence of NB to detect all chromosomal anomalies (including T21)

Table 16 details the relevant characteristics of the included studies measuring absence of nasal bone to detect all chromosomal anomalies (including T21). It also provides sensitivity, specificity, DOR, positive LR, and negative LR, with their respective 95% confidence intervals for each study. Pooled data where presented only if the level of inconsistency was below 75%. All analyses were performed using fixed effect model in the MetaDisc Software

(version 1.1.4) [5].

Six studies reported 6 data sets on the diagnostic accuracy of absence of nasal bone used as a softmarker to detect chromosomal abnormalities (including T21). All except one retrospective study (Monni_Italy 2001) were prospective diagnostic cohort studies. Three studies used karyotyping for all women as the reference standards, and three studies used karyotyping for women with a positive finding and pregnancy outcome for those with a negative finding in the ultrasound. All studies included women in the first trimester of pregnancy; four of them specifically expressed that women were between 9 to 14 weeks of pregnancy. The softmarker used in these studies was the absence of nasal bone (Table 16).

Sensitivities (Figure 12) among the included studies ranged from 9 (specificity 100%) to 77% (specificity 99%). Specificities (Figure 13) ranged from 100% (sensitivity 9%) to 97% (sensitivity 52%). The DOR (Figure 14) ranged from 22.71 to 1466.7. Positive likelihood ratios ranged from 18.81 (-LR = 0.49) to 334.1 (-LR = 0.23). Negative likelihood ratios ranged from 0.23 (+LR = 334.1) to 0.91 (+LR = 20.74). Significant heterogeneity (I^2 above 75%) was observed for all pooled data and these results were not presented. Results for each one of the included studies are presented in Table 16.

In addition, Figure 15 shows the dispersion of sensitivity and specificity plotted in a symmetric ROC space (AUC: 0.98, SE: 0.03). As it is customary with meta-analysis of diagnostic tests, due to significant statistical heterogeneity results should be interpreted with caution.

Multivariate regression analyses

A multivariate regression analysis to explore sources of heterogeneity among diagnostic accuracy studies of NB measurement for the detection of all chromosomal anomalies (including T21) was not carried out given the small number of studies included.

Studies measuring absence of NB to detect chromosomal anomalies (excluding T21)

Table 17 details the relevant characteristics of the included studies measuring absence of nasal bone to detect chromosomal anomalies (excluding T21). It also provides sensitivity, specificity, DOR, positive LR, and negative LR, with their respective 95% confidence intervals for each study. Pooled data where presented only if the level of inconsistency was below 75%. All analyses were performed using fixed effect model in the MetaDisc Software (version 1.1.4) [5].

Six studies reported 6 data sets on the diagnostic accuracy of absence of nasal bone used as a softmarker to detect chromosomal anomalies (including T21). All studies were prospective diagnostic cohort studies. Three studies used karyotyping for all women as the reference standards, and three studies used karyotyping for women with a positive finding and pregnancy outcome for those with a negative finding in the ultrasound. All studies included women in the first trimester of pregnancy; four of them specifically said that women were between 9 to 14 weeks of pregnancy. The softmarker used in these studies was the absence of nasal bone (Table 17).

Sensitivities (Figure 16) among the included studies ranged from 30 (specificity 99%) to 88% (specificity 99%). Specificities (Figure 17) ranged from 100% (sensitivity 50%) to 97% (sensitivity 33%). The DOR (Figure 18) ranged from 17.27 to 3235.3. Positive likelihood ratios ranged from 11.91 (-LR = 0.69) to 381.5 (-LR = 0.12). Negative likelihood ratios ranged from 0.12 (+LR = 381.5) to 0.71 (+LR = 25.05). Significant heterogeneity (I^2 above 75%) was observed for all pooled data and these results were not presented. Results for each one of the included studies are presented in Table 17.

In addition, Figure 19 shows the dispersion of sensitivity and specificity plotted in a symmetric ROC space (AUC: 0.99, SE: 0.03). As it is customary with meta-analysis of diagnostic tests, due to significant statistical heterogeneity results should be interpreted with

caution.

Multivariate regression analyses

A multivariate regression analysis to explore sources of heterogeneity among diagnostic accuracy studies of NB measurement for the detection of chromosomal anomalies (excluding T21) was not carried out given the small number of studies included.

Studies measuring a calculation of risk based on maternal age, and NT to detect all chromosomal anomalies (including T21)

Table 18 details the relevant characteristics of the included studies measuring a calculation of risk based on maternal age and NT to detect all chromosomal anomalies (including T21). It also provides sensitivity, specificity, DOR, positive LR, and negative LR, with their respective 95% confidence intervals for each study. Pooled data where presented only if the level of inconsistency was below 75%. All analyses were performed using fixed effect model in the MetaDisc Software (version 1.1.4) [5].

Five studies reported 5 data sets on the diagnostic accuracy of the use of a calculation of risk based on maternal age, biochemistry and NT to detect all chromosomal abnormalities (including T21). All studies were prospective diagnostic cohort studies. Two studies used karyotyping for all women as the reference standards, and three studies used karyotyping for women with a positive finding and pregnancy outcome for those with a negative finding in the ultrasound. All studies included women in the first trimester of pregnancy; specifically between 9 to 14 weeks of pregnancy. Risk of a chromosomal abnormality was calculated based on maternal age, biochemistry and NT, but only one study (Tercanli_Germany 2001) explicitly mentioned the cut-off used to define risk (1:400) (Table 18).

Sensitivities (Figure 20) among the included studies ranged from 57 (specificity 94%) to 89% (specificity 90%); pooled sensitivity was 77% (95%CI: 69-83%, I^2 = 64.3%). Specificities

(Figure 21) ranged from 98% (sensitivity 59%) to 94% (sensitivity 57%). The DOR (Figure 22) ranged from 21.84 to 94.07; pooled DOR was 68.58 (95% CI: 43.26-108.72, I^2 = 22.4%). Positive likelihood ratios ranged from 9.31 (-LR = 0.12) to 34.65 (-LR = 0.41). Negative likelihood ratios ranged from 0.12 (+LR = 9.31) to 0.45 (+LR = 9.93). Significant heterogeneity (I^2 above 75%) was observed for the pooled specificity and positive likelihood ratio and these results were not presented. Results for each one of the included studies are presented in Table 18.

In addition, Figure 23 shows the dispersion of sensitivity and specificity plotted in a symmetric ROC space (AUC: 0.95, SE: 0.02). As it is customary with meta-analysis of diagnostic tests, due to significant statistical heterogeneity results should be interpreted with caution.

Multivariate regression analyses

A multivariate regression analysis to explore sources of heterogeneity among diagnostic accuracy studies of evaluating the detection of all chromosomal anomalies (including T21) by measuring a calculation of risk based on maternal age, and NT was not carried out given the small number of studies included.

Studies measuring a calculation of risk based on maternal age, and NT to detect chromosomal anomalies (excluding T21)

Table 19 details the relevant characteristics of the included studies measuring a calculation of risk based on maternal age, and NT to detect chromosomal anomalies (excluding T21). It also provides sensitivity, specificity, DOR, positive LR, and negative LR, with their respective 95% confidence intervals for each study. Pooled data where presented only if the level of inconsistency was below 75%. All analyses were performed using fixed effect model in the MetaDisc Software (version 1.1.4) [5].

Four studies reported 4 data sets on the diagnostic accuracy of the use of a calculation of risk based on maternal age, and NT to detect chromosomal anomalies (excluding T21). All studies were prospective diagnostic cohort studies. One study used karyotyping for all women as the reference standard, and three studies used karyotyping for women with a positive finding and pregnancy outcome for those with a negative finding in the ultrasound. All studies included women in the first trimester of pregnancy; specifically between 10 to 14 weeks of pregnancy. Risk of a chromosomal abnormality was calculated based on maternal age, and NT, but only one study (Tercanli_Germany 2001) explicitly mentioned the cut-off used to define risk (1:400) (Table 19).

Sensitivities (Figure 24) among the included studies ranged from 53 (specificity 98%) to 86% (specificity 90%); pooled sensitivity was 77% (95%CI: 66-86%, I^2 = 53.7%). Specificities (Figure 25) ranged from 98% (sensitivity 53%) to 90% (sensitivity 86%). The DOR (Figure 26) ranged from 53.94 to 127.0; pooled DOR was 74.28 (95% CI: 41.91-131.64, I^2 = 0.0%). Positive likelihood ratios ranged from 8.95 (-LR = 0.16) to 31.18 (-LR = 0.47). Negative likelihood ratios ranged from 0.15 (+LR = 19.0) to 0.47 (+LR = 31.18). Significant heterogeneity (I^2 above 75%) was observed for the pooled specificity and positive likelihood

ratio and these results were not presented. Results for each one of the included studies are presented on Table 19.

In addition, Figure 27 shows the dispersion of sensitivity and specificity plotted in a symmetric ROC space (AUC: 0.96, SE: 0.02). As it is customary with meta-analysis of diagnostic tests, due to significant statistical heterogeneity results should be interpreted with caution.

Multivariate regression analyses

A multivariate regression analysis to explore sources of heterogeneity among diagnostic accuracy studies of evaluating the detection of chromosomal anomalies (excluding T21) by measuring a calculation of risk based on maternal age, and NT was not carried out given the small number of studies included.

Studies describing different ultrasound measurements to detect all chromosomal anomalies (including T21)

Table 20 details the relevant characteristics of the included studies describing different ultrasound measurements to detect all chromosomal anomalies (including T21). It also provides sensitivity, specificity, DOR, positive LR, and negative LR, with their respective 95% confidence intervals for each study. Pooled data were not presented even when the level of inconsistency was below 75% because these studies used different markers to detect chromosomal anomalies and should not be combined. All analyses were performed using fixed effect model in the MetaDisc Software (version 1.1.4) [5].

Four studies reported 4 data sets on the diagnostic accuracy of different ultrasound measurements to detect all chromosomal anomalies (including T21). Three studies were prospective diagnostic cohort studies and one study was a retrospective cohort (Monni_Italy 2005). All studies used karyotyping for women with a positive finding and pregnancy

outcome for those with a negative finding in the ultrasound. All studies included women in the first trimester of pregnancy; except for one that included women between 12 to 16 weeks of pregnancy (Blazer_Israel 2004). Risk of a chromosomal abnormality was calculated using different markers (omphalocele, placental volume, NT or NB) combined with a transabdominal and/or transvaginal ultrasound scan (Table 20).

Sensitivities (Figure 28) among the included studies ranged from 53 (specificity 90%) to 100% (specificity 100%). Specificities (Figure 29) ranged from 100% (sensitivity 100%) to 90% (sensitivity 53%). The DOR (Figure 30) ranged from 10.35 to 28247.8. Positive likelihood ratios ranged from 5.40 (-LR = 0.52) to 1413.3 (-LR = 0.05). Negative likelihood ratios ranged from 0.05 (+LR = 1413.3) to 0.52 (+LR = 5.40). Results for each one of the included studies are presented in Table 20.

In addition, Figure 31 shows the dispersion of sensitivity and specificity plotted in a symmetric ROC space (AUC: 0.52, SE: 0.40). As it is customary with meta-analysis of diagnostic tests, due to significant statistical heterogeneity results should be interpreted with caution.

Additionally, no multivariate regression analysis to explore sources of heterogeneity was carried out.

Studies describing different ultrasound measurements to detect chromosomal anomalies (excluding T21)

Table 21 details the relevant characteristics of the included studies describing different ultrasound measurements to detect chromosomal anomalies (excluding T21). It also provides sensitivity, specificity, DOR, positive LR, and negative LR, with their respective 95% confidence intervals for each study. Pooled data were not presented even when the level of inconsistency was below 75% because these studies used different markers to detect chromosomal anomalies and should not be combined. All analyses were performed using fixed effect model in the MetaDisc Software (version 1.1.4) [5].

Four studies reported 4 data sets on the diagnostic accuracy of different ultrasound measurements to detect chromosomal anomalies (excluding T21). Three studies were prospective diagnostic cohort studies and one study was a retrospective cohort (Monni_Italy 2005). All studies used karyotyping for women with a positive finding and pregnancy outcome for those with a negative finding in the ultrasound. All studies included women in the first trimester of pregnancy; except for one that included women between 12 to 16 weeks of pregnancy (Blazer_Israel 2004). Risk of a chromosomal abnormality was calculated using different markers (omphalocele, placental volume, NT or NB) combined with a transabdominal and/or transvaginal ultrasound scan (Table 21).

Sensitivities (Figure 32) among the included studies ranged from 62 (specificity 90%) to 100% (specificity 100%). Specificities (Figure 33) ranged from 100% (sensitivity 100%) to 90% (sensitivity 62%). The DOR (Figure 34) ranged from 15.33 to 22301. Positive likelihood ratios ranged from 6.37 (-LR = 0.42) to 1394.7 (-LR = 0.06). Negative likelihood ratios ranged from 0.06 (+LR = 1394.7) to 0.42 (+LR = 6.37). Results for each one of the included studies are presented in Table 21.

In addition, Figure 35 shows the dispersion of sensitivity and specificity plotted in a symmetric ROC space (AUC: 0.59, SE: 0.31). As it is customary with meta-analysis of diagnostic tests, due to significant statistical heterogeneity results should be interpreted with caution.

In addition, no multivariate regression analysis to explore sources of heterogeneity was carried out.

An overview of all our results for chromosomal analysis is shown in Table 22.

Diagnostic accuracy of chorionicity in twin pregnancies using first trimester foetal ultrasound

In discussions about our objectives, Austrian experts pointed out that the chorionicity in twin pregnancies was important to be diagnosed in the first trimester. In the first trimester chorionicity can be detected reliably which would lead to identify twin pregnancies at high risk for twin-to-twin-transfusion-syndrome. TTTS leads to up to 50% higher mortality rates in monochorionic-monoamniotic twins.

Thus, it was decided to include chorionicity as an objective into the review. Screening for chorionicity should be done in a specialised centre where the high risk pregnancy will be observed until delivery. Part of the screening examination will be detecting the twin pregnancy. It was concluded in the discussion that the detection of twin pregnancy can be done in outpatient setting very easily and does not have to be evaluated for accuracy. But screening for twin pregnancy can only be valid if the following ultrasound detection of chorionicity is accurate. So the PICO question was to determine whether accuracy of ultrasound for detection of chorionicity in twin pregnancies is higher in the first than in the second trimester.

One retrospective study reporting on the diagnostic accuracy of first trimester ultrasound to determine chorionicity in a sample of 463 twin pregnancies delivered over a 6-year period was identified (Menon_Austria 2005).

This study about chorionicity describes the detection with vaginal ultrasound at 10-14 weeks [7] with a sensitivity of 100% and specificity of 98% using the lambda sign, or 92% sensitivity and specificity using the inter-twin-membrane thickness (Carroll 2002). The Menon study used a retrospective diagnostic cohort design to compare the detection of chorionicity by sonographers using the results of a transvaginal ultrasound performed in the first 14 weeks of pregnancy or after week 15. The reference standard used was the post partum pathological diagnosis of chorionicity. Details of this study are presented in Table 9.

The authors described that 428 out of 436 twin pregnancies were correctly diagnosed for chorionicity as confirmed by pathology reports. This results in a sensitivity of 100%, a specificity of 97.9% and a positive predictive value of 88.2% for sonography as a screening tool for monochorionic twin pregnancies.

As no other study was found, no meta-analysis was performed.

Diagnostic accuracy of gestational age using first trimester foetal ultrasound

One of the scientific questions related to the accuracy and effects of early ultrasound scan to ascertain gestational age. Various recommendations in guidelines [8] for pregnancy care include the measurement of CRL for gestational age and adjustment of the expected date of delivery, if necessary with the aim of avoiding or reducing unnecessary induction of labour.

In Austria, there is an additional legal issue concerning the EDD: the claim for maternity allowance is linked to the calculation of EDD. For normal pregnancy the claim starts eight weeks before EDD, in case of preterm birth, multiple pregnancy or caesarean delivery there is a claim of twelve weeks, prospective or retrospective (§162 ASVG).

In this review we found two randomised controlled studies matching our objective about gestational age, which met the inclusion criteria of screening population, high level study design (RCT, Cohort), published after 1.1.1996, and ultrasound in first trimester (inclusive 14^{th} week of pregnancy). Both studies were considered to be of good quality. In particular, they reported an adequate procedure to generate randomisation and allocation concealment, clearly described sample size calculation and eligible criteria, and patient's characteristics at baseline were similar. Results were presented indicating the uncertainty of the main measurement, and data were analysed based on an intention-to-treat approach. No blinding of care providers, assessor and patients was attempted. Although both studies attempted to measure the accuracy of a first trimester ultrasound dating scan in predicting gestational age, they have done so in a different manner, therefore, we chose to summarise each study separately.

Bennett et al. [9] compared estimation of gestational age in the first trimester (8-12 weeks) by

CRL measurement versus second trimester (19 weeks) biometry alone in an RCT. Two different teams performed the ultrasound examination in the first and in the second trimester. The mean maternal age of the 196 women was 29,3 years (range 16-40), the study was performed in Canada, most of the pregnancies were singleton. The endpoints were induction of labour (all reasons), adjustment for gestational age, post-term labour, delivery at gestational age >287 days, and caesarean birth.

They found differences in adjustment of gestational age (41,3% in the first trimester scan group versus 10,9% in the second trimester scan group; RR 0.26, 95% CI 0.15-0.46, P<0.001), in labour induction (4,8% in the first trimester scan group versus 13,0% in the second trimester scan group; RR 0.37, 95% CI 0.14-0.96, P=0.04) and delivery on or after 287 days of pregnancy (6,7% in the first trimester scan group versus 16,3% in the second trimester scan group) (Table 23). No differences between the two groups were found in kind of delivery (vaginal, caesarean) and in neonatal outcome.

Bennett et al. compared the advantages of first trimester and those of second trimester ultrasound in the discussion. (Table 24)

The authors also report a Medline search from 1970 – 2002 and found no studies which addressed the question of first trimester ultrasound (with CRL measurement) as a strategy to reduce labour induction rates.

Nine of 12 possible quality criteria were fulfilled in this study according to our review method. The three quality criteria with negative answers focus on blinding of the outcome assessor, the provided care and the patient, which were not reported clearly.

In general it is a good study. There is maybe a bias in the cut off for dating a prolonged pregnancy with 287 days, knowing that first trimester ultrasound allows an accurate dating within 4-5 days, and second trimester dating an accuracy of 7-14 days. The reasons for labour

inductions were not reported, so other indications than just date (like gestational diabetes or big baby) may contribute to the results. The authors suggested that a program of first trimester ultrasound screening to a low-risk obstetric population may result in a reduction in the rate of labor induction for post term pregnancy.

Harrington et al. [10] reported a multi centre RCT comparing two groups of pregnant women with a mean age of 30. One group received first trimester ultrasound screening (8^{th} – 12^{th} pregnancy week) measuring CRL; the second group without first trimester screening had gestational age calculated from the last menstrual period. For the scan group, the EDD was changed if there was a discrepancy of more than 5 days from the gestation, calculated from the last menstrual period. The primary endpoints were correction of EDD, rate of induction for prolonged pregnancy, overall induction, and kind of delivery. Suspected and proven foetal growth rate were used as secondary endpoints.

They found no difference in the rate of induction for prolonged pregnancy (scan group 19/233; 8,15%; no scan group 17/230; 7,39%), the EDD was adjusted in 13/233; 5,6% (scan group) versus 2/230; 0,87% (no scan group). The proportion of spontaneous labour was greater in the scan group (154/233; 66,5%) than in the no-scan group (132/230; 57,4%). Neither of these findings were statistically significant, there were also no significant differences in the number of assisted deliveries or caesarean sections. Eight women in the no-scan group (3,4%) and three women in the scan group (1,2%) were suspected of FGR; low birth weight was confirmed in seven of the eight in the no-scan group and in two of the three in the scan group. (Table 25)

Diagnostic accuracy of gestational diabetes using first trimester foetal ultrasound

Kelecky et al. [11] reported a prospective study in an unselected population, excluding all women with previous risk factors for gestational diabetes. NT was used as the assessment parameter (cut off at the 95th percentile); a proposed discussed pathophysiological mechanism was that hyperglycemia could increase the microvascular permeability, and there is also a direct relation between hyperglycemia and congenital cardiac malformations (Kelecky refer to [12], [13]). Two groups of pregnant women were compared after NT measurement: 389 women with NT greater than the 95th percentile, and 386 age-matched consecutive women with NT in the normal range. At 24 – 28 weeks of pregnancy the women underwent a 50 g glucose screening test; if it was positive then a 100 g oral glucose tolerance test was performed. The main outcome measures were the prevalence of gestational diabetes mellitus, impaired glucose tolerance and the number of macrosomic infants. Two reviewers defined the quality of the study as good with 9 positive ratings in QUADAS.

The authors reported significant differences between the two groups for impaired glucose tolerance in the 100g oral glucose tolerance test (3.6% versus 2.1%, Table 26) and macrosomic infants. Macrosomia was more common in the group with NT >95th percentile (6.2% versus 4.4%). Kelecki et al. conclude that NT> 95th percentile is associated with impaired glucose tolerance. However, p-values are close to the cut-off level for significance, there was no difference in the 50g OGTT and more importantly, no difference in the prevalence of gestational diabetes: nine women in the NT normal group and ten women in the group with NT >95th percentile were diagnosed with gestational diabetes (p=0,795).

Diagnostic accuracy of preterm birth using first trimester foetal ultrasound

No study investigating the role of first trimester ultrasound in predicting preterm birth fulfilled our inclusion criteria.

Experience of the assessor and ultrasound equipment

From the 25 studies included in our review the assessor of the test was an obstetrician (7 studies), a sonographer/technician (10 studies), a GP (one study), a midwife/nurse (one study) and other/ unclear (6 studies).

The assessors' experience was reported with the FMF certificate (8 studies), with uniform training (and details about that) in one study, with more than 2 years experience (one study) and unclear in 15 studies.

The type of measurement was transabdominal in 4 studies, transvaginal in two studies, both, transabdominal and /or transvaginal in 11 studies, and unclear in 8 studies.

Technical quality was reported with levels of megahertz (MHz) in 10 of the 25 included studies. In three of these 10 studies 6 MHz were used, in three 3,5-5 MHz were used, in two 5/8 MHz were used, in one study they used 3,5-7,5 and in another study 6,5-7,5 MHz.

In five of the 25 included studies different instruments were used, in 10 studies the instrument name was not reported, ten studies reported six different instruments for their ultrasound assessments. The extracted details are listed in table 27.

Due to a lack of reported quality criteria no pooled data were calculated.

Update

In order to keep our review up to date, we repeated the same search strategy in Pubmed as at the start of the review and limited the time of publication from 2006-10-01 to 2008-02-19. 118 studies were detected for accuracy and 45 for comparison studies. 38 Studies were overlapping in the two results. Reviewed by title and abstract 109 studies were excluded (1 bibliography, 18 combined with biochemistry, 18 case studies, 1 comment, 6 studies including Doppler measurement, 1 editorial, 1 guideline, 1 foreign language which we could not read – Slowak, 2 letters, 15 reviews, 39 not fitting our aims, 4 selected populations, 3 not first trimester) and 9 full papers were ordered (table 28). All of these studies were excluded when applying our inclusion/exclusion criteria:

No data for 2x2 table (Benoit, Czuba, Pons, Evans, Scott, Watson)

Selected population (Benoit, Watson)

No scan in first trimester (Benoit, Scott)

No confirmation of the scan results (Czuba)

Data combined with biochemistry (Breathnach)

One study did not include our aims (Westin).

The study of El Kateb et al. 2007 examined the outcome of monochorionic twin pregnancies. In this paper only pregnancies with already detected chorionicity were studied.

No additional studies were found to include in our review.

Discussion

Chromosomal anomalies

In our review we found a pooled sensitivity of 71% (range 41-85%) for NT measurement including T21 with a pooled specificity of 96% (range 87-100%) and also a pooled sensitivity of 71% (range 50-94%) for NT measurement excluding T21 with a pooled specificity of 96% (range 87-100%). Studies, measuring the absence of nasal bone, had sensitivity ranges of 9-77% including T21 with specificity ranges of 97-100%, and sensitivity ranges of 30-88% excluding T21 with specificity ranges of 97-100%. For calculation of risk based on maternal age and NT the pooled sensitivity was 77% (range 57-89%) including T21 with a pooled specificity of 96% (range 90-98%) and 77% (range 53-86%) excluding T21 with a pooled specificity of 96% (range 90-98%).

For studies describing different ultrasound measurements the sensitivity ranges are 53-100% including T21 with specificity ranges of 94-100%, and sensitivity ranges are 63-100% excluding t21 with specificity ranges of 94-100%: In general we found no major differences when excluding T21 and just focussing on all other chromosomal anomalies.

In multivariate regression analysis for the studies with NT measurement there was not any significant influence of maternal age, NT cut off or QUADAS results detectable.

The results represent generally reasonable sensitivity and moderate specificity for detecting chromosomal anomalies by first trimester ultrasound, with false positive rates of up to 12%, with a mean FPR of about 5%. The results are similar when including or excluding T21, slightly better for chromosomal defects exclusive T21. Studies focusing just on T21 were not included.

Different reviews and their included studies detecting chromosomal anomalies with ultrasound examination overlap with the included studies in our review. Differences were

found in terms of pregnancy weeks (8-16 weeks vs. up to 12weeks + 2), included population and publication date of the included studies. NT cut off points vary, but in most of the studies 2.5mm, 3mm, 95th percentile or 99th percentile were used. Most of the reviews report relatively wide ranges of sensitivity, but pooled data show almost all about 70%. Specificity rates were not reported in all reviews, but if, they were about 90-100%. FPR are reported in most of the reviews of about 4-5% [14], [15], [16], [17], [18], [19].

Beke [20] report in their review 12 studies with measurement of nuchal translucency (subcutaneous edema). The cut off points varied between 2.5mm, 3mm, 5-6mm, 95th percentile, 99th percentile and adjusted to CRL, gestational weeks range from 8 to 16 weeks, it was not reported whether the populations were selected or unselected. The detection rate of chromosomal anomalies varies in the different papers, ranging between wide limits (2,86%-48,15%). The abnormal karyotype was mainly T21.

Brambati et al. [21] included 10 studies measuring NT with cut off levels >2.5mm in gestational week 8-15 in unselected populations. Ranges of sensitivity from 33% to 93.5% and ranges of specificity from 90.5% to 99% are reported in these studies.

Bindra et al. [14] show the results of 13 studies, screening by NT with cut offs of 2.5mm, 3mm, 95th percentile and 99th percentile in gestational week 9-14. They report the combined results out of 170343 pregnancies included with 77% sensitivity (range 57%-100%), and a false positive rate of 4,7% (range 0,4%-8%) for T21.

Taipale et al. [15] included 14 studies with NT measurement and cut off points of 2.5mm, 3mm, 95thpercentile and 99th percentile in weeks 8-14 and report a sensitivity of 72% for T21 and for any aneuploidy of 69% with a false positive rate of 4,2% (total rates).

Sherer et al. [18] report about seven studies about NT measurements in unselected populations including cut off ranges between 2.5mm and 4mm, including pregnancy weeks 8

to 16. They report sensitivity ranges between 0 to 83%.

Nicolaides 2003 [16] show 14 studies examining NT with cut off from 2.5mm, 3mm, 95thpercentile and 99th percentile in week 9-14 and report detection rates for T21 of 77% (range 57-100%) and false positive rates of total 4.7% (range 0.4-8%).

Stewart and Malone [17] report 8 studies with NT in unselected population with cut off points of 2.5mm, 3mm, 95th percentile and 99th percentile in gestational week 8-15.9. These studies reported a of sensitivity 70% (range 20-91%) for all aneuploidies, 70% for Down syndrome (range 29-91%) and a positive predictive value of 10% (range 3-24%).

Chitty et al. [22] discuss six studies in unselected populations with NT measurement and cut off points of 2.5mm and 3mm in week 8-13. They report a total sensitivity for detection of aneuploidy of 70% (range 40-100%) and for detection of T21 of 62% (range 33-100%) with a false positive rate of 4% (range 0.9-6.3%).

Snijders et al. [23] report about 8 studies with NT measurement in unselected populations, cut off points of 2.5mm, 3mm, 4mm, in week 8-15. The detection rate for T21 ranges from 33% to 90% with false positive rates from 0.6% to 6.3%.

Nicolaides 2004 [19] reports 19 studies with an overall sensitivity of 76,8% with 4.2% FPR for detection of T21 by NT measurement, and 69% sensitivity with 1.4% FPR for measurement of NB out of four post mortem radiologic studies.

Most of the authors point out the need for appropriate training of the examiners.

Chorionicity

Monteagudo [24] reports an accuracy of 50-100% for detection of chorionicity by transvaginal ultrasound at an early gestational age (8 weeks postmenstrual) in a review about different methods (number of chorionic sacs, number of amniotic sacs, gender

differentiation).

In the second and third trimester of pregnancy chorionicity can be detected by gender difference with a sensitivity of 51% and a specificity of 100% (different gender), and with a sensitivity of 100% and specificity of 51% (same gender). For the number of placental sites the sensitivities measured were 32% and 95,8% and the specificities 100% and 57,9%. The Lambda sign was measured with sensitivities of 7% and of 83% (no specificities reported), and for intertwin membranes counting sensitivities of 94% (for monochorionic multiple pregnancies) and 100% (for dichorionic multiple pregnancies). The absence of a thick membrane (>=2mm) is reported with a predicitve value of *up to 82%* for monochorionicity.

If TTTS occurs, it can be detected by symptoms like a rapidly growing womb [25] and/or the image of two differently growing foetuses in ultrasound. TTTS is associated with maternal hyperaldosteronism dissociated from renin-angiotensin changes.

Laser fetoscopy is the "gold-standard" treatment of TTTS. However, it is a sophisticated technique that relies on proper training [26]. Laser surgery of the chorionic plate inter-twin anastomoses is the best first-line treatment when the syndrome develops before 26 weeks' gestation [27].

Survival rates after laser coagulation are reported to be 57% [28], 70% [29], 71.5% [30], 77.4% [31], and between 55 and 69% [32] to 100% [33] survival; and for at least one survivor in 74% [28],81% [29], 83.5% [30]. In preterm TTTS cases, neonatal morbidity decreases independently with gestational age and after successful fetoscopic laser surgery. Neonatal morbidity due to TTTS was higher in the amnioreduction group and in cases with failed laser therapy [34]. Although perinatal outcome in TTTS has improved after laser therapy, neonatal mortality and morbidity rates remain high.

In 2005 587 multiple deliveries (ICD 10 O84) were documented in public hospitals. A differentiation into twin- and multiple deliveries is not possible from the documentation level

used (PEGASUS) and our calculations are done assuming that all are twin pregnancies. 67.424 deliveries were documented for 2005 in Austria. From the literature 13,7% of twin pregnancies are monochorionic (Menon 2005) and 20% of monochorionic twins develop a TTTS [49]. Monochorionic twins have a 5-8 times higher morbidity and mortality. (Menon 2005). If 13,7% of all twin pairs are monochorionic, then there were 80 monochorionic twin cases (=160 babies) affected in Austria in 2005. (587x0,137). If out of the monochorionic pairs 20% are affected with TTTS, there would be 16 estimated cases in 2005 in Austria. A screening examination for TTTS would reach 0,02% of pregnancies (16/67.424). For such a rare condition an overall screening is not reasonable.

Gestational age

Harrington et al. [10] report the induction of labour after 41 weeks of gestation as a routine procedure in most of the obstetric units in the UK. Induction of labour is associated with longer inpatient stay, a shorter but more intensively monitored time on the delivery ward and a higher intervention rate (Harrington refer to [35]). The estimation of gestational age based on LMP has been shown to be unreliable [36], [37], [38], dating by LMP leads to a too early estimation of EDD. Ultrasound dating lead to up to a 70% reduction in the number of pregnancies considered post-term [39], [40]. Crowther et al. [41] found that 24% of women had adjustment of EDD because of a discrepancy of ten or more days from based on estimation LMP. Ewigman et al. found no benefit from routine ultrasound in relation to labour-induction in a study from 1990 (n=2171) [42]. Two Cochrane reviews [43], [44], show that accurate calculation of gestational age by early ultrasound and subsequent adjustment of EDD reduces the incidence of women requiring induction of labour for apparently post-term or prolonged pregnancy. The NICE Guidance for *Routine care for the Healthy Pregnant Woman* [8] recommends *early ultrasound* for ascertainment of gestational age based on *Category I Evidence*. The cited references (Chapter 4.6 Gestational age

assessment, Page 144ff) are based on ultrasound scans at <24 weeks (Neilson 1999), <17 weeks (Crowther 1999), 24 to 29 weeks (Savitz et al. 2003), 18 weeks (Tunon et al. 1996), <20 weeks (Backe and Nakling 1994) and 16-18 weeks (Blondel et al. 2002).

Kalish et al. [45] report in their review that foetal assessment by gestation sac measurement has a prediction error up to two weeks (Warren 1989) CRL allows a precise and rapid gestational age calculation (Daya 1993; Wisser 1994) with small systematic and random error components and an absolute error of less than three days (Kalish 2004). For second trimester gestational age measurement Kalish describe ultrasound as a useful and reliable tool. Multiple parameters have been shown to improve the accuracy of gestational age assessment (Chervenak 1998; Kalish 2004). The accuracy in the third trimester is not as reliable.

Demianczuk et al. [46] report that accurate dating has been the strongest argument for routine early ultrasound (referring to Neilson et al. [43]; ultrasound before 24 weeks). Crown-rump length at 8 to 12 weeks is the most accurate method to date pregnancy; it will predict the expected date of birth within 5 days (2 standard deviations) (referring to Selbing [47]). Accurate dating decreases the number of labour inductions for post-term pregnancy and is important to determine the timing of planned Caesarean sections to prevent iatrogenic prematurity (referring to Crowley [44] and Mongelli [48]). Accurate dating is also important to assess foetal growth and interpret maternal serum screening (referring to SOGC 1997 [49]). For women who have regular menstrual cycles and who have not used oral contraceptives just prior to pregnancy, ultrasound dating may be less important (referring to Olsen [50]). In these circumstances, ultrasound at 18 to 20 weeks will allow for gestational age confirmation. Demianczuk et al. (SOGC Clinical Practice Guidelines) do not recommend first trimester ultrasound to diagnose pregnancy, to date pregnancy when last normal menstrual period and physical examination are concordant, or to investigate an inevitable abortion (II-2B). First trimester ultrasound is indicated when last menstrual period date is uncertain (I-A).

Out of 67.424 deliveries registered in inpatient data from PEGASUS in 2005 228 prolonged pregnancies were documented (ICD 10 O 48) and 165 disorders in relation to prolonged pregnancy and high birth weight (ICD 10 P08). The rate of coded prolonged pregnancies out of these data is 0,58% (228+165 of 67.424). As small for date babies (ICD 10 P07) n=3.814 were registered, which gives a rate of 5,6% (3814/67.424). These data lead to the assumption that prolonged pregnancies seem to create a certain fear and therefore are more likely to be induced rather than being awaited.

Gestational diabetes

An Austrian study published in the year 2005 from Leipold et al. [51] examined the correlation between nuchal translucency in first trimester and development of gestational diabetes. NT was significantly associated with CRL, abnormal karyotype, and GDM, women with GDM were significantly older (on average 5 years) and had a higher BMI and a higher HbA_1C compared with the normal glucose tolerance group. Corrected for CRL, GDM/NGT was no longer associated with NT. Interestingly, the study from Narchi et al. [52] found a significant correlation between GDM and the incidence of chromosomal disorders, the incidence of T21 was 3.75 per 1000 infants of mothers with GDM and 1.36 per 1000 infants of mothers with NGT.

Preterm birth

A review from 2006 [53] summarised the evidence for preventive interventions to reduce preterm delivery. PTD is defined in the UK as delivery after 24 completed weeks gestation and before the onset of 37 weeks gestation. PTD affects 6–15% of deliveries and represents a major worldwide health concern [54]. The causes and subgroups associated with PTD include spontaneous preterm labour (31–50%), multiple pregnancy (12–28%), preterm premature rupture of membranes (6–40%), medically indicated, e.g. hypertensive disorders of pregnancy, intrauterine growth restriction, ante partum haemorrhage and chorioamnionitis,

(20–25%), and miscellaneous causes like cervical incompetence or uterine malformation (10%) [55]. PTD accounts for 50–70% of all neonatal morbidity and mortality. Importantly, the earlier the gestation at delivery, the greater the risk of adverse perinatal outcome [56].

Varma et al. [53] have conducted a systematic search and critical appraisal of the literature to identify the evidence for this approach to reduce the rate of PTD and related perinatal morbidity and mortality. Moreover, this review considers health approaches that address all risk factors that affect the entire population of pregnant women, as well as those screening preventative strategies directed only at high-risk asymptomatic women. They found no direct evidence that early pregnancy booking and ultrasound dating (10–13 weeks) would decrease PTD, insufficient evidence of a beneficial reduction in PTD following increased psychosocial support and home visits, preterm delivery education, bed rest, hydration, reducing excess manual labour and psychological stress, and ensuring that BMI is greater than 20 before conception [57], [58].

Varma et al. cite two meta-analyses [59], [60] showing that increased antenatal attendance without specific specialist investigations, such as foetal biophysical or microbiological surveillance, does not reduce the risk of PTD, low birth weight or perinatal mortality in low-risk women.

The largest cervical cerclage trial [61] showed that elective cervical cerclage performed between 12 and 16 weeks gestation, in women at risk of cervical incompetence based on clinical history, reduced the risk of PTD (<34 weeks) but did not reduce perinatal mortality.

Sociological / societal view

In the current Austrian review about *first trimester* screening the included studies for detection of chromosomal anomalies (regardless the ultrasound marker) range from a positive predictive value between 1,7% and 80%, with a mean of 18% (not weighted).

Doing the same calculation as Smith-Bindman et al. [62] this would lead to a risk of 1:1,25 – 1:59 (mean 1:5,5) that a fetus will actually have chromosomal abnormality. Due to the risk of abortion associated with invasive diagnostic testing (1%) 1 out of 5.900 to 1 out of 125 (mean 1 out of 1.800) will lose a healthy foetus. Assuming a risk for invasive testing of 0,5% it will be 1 out of 12.000 to 1 out of 250 (mean 1 out of 3600) losing a healthy foetus with invasive testing.

Using the same cost-analysis as Ritchie et al. [63] and estimating post-delivery costs for undetected anomaly would increase the cost-benefit of the screening 1 strategy (NT scan in first trimester and anomaly scan in second trimester). Costs were defined as monetary costs for the health system, benefit as the best detection rate of disorders. Societal costs were not calculated.

Using anomaly screening in the first trimester as a screening tool in mother-child-booklet contains the threat of lack of information and communication because “it is usual to do the screening”. The women (parents) have to know about the consequence of their decision of termination or not termination after a positive anomaly screening. Difficult decisions would be faced by the persons after an implemented screening examination.

General discussion

Most of the included studies, even if they included a consecutive unselected population of pregnant women, were carried out in hospital centers and not in the outpatient sector, which is the usual way of pregnancy care in first trimester in Austria. Therefore the study populations

may not have been well comparable to the *outpatient sector*, and the effect of training in ultrasound can vary with the number of examined women – a higher number of examinations per doctor can be assumed in a clinical setting compared to outpatient care. Outpatient care in Austria means a specialist in gynaecology and obstetrics working in a practice like a GP.

The variation in the results of different studies for sensitivity and false positive rate can be related to different levels of provider quality, provider experience and provider training. Our results for the extraction of quality details like assessors' experience and technical details of the equipment were not comparable because of a lack in reporting. Another review about ultrasound in pregnancy related to provider experience and equipment quality has been published in Germany recently [64]. The authors found just a few studies which examine the influence of qualification and experience in a direct comparison, and just few studies (primary studies or reviews) that compare the accuracy of different machines directly. It is reported that the use of dynamic cut off values calculated from maternal age and crown-rump-length for NT measurement seem to result in higher detection rates than fixed cut off points (like 3mm). For NT measurement a higher detection rate correlates with higher quality/ experience of the examiner and higher quality of the ultrasound machines. The Germans recommend that ultrasound machines should exceed at least 256 levels of grey, and a quality management for screening programs although a precise method cannot be detected out of the studies included in their review.

One possible bias in our review could be due to publication bias, its effect on the estimated test accuracy data would probably make those less favourable [65].

As another possible bias should be addressed, that three of the authors of this review are employed by the Social Insurance Organisation in Austria which is the main payer in the Austrian outpatient health care system and also one out of three payers for screening like mother-child-booklet examinations. Nowadays ultrasound examinations in the first trimester

are done very often in the outpatient sector *by indication.* If the ultrasound examination in the first trimester is not included into the mother-child-booklet, the Social Insurance pays each ultrasound provided in outpatient sector which is done beside the mother-child-booklet separately. Therefore it is not a question of costs, but a question of evidence to implement a first trimester ultrasound screening. We are confident that this bias would be negligible because the other authors have no direct interests relevant to the Austrian health care system.

The results of this review show the best evidence out of the five PICO questions for the detection of an increased risk for chromosomal anomalies. Further research will be necessary about how to provide diagnostic procedures with lower false positive rates. It seems to be essential to integrate valid patient information about the consequences before providing the examination, and that the decision about the examination has some ethical implications, and therefore the examination should be clearly presented as being optional. Before introducing a first trimester ultrasound examination it should be ensured that there are enough resources to provide the CVS in an appropriate time frame after the positive ultrasound diagnosis to avoid unnecessary time of psychical distress.

Conclusion

First trimester ultrasound may have adequate test accuracy in screening for chromosomal anomalies, with the limitation of possibly high false positive rates and the need for karyotyping as a definitive diagnostic tool.

Although TTTS can theoretically occur at each time in multiple pregnancies, most of them are detected and treated around the 20th gestational week. Ultrasound examination for chorionicity in pregnancies in the first trimester would be relevant only for an extremely small number of pregnancies, and is therefore not a screening matter.

Gestational age should be ascertained before week 20, a recommendation for the ultrasound with this specific aim can not be fixed to the first trimester.

Considering the fact that claim for maternity allowance in Austria needs a calculation of EDD eight to twelve weeks before delivery, a valid estimation of gestational age should be done at least at pregnancy week 28.

Most of the studies addressing the question whether gestational diabetes can be predicted by nuchal fold assessment in first trimester ultrasound dealt with selected populations, which means that just women with diabetes were included. We found only one study fulfilling our inclusion criteria. In our review we could not find any evidence that NT measurement can predict the development of gestational diabetes mellitus in an unselected population.

Screening for prevention of PTD seems to be an ongoing challenge in prenatal care. Different strategies were evaluated, but few of them showed evidence to work. Ultrasound in the first trimester does not have such evidence, it plays some role for cervical length assessment in the second and third trimester, but not in the first trimester.

First trimester ultrasound as a screening tool is

- To be considered for the detection of increased risk of chromosomal anomalies with a pooled sensitivity of 71%, a specificity of 98%, and false positive rates between 5 and 12%
- Useful for the detection of chorionicity in multiple pregnancies, but not for screening
- Useful for the estimation of gestational age, but this can be done up to week 20 and does not have to be provided until week 14
- Not evidence based for detection of gestational diabetes
- Not studied in connection with preterm birth

An additional ultrasound screening in the first trimester of pregnancy should be aimed at avoiding complications for mother and/ or child during pregnancy and delivery.

Only complications due to late termination because of severe chromosomal anomalies could be reduced with an additional ultrasound screening in the first trimester of pregnancy.

For a first trimester ultrasound screening using parameters like NT or nasal bone the examiner needs to be adequately trained.

Recommendations

1. First trimester ultrasound can not displace second trimester organ screening
2. First trimester ultrasound screening needs the informed consent of the woman (parents). Women have to understand the limits and risks
3. If first trimester ultrasound screening is provided, CVS has to be available in an adequate way
4. First trimester ultrasound for detection of chorionicity is not a screening aim
5. First trimester screening should be presented as an option and not an obligation to all women
6. First trimester ultrasound needs to be included into a quality management system to ascertain an adequate training of the examiners

List of abbreviations

AC	Amniocentesis
ASVG	Allgemeines Sozialversicherungsgesetz (General Social Insurance Act)
AUC	Area under the ROC curve
BMI	Body mass index
BUN	Biochemistry and Foetal Nuchal Translucency Screening
CI	Confidence Interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CRD	Centre for Reviews and Dissemination
CRL	Crown rump length
CVS	Chorion villus sampling
DARE	Database of Abstracts of Reviews of Effects
DOR	Diagnostic Odds Ratio
EBHVB	Department for Evidence Based Health Care in Federation of Austrian Social Insurance Institutions
EBM	Evidence Based Medicine
EDD	Expected date of delivery
Et al.	Et altera
Excl.	Exclusive
FGR	Foetal growth rate
FMF	Foetal Medicine Foundation
FN	False negative
FP	False positive
FPR	False positive rate
GDM	gestational diabetes mellitus
GP	General practitioner
HTA	Health Technology Assessment
I^2	Inconsistency
I-A	Evidence recommendation based on required – at least one randomised controlled trial as part of the body of literature of overall good quality and consistency addressing specific recommendation

ICD	International Classification of Diseases
II-2B	Evidence recommendation based on required – availability of well-conducted clinical studies but no randomised clinical trials on the topic of recommendation
Incl.	Inclusive
IQWIG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
IVF	In vitro fertilisation
LILACS	Latin American and Caribbean Health Sciences Literature
LMP	Last menstrual period
LR	Likelihood
MTOP	Medical termination of pregnancy
NB	Nasal bone
NICE	National Institute of Clinical Excellence (England)
NT	Nuchal translucency
OGTT	Oral glucose tolerance test
P	Probability
PAPP-A	pregnancy-associated protein A
PEGASUS	Projekt Elektronisches Gesundheits- Auskunftssystem und Service – Datawarehouse of the Austrian Health Data in Social Insurance
PICO	Patient/ Problem – Intervention – Control/ Comparison – Outcome
PPROM	Preterm premature rupture of membranes
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTD	Preterm delivery
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
QUOROM	Quality of Reporting of Meta-analyses
RCT	Randomised controlled trial
RDOR	Relative diagnostic Odds ratio
ROC	Receiver operating characteristic
RR	Relative risk
SE	Standard error

SOGC	Society of Obstetricians and Gynaecologists of Canada
ß-HCG	Beta - Human Chorionic Gonatotropin
STOP	Surgical termination of pregnancy
TA	Transabdominal
TN	True negative
TP	True positive
T21	Trisomy 21
TTTS	Twin to twin transfusion syndrome
TV	Transvaginal
UK	United Kingdom
US	Ultrasound
USA	United States of America

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

GE participated in the design of the study, created and was responsible for maintenance of the study database. ISF participated in the review process and drafted the abstract. IW was project leader, participated in the design of the study and in the review process, and drafted the protocol and the report together with JK. KSW participated in the review process. JK participated in the design of the study, in the review process and drafted the protocol and the report together with IW.

SP participated in the review process. DB participated in the review process.

All authors read and approved the final manuscript.

Acknowledgements

Andrea Weisser employed by the *Federation of Austrian Social Insurance Institutions,* Andrea.Weisser@hvb.sozvers.at - *no extra funding*

Sources of Funding - Study:

The Federation of Austrian Social Insurance Institutions.

Sources of Funding - Authors:

Dr. Gottfried Endel, employed by the *Federation of Austrian Social Insurance Institutions, no extra funding*

Mag. Ingrid Wilbacher, employed by the *Federation of Austrian Social Insurance Institutions, no extra funding*

Dr. Irmgard Schiller-Fruehwirth, employed by the *Federation of Austrian Social Insurance Institutions, no extra funding*

Dagmar Bernardis employed by the *Federation of Austrian Social Insurance Institutions,* Dagmar.Bernardis@hvb.sozvers.at - *no extra funding*

Dr. Karla Soares-Weiser, *Kleijnen Systematic Reviews Ltd.*

Dr. Jos Kleijnen, *Kleijnen Systematic Reviews Ltd.*

Kleijnen Systematic Reviews Ltd received funding from the Federation of Austrian Social Insurance Institutions according to a contract.

Sources of Founding:

The Federation of Austrian Social Insurance Institutions.

The Federation of Austrian Social Insurance Institutions did not influence the study design, the collection, analysis and interpretation of data, the manuscript writing nor the decision to submit the article.

References

[1] Wilbacher I. Protocol: Diagnostic accuracy and outcomes of ultrasound in the first trimester of pregnancy for the detection of complications relevant for Austrian Population, exclusive of screening for Down syndrome. 2007. Hauptverband der Österreichischen Sozialversicherungsträger, sector Evidence Based Medicine [http://www.hauptverband.at/mediaDB/119419.PDF]
Ref Type: Generic
Ref ID: 1318

[2] Wilbacher I. Report PICO. 2007. Hauptverband der Österreichischen Sozialversicherungsträger, sector Evidence Based Medicine [http://www.hauptverband.at/portal/index.html?ctrl:cmd=render&ctrl:window=hvbportal.channel_content.cmsWindow&p_menuid=65859&p_tabid=5&p_pubid=136131].
Ref Type: Generic
Ref ID: 1316

[3] Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J: **The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews.** *BMC Med Res Methodol* 2003, **3:** 25.
Ref ID: 5

[4] Khan KS, ter Riet G, Popay J, Nixon J, Kleijnen J. Study quality assessment. **The University of York** [http://www.york.ac.uk/inst/crd/pdf/crd4_ph5.pdf]
Ref Type: Generic
Ref ID: 1321

[5] Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A: **Meta-DiSc: a software for meta-analysis of test accuracy data.** *BMC Med Res Methodol* 2006, **6:** 31.
Ref ID: 1

[6] Moses LE, Shapiro D, Littenberg B: **Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations.** *Stat Med* 1993, **12:** 1293-1316.
Ref ID: 10

[7] Menon DK: **A retrospective study of the accuracy of sonographic chorionicity determination in twin pregnancies.** *Twin Research & Human Genetics: the Official Journal of the International Society for Twin Studies Vol 8(3)()(pp 259-261), 2005* 2005, 259-261.
Ref ID: 1282

[8] Antenatal Care: Routine Care for the Healthy Prgenant Woman. National Collaborating Centre for Women's and Children's Health. 2003. National Institute for Clinical Excellence.
Ref Type: Generic
Ref ID: 1319

[9] Bennett KA, Crane JM, O'shea P, Lacelle J, Hutchens D, Copel JA: **First trimester ultrasound screening is effective in reducing postterm labor induction rates: a randomized controlled trial.[see comment].** *American Journal of Obstetrics & Gynecology* 2004, **190:** 1077-1081.

Ref ID: 95

[10] Harrington DJ, MacKenzie IZ, Thompson K, Fleminger M, Greenwood C: **Does a first trimester dating scan using crown rump length measurement reduce the rate of induction of labour for prolonged pregnancy? An uncompleted randomised controlled trial of 463 women.** *BJOG: An International Journal of Obstetrics & Gynaecology Vol 113(2)()(pp 171-176), 2006* 2006, 171-176.
Ref ID: 49

[11] Kelekci S, Yilmaz B, Savan K, Sonmez S: **Can increased nuchal translucency in the first trimester of pregnancy predict gestational diabetes mellitus.** *Journal of Obstetrics & Gynaecology Vol 25(6)()(pp 579-582), 2005* 2005, 579-582.
Ref ID: 57

[12] Antonetti DA, Barber AJ, Khin S, Lieth E, Tarbell JM, Gardner TW: **Vascular permeability in experimental diabetes is associated with reduced endothelial occludin content: vascular endothelial growth factor decreases occludin in retinal endothelial cells. Penn State Retina Research Group.** *Diabetes* 1998, **47:** 1953-1959.
Ref ID: 1314

[13] Brausewetter F, Jehle PM, Jung MF, Boehm BO, Brueckel J, Hombach V *et al.*: **Microvascular permeability is increased in both types of diabetes and correlates differentially with serum levels of insulin-like growth factor I (IGF-I) and vascular endothelial growth factor (VEGF).** *Horm Metab Res* 2001, **33:** 713-720.
Ref ID: 1312

[14] Bindra R, Heath V, Nicolaides KH: **Screening for chromosomal defects by fetal nuchal translucency at 11 to 14 weeks. [Review] [69 refs].** *Clinical Obstetrics & Gynecology* 730, **45:** 661-670.
Ref ID: 429

[15] Taipale P, Hiilesmaa V: **Ultrasonographic screening for fetal chromosomal abnormalities in early pregnancy.** *Acta Obstetricia et Gynecologica Scandinavica Vol 78(7)()(pp 563-572), 1999* 1999, 563-572.
Ref ID: 1145

[16] Nicolaides KH: **Screening for chromosomal defects.** *Ultrasound in Obstetrics & Gynecology* 2003, **21:** 313-321.
Ref ID: 364

[17] Stewart TL, Malone FD: **First trimester screening for aneuploidy: nuchal translucency sonography. [Review] [55 refs].** *Seminars in Perinatology* 1999, **23:** 369-381.
Ref ID: 753

[18] Sherer DM, Manning FA: **First-trimester nuchal translucency screening for fetal aneuploidy. [Review] [85 refs].** *American Journal of Perinatology* 1999, **16:** 103-120.
Ref ID: 703

[19] Nicolaides KH: **Nuchal translucency and other first-trimester sonographic markers of chromosomal abnormalities. [Review] [218 refs].** *American Journal of Obstetrics & Gynecology* 2004, **191:** 45-67.
Ref ID: 211

[20] Beke A: **Ultrasound markers which indicate chromosome analysis.** *Ultrasound Review of Obstetrics & Gynecology Vol 3(4)()(pp 259-272), 2003* 2003, 259-272.
Ref ID: 557

[21] Brambati B, Tului L, Alberti E: **Sonography in the first trimester screening of trisomy 21 and other fetal aneuploidies. [Review] [73 refs].** *Early Pregnancy* 1996, **2:** 155-167.
Ref ID: 925

[22] Chitty LS, Pandya PP: **Ultrasound screening for fetal abnormalities in the first trimester.** *Prenatal Diagnosis Vol 17(13)()(pp 1269-1281), 1997* 1997, 1269-1281.
Ref ID: 1281

[23] Snijders R: **First-trimester ultrasound.** *Clin Perinatol* 2001, **28:** 333-52, viii.
Ref ID: 89

[24] Monteagudo A: **Sonographic assessment of chorionicity and amnionicity in twin pregnancies: How, when and why?** *Croatian Medical Journal Vol 39(2)()(pp 191-196), 1998* 1998, 191-196.
Ref ID: 1237

[25] Huber A, Hecher K: **How can we diagnose and manage twin-twin transfusion syndrome?** *Best Pract Res Clin Obstet Gynaecol* 2004, **18:** 543-556.
Ref ID: 1301

[26] Pedreira DA, Acacio GL, Drummond CL, Oliveira RC, Deustch AD, Taborda WG: **Laser for the treatment of twin to twin transfusion syndrome.** *Acta Cir Bras* 2005, **20:** 478-481.
Ref ID: 1290

[27] Robyr R, Quarello E, Ville Y: **Management of fetofetal transfusion syndrome.** *Prenat Diagn* 2005, **25:** 786-795.
Ref ID: 1289

[28] Ierullo AM, Papageorghiou AT, Bhide A, Fratelli N, Thilaganathan B: **Severe twin-twin transfusion syndrome: outcome after fetoscopic laser ablation of the placental vascular equator.** *BJOG* 2007, **114:** 689-693.
Ref ID: 1300

[29] Middeldorp JM, Sueters M, Lopriore E, Klumper FJ, Oepkes D, Devlieger R *et al*.: **Fetoscopic laser surgery in 100 pregnancies with severe twin-to-twin transfusion syndrome in the Netherlands.** *Fetal Diagn Ther* 2007, **22:** 190-194.
Ref ID: 1298

[30] Huber A, Diehl W, Bregenzer T, Hackeloer BJ, Hecher K: **Stage-related outcome in twin-twin transfusion syndrome treated by fetoscopic laser coagulation.** *Obstet Gynecol* 2006, **108:** 333-337.
Ref ID: 1302

[31] Gray PH, Cincotta R, Chan FY, Soong B: **Perinatal outcomes with laser surgery for twin-twin transfusion syndrome.** *Twin Res Hum Genet* 2006, **9:** 438-443.
Ref ID: 1303

[32] Duncan KR: **Twin-to-twin transfusion: update on management options and outcomes.** *Curr Opin Obstet Gynecol* 2005, **17:** 618-622.
Ref ID: 1307

[33] Middeldorp JM, Lopriore E, Sueters M, Klumper FJ, Kanhai HH, Vandenbussche FP *et al*.: **Twin-to-twin transfusion syndrome after 26 weeks of gestation: is there a role for fetoscopic laser surgery?** *BJOG* 2007, **114:** 694-698.
Ref ID: 1297

[34] Lenclen R, Paupe A, Ciarlo G, Couderc S, Castela F, Ortqvist L, Ville Y: **Neonatal outcome in preterm monochorionic twins with twin-to-twin transfusion syndrome after intrauterine treatment with amnioreduction or fetoscopic laser surgery: comparison with dichorionic twins. [see comment]**. *Am J Obstet Gynecol* 2007, **196(5):** 450.e1-7.

[35] Yudkin P, Frumar AM, Anderson AB, Turnbull AC: **A retrospective study of induction of labour.** *Br J Obstet Gynaecol* 1979, **86:** 257-265.
Ref ID: 1283

[36] Campbell S, Warsof SL, Little D, Cooper DJ: **Routine ultrasound screening for the prediction of gestational age.** *Obstet Gynecol* 1985, **65:** 613-620.
Ref ID: 1311

[37] Waldenstrom U, Axelsson O, Nilsson S: **A comparison of the ability of a sonographically measured biparietal diameter and the last menstrual period to predict the spontaneous onset of labor.** *Obstet Gynecol* 1990, **76:** 336-338.
Ref ID: 18

[38] Geirsson RT, Busby-Earle RM: **Certain dates may not provide a reliable estimate of gestational age.** *Br J Obstet Gynaecol* 1991, **98:** 108-109.
Ref ID: 1304

[39] Gardosi J, Vanner T, Francis A: **Gestational age and induction of labour for prolonged pregnancy.** *Br J Obstet Gynaecol* 1997, **104:** 792-797.
Ref ID: 1305

[40] Mongelli M, Wilcox M, Gardosi J: **Estimating the date of confinement: ultrasonographic biometry versus certain menstrual dates.** *Am J Obstet Gynecol* 1996, **174:** 278-281.
Ref ID: 1296

[41] Crowther CA, Kornman L, O'Callaghan S, George K, Furness M, Willson K: **Is an ultrasound assessment of gestational age at the first antenatal visit of value? A randomised clinical trial.[see comment].** *British Journal of Obstetrics & Gynaecology* 1999, **106:** 1273-1279.
Ref ID: 267

[42] Ewigman B, LeFevre M, Hesser J: **A randomized trial of routine prenatal ultrasound.** *Obstet Gynecol* 1990, **76:** 189-194.
Ref ID: 1306

[43] Neilson JP: **Ultrasound for fetal assessment in early pregnancy.** *Cochrane Database Syst Rev* 2000, CD000182.

Ref ID: 1292

[44] Crowley P: **Interventions for preventing or improving the outcome of delivery at or beyond term.** *Cochrane Database Syst Rev* 2000, CD000170.
Ref ID: 1308

[45] Kalish RB, Chervenak FA: **Sonographic determination of gestational age.** *Ultrasound Review of Obstetrics & Gynecology Vol 5(4)()(pp 254-258), 2005* 2005, 254-258.
Ref ID: 68

[46] Demianczuk NN, Van den Hof MC, Farquharson D, Lewthwaite B, Gagnon R, Morin L *et al*.: **The use of first trimester ultrasound.[see comment].** *Journal of Obstetrics & Gynaecology Canada: JOGC* 2003, **25:** 864-875.
Ref ID: 316

[47] Selbing A, Fjallbrant B: **Accuracy of conceptual age estimation from fetal crown-rump length.** *J Clin Ultrasound* 1984, **12:** 343-346.
Ref ID: 1286

[48] Mongelli M, Wong YC, Venkat A, Chua TM: **Induction policy and missed post-term pregnancies: a mathematical model.** *Aust N Z J Obstet Gynaecol* 2001, **41:** 38-40.
Ref ID: 1295

[49] Members of the SOGC Genetics Committee and Diagnostic Imaging Committee.. The use of ultrasound in conjunction with maternal serum screening for aneuploidy. Committee Opinion. No 17. 1997.
Ref Type: Generic
Ref ID: 1320

[50] Olsen O, Aaroe CJ: **Routine ultrasound dating has not been shown to be more accurate than the calendar method.** *Br J Obstet Gynaecol* 1997, **104:** 1221-1222.
Ref ID: 1291

[51] Leipold H, Worda C, Ozbal A, Husslein P, Krampl E: **First-trimester nuchal translucency screening in pregnant women who subsequently developed gestational diabetes.** *Journal of the Society for Gynecologic Investigation* 2005, **12:** 529-532.
Ref ID: 30

[52] Narchi H, Kulaylat N: **High incidence of Down's syndrome in infants of diabetic mothers.** *Arch Dis Child* 1997, **77:** 242-244.
Ref ID: 1293

[53] Varma R, Gupta JK, James DK, Kilby MD: **Do screening-preventative interventions in asymptomatic pregnancies reduce the risk of preterm delivery-A critical appraisal of the literature.** *European Journal of Obstetrics, Gynecology, & Reproductive Biology Vol 127(2)()(pp 145-159), 2006* 2006, 145-159.
Ref ID: 3

[54] Moutquin JM: **Classification and heterogeneity of preterm birth.** *BJOG* 2003, **110 Suppl 20:** 30-33.
Ref ID: 4

[55] Slattery MM, Morrison JJ: **Preterm delivery.** *Lancet* 2002, **360:** 1489-1497.
Ref ID: 1285

[56] Ward RM, Beachy JC: **Neonatal complications following preterm birth.** *BJOG* 2003, **110 Suppl 20:** 8-16.
Ref ID: 1284

[57] Berkman ND, Thorp JM, Jr., Hartmann KE, Lohr KN, Idicula AE, McPheeters M *et al*.: **Management of preterm labor.** *Evid Rep Technol Assess (Summ)* 2000, 1-6.
Ref ID: 1313

[58] Lu MC, Tache V, Alexander GR, Kotelchuck M, Halfon N: **Preventing low birth weight: is prenatal care the answer?** *J Matern Fetal Neonatal Med* 2003, **13:** 362-380.
Ref ID: 1299

[59] Carroli G, Villar J, Piaggio G, Khan-Neelofur D, Gulmezoglu M, Mugford M *et al*.: **WHO systematic review of randomised controlled trials of routine antenatal care.** *Lancet* 2001, **357:** 1565-1570.
Ref ID: 1309

[60] Villar J, Carroli G, Khan-Neelofur D, Piaggio G, Gulmezoglu M. Patterns of routine antenatal care for low-risk pregnancy. Cochrane Database Syst Rev 2005.
Ref Type: Generic
Ref ID: 1315

[61] **Final report of the Medical Research Council/Royal College of Obstetricians and Gynaecologists multicentre randomised trial of cervical cerclage. MRC/RCOG Working Party on Cervical Cerclage.** *Br J Obstet Gynaecol* 1993, **100:** 516-523.
Ref ID: 1294

[62] Smith-Bindman R, Hosmer W, Feldstein VA, Deeks JJ, Goldberg JD: **Second-trimester ultrasound to detect fetuses with Down syndrome: a meta-analysis.** *JAMA* 2001, **285:** 1044-1055.
Ref ID: 2

[63] Ritchie K, Bradbury I, Slattery J, Wright D, Iqbal K, Penney G: **Economic modelling of antenatal screening and ultrasound scanning programmes for identification of fetal abnormalities.** *BJOG: An International Journal of Obstetrics & Gynaecology* 2005, **112:** 866-874.
Ref ID: 109

[64] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Ultraschallscreening in der Schwangerschaft: Testgüte hinsichtlich der Entdeckungsrate fetaler Anomalien. Vorbericht S05-03. Köln: IQWiG; 2007.

[65] Scherer RW, Langenberg P, von Elm E: **Full publication of results initially presented in abstracts.** *Cochrane Database Syst Rev* 2007, MR000005.
Ref ID: 1287

List of excluded potential relevant studies

[66] Beke A, Papp C, Toth-Pal E, Mezei G, Joo JG, Csaba A *et al*.: **Trisomies and other chromosome abnormalities detected after positive sonographic findings.** *J Reprod Med* 2005, **50:** 675-691.
Ref ID: 33

[67] Cheng CC, Bahado-Singh RO, Chen SC, Tsai MS: **Pregnancy outcomes with increased nuchal translucency after routine Down syndrome screening.** *Int J Gynaecol Obstet* 2004, **84:** 5-9.
Ref ID: 26

[68] Chitty LS, Kagan KO, Molina FS, Waters JJ, Nicolaides KH: **Fetal nuchal translucency scan and early prenatal diagnosis of chromosomal abnormalities by rapid aneuploidy screening: observational study.** *BMJ* 2006, **332:** 452-455.
Ref ID: 12

[69] Conoscenti G, Meir YJ, D'Ottavio G, Rustico MA, Pinzano R, Fischer-Tamaro L *et al*.: **Does cervical length at 13-15 weeks' gestation predict preterm delivery in an unselected population?** *Ultrasound Obstet Gynecol* 2003, **21:** 128-134.
Ref ID: 8

[70] Crowther CA, Kornman L, O'Callaghan S, George K, Furness M, Willson K: **Is an ultrasound assessment of gestational age at the first antenatal visit of value? A randomised clinical trial.** *Br J Obstet Gynaecol* 1999, **106:** 1273-1279.
Ref ID: 22

[71] 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.** *J Obstet Gynaecol* 2002, **22:** 159-165.
Ref ID: 17

[72] Economides DL, Whitlow BJ, Kadir R, Lazanakis M, Verdin SM: **First trimester sonographic detection of chromosomal abnormalities in an unselected population.** *Br J Obstet Gynaecol* 1998, **105:** 58-62.
Ref ID: 14

[73] Ghezzi F, Raio L, Di Naro E, Franchi M, Buttarelli M, Schneider H: **First-trimester umbilical cord diameter: a novel marker of fetal aneuploidy.** *Ultrasound Obstet Gynecol* 2002, **19:** 235-239.
Ref ID: 21

[74] Gonzalez R C. Diagnóstico prenatal invasivo al final del primer trimestre: resultados de un ensayo internacional randomizado. Vera C, Carvajal C J. Rev.chil.obstet.ginecol 69(3): 263-265. 2004.
Ref Type: Generic
Ref ID: 36

[75] Hewitt BG, de Crespigny L, Sampson AJ, Ngu AC, Shekleton P, Robinson HP: **Correlation between nuchal thickness and abnormal karyotype in first trimester fetuses.** *Med J Aust* 1996, **165:** 365-368.

Ref ID: 7

[76] Krantz DA, Hallahan TW, Orlandi F, Buchanan P, Larsen JW, Jr., Macri JN: **First-trimester Down syndrome screening using dried blood biochemistry and nuchal translucency.** *Obstet Gynecol* 2000, **96:** 207-213.
Ref ID: 15

[77] Lewis PR, Pasalio R: **First-trimester tests for trisomies 21 and 18 as sensitive as triple screen.** *J Fam Pract* 2004, **53:** 184-186.
Ref ID: 20

[78] Malone FD, Ball RH, Nyberg DA, Comstock CH, Saade G, Berkowitz RL *et al*.: **First-trimester nasal bone evaluation for aneuploidy in the general population.** *Obstet Gynecol* 2004, **104:** 1222-1228.
Ref ID: 16

[79] Malone FD, Ball RH, Nyberg DA, Comstock CH, Saade GR, Berkowitz RL *et al*.: **First-trimester septated cystic hygroma: prevalence, natural history, and pediatric outcome.** *Obstet Gynecol* 2005, **106:** 288-294.
Ref ID: 19

[80] Mustafa SA, Brizot ML, Carvalho MH, Watanabe L, Kahhale S, Zugaib M: **Transvaginal ultrasonography in predicting placenta previa at delivery: a longitudinal study.** *Ultrasound Obstet Gynecol* 2002, **20:** 356-359.
Ref ID: 32

[81] O'Leary P, Breheny N, Dickinson JE, Bower C, Goldblatt J, Hewitt B *et al*.: **First-trimester combined screening for Down syndrome and other fetal anomalies.** *Obstet Gynecol* 2006, **107:** 869-876.
Ref ID: 3

[82] Orlandi F, Damiani G, Hallahan TW, Krantz DA, Macri JN: **First-trimester screening for fetal aneuploidy: biochemistry and nuchal translucency.** *Ultrasound Obstet Gynecol* 1997, **10:** 381-386.
Ref ID: 18

[83] Panburana P, Ajjimakorn S, Tungkajiwangoon P: **First trimester Down Syndrome screening by nuchal translucency in a Thai population.** *Int J Gynaecol Obstet* 2001, **75:** 311-312.
Ref ID: 2

[84] Papp C, Beke A, Mezei G, Szigeti Z, Ban Z, Papp Z: **Prenatal diagnosis of Turner syndrome: report on 69 cases.** *J Ultrasound Med* 2006, **25:** 711-717.
Ref ID: 27

[85] Peralta CF, Falcon O, Wegrzyn P, Faro C, Nicolaides KH: **Assessment of the gap between the fetal nasal bones at 11 to 13 + 6 weeks of gestation by three-dimensional ultrasound.** *Ultrasound Obstet Gynecol* 2005, **25:** 464-467.
Ref ID: 5

[86] Rosati P, Guariglia L: **Prognostic value of ultrasound findings of fetal cystic hygroma detected in early pregnancy by transvaginal sonography.** *Ultrasound Obstet*

Gynecol 2000, **16:** 245-250.
Ref ID: 28

[87] 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* 2006, **113:** 664-674.
Ref ID: 4

[88] Schouwink MH, Fong BF, Mol BW, van d, V: **Ultrasonographic criteria for non-viability of first trimester intra-uterine pregnancy.** *Early Pregnancy* 2000, **4:** 203-213.
Ref ID: 34

[89] Sepulveda W, Sebire NJ, Hughes K, Odibo A, Nicolaides KH: **The lambda sign at 10-14 weeks of gestation as a predictor of chorionicity in twin pregnancies.** *Ultrasound Obstet Gynecol* 1996, **7:** 421-423.
Ref ID: 30

[90] Sladkevicius P, Saltvedt S, Almstrom H, Kublickas M, Grunewald C, Valentin L: **Ultrasound dating at 12-14 weeks of gestation. A prospective cross-validation of established dating formulae in in-vitro fertilized pregnancies.** *Ultrasound Obstet Gynecol* 2005, **26:** 504-511.
Ref ID: 35

[91] Souka AP, Krampl E, Bakalis S, Heath V, Nicolaides KH: **Outcome of pregnancy in chromosomally normal fetuses with increased nuchal translucency in the first trimester.** *Ultrasound Obstet Gynecol* 2001, **18:** 9-17.
Ref ID: 24

[92] Spencer K, Spencer CE, Power M, Dawson C, Nicolaides KH: **Screening for chromosomal abnormalities in the first trimester using ultrasound and maternal serum biochemistry in a one-stop clinic: a review of three years prospective experience.** *BJOG* 2003, **110:** 281-286.
Ref ID: 29

[93] Srisupundit K, Tongsong T, Sirichotiyakul S, Chanprapaph P: **Fetal structural anomaly screening at 11-14 weeks of gestation at Maharaj Nakorn Chiang Mai Hospital.** *J Med Assoc Thai* 2006, **89:** 588-593.
Ref ID: 13

[94] Taipale P, Hiilesmaa V: **Predicting delivery date by ultrasound and last menstrual period in early gestation.** *Obstet Gynecol* 2001, **97:** 189-194.
Ref ID: 25

[95] Taipale P, Ammala M, Salonen R, Hiilesmaa V: **Learning curve in ultrasonographic screening for selected fetal structural anomalies in early pregnancy.** *Obstet Gynecol* 2003, **101:** 273-278.
Ref ID: 23

[96] Thilaganathan B, Slack A, Wathen NC: **Effect of first-trimester nuchal translucency on second-trimester maternal serum biochemical screening for Down's syndrome.** *Ultrasound Obstet Gynecol* 1997, **10:** 261-264.

Ref ID: 11

[97] van Bogaert LJ: **Accuracy of menstrual history in early pregnancy.** *S Afr Med J* 2003, **93:** 678-679.
Ref ID: 1

[98] Whitlow BJ, Lazanakis ML, Kadir RA, Chatzipapas I, Economides DL: **The significance of choroid plexus cysts, echogenic heart foci and renal pyelectasis in the first trimester.** *Ultrasound Obstet Gynecol* 1998, **12:** 385-390.
Ref ID: 31

[99] Zoppi MA, Ibba RM, Floris M, Manca F, Axiana C, Monni G: **Changes in nuchal translucency thickness in normal and abnormal karyotype fetuses.** *BJOG* 2003, **110:** 584-588.
Ref ID: 6

List of included studies

[99] Acacio GL, Barini R, Pinto JW, Ximenes RL, Pettersen H, Faria M: **Nuchal translucency: an ultrasound marker for fetal chromosomal abnormalities.** *Sao Paulo Med J* 2001, **119:** 19-23.
Ref ID: 1

[100] Bennett KA, Crane JM, O'shea P, Lacelle J, Hutchens D, Copel JA: **First trimester ultrasound screening is effective in reducing postterm labor induction rates: a randomized controlled trial.** *Am J Obstet Gynecol* 2004, **190:** 1077-1081.
Ref ID: 2

[101] Blazer S, Zimmer EZ, Gover A, Bronshtein M: **Fetal omphalocele detected early in pregnancy: associated anomalies and outcomes.** *Radiology* 2004, **232:** 191-195.
Ref ID: 3

[102] Brizot ML, Carvalho MH, Liao AW, Reis NS, Armbruster-Moraes E, Zugaib M: **First-trimester screening for chromosomal abnormalities by fetal nuchal translucency in a Brazilian population.** *Ultrasound Obstet Gynecol* 2001, **18:** 652-655.
Ref ID: 4

[103] Cheng PJ, Shaw SW, Soong YK: **Association of fetal choroid plexus cysts with trisomy 18 in a population previously screened by nuchal translucency thickness measurement.** *J Soc Gynecol Investig* 2006, **13:** 280-284.
Ref ID: 5

[104] Cicero S, Longo D, Rembouskos G, Sacchini C, Nicolaides KH: **Absent nasal bone at 11-14 weeks of gestation and chromosomal defects.** *Ultrasound Obstet Gynecol* 2003, **22:** 31-35.
Ref ID: 6

[105] D'Alton M, Cleary-Goldman J: **First and second trimester evaluation of risk for fetal aneuploidy: the secondary outcomes of the FASTER Trial.** *Semin Perinatol* 2005, **29:**

240-246.
Ref ID: 7

[106] D'Ottavio G, Meir YJ, Rustico MA, Pecile V, Fischer-Tamaro L, Conoscenti G *et al*.: **Screening for fetal anomalies by ultrasound at 14 and 21 weeks.** *Ultrasound Obstet Gynecol* 1997, **10:** 375-380.
Ref ID: 9

[107] D'Ottavio G, Mandruzzato G, Meir YJ, Rustico MA, Fischer-Tamaro L, Conoscenti G *et al*.: **Comparisons of first and second trimester screening for fetal anomalies.** *Ann N Y Acad Sci* 1998, **847:** 200-209.
Ref ID: 8

[108] Hafner E, Schuchter K, Liebhart E, Philipp K: **Results of routine fetal nuchal translucency measurement at weeks 10-13 in 4233 unselected pregnant women.** *Prenat Diagn* 1998, **18:** 29-34.
Ref ID: 10

[109] Harrington DJ, MacKenzie IZ, Thompson K, Fleminger M, Greenwood C: **Does a first trimester dating scan using crown rump length measurement reduce the rate of induction of labour for prolonged pregnancy? An uncompleted randomised controlled trial of 463 women.** *BJOG* 2006, **113:** 171-176.
Ref ID: 11

[110] Kelekci S, Yilmaz B, Savan K, Sonmez S: **Can increased nuchal translucency in the first trimester of pregnancy predict gestational diabetes mellitus.** *J Obstet Gynaecol* 2005, **25:** 579-582.
Ref ID: 12

[111] Menon DK: **A retrospective study of the accuracy of sonographic chorionicity determination in twin pregnancies.** *Twin Res Hum Genet* 2005, **8:** 259-261.
Ref ID: 13

[112] Metzenbauer M, Hafner E, Schuchter K, Philipp K: **First-trimester placental volume as a marker for chromosomal anomalies: preliminary results from an unselected population.** *Ultrasound Obstet Gynecol* 2002, **19:** 240-242.
Ref ID: 14

[113] Monni G, Zoppi MA, Ibba RM, Floris M, Manca F, Axiana C: **Nuchal translucency and nasal bone for trisomy 21 screening: single center experience.** *Croat Med J* 2005, **46:** 786-791.
Ref ID: 15

[114] O'Callaghan SP, Giles WB, Raymond SP, McDougall V, Morris K, Boyd J: **First trimester ultrasound with nuchal translucency measurement for Down syndrome risk estimation using software developed by the Fetal Medicine Foundation, United Kingdom--the first 2000 examinations in Newcastle, New South Wales, Australia.** *Aust N Z J Obstet Gynaecol* 2000, **40:** 292-295.
Ref ID: 16

[115] Orlandi F, Bilardo CM, Campogrande M, Krantz D, Hallahan T, Rossi C *et al*.: **Measurement of nasal bone length at 11-14 weeks of pregnancy and its potential role in**

Down syndrome risk assessment. *Ultrasound Obstet Gynecol* 2003, **22:** 36-39.
Ref ID: 17

[116] Schwarzler P, Carvalho JS, Senat MV, Masroor T, Campbell S, Ville Y: **Screening for fetal aneuploidies and fetal cardiac abnormalities by nuchal translucency thickness measurement at 10-14 weeks of gestation as part of routine antenatal care in an unselected population.** *Br J Obstet Gynaecol* 1999, **106:** 1029-1034.
Ref ID: 18

[117] Snijders RJ, Noble P, Sebire N, Souka A, Nicolaides KH: **UK multicentre project on assessment of risk of trisomy 21 by maternal age and fetal nuchal-translucency thickness at 10-14 weeks of gestation. Fetal Medicine Foundation First Trimester Screening Group.** *Lancet* 1998, **352:** 343-346.
Ref ID: 19

[118] Tercanli S, Holzgreve W, Batukan C, Gerber A, Ermis H, Miny P: **[Screening for aneuploidy by first trimester nuchal translucency measurement: results from a prospective trial including 1980 cases in a single center in Switzerland].** *Ultraschall Med* 2002, **23:** 22-26.
Ref ID: 20

[119] Thilaganathan B, Sairam S, Michailidis G, Wathen NC: **First trimester nuchal translucency: effective routine screening for Down's syndrome.** *Br J Radiol* 1999, **72:** 946-948.
Ref ID: 21

[120] Viora E, Masturzo B, Errante G, Sciarrone A, Bastonero S, Campogrande M: **Ultrasound evaluation of fetal nasal bone at 11 to 14 weeks in a consecutive series of 1906 fetuses.** *Prenat Diagn* 2003, **23:** 784-787.
Ref ID: 22

[121] Wayda K, Kereszturi A, Orvos H, Horvath E, PAl A, Kovacs L *et al*.: **Four years experience of first-trimester nuchal translucency screening for fetal aneuploidies with increasing regional availability.** *Acta Obstet Gynecol Scand* 2001, **80:** 1104-1109.
Ref ID: 23

[122] Wojdemann KR, Shalmi AC, Christiansen M, Larsen SO, Sundberg K, Brocks V *et al*.: **Improved first-trimester Down syndrome screening performance by lowering the false-positive rate: a prospective study of 9941 low-risk women.** *Ultrasound Obstet Gynecol* 2005, **25:** 227-233.
Ref ID: 24

[123] Zoppi MA, Ibba RM, Axiana C, Floris M, Manca F, Monni G: **Absence of fetal nasal bone and aneuploidies at first-trimester nuchal translucency screening in unselected pregnancies.** *Prenat Diagn* 2003, **23:** 496-500.
Ref ID: 25

Figure legends

Figure 1

Flow chart of studies through review process

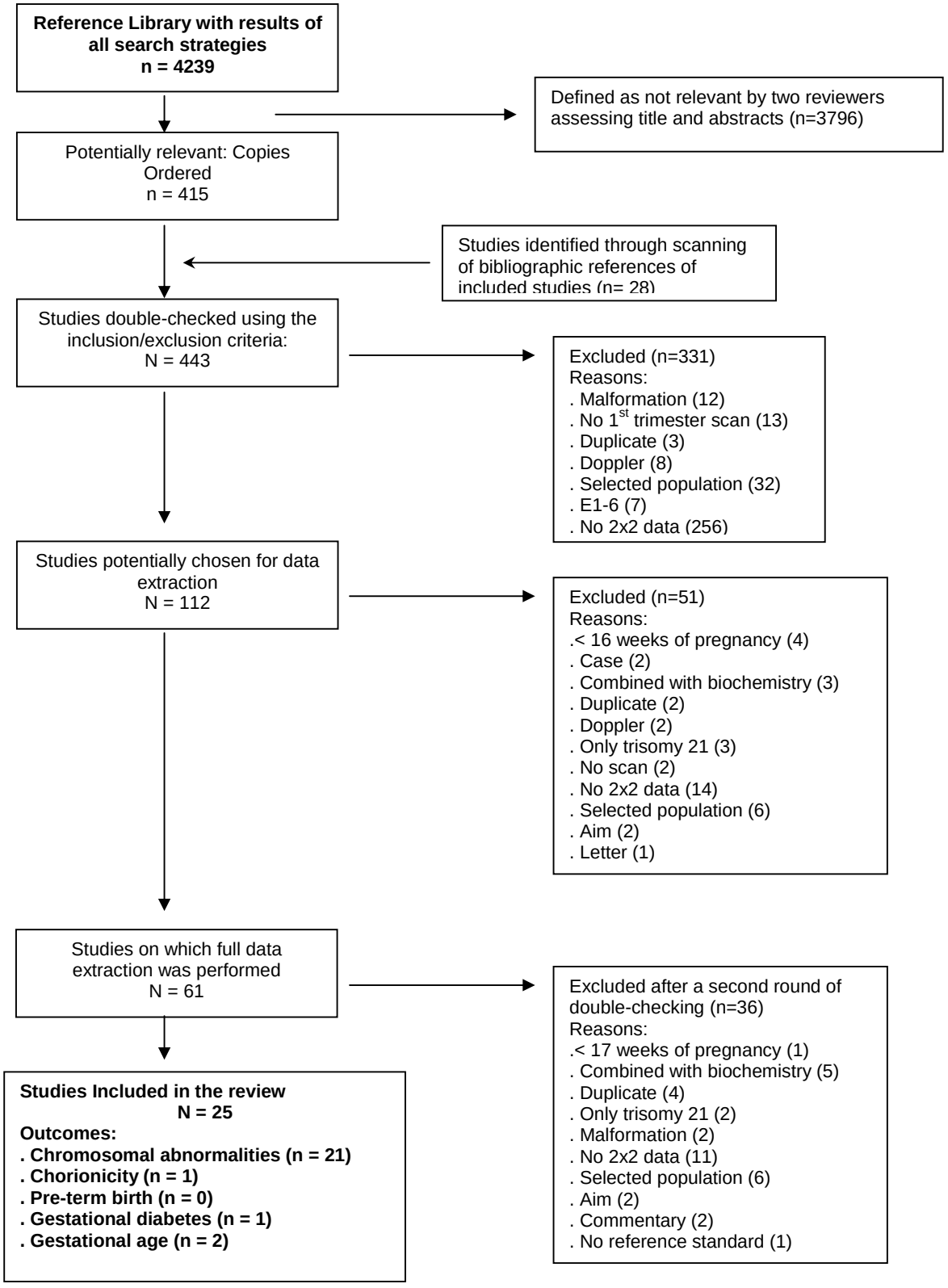


Figure 2

Proportion of studies rated as yes, no, or unclear for each of the QUADAS items

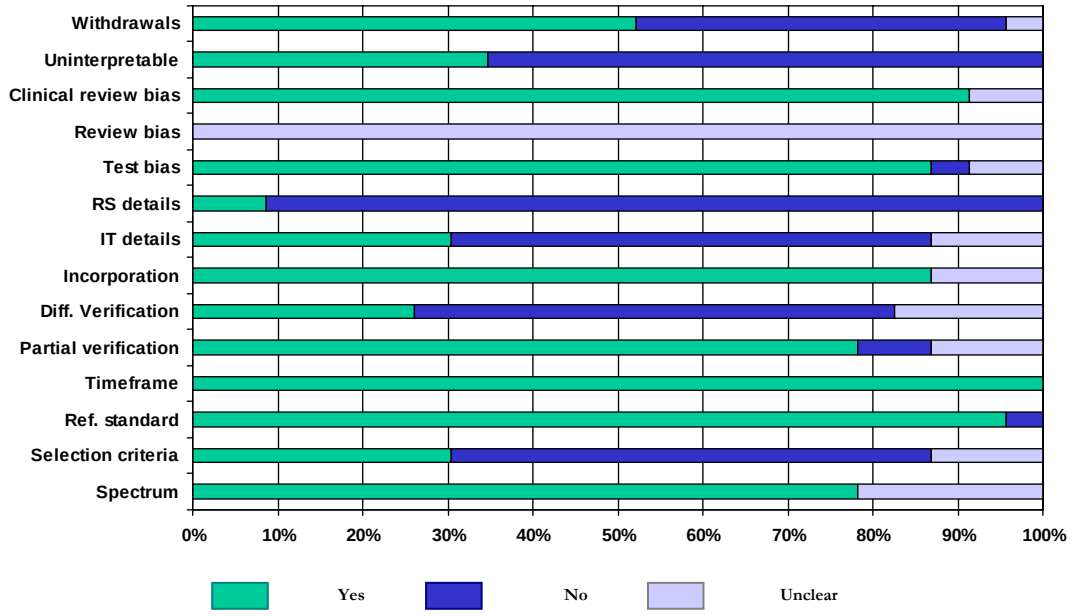


Figure 3

Proportion of controlled studies rated as yes, no, or unclear for each of the RCT items

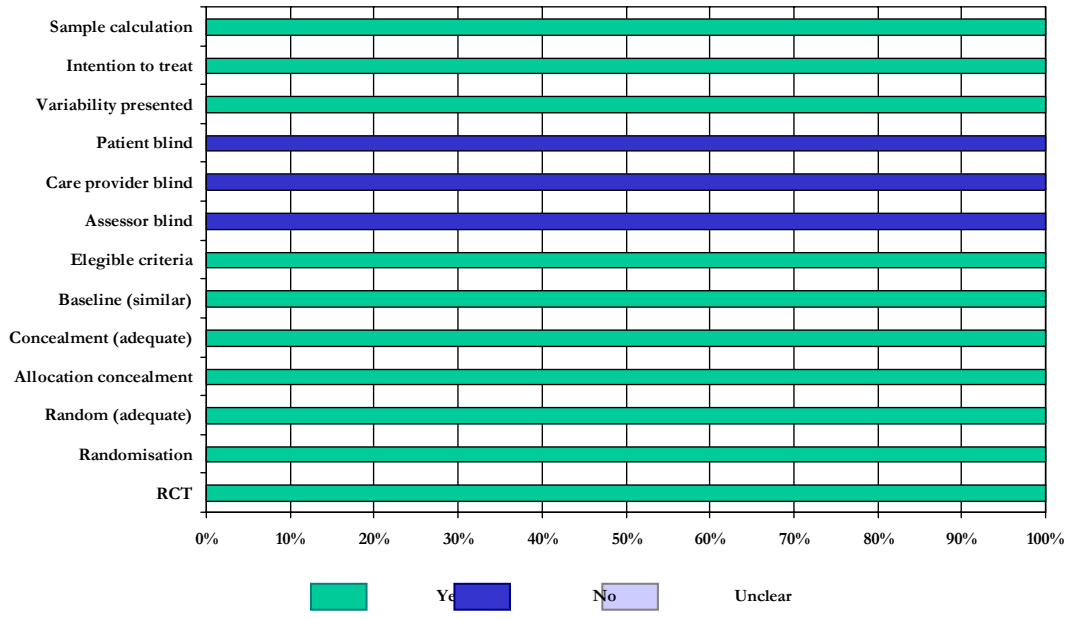


Figure 4

Sensitivities for diagnostic cohort studies evaluating the detection of all chromosomal abnormalities (including T21) by measuring the nuchal translucency thickness

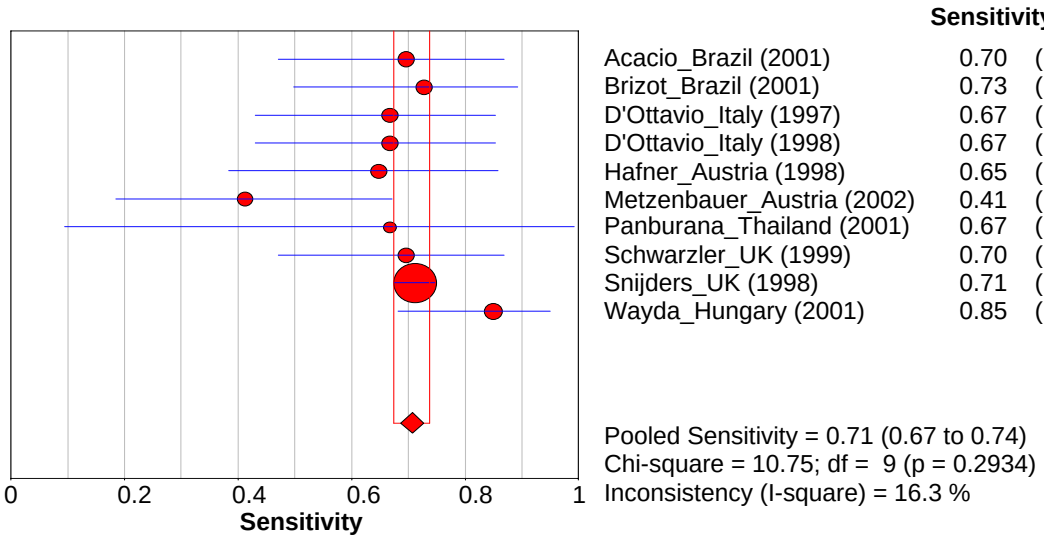


Figure 5

Specificities for diagnostic cohort studies evaluating the detection of all chromosomal abnormalities (including T21) by measuring the nuchal translucency thickness

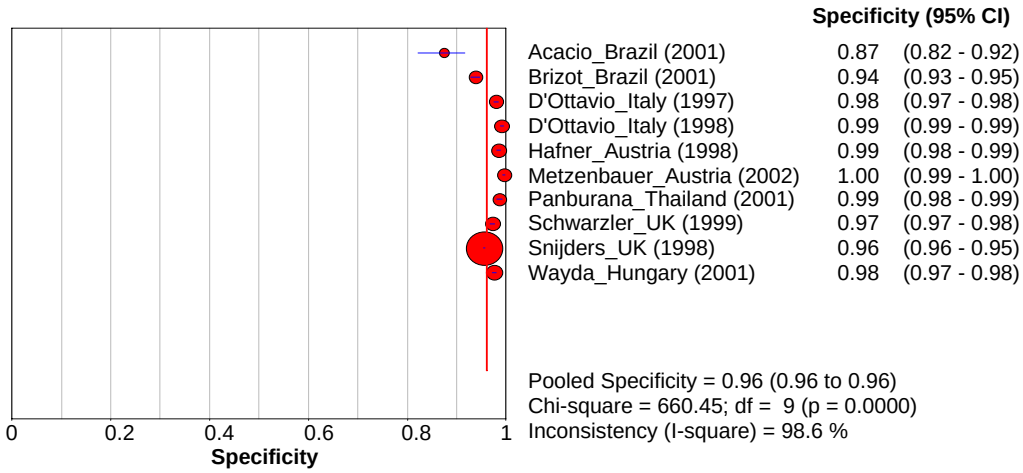


Figure 6

DOR for diagnostic cohort studies evaluating the detection of all chromosomal abnormalities (including T21) by measuring the nuchal translucency thickness

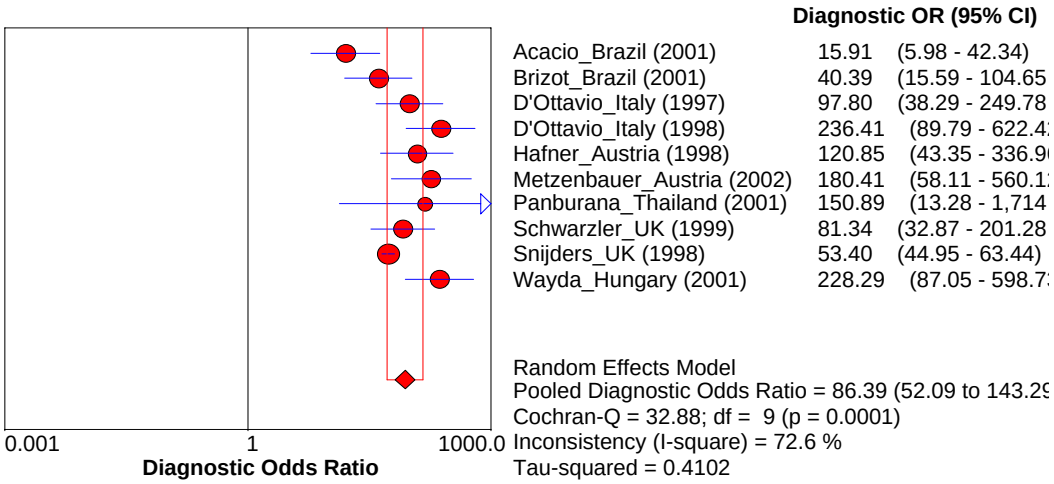


Figure 7

Detection of all chromosomal abnormalities (including T21) by measuring the nuchal translucency thickness: sensitivity and 1-specificity plotted in the ROC space

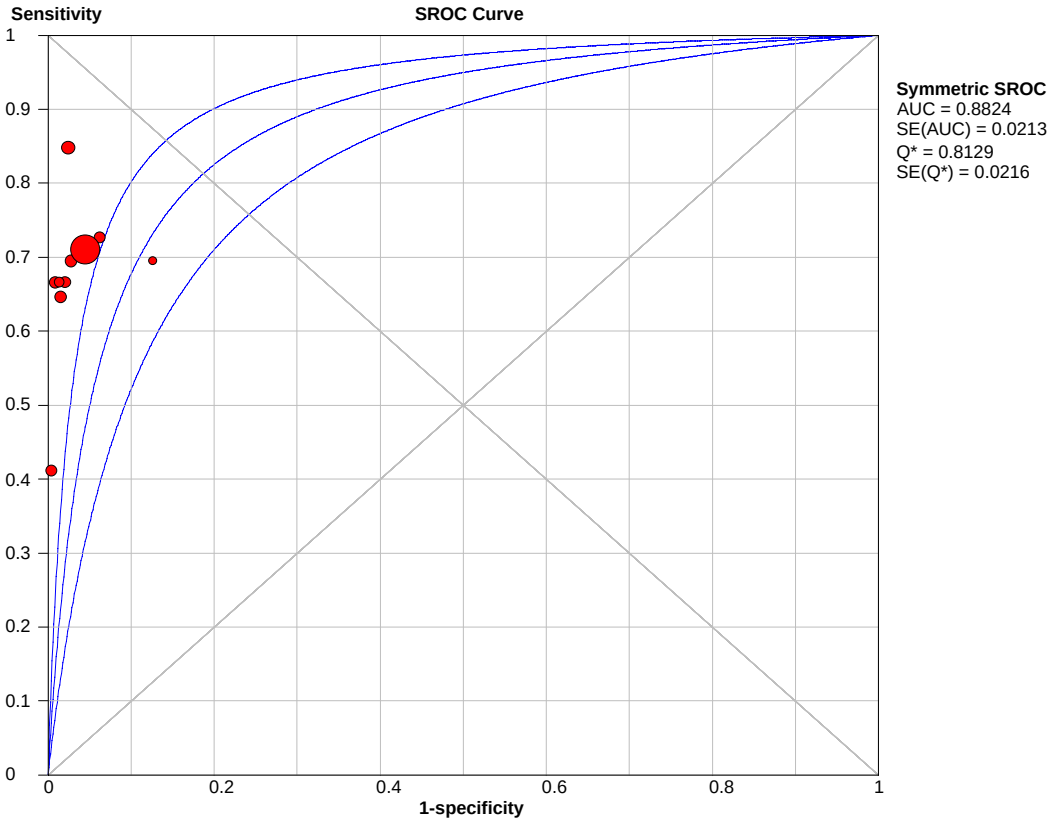


Figure 8

Sensitivities for diagnostic cohort studies evaluating the detection of chromosomal abnormalities (excluding T21) by measuring the nuchal translucency thickness

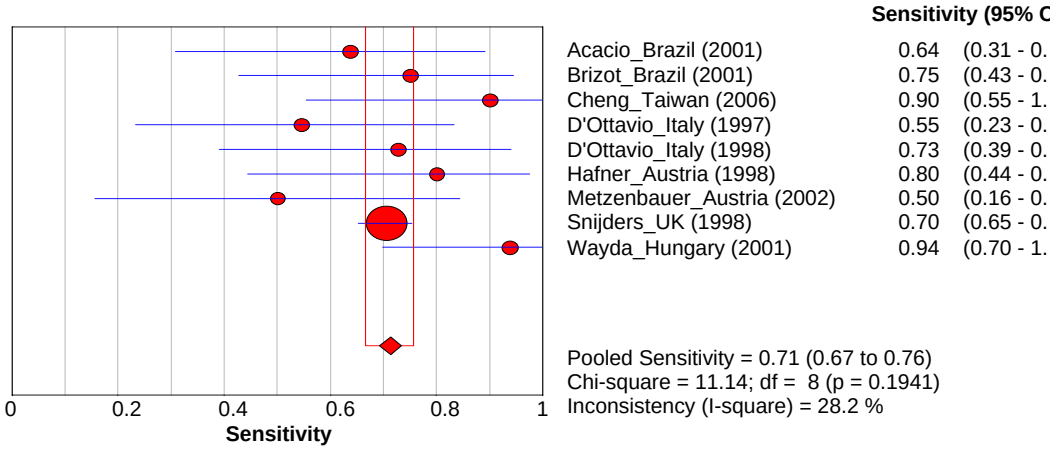


Figure 9

Specificities for diagnostic cohort studies evaluating the detection of chromosomal abnormalities (excluding T21) by measuring the nuchal translucency thickness

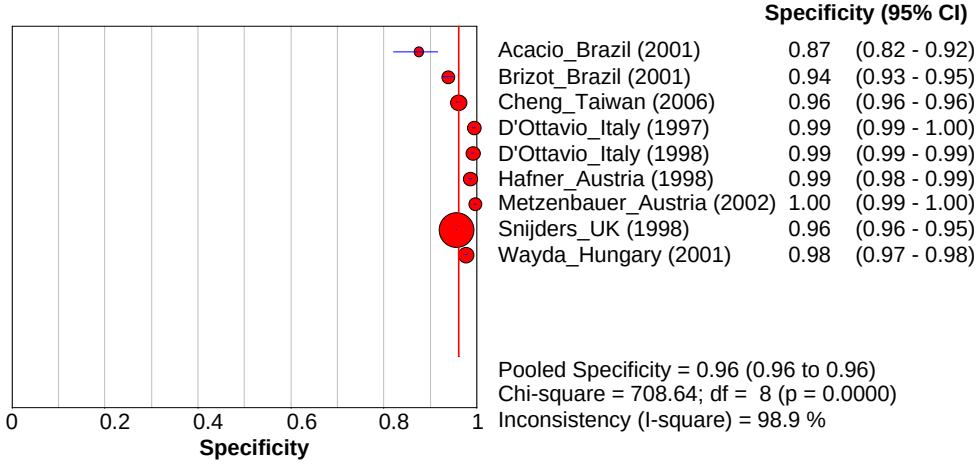


Figure 10

DOR for diagnostic cohort studies evaluating the detection of chromosomal abnormalities (excluding T21) by measuring the nuchal translucency thickness

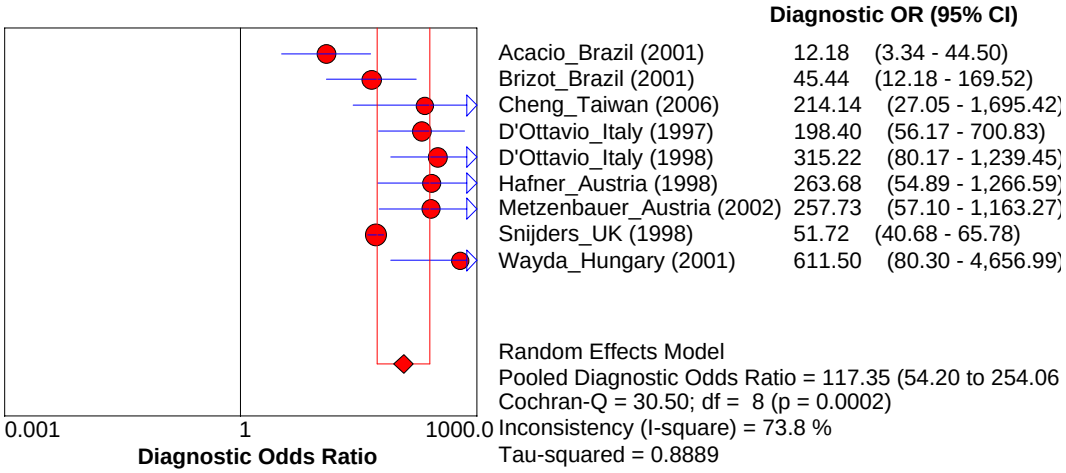


Figure 11

Detection of chromosomal abnormalities (excluding T21) by measuring the nuchal translucency thickness: sensitivity and 1-specificity plotted in the ROC space

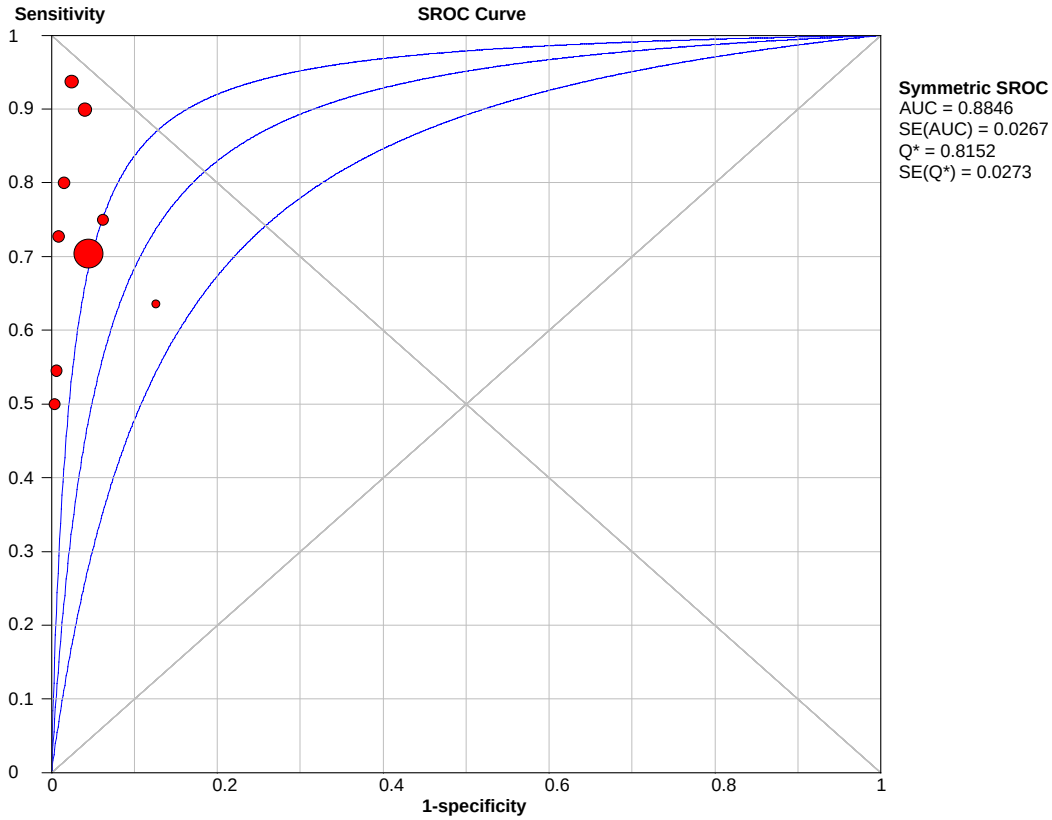


Figure 12

Sensitivities for diagnostic cohort studies evaluating the detection of all chromosomal abnormalities (including T21) by measuring the absence of nasal bone

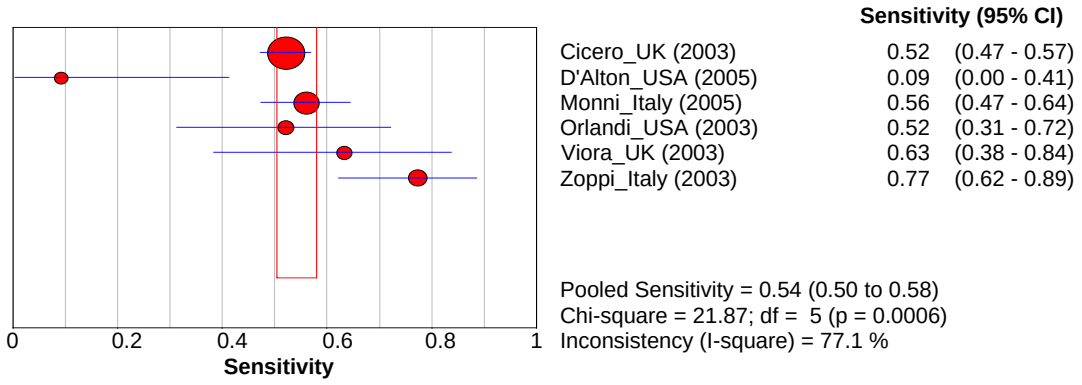


Figure 13

Specificities for diagnostic cohort studies evaluating the detection of all chromosomal abnormalities (including T21) by measuring the absence of nasal bone

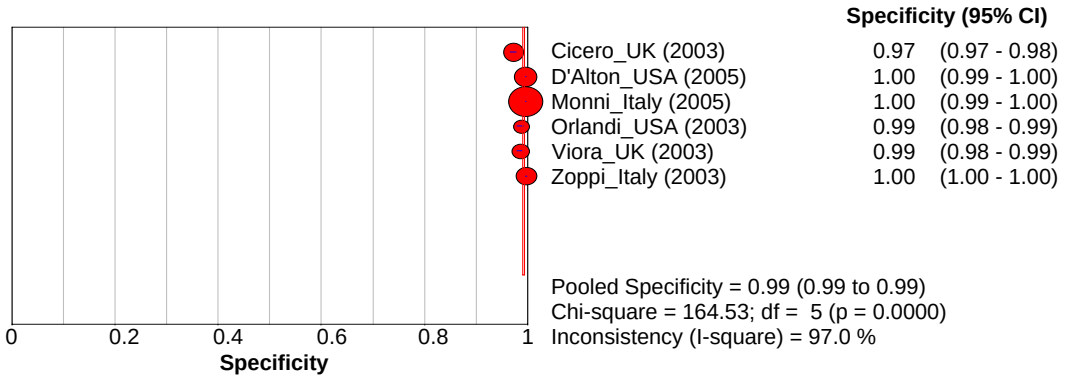


Figure 14

DOR for diagnostic cohort studies evaluating the detection of all chromosomal abnormalities by measuring the absence of nasal bone

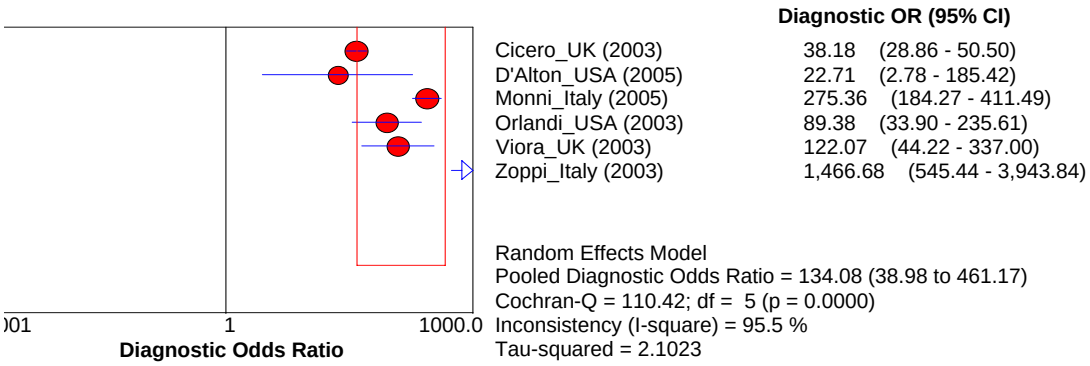


Figure 15

Detection of all chromosomal abnormalities (including T21) by measuring the absence of nasal bone: sensitivity and 1-specificity plotted in the ROC space

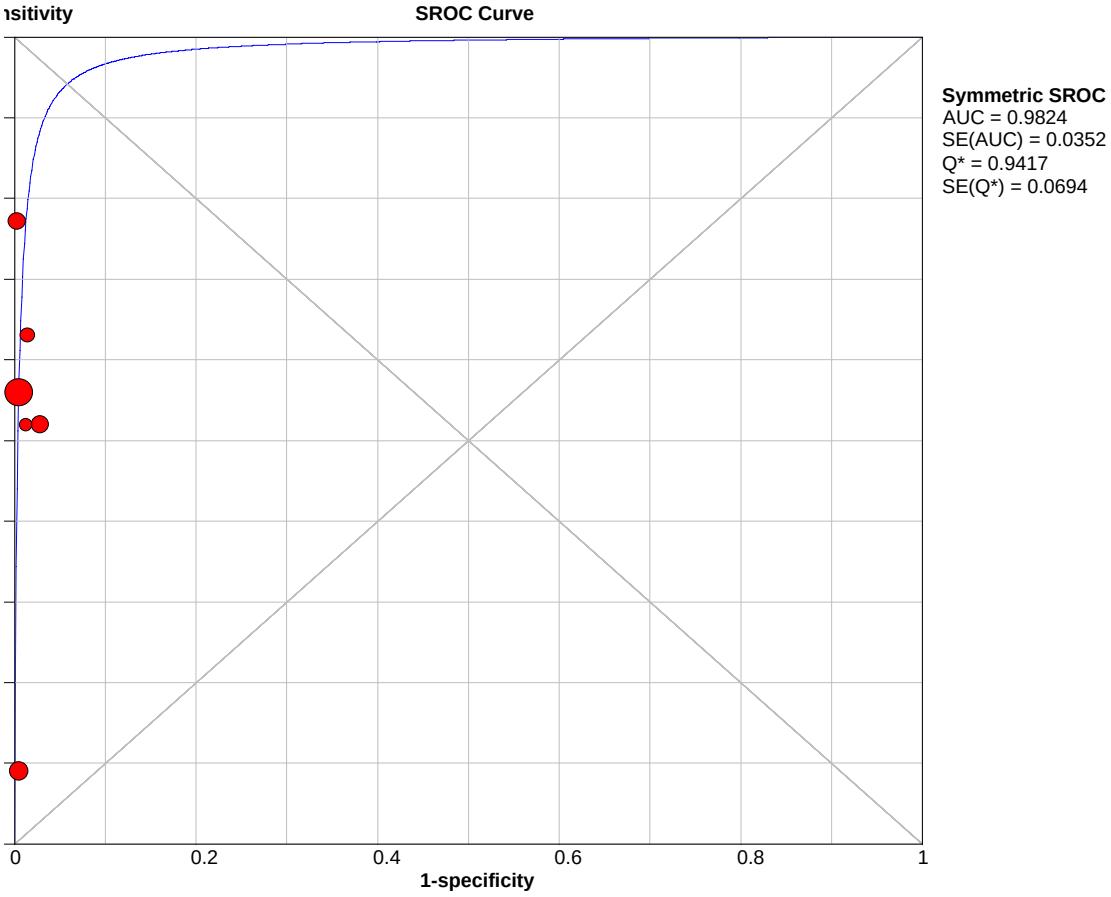


Figure 16

Sensitivities for diagnostic cohort studies evaluating the detection of chromosomal abnormalities (excluding T21) by measuring the absence of nasal bone

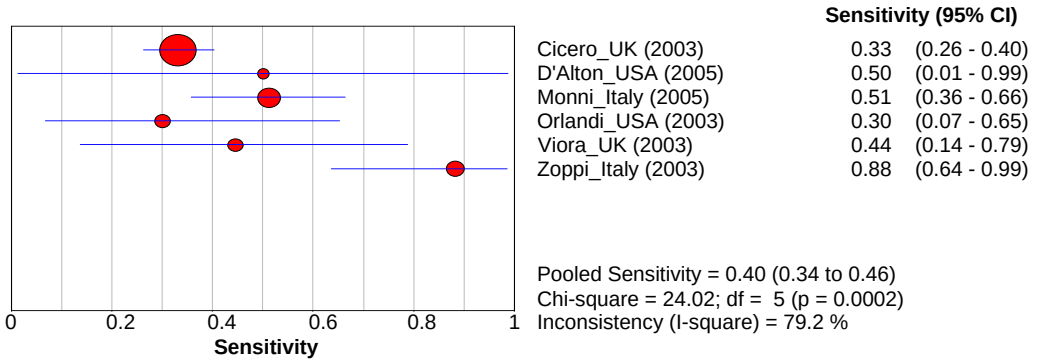


Figure 17

Specificities for diagnostic cohort studies evaluating the detection of chromosomal abnormalities (excluding T21) by measuring the absence of nasal bone

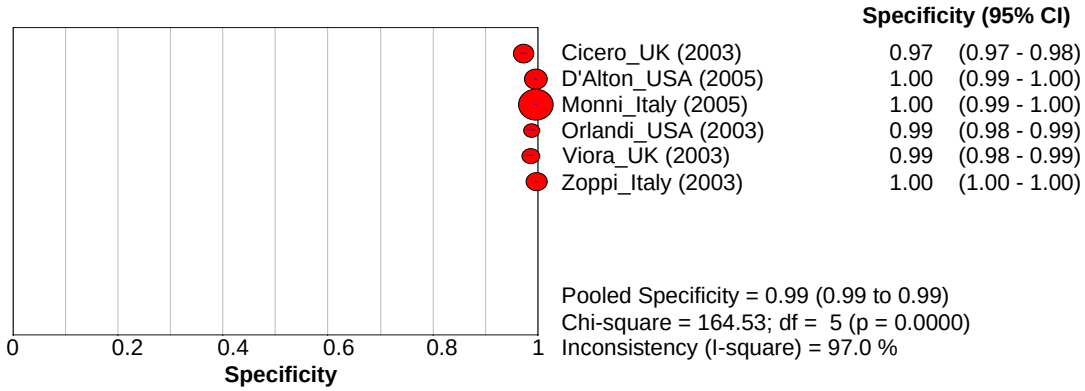


Figure 18

DOR for diagnostic cohort studies evaluating the detection of chromosomal abnormalities (excluding T21) by measuring the absence of nasal bone

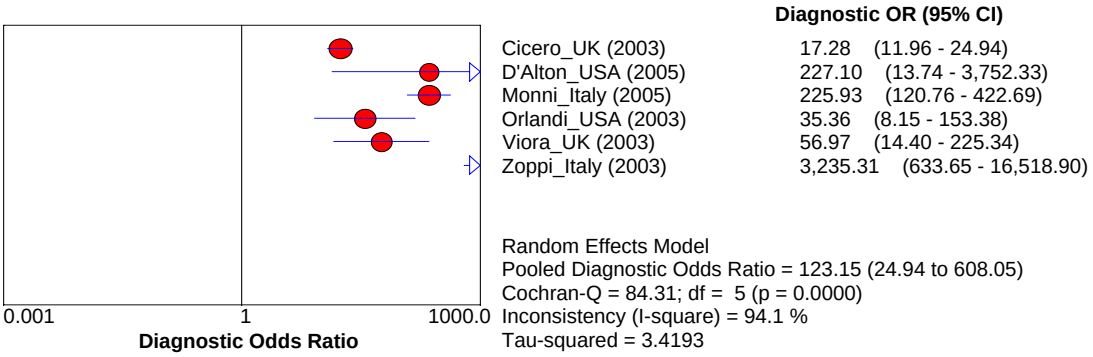


Figure 19

Detection of chromosomal abnormalities (excluding T21) by measuring the absence of nasal bone: sensitivity and 1-specificity plotted in the ROC space

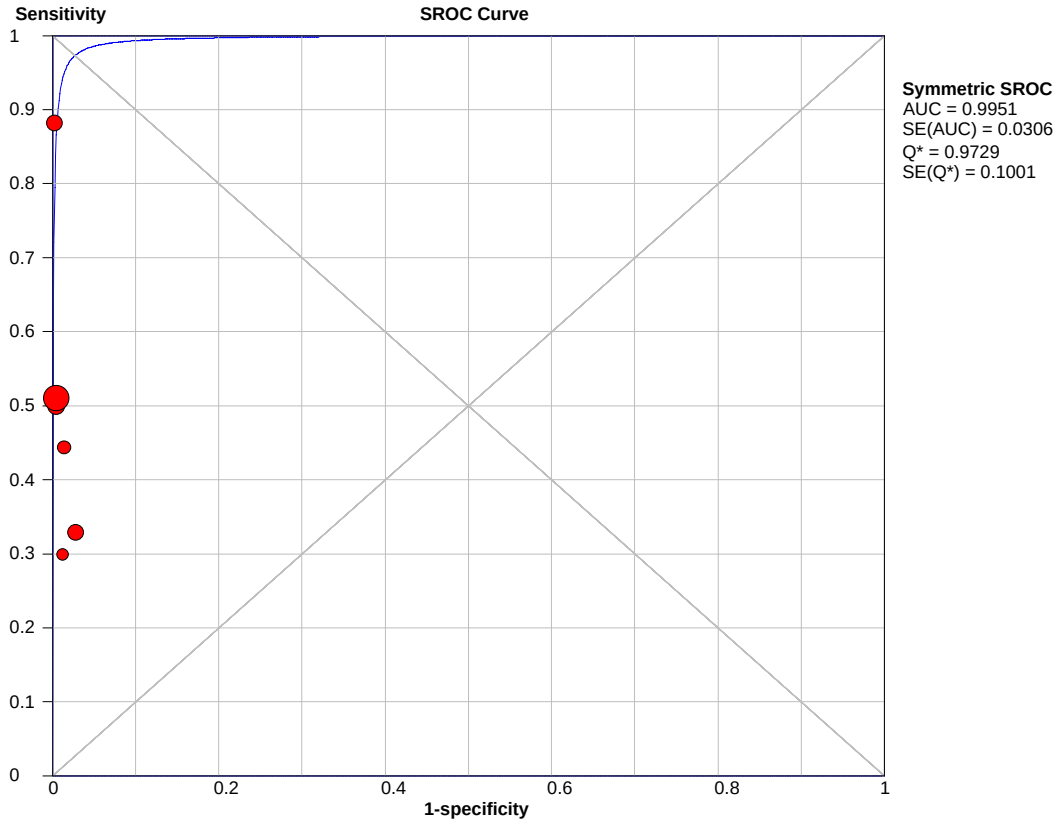


Figure 20

Sensitivities for diagnostic cohort studies evaluating the detection of all chromosomal abnormalities (including T21) by measuring a calculation of risk based on maternal age, and
NT

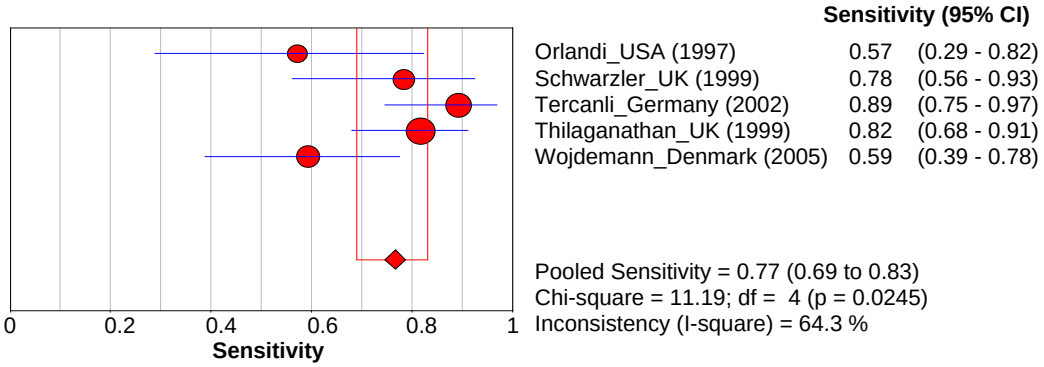


Figure 21

Specificities for diagnostic cohort studies evaluating the detection of all chromosomal abnormalities (including T21) by measuring a calculation of risk based on maternal age, and NT

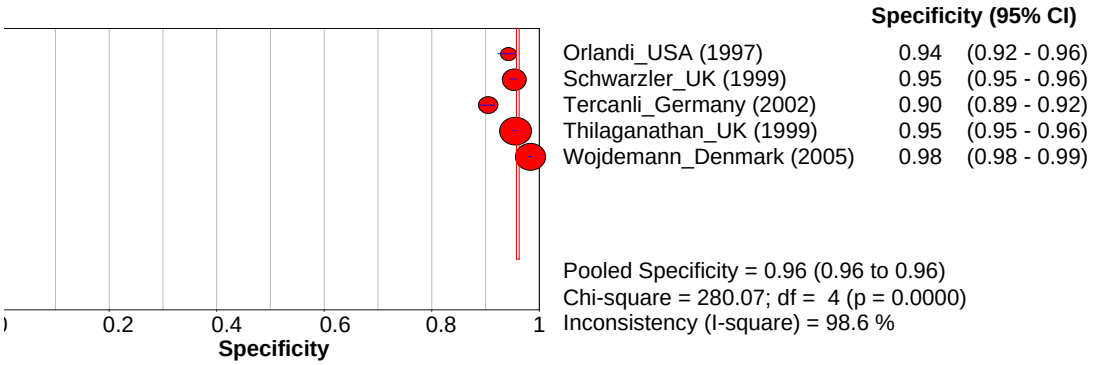


Figure 22

DOR for diagnostic cohort studies evaluating the detection of all chromosomal abnormalities (including T21) by measuring a calculation of risk based on maternal age, and NT

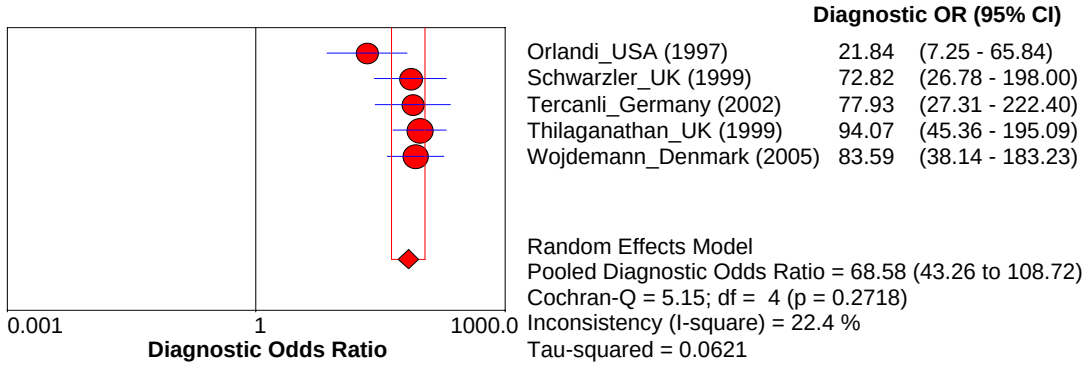


Figure 23

Detection of all chromosomal abnormalities (including T21) by measuring a calculation of risk based on maternal age, and NT: sensitivity and 1-specificity plotted in the ROC space

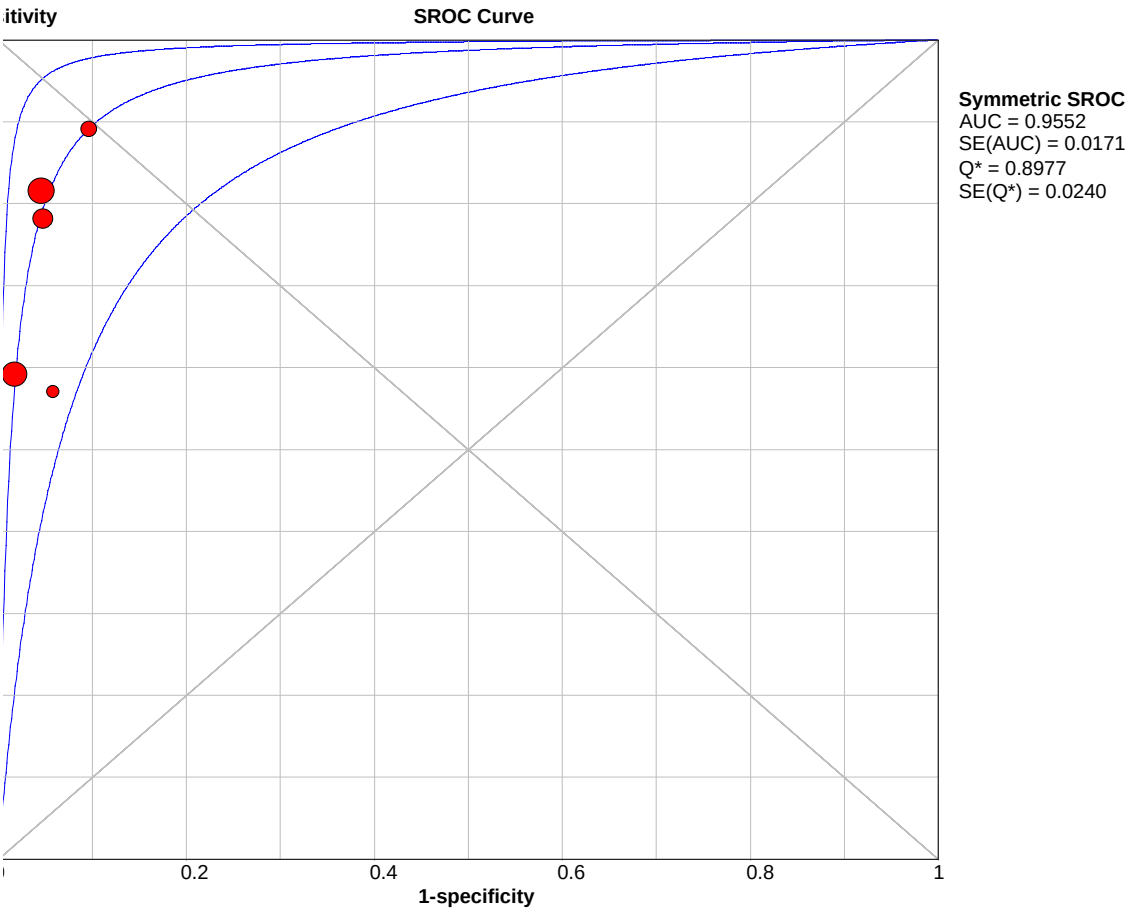


Figure 24

Sensitivities for diagnostic cohort studies evaluating the detection of chromosomal abnormalities (excluding T21) by measuring a calculation of risk based on maternal age, and NT

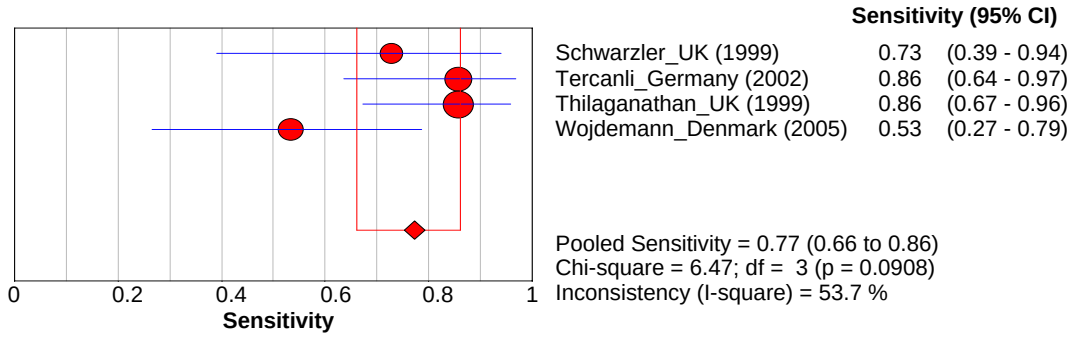


Figure 25

Specificities for diagnostic cohort studies evaluating the detection of chromosomal abnormalities (excluding T21) by measuring a calculation of risk based on maternal age, and NT

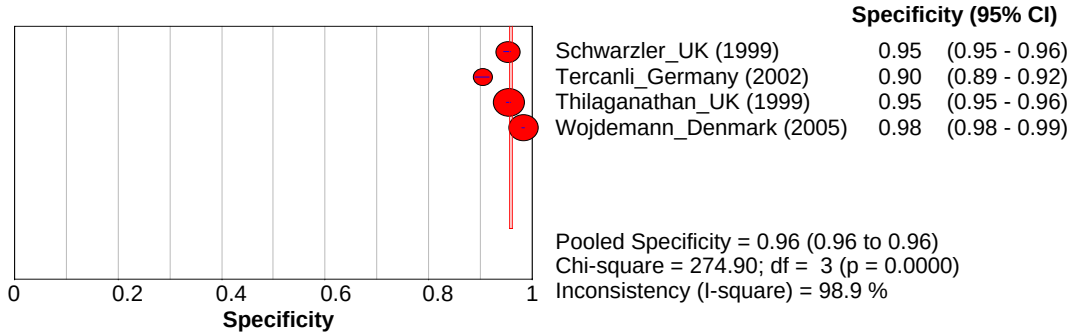


Figure 26

DOR for diagnostic cohort studies evaluating the detection of chromosomal abnormalities (excluding T21) by measuring a calculation of risk based on maternal age, and NT

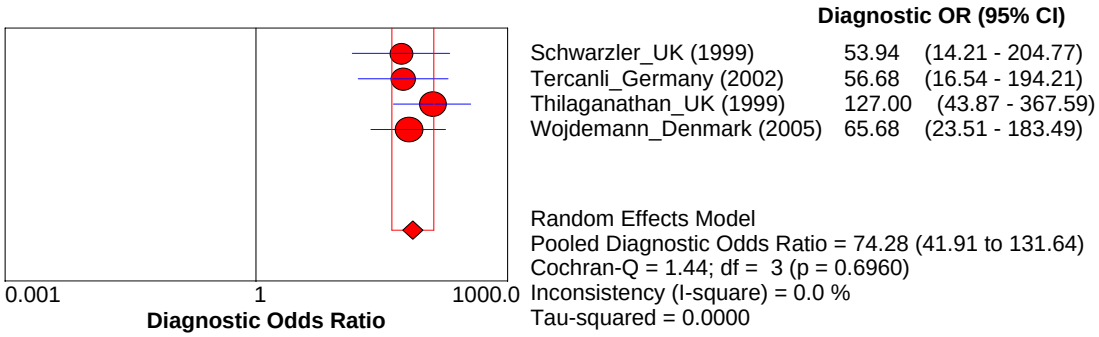


Figure 27

Detection of chromosomal abnormalities (excluding T21) by measuring a calculation of risk based on maternal age, and NT: sensitivity and 1-specificity plotted in the ROC space

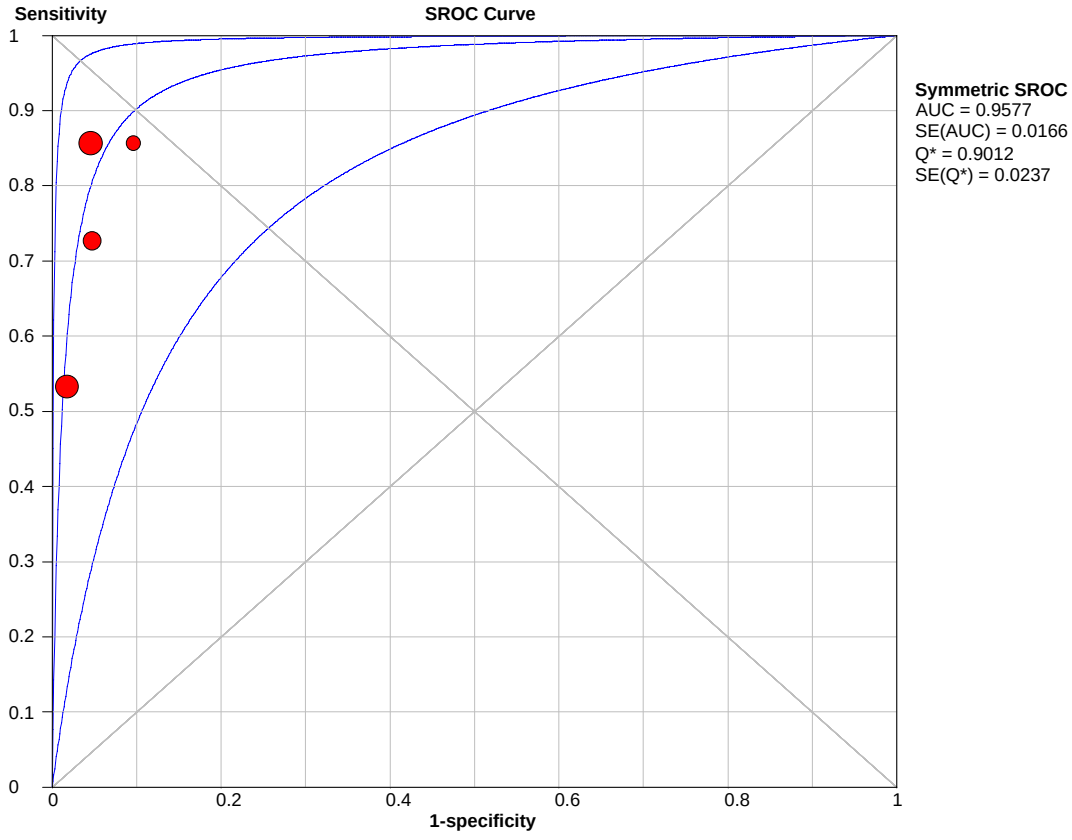


Figure 28

Sensitivities for diagnostic cohort studies describing different ultrasound measurements to detect all chromosomal abnormalities (including trisomy 21)

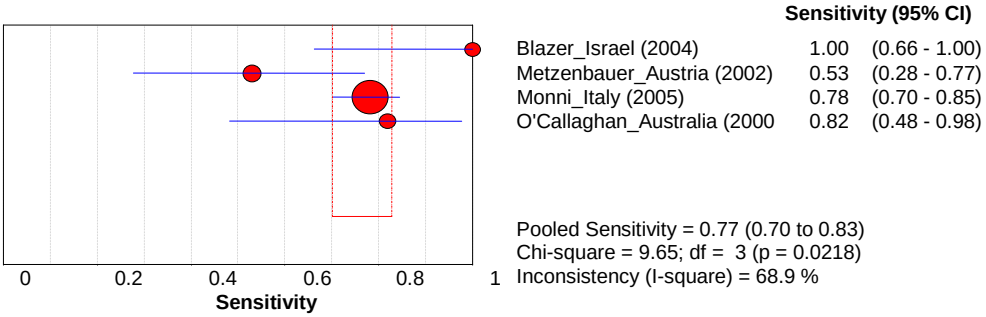


Figure 29

Specificities for diagnostic cohort studies describing different ultrasound measurements to detect all chromosomal abnormalities (including trisomy 21)

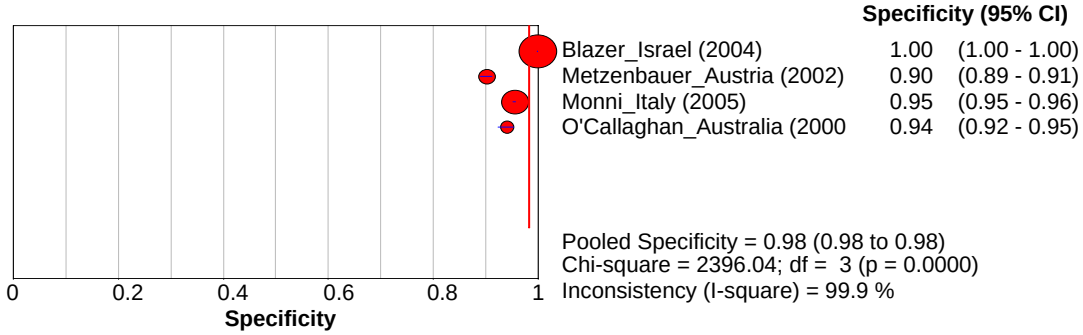


Figure 30

DOR for diagnostic cohort studies describing different ultrasound measurements to detect all chromosomal abnormalities (including trisomy 21)

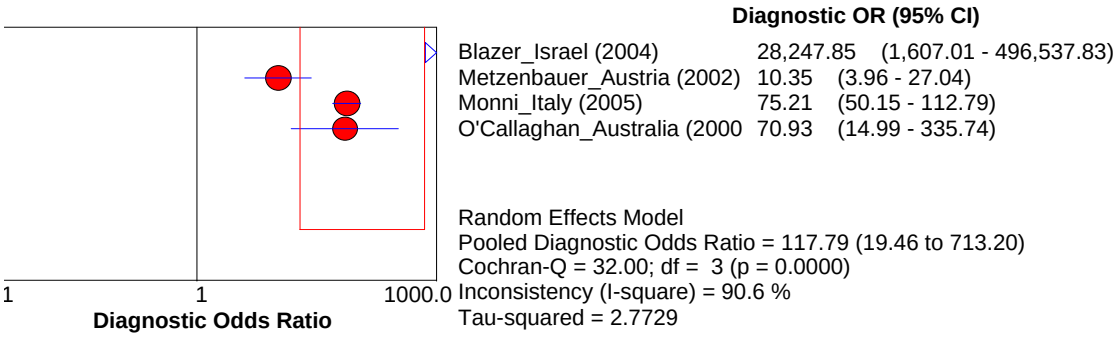


Figure 31

Detection of chromosomal abnormalities (excluding T21) by different ultrasound measurements: sensitivity and 1-specificity plotted in the ROC space

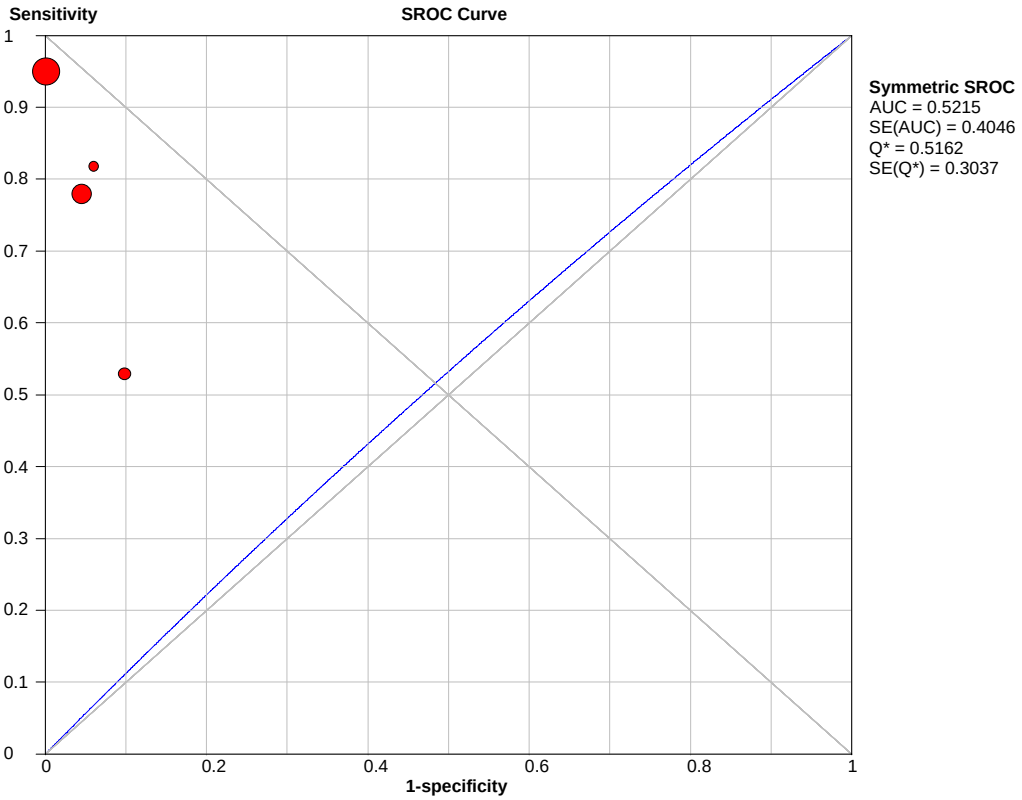


Figure 32

Sensitivities for diagnostic cohort studies describing different ultrasound measurements to detect chromosomal abnormalities (excluding trisomy 21)

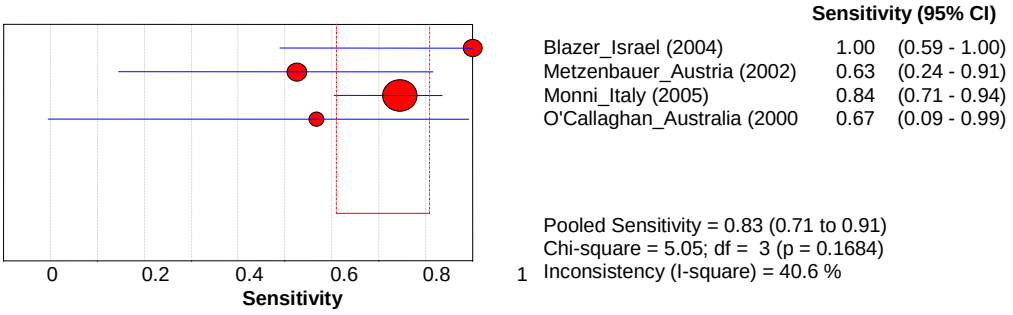


Figure 33

Specificities for diagnostic cohort studies describing different ultrasound measurements to detect chromosomal abnormalities (excluding trisomy 21)

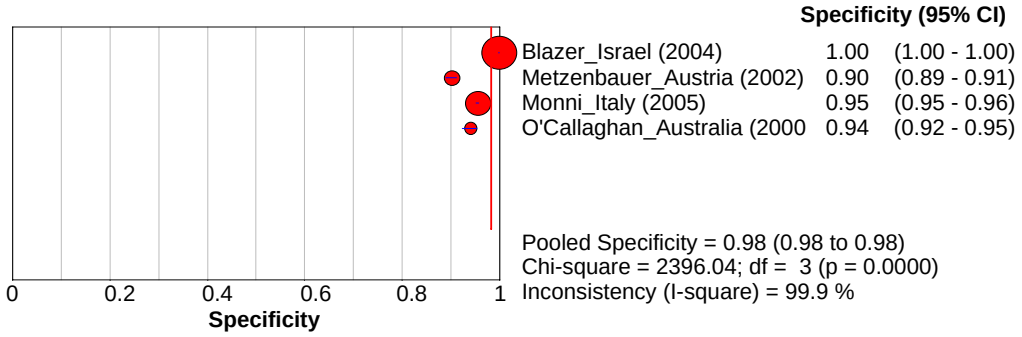


Figure 34

DOR for diagnostic cohort studies describing different ultrasound measurements to detect chromosomal abnormalities (including trisomy 21)

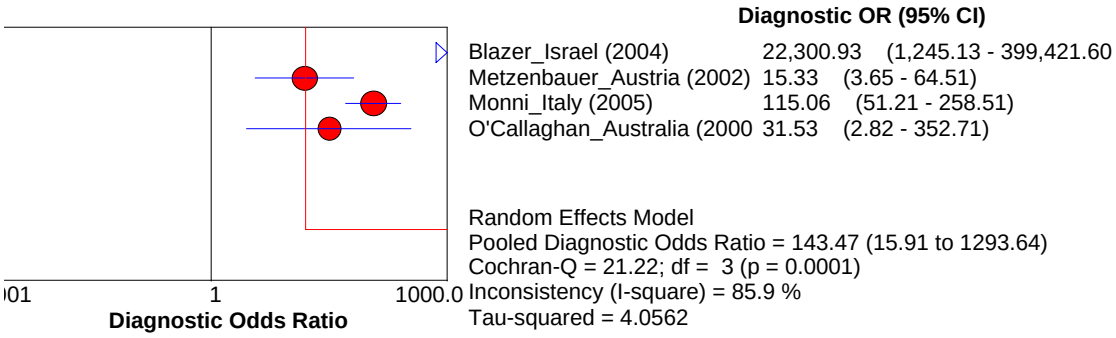
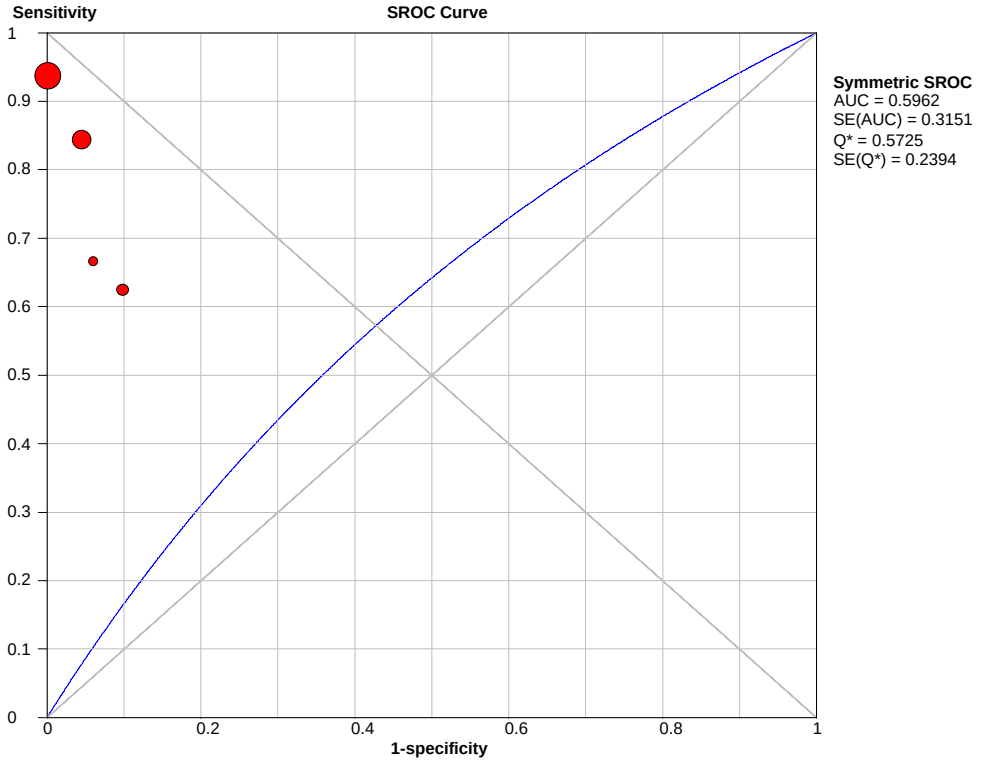


Figure 35

Detection of chromosomal abnormalities (excluding T21) by different ultrasound measurements: sensitivity and 1-specificity plotted in the ROC space



Tables and captions

Table 1 Medline Search Results - Outcomes

	Search History	Results
1	exp Pregnancy Trimester, First/	9170
2	exp Ultrasonography, Prenatal/	16162
3	exp Clinical Trials/	193918
4	exp Research Design/	216276
5	exp Treatment Outcome/	291062
6	exp Double-Blind Method/	90532
7	exp Single-Blind Method/	10558
8	((single or double or triple) adj3 blind$3).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	120481
9	random$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	486959
10	controlled clinical trial.pt.	74768
11	clinical trial.pt.	455937
12	(clinical adj trial$1).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	586148
13	exp Epidemiologic Research Design/	472717
14	(control$3 adj trial$1).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	292591
15	randomi#ed controlled trial.pt.	233178
16	comparative study/	1343564
17	or/3-16	2381580
18	1 and 2 and 17	390
19	limit 18 to humans	390
20	limit 19 to yr="1996 - 2006"	334

Table 2 Medline Search Results - Accuracy

	Search History	Results
1	exp Pregnancy Trimester, First/	9160
2	exp Ultrasonography, Prenatal/	16148
3	exp "Sensitivity and Specificity"/	216138
4	exp Diagnosis/	3914349
5	diagnos$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	1206136
6	sensitiv$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	732282
7	predict$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	471803
8	accura$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	229300
9	or/3-8	5076163
10	1 and 2 and 9	1257
11	limit 10 to humans	1256
12	limit 11 to yr="1996 - 2006"	987

Table 3 Embase Search Results - Outcomes

	Search History	Results
1	exp Pregnancy Trimester, First/	7387
2	exp Ultrasonography, Prenatal/	191297
3	exp Clinical Trials/	402678
4	exp Research Design/	1057970
5	exp Treatment Outcome/	337526
6	exp Double-Blind Method/	61061
7	exp Single-Blind Method/	6068
8	((single or double or triple) adj3 blind$3).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	100919
9	random$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	343722
10	controlled clinical trial.pt.	0
11	clinical trial.pt.	0
12	(clinical adj trial$1).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	447864
13	exp Epidemiologic Research Design/	622451
14	(control$3 adj trial$1).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	135023
15	randomi#ed controlled trial.pt.	0
16	comparative study/	81848
17	or/3-16	2141891
18	1 and 2 and 17	570
19	limit 18 to humans	565
20	limit 19 to yr="1996 - 2006"	487

Table 4 Medline Search Results - Accuracy

	Search History	Results
1	exp Pregnancy Trimester, First/	7387
2	exp Ultrasonography, Prenatal/	191297
3	exp "Sensitivity and Specificity"/	28907
4	exp Diagnosis/	1709819
5	diagnos$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	1333466
6	sensitiv$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	573444
7	predict$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	404802
8	accura$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]	286509
9	or/3-8	2798767
10	1 and 2 and 9	1727
11	limit 10 to humans	1699
12	limit 11 to yr="1996 - 2006"	1394

Table 5 Cinahl Search Results - Outcomes

	Search History	Results
1	exp Pregnancy Trimester, First/	424
2	exp Ultrasonography, Prenatal/	1144
3	exp Clinical Trials/	38903
4	exp Research Design/	174536
5	exp Treatment Outcome/	30633
6	exp Double-Blind Method/	0
7	exp Single-Blind Method/	0
8	((single or double or triple) adj3 blind$3).mp. [mp=title, subject heading word, abstract, instrumentation]	11110
9	random$.mp. [mp=title, subject heading word, abstract, instrumentation]	50991
10	controlled clinical trial.pt.	0
11	clinical trial.pt.	18196
12	(clinical adj trial$1).mp. [mp=title, subject heading word, abstract, instrumentation]	34432
13	exp Epidemiologic Research Design/	0
14	(control$3 adj trial$1).mp. [mp=title, subject heading word, abstract, instrumentation]	11503
15	randomi#ed controlled trial.pt.	0
16	comparative study/	35719
17	or/3-16	220517
18	1 and 2 and 17	35
19	limit 18 to humans [Limit not valid in: CINAHL; records were retained]	35
20	limit 19 to yr="1996 - 2006"	35

Table 6 Cinahl Search Results - Accuracy

	Search History	Results
1	exp Pregnancy Trimester, First/	424
2	exp Ultrasonography, Prenatal/	1144
3	exp "Sensitivity and Specificity"/	8634
4	exp Diagnosis/	236756
5	diagnos$.mp. [mp=title, subject heading word, abstract, instrumentation]	76030
6	sensitiv$.mp. [mp=title, subject heading word, abstract, instrumentation]	24173
7	predict$.mp. [mp=title, subject heading word, abstract, instrumentation]	34556
8	accura$.mp. [mp=title, subject heading word, abstract, instrumentation]	13889
9	or/3-8	293881
10	1 and 2 and 9	108
11	limit 10 to humans [Limit not valid in: CINAHL; records were retained]	108
12	limit 11 to yr="1996 - 2006"	106

Table 7 Lilacs Search Results – Accuracy

	Search History	Results
1	("gravidez/" and primeiro trimestre) or ("embarazo/" and primero trimestre) or "pregnancy tests" or "pregnancy trimester, first/" or "pregnancy, first trimester/"	139
2	"ultrasonografia fetal/" or "ultrasonografia pre-natal/" or "ultrasonografia prenatal/" or "ultrasonography, fetal/" or "ultrasonography, prenatal/"	346
3	"diagnosis" or "diagnosis, prenatal" or "diagnosis, prenatal/" or "diagnostico intra-uterino/" or "diagnostico intrauterino/" or "diagnostico por imagem/" or "diagnostico por ultra-som/" or "diagnostico por ultrasonido/" or "diagnostico pre-natal" or "diagnostico pre-natal por ultra-som/" or "diagnostico pre-natal ultra-sonico/" or "diagnostico pre-natal/"	62947
4	1 and 2 and 3	17

Table 8 Study characteristics and reason for excluded potential relevant studies

Study ID	Title (first 3 words)	Design	Weeks of preg	Reference standard	Cut-off	Reason for exclusion
Beke (2005)	Trisomies and other	Prospective (Cohort study)	first trimester	Karyotyping	NT≥3mm	Selected population
Cheng (2004)	Pregnancy outcomes with	Retrospective (cohort study)	11 to 14	Karyotyping	NT≥3mm	No data for 2x2
Chitty (2006)	Fetal nuchal translucency	Observational study	11 to 13	Karyotyping	NT≥3,5mm	Selected population
Conoscenti (2003)	Does cervical length	Prospective (Cohort study)	13 to 15	Birth < 37, 34 weeks	50th percentile	No data for 2x2
Crowther (1999)	Is an ultrasound	Prospective (RCT)	up to 17	Not reported	CRL>=10 diff	Up to 17 weeks
Drysdale (2002)	First-trimester pregnancy	Prospective (Cohort study)	before 12	Karyotyping	risk > 1:300	Just T21
Economides (1998)	First trimester	Prospective (Cohort study)	11 to 14	Karyotyping	NT≥ 99th percentile, NT>4mm	No data for 2x2
Ghezzi (2002)	First-trimester umbilical	Prospective (Cohort study)	10 to 14	Not reported		No data for 2x2
Gonzalez (2004)	Diagnostico prenatal invasivo					Comment
Hewitt 1996	Correlation between nuchal	Prospective (Cohort study)	up to 14	Karyotyping	NT≥3mm	Selected population
Krantz (2000)	First Trimester Down	Prospective (Cohort study)	9 to 13+6		Risk calculation	Combined with serum
Lewis (2003)	First trimester tests	Prospective (Cohort study)	10 to 13+6	Karyotyping		Combined with serum
Malone (2005)	First-Trimester Septated	Prospective (Cohort study)	10 to 13+6	Karyotyping	NT≥3mm	No data for 2x2
Malone (2004)	First trimester nasal	Prospective (Cohort study)	10 to 13+6	Karyotyping	Nasal bone	Double with D'Alton
Mustafa (2002)	Transvaginal ultrasonography in	Prospective (Cohort study)	11 to 14,	Delivery date	Plazenta praevia	Not the aim
O'Leary (2006)	First-Trimester Combined	Prospective (Cohort study)	11 to 13+6	Karyotyping		Just T21
Orlandi (1997)	First trimester screening	Prospective (Cohort study)	10 to 13+4			Combined with serum
Panburana (2001)	First trimester down	Prospective (Cohort study)	10 to 13	Karyotyping	NT≥3mm	No data for 2x2
Papp (2006)	Prenatal Diagnosis of	Retrospective (case-control)	10 to 22	Karyotyping	NT≥3mm	No data for 2x2
Peralta (2005)	Gap between fetal	Prospective (Cohort study)	10 to 14	Karyotyping	Nasal bone	Selected population
Rosati (2000)	Prognostic value of	Prospective (Cohort study)	10 to 15		NT≥3mm	Selected population
Saltvedt (2006)	Detection of malformations	Prospective (RCT)	12 to 14	Extended anomaly scan	NT≥3.5mm	Malformations
Schouwink (2000)	Ultrasonographic criteria for	Prospective (Cohort study)	> 12 weeks	Second scan		No data for 2x2
Sepulveda (1996)	The lambda sign	Prospective (Cohort study)	10 to 14	Karyotyping		No Gold Standard
Sladkevicius (2005)	Ultrasound dating at	Prospective (Cohort study)	12 to 14+6	Date of oocyte retrieval	Not reported	Combined with serum
Souka (2001)	Outcome of pregnancy	Retrospective (case-control)	11 to 14	Delivery	NT>3,5mm	Double with Snijders 1998
Spencer (2003)	Screening for chromosomal	Prospective (Cohort study)	10 to 14	Karyotyping	1:300 risk	Combined with serum
Srisupundit (2006)	Fetal structural anomaly	Prospective (Cohort study)	11 to 14	Karyotyping	NT≥95th percentile	Malformations
Taipale (2001)	Predicting delivery date	Prospective (Cohort study)	8 to 16+6	Delivery date	CRL>15mm, BPD<36mm,	Selected population
Taipale (2003)	Learning Curve in	Prospective (Cohort study)	13 to 14	Karyotyping	NT>3mm	Not the aim
Taliganathan (1997)	First trimester nuchal	Retrospective (cohort study)	11 to 14	Karyotyping	NT>3mm	Double with Thilaganathan 1999
van Bogaert (2003)	Accuracy of menstrual	Prospective (Cohort study)	early pregnancy			Comment
Whitlow (1998)	The significance of	Prospective (Cohort study)	11 to 14+6	Karyotyping		No data for 2x2
Zoppi (2003)	Changes in nuchal	Prospective (Cohort study)		Not reported	NT≥95th percentile	No data for 2x2

Table 9 List of all included studies with study details

Study details[1]	Titel, first words	Study design	Weeks of pregnancy	Reference standard	Database
Acacio_Brazil (2001)	Nuchal translucency: an ultrasound marker	Prospective cohort	Okt.14	karyotyping	Ovid
Bennet_USA (2004)	First trimester ultrasound with nuchal translucency	RCT		age	Medline
Blazer_Israel (2004)	Fetal omphalocele detected early	Prospective cohort	Dez.16	karyotyping + pregnancy outcome	Medline
Brizot_Brazil (2001)	First-trimester screening for chromosomal	Retrospective cohort	Okt.14	karyotyping + pregnancy outcome	Medline
Cheng_Taiwan (2006)	Association of fetal choroid plexus cysts	Prospective cohort	Okt.14	karyotyping + pregnancy outcome	Embase, Medline
D'Alton_USA (2005)	First and second trimester evaluation of risk	Prospective cohort	Okt.14	karyotyping + pregnancy outcome	Embase, Medline
D'Ottavio_Italy (1997)	Screening for fetal anomalies	Prospective cohort	13-15	karyotyping	Medline
D'Ottavio_Italy (1998)	Comparison of first and second trimester	Prospective cohort	14	karyotyping + pregnancy outcome	Embase, Medline
Harrington_UK (2006)	Does a first trimester	RCT		age	Embase
Kelecki_Turkey (2005)	Can increased nuchal translucency	Prospective cohort		diabetes	Embase, Medline
Menon_Malaysia (2005)	A retrospective study of the accuracy	Retrospective cohort		twins	Embase
Monni_Italy (2005)	Nuchal translucency and nasal bone	Retrospective cohort	Not specified	karyotyping + pregnancy outcome	Embase, Medline
O'Callaghan_Australia (2000)	First trimester ultrasound with nuchal translucency	Prospective cohort	Nov.14	karyotyping + pregnancy outcome	Embase, Medline
Orlandi_USA (2003)	Measurement of nasal bone length	Prospective cohort	Sep.14	karyotyping + pregnancy outcome	Embase, Medline
Schwarzler_UK (1999)	Screening for fetal aneuploidies	Prospective cohort	Okt.14	karyotyping + pregnancy outcome	Embase, Medline
Snijders_UK (1998)	UK multicentre project on assessment of risk	Prospective cohort	Okt.14	karyotyping	Embase
Tercanli_Germany (2003)	[Screening for aneuploidy	Prospective cohort	Nov.14	karyotyping	Medline
Thilaganathan_UK (1999)	First trimester nuchal translucency	Prospective cohort	Okt.14	karyotyping + pregnancy outcome	Embase, Medline
Viora_UK (2003)	Ultrasound evaluation of fetal nasal bone	Prospective cohort	Nov.13	karyotyping	Embase
Wayda_Hungary (2001)	Four years experience of first-trimester	Prospective cohort	10.Dez	karyotyping + pregnancy outcome	Embase
Wojdemann_Denmark (2005)	Improved first-trimester Down syndrome	Prospective cohort	Nov.14	karyotyping + pregnancy outcome	Medline
Zoppi_Italy (2003)	Absence of fetal nasal bone and aneuploidies	Prospective cohort	Not specified	karyotyping	Embase, Medline

Table 10 Quality assessment questions

	Questions for quality assessment
Q1	Was the spectrum of patients representative of the patients who will receive the test in practice in Austria?
Q2	Were selection criteria clearly described?
Q3	Is the reference standard likely to correctly classify the target condition?
Q4	Is the time period between reference standard and index test short enough to be reasonably
	sure that the target condition did not change between the two tests?
Q5	Did the whole sample or a random selection of the sample, receive verification using a reference standard?
Q6	Did patients receive the same reference standard regardless of the index test result?
Q7	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference
Q8a	Was the execution of the index test described in sufficient detail to permit its replication?
Q9	Was the execution of the reference standard described in sufficient detail to permit its replication?
Q10	Were the index test results interpreted without knowledge of the results of the reference standard?
Q11	Were the reference standard results interpreted without knowledge of the results of the index test?
Q12	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
Q13	Were uninterpretable/ intermediate test results reported?
Q14	Were withdrawals from the study explained?

Table 11 Details of the quality assessment of each included study

STUDY ID (First author + year)	Q1: Was the spectrum of patients representative of the patients who will receive the test in practice in Austria?	Q2. Were selection criteria clearly described?	Q3. Is the reference standard likely to correctly classify the target condition?	Q4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	Q5. Did the whole sample or a random selection of the sample, receive verification using a reference standard?	Q6. Did patients receive the same reference standard regardless of the index test result?	Q7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	Q8a. Was the execution of the index test described in sufficient detail to permit replication of the test?	Q9. Was the execution of the reference standard described in sufficient detail to permit its replication?	Q10. Were the index test results interpreted without knowledge of the results of the reference standard?	Q11. Were the reference standard results interpreted without knowledge of the results of the index test?	Q12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	Q13. Were uninterpretable/ intermediate test results reported?	Q14. Were withdrawals from the study explained?	TOTAL (number of cells with "YES")
Acacio 2001	unclear	no	yes	yes	yes	yes	yes	no	no	yes	unclear	yes	no	no	7
Blazer 2004	yes	no	yes	yes	yes	yes	yes	no	no	yes	unclear	yes	no	no	8
Brizot 2001	yes	no	yes	yes	yes	no	yes	yes	no	yes	unclear	yes	no	no	8
Cheng 2006	yes	no	yes	yes	yes	no	yes	yes	no	yes	unclear	yes	no	no	8
Cicero 2003	unclear	no	yes	yes	yes	yes	yes	no	no	yes	unclear	yes	no	no	7
D'Alton 2005	yes	yes	yes	yes	yes	no	yes	no	no	yes	unclear	yes	yes	yes	10
D'Ottavio 1997	yes	unclear	yes	yes	yes	no	unclear	no	no	yes	unclear	yes	no	no	6
D'Ottavio 1998	yes	unclear	yes	yes	yes	no	unclear	yes	no	no	unclear	unclear	no	no	5
Hafner 1998	yes	yes	yes	yes	unclear	unclear	yes	unclear	no	unclear	unclear	yes	yes	yes	8
Kelecki 2005	unclear	no	yes	yes	yes	yes	yes	no	no	yes	unclear	yes	no	no	7
Metzenbauer 2002	yes	yes	no	yes	unclear	unclear	yes	unclear	no	yes	unclear	yes	yes	yes	8
Monni 2005	yes	no	yes	yes	yes	yes	yes	yes	no	yes	unclear	yes	yes	yes	11
O'Callaghan 2000	yes	no	yes	yes	yes	no	yes	no	no	yes	unclear	yes	no	no	7
Orlandi 2003	unclear	yes	yes	yes	yes	no	yes	yes	yes	yes	unclear	yes	yes	yes	11
Schwarzler 1999	yes	unclear	yes	yes	yes	no	unclear	no	no	yes	unclear	yes	no	yes	7
Snijders 1998	yes	no	yes	yes	yes	no	yes	no	no	yes	unclear	yes	no	yes	8
Tercalini 2002	yes	yes	yes	yes	no	unclear	yes	no	yes	yes	unclear	yes	no	yes	9
Thilaganathan 1999	yes	no	yes	yes	yes	no	yes	no	no	yes	unclear	yes	yes	yes	9
Viora 2003	yes	yes	yes	yes	unclear	unclear	yes	unclear	no	unclear	unclear	yes	yes	yes	8
Wayda 2001	yes	no	yes	yes	yes	no	yes	no	no	yes	unclear	yes	no	yes	8
Wojdemann 2005	yes	yes	yes	yes	yes	no	yes	yes	no	yes	unclear	yes	no	yes	10
Zoppi 2003	yes	no	yes	yes	no	no	yes	no	no	yes	unclear	yes	yes	unclear	7
Menon 2005	unclear	no	yes	yes	yes	yes	yes	yes	no	yes	unclear	unclear	no	no	7
yes	**18**	**7**	**22**	**23**	**18**	**6**	**20**	**7**	**2**	**20**	**0**	**21**	**8**	**12**	**0**
no	**0**	**13**	**1**	**0**	**2**	**13**	**0**	**13**	**21**	**1**	**0**	**0**	**15**	**10**	**0**
unclear	**5**	**3**	**0**	**0**	**3**	**4**	**3**	**3**	**0**	**2**	**23**	**2**	**0**	**1**	**0**

Table 12 Studies measuring NT, including T21

Study details[1]	Study design	Weeks of pregnancy	Reference standard	Softmarker (cut-off)	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	+LR (95% CI)	-LR (95% CI)
Acacio_Brazil (2001)	Prospective cohort	10-14	karyotyping	NT ≥ 2.5mm	16	26	7	181	0.70 (0.47; 0.87)	0.87 (0.82; 092)	15.91 (5.98; 42.32)	5.54 (3.53; 8.68)	0.35 (0.19; 0.65)
Brizot_Brazil (2001)	Retrospective cohort	10-14	karyotyping + pregnancy outcome	NT ≥ 95% percentile	16	157	6	2378	0.73 (0.50; 0.89)	0.94 (0.93; 0.95)	40.39 (15.59; 104.6)	11.74 (8.72; 15.81)	0.29 (0.15; 0.57)
D'Ottavio_Italy (1997)	Prospective cohort	13-15	karyotyping	NT ≥ 4mm	14	70	7	3423	0.67 (0.43; 0.85)	0.98 (0.97; 0.98)	97.8 (38.29; 249.8)	33.27 (22.7; 48.7)	0.34 (0.19; 0.62)
D'Ottavio_Italy (1998)	Prospective cohort	14	karyotyping + pregnancy outcome	NT ≥ 4mm	14	34	7	4019	0.67 (0.43; 0.85)	0.99 (0.99; 0.99)	236.4 (89.79; 622.4)	79.47 (50.6; 124.7)	0.35 (0.18; 0.61)
Hafner_Austria (1998)	Prospective cohort	10-13	karyotyping + pregnancy outcome	NT ≥ 2.5mm	11	63	6	4153	0.65 (0.38; 0.86)	0.98 (0.98; 0.99)	120.8 (43.35; 336.9)	43.3 (28.2; 66.4)	0.36 (0.19; 0.68)
Metzenbauer_Austria (2002)	Prospective cohort	10-13	karyotyping + pregnancy outcome	NT ≥ 3.5mm	7	11	10	2835	0.41 (0.18; 0.67)	1.00 (0.99; 1.00)	180.4 (58.11; 560.1)	106.5 (46.9; 241.6)	0.59 (0.40; 0.88)
Panburana_Thailand (2001)	Prospective cohort	10-13	karyotyping	NT ≥ 2.5mm	2	27	1	2037	0.67 (0.09; 0.99)	0.99 (0.98; 0.99)	150.9 (13.28; 1714.5)	50.96 (21.1; 123.3)	0.34 (0.07; 1.67)
Schwarzler_UK (1999)	Prospective cohort	10-14	karyotyping + pregnancy outcome	NT ≥ 95% percentile	16	123	7	4377	0.70 (0.47; 0.87)	0.97 (0.97; 0.98)	81.34 (32.89; 201.2)	25.45 (18.4; 35.1)	0.31 (0.17; 0.58)
Snijders_UK (1998)	Prospective cohort	10-14	karyotyping	NT ≥ 2.5mm	463	4209	188	91267	0.71 (0.67; 0.74)	0.96 (0.96; 0.96)	53.40 (44.95; 63.44)	16.13 (15.2; 17.1)	0.30 (0.27; 0.34)
Wayda_Hungary (2001)	Prospective cohort	10-12	karyotyping + pregnancy outcome	NT ≥ 2.5mm	28	163	5	6645	0.85 (0.68; 0.95)	0.98 (0.97; 0.98)	228.3 (87.05; 598.7)	35.44 (28.7; 43.7)	0.15 (0.07; 0.35)
Pooled results (only for $I^2 \leq 75\%$)									***0.71 (0.67; 0.74)***	***I^2 = 98.6%***	***86.39 (52.09; 143.29)***		

Table 13 Regression analysis for studies measuring NT, including T21 (10 studies)

Variables	Coefficient	SE	p-value	RDOR	95% CI
Constant	Jän.42	02.Jän	0.51	---	---
S	-0.35	0.20	0. 16	---	---
Mean age	0.05	0.21	0.81	01.Mai	0.62; 1.79
NT cut-off	0.54	0.57	0.39	Jän.71	0.40; 7.35
QUADAS	0.55	0.88	0.56	Jän.73	(0.18; 16.80)

Table 14 Studies measuring NT, excluding T21

Study details[1]	Study design	Weeks of pregnancy	Reference standard	Softmarker (cut-off)	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	+LR (95% CI)	-LR (95% CI)
Acacio_Brazil (2001)	Prospective cohort	10-14	karyotyping	NT ≥ 2.5mm	7	26	4	181	0.63 (0.31; 0.89)	0.87 (0.82; 092)	11.18 (3.33; 45.5)	5.06 (2.85; 8.99)	0.42 (0.19; 0.91)
Brizot_Brazil (2001)	Retrospective cohort	10-14	karyotyping + pregnancy outcome	NT ≥ 95% percentile	9	157	3	2378	0.75 (0.43; 0.94)	0.94 (0.93; 0.95)	44.43 (12.18; 169.5)	12.11 (8.45; 17.36)	0.27 (0.10; 0.71)
Cheng_Taiwan (2006)	Prospective cohort	10-14	karyotyping + pregnancy outcome	NT ≥ 3mm	9	314	1	7471	0.90 (0.55; 1.00)	0.96 (0.95; 0.96)	214.1 (27.05; 1695.4)	22.31 (17.67; 28.2)	0.10 (0.02; 0.67)
D'Ottavio_Italy (1997)	Prospective cohort	13-15	karyotyping	NT ≥ 4mm	6	21	5	3472	0.54 (0.23; 0.83)	0.99 (0.99; 1.00)	198.4 (56.16; 700.8)	90.73 (45.6; 180.5)	0.46 (0.24; 0.87)
D'Ottavio_Italy (1998)	Prospective cohort	14	karyotyping + pregnancy outcome	NT ≥ 4mm	8	34	3	4019	0.73 (0.39; 0.94)	0.99 (0.99; 0.99)	315.2 (80.16; 1239.5)	86.7 (52.9; 141.9)	0.27 (0.10; 0.72)
Hafner_Austria (1998)	Prospective cohort	10-13	karyotyping + pregnancy outcome	NT ≥ 2.5mm	8	63	2	4153	0.80 (0.44; 0.97)	0.98 (0.98; 0.99)	263.6 (58.11; 1266.6)	53.5 (36.1; 79.48)	0.20 (0.06; 0.70)
Metzenbauer_Austria (2002)	Prospective cohort	10-13	karyotyping + pregnancy outcome	NT ≥ 3.5mm	4	11	4	2835	0.50 (0.16; 0.84)	1.00 (0.99; 1.00)	257.7 (57.1; 1163.3)	129.4 (52.1; 321.4)	0.50 (0.25; 1.00)
Snijders_UK (1998)	Prospective cohort	10-14	karyotyping	NT ≥ 2.5mm	229	4209	96	91267	0.70 (0.65; 0.75)	0.96 (0.96; 0.96)	51.7 (40.68; 65.77)	15.98 (14.8; 17.25)	0.31 (0.26; 0.37)
Wayda_Hungary (2001)	Prospective cohort	10-12	karyotyping + pregnancy outcome	NT ≥ 2.5mm	15	163	1	6645	0.94 (0.70; 1.00)	0.98 (0.97; 0.98)	611.5 (80.3; 4657.0)	39.15 (32.1; 47.71)	0.06 (0.01; 0.42)
Pooled results (only for $I^2 \leq 75\%$)									***0.71 (0.67; 0.76)***	***I^2 = 98.9%***	***117.35 (54.20; 254.06)***		

Table 15 Regression analysis for studies measuring NT, excluding T 21 (9 studies)

Variables	Coefficient	SE	p-value	RDOR	95% CI
Constant	3.57	1.60	0.09	---	---
S	-0.48	0.16	0.04	---	---
Mean age	0.18	0.18	0.38	Jän.19	0.72; 1.98
NT cut-off	0.24	0.54	0.68	Jän.27	0.28; 5.64
QUADAS	-0.64	0.60	0.35	0.53	0.10; 2.78

Table 16 Studies measuring nasal bone, including T21

Study details[1]	Study design	Weeks of pregnancy	Reference standard	Softmarker (cut-off)	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	+LR (95% CI)	-LR (95% CI)
Cicero_UK (2003)	Prospective cohort	11-14	karyotyping	absence of NB	224	93	206	3265	0.52 (0.47; 0.57)	0.97 (0.97; 098)	38.17 (28.86; 50.5)	18.81 (15.1; 23.44)	0.49 (0.45; 0.54)
D'Alton_USA (2005)	Prospective cohort	10-14	karyotyping + pregnancy outcome	absence of NB	1	21	10	4769	0.09 (0.00; 0.41)	1.00 (0.99; 1.00)	22.71 (2.78; 185.4)	20.74 (3.05; 141.0)	0.91 (0.76; 1.10)
Monni_Italy (2005)	Retrospective cohort	Not specified	karyotyping + pregnancy outcome	absence of NB	79	76	62	16424	0.56 (0.47; 0.64)	0.99 (0.99; 1.00)	275.4 (184.3; 411.5)	121.6 (93.1; 159.0)	0.44 (0.37; 0.53)
Orlandi_USA (2003)	Prospective cohort	9-14	karyotyping + pregnancy outcome	absence of NB	13	12	12	990	0.52 (0.31; 0.72)	0.99 (0.98; 0.99)	89.37 (33.90; 235.61)	43.42 (22.1; 85.44)	0.49 (0.32; 0.73)
Viora_UK (2003)	Prospective cohort	11-13	karyotyping	absence of NB	12	24	7	1709	0.63 (0.38; 0.84)	0.99 (0.98; 0.99)	122.1 (44.22; 337.0)	45.6 (26.97; 77.11)	0.37 (0.21; 0.67)
Zoppi_Italy (2003)	Prospective cohort	Not specified	karyotyping	absence of NB	34	8	10	3451	0.77 (0.62; 0.88)	0.99 (0.99; 1.00)	1466.7 (545.4; 3943.8)	334.1 (164.2; 680)	0.23 (0.13; 0.40)
Pooled results (only for $I^2 \leq 75\%$)									***I^2 = 77.1%***	***I^2 = 97.0%***	***I^2 = 95.5%***		

Table 17 Studies measuring nasal bone, excluding T 21

Study details[1]	Study design	Weeks of pregnancy	Reference standard	Softmarker (cut-off)	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	+LR (95% CI)	-LR (95% CI)
Cicero_UK (2003)	Prospective cohort	11-14	karyotyping	Absence of NB	224	93	206	3265	0.52 (0.47; 0.57)	0.97 (0.97; 098)	38.17 (28.86; 50.5)	18.81 (15.1; 23.44)	0.49 (0.45; 0.54)
D'Alton_USA (2005)	Prospective cohort	10-14	karyotyping + pregnancy outcome	Absence of NB	1	21	10	4769	0.09 (0.00; 0.41)	1.00 (0.99; 1.00)	22.71 (2.78; 185.4)	20.74 (3.05; 141.0)	0.91 (0.76; 1.10)
Monni_Italy (2005)	Retrospective cohort	Not specified	karyotyping + pregnancy outcome	Absence of NB	79	76	62	16424	0.56 (0.47; 0.64)	0.99 (0.99; 1.00)	275.4 (184.3; 411.5)	121.6 (93.1; 159.0)	0.44 (0.37; 0.53)
Orlandi_USA (2003)	Prospective cohort	9-14	karyotyping + pregnancy outcome	Absence of NB	13	12	12	990	0.52 (0.31; 0.72)	0.99 (0.98; 0.99)	89.37 (33.90; 235.61)	43.42 (22.1; 85.44)	0.49 (0.32; 0.73)
Viora_UK (2003)	Prospective cohort	11-13	karyotyping	Absence of NB	12	24	7	1709	0.63 (0.38; 0.84)	0.99 (0.98; 0.99)	122.1 (44.22; 337.0)	45.6 (26.97; 77.11)	0.37 (0.21; 0.67)
Zoppi_Italy (2003)	Prospective cohort	Not specified	karyotyping	Absence of NB	34	8	10	3451	0.77 (0.62; 0.88)	0.99 (0.99; 1.00)	1466.7 (545.4; 3943.8)	334.1 (164.2; 680)	0.23 (0.13; 0.40)
Pooled results (only for $I^2 \leq 75\%$)									***I^2 = 77.1%***	***I^2 = 97.0%***	***I^2 = 95.5%***		

Table 18 Studies measuring risk, based on maternal age and NT, including T 21

Study details[1]	Study design	Weeks of pregnancy	Reference standard	Softmarker (cut-off)	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	+LR (95% CI)	-LR (95% CI)
Orlandi_USA (1997)	Prospective cohort	9-14	karyotyping	NT + chemistry	8	42	6	688	0.57 (0.29; 0.82)	0.94 (0.92; 0.96)	21.84 (7.24; 65.84)	9.93 (5.79; 17.05)	0.45 (0.25; 0.83)
Schwarzler_UK (1999)	Prospective cohort	10-14	karyotyping + pregnancy outcome	NT + maternal age	18	212	5	4288	0.78 (0.56; 0.92)	0.95 (0.94; 0.96)	72.81 (26.78; 198.0)	16.61 (12.91; 21.38)	0.23 (0.10; 0.49)
Tercanli_Germany (2003)	Prospective cohort	11-14	karyotyping	NT + maternal age (Risk 1:400)	33	186	4	1757	0.89 (0.75; 0.97)	0.90 (0.89; 0.92)	77.93 (27.31; 222.4)	9.31 (7.81; 11.12)	0.12 (0.05; 0.30)
Thilaganathan_UK (1999)	Prospective cohort	10-14	karyotyping + pregnancy outcome	NT + maternal age	40	440	9	9313	0.82 (0.68; 0.91)	0.95 (0.95; 0.96)	94.07 (45.36; 195.1)	18.09 (15.4; 21.26)	0.19 (0.11; 0.35)
Wojdemann_Denmark (2003)	Prospective cohort	11-14	karyotyping + pregnancy outcome	NT + maternal age	16	147	11	8448	0.59 (0.39; 0.78)	0.98 (0.98; 0.99)	83.59 (38.14; 183.23)	34.65 (24.38; 49.24)	0.41 (0.26; 0.65)
Pooled results (only for $I^2 \leq 75\%$)									***0.77 (0.69; 0.83)***	***I^2 = 98.6%***	***68.58 (43.26; 108.72)***		

Table 19 Studies measuring risk, based on maternal age and NT, excluding T 21

Study details[1]	Study design	Weeks of pregnancy	Reference standard	Softmarker (cut-off)	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	+LR (95% CI)	-LR (95% CI)
Schwarzler_UK (1999)	Prospective cohort	10-14	karyotyping + pregnancy outcome	NT + maternal age	8	212	3	4288	0.73 (0.39; 0.94)	0.95 (0.94; 0.96)	53.94 (14.21; 204.8)	15.44 (10.54; 22.69)	0.29 (0.11; 0.75)
Tercanli_Germany (2003)	Prospective cohort	11-14	karyotyping	NT + maternal age (Risk 1:400)	18	186	3	1757	0.86 (0.64; 0.97)	0.90 (0.89; 0.92)	56.68 (16.54; 194.2)	8.95 (7.17; 11.18)	0.16 (0.05; 0.45)
Thilaganathan_UK (1999)	Prospective cohort	10-14	karyotyping + pregnancy outcome	NT + maternal age	24	440	4	9313	0.86 (0.67; 0.96)	0.95 (0.95; 0.96)	127.0 (43.87; 367.6)	19.0 (15.92; 22.67)	0.15 (0.06; 0.37)
Wojdemann_Denmark (2003)	Prospective cohort	11-14	karyotyping + pregnancy outcome	NT + maternal age	8	147	7	8448	0.53 (0.27; 0.79)	0.98 (0.98; 0.99)	65.68 (23.51; 183.49)	31.18 (18.92; 51.40)	0.47 (0.28; 0.82)
Pooled results (only for $I^2 \leq 75\%$)									***0.77 (0.66; 0.86)***	***I^2 = 98.9%***	***74.28 (41.91; 131.64)***		

Table 20 Studies including different measurement parameters, including T 21

Study details[1]	Study design	Weeks of pregnancy	Reference standard	Softmarker (cut-off)	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	+LR (95% CI)	-LR (95% CI)
Blazer_Israel (2004)	Prospective cohort	12-16	karyotyping + pregnancy outcome	US + omphalocele	9	29	0	43858	1.00 (0.66; 1.00)	1.00 (1.00; 1.00)	28247 (1607; 496537)	1413.3 (959.1; 2082.8)	0.05 (0.00; 0.75)
Metzenbauer_Austria (2002)	Prospective cohort	10-13	karyotyping + pregnancy outcome	Placental volume + NT	9	279	8	2567	0.53 (0.28; 0.77)	0.90 (0.89; 0.91)	10.35 (3.96; 27.04)	5.40 (3.40; 8.57)	0.52 (0.31; 0.86)
Monni_Italy (2005)	Retrospective cohort	11-14	karyotyping + pregnancy outcome	NT + NB	110	744	31	15769	0.78 (0.70; 0.84)	0.95 (0.95; 0.96)	75.21 (50.15; 112.8)	17.31 (15.48; 19.37)	0.23 (0.17; 0.31)
O'Callaghan_Australia (2000)	Prospective cohort	11-14	karyotyping + pregnancy outcome	US + NT	9	59	2	930	0.82 (0.48; 0.98)	0.94 (0.92; 0.95)	70.93 (14.99; 335.74)	13.71 (9.44; 19.91)	0.19 (0.05; 0.68)
Pooled results (only for $I^2 \leq 75\%$)									***I^2 = 68.9%***	***I^2 = 99.9%***	***I^2 = 90.6%***		

Table 21 Studies including different measurement parameters, excluding T 21

Study details[1]	Study design	Weeks of pregnancy	Reference standard	Softmarker (cut-off)	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	+LR (95% CI)	-LR (95% CI)
Blazer_Israel (2004)	Prospective cohort	12-16	karyotyping + pregnancy outcome	US + omphalocele	7	29	0	43858	1.00 (0.59; 1.00)	1.00 (1.00; 1.00)	22301 (1245; 399421)	1394.7 (932.4; 2086.3)	0.06 (0.00; 0.92)
Metzenbauer_Austria (2002)	Prospective cohort	10-13	karyotyping + pregnancy outcome	Placental volume + NT	5	279	3	2567	0.62 (0.24; 0.91)	0.90 (0.89; 0.91)	15.33 (3.64; 64.51)	6.37 (3.68; 11.03)	0.42 (0.17; 1.02)
Monni_Italy (2005)	Retrospective cohort	11-14	karyotyping + pregnancy outcome	NT + NB	38	744	7	15769	0.84 (0.70; 0.93)	0.95 (0.95; 0.96)	115.1 (51.21; 258.5)	18.74 (16.23; 21.64)	0.16 (0.08; 0.32)
O'Callaghan_Australia (2000)	Prospective cohort	11-14	karyotyping + pregnancy outcome	US + NT	2	59	1	930	0.67 (0.09; 0.99)	0.94 (0.92; 0.95)	31.52 (2.81; 352.71)	11.17 (4.84; 25.82)	0.35 (0.07; 1.75)
Pooled results (only for $I^2 \leq 75\%$)									***I^2 = 40.6%***	***I^2 = 99.9%***	***I^2 = 85.9%***		

Table 22 Results - Overview for DOR, sensitivity and specificity ranges

Measurement parameter	DOR ranges	Sensitivity ranges	Specifity ranges
NT incl. T21	15-236	41-85%	87-100%
NT excl. T21	12-611	50-94%	87-100%
Nb incl. T21	22-1466	9-77%	97-100%
Nb excl. T21	17-3235	33-88%	97-100%
Risk	21-94	57-89%	90-98%

Table 23 Differences in the two groups in the study of Bennett et al.

Endpoint	First trimester scan group	Second trimester scan group
Adjustment of gestational age	41,3%	10,9%
Labour induction	4,8%	13,0%
Delivery at $\geq$ 287days	6,7%	16,3%

Table 24 Comparison of advantages of first versus second trimester ultrasound in Bennett et al.

First trimester scan	Second trimester scan
Dating within days ($\pm$ 4-5 days)	Dating within weeks ($\pm$ 7-14 days)
Early diagnosis of missed abortion and ectopic pregnancy	Diagnosis of congenital anomalies
Early diagnosis of multiple pregnancy	

Table 25 Differences in Scan group and no scan group in Harrington et al.

Endpoint	Scan group	No scan group
Induction rate for prolonged pregnancy	8,15%	7,39%
Correction of EDD	5,5%	0,87%
Spontaneuous labour	66,5%	57,4%
Suspected FGR	1,2%	3,4%

Table 26 Reported results from Kelecky et al.

	NT normal	NT >95 percentile	p-value
Abnormal 50g OGTT	56 (14,5%)	54(13,9%)	0,626
Abnormal 100g OGTT	8 (2,1%)	14 (3,6%)	0,048
Gestational diabetes (prevalence)	9 (2,3%)	10 (2,6%)	0,795
Macrosomia	17 (4,4%)	24 (6,2%)	0,045

Table 27 Assessors' experience and equipment details

	assessor of test	assessor experience	type of measurement	technical quality	instrument
Acacio(2001)	Obstetrician	FMF certificate	TA and/or TV*	unclear/not reported	different instruments
Bennett(2004)	Obstetrician	unclear	TA and/or TV	unclear/not reported	not reported
Blazer(2004)	Sonographer/Technician	>2 years experience	TA and/or TV	6 MHz	ESI 3000
Brizot(2001)	Sonographer/Technician	FMF certificate	TA and/or TV	unclear/not reported	not reported
Cheng(2006)	Obstetrician	unclear	TA	6 MHz	Aloka 2100
Cicero(2003)	Sonographer/Technician	unclear	unknown	unclear/not reported	not reported
D'Alton(2005)	Sonographer/Technician	uniform training	unknown	unclear/not reported	not reported
D'Ottavio(1997)	other or unclear	unclear	TV	unclear/not reported	not reported
D'Ottavio(1998)	other or unclear	unclear	TA and/or TV	3,5-5MHz	Accuson 128XP
Hafner(19989	Obstetrician	unclear	TA and/or TV	5/8MHz	Accuson 128XP
Harrington(2006)	Sonographer/Technician	unclear	TA and/or TV	6 MHz	Toshiba SSA250 or 270
Kelecki(2005)	Obstetrician	FMF certificate	unknown	unclear/not reported	not reported
Menon(2005)	Sonographer/Technician	unclear	TA and/or TV	3,5-5MHz	Toshiba SSA250 or 271
Metzenbauer(2002)	Sonographer/Technician	unclear	unknown	unclear/not reported	Voluson 530 or 730
Monni(2005)	other or unclear	unclear	TA	unclear/not reported	not reported
O'Callaghan(2000)	GP	FMF certificate	unknown	unclear/not reported	different instruments
Orlandi(2003)	other or unclear	unclear	unknown	unclear/not reported	not reported
Schwarzler(1999)	Sonographer/Technician	unclear	TA and/or TV	unclear/not reported	not reported
Snijders(1998)	Sonographer/Technician	FMF certificate	TA and/or TV	unclear/not reported	different instruments
Tercanli(2002)	Obstetrician	FMF certificate	unknown	unclear/not reported	different instruments
Thilaganathan(1999)	Midwife/Nurse	FMF certificate	TA	3,5-7,5MHz	different instruments
Viora(2003)	Sonographer/Technician	unclear	TA	3,5-5MHz	Aloka 2100
Wayda(2001)	Obstetrician	unclear	TV	6,5-7,5MHz	Combison 530
Wojdemann(2005)	other or unclear	FMF certificate	TA and/or TV	5/8MHz	Three Logic 700 MR
Zoppi(2003)	other or unclear	unclear	unknown	unclear/not reported	not reported

Table 28 Citation list of the studies identified for the Update

Benoit B, Levaillant JM. [Ultrasonography in twin pregnancies]. Rev Prat. 2006 Dec 31;56(20):2236-7.
Breathnach FM, Malone FD, Lambert-Messerlian G, Cuckle HS, Porter TF, Nyberg DA, Comstock CH, Saade GR, Berkowitz First- and second-trimester screening: detection of aneuploidies other than Down syndrome. Obstet Gynecol. 2007 Sep;110(3):651-7.
Czuba B, Borowski D, Cnota W, Sieroszewski P, Grettka K, Pietryga M, Wyrwas D, Czekierdowski A, Wloch A, Wielgos M, Ultrasonographic assessment of fetal nuchal translucency (NT) at 11th and 14th week of gestation--Polish multicentre study. Neuro Endocrinol Lett. 2007 Apr;28(2):175-81.
El Kateb A, Nasr B, Nassar M, Bernard JP, Ville Y. First-trimester ultrasound examination and the outcome of monochorionic twin pregnancies. Prenat Diagn. 2007 Oct;27(10):922-5.
Pons JC, Hoffmann P, Bringer S, Deutsch V, Lisik F, Schaal JP. [Management of twin pregnancy]. Rev Prat. 2006 Dec 31;56(20):2227-35.
Evans MI, Van Decruyes H, Nicolaides KH. Nuchal translucency measurements for first-trimester screening: the 'price' of inaccuracy. Fetal Diagn Ther. 2007;22(6):401-4. Epub 2007 Jul 24.
Scott A. Nuchal translucency measurement in first trimester Down syndrome screening. Issues Emerg Health Technol. 2007 Jun;(100):1-6.
Watson WJ, Miller RC, Wax JR, Hansen WF, Yamamura Y, Polzin WJ. Sonographic detection of trisomy 13 in the first and second trimesters of pregnancy. J Ultrasound Med. 2007 Sep;26(9):1209-14.
Westin M, Saltvedt S, Almstrom H, Grunewald C, Valentin L. By how much does increased nuchal translucency increase the risk of adverse pregnancy outcome in chromosomally normal Ultrasound Obstet Gynecol. 2007 Feb;29(2):150-8.