Received Date : 07-Jun-2016

Revised Date : 17-Oct-2016

Accepted Date : 19-Oct-2016

Article type : Systematic review

Analysis of cell-free fetal DNA in maternal blood for detection of trisomy 21, 18 and 13 in a general pregnant population and in a high risk population - a systematic review and meta-analysis

Running headline: Detection of trisomy 21, 18 and 13 using cfDNA

Erik IWARSSON[1,*]; Bo JACOBSSON[2,3,*]; Jessica DAGERHAMN[4]; Thomas DAVIDSON[4,5]; Eduardo BERNABÉ[6] & Marianne HEIBERT ARNLIND[4,7]

* These authors contributed equally.

[1]Department of Molecular Medicine and Surgery, Clinical genetics unit, Karolinska Institute, Karolinska University,Stockholm, [2]Department of Obstetrics and Gynecology, Sahlgrenska Academy, Gothenburg University, Gothenburg, Sweden, [3]Department of Genetics and Bioinformatics, Area of Health Data and Digitalisation, Institute of Public Health, Oslo, Norway, [4]Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU), Stockholm, [5]Division of Health Care Analysis, Department of Medical and Health Sciences, Linköping University, Linköping, Sweden, [6]Division of Population and Patient Health, King's College London Dental Institute at Guy's, King's College and St. Thomas Hospitals, London, UK, [7]

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/aogs.13047

Medical Management Centre/LIME, Karolinska Institute, Stockholm, Sweden

Corresponding author:

Erik Iwarsson

Department of Molecular Medicine and Surgery, Karolinska Institutet, Karolinska University Hospital, L5:03, 171 76 Stockholm, Sweden

E-mail: Erik.Iwarsson@ki.se

Conflict of interest

BJ has participated in four studies funded by Ariosa Diagnostics. Ariosa Diagnostics paid a fee per recruited patient. The rest of authors declare that they have no conflicts of interest.

Abstract

Introduction: To review the performance of non-invasive prenatal testing (NIPT) for detection of trisomy 21, 18 and 13 (T21, T18 and T13) in a general pregnant population as well as to update the data on high-risk pregnancies. *Material and methods:* Systematic Review and Meta-Analysis. PubMed, Embase and the Cochrane Library were searched. Methodological quality was rated using QUADAS and scientific evidence using GRADE. Summary measures of diagnostic accuracy were calculated using a bivariate random-effects model. *Results:* In a general pregnant population, there is moderate evidence that the pooled sensitivity is 0.993 (95% 0.955 to 0.999) and specificity was 0.999 (95% 0.998 to 0.999) for the analysis of T21. Pooled sensitivity and specificity for T13 and T18 was not calculated in this population due to the low number of studies. In a high-risk pregnant population, there is moderate evidence that the pooled sensitivities for T21 and T18 are 0.998 (95% CI 0.981 to 0.999) and 0.977 (95% CI 0.958 to 0.987) respectively, and low evidence that the pooled sensitivity for T13 is 0.975 (95% CI 0.819 to 0.997). The pooled specificity for all three trisomies is 0.999 (95% 0.998 to 0.999). *Conclusions:* This is the first meta-analysis using GRADE that shows that NIPT performs well as a screen for trisomy 21 in a general pregnant population. Although the false positive rate is low compared to First Trimester Combined Screening, women should still be advised to confirm a positive result by invasive testing if termination of pregnancy is under consideration.

Key words

trisomy, prenatal diagnosis, non-invasive prenatal testing, NIPT, cell-free DNA, cfDNA, fetus, chromosome aberration, aneuploidy, pregnancy

Abbreviations

AUC	area under the curve
cfDNA	cell-free DNA
DOR	diagnostic odds ratios
FN	false negatives
FP	false positive
FTS	first trimester combined screening
LR-	negative likelihood ratio
LR+	positive likelihood ratio
MPSS	massive parallel shotgun sequencing
NIPT	non-invasive prenatal testing
PICO	P – population, I – index test, C – control/reference test and O – outcome
SNP	single nucleotide polymorphisms
SROC	summary receiver operating characteristics
T13	trisomy 13
T18	trisomy 18
T21	trisomy 21
t-MPS	targeted massive parallel sequencing
TN	true negatives
TP	true positives

Key message

Non-invasive prenatal testing performs well as a screen for trisomy 21 in a general pregnant population.

Introduction

Prenatal diagnosis, including screening and diagnosis of chromosome aberrations, has been offered in various forms as part of prenatal care during the last 40 years (1). Diagnosis of chromosome aberrations requires either first-trimester chorionic villus sampling (CVS) or second-trimester amniocentesis. However, these invasive procedures entail a miscarriage risk of 0.1 – 0.5 % (2, 3).

The presence of fetal cell-free DNA (cfDNA) in maternal circulation was first demonstrated by Lo et al. (4). This finding led to the discovery that cfDNA obtained from maternal plasma could be used for fetal aneuploidy analysis, termed non-invasive prenatal testing (NIPT), a long-awaited improvement to reduce invasive procedures and accompanying miscarriage risk (5, 6). NIPT for fetal aneuploidy analysis was introduced clinically in 2011 and is implemented in many countries worldwide; more than 2 million procedures have hitherto been performed (7). Many studies have investigated analysis of cfDNA in maternal blood for detection of trisomies 21, 18 and 13 in women at high risk of aneuploidy, finding weighted pooled detection rates of 99.2% for trisomy 21 (T21), 96.3% for trisomy 18 (T18) and 91.0% for trisomy 13 (T13), as well as false positive (FP) rates of 0.09% for T21 and 0.13% for T18 and T13 (8). However, significant data from large studies on test performance in a general (i.e. average-risk) pregnant population have, until recently, been lacking (9, 10). The objective of this meta-analysis was to update the data on high-risk pregnancies, including studies published up until April 2015, and, more importantly, present data concerning test performance in a general pregnant population at average risk of aneuploidy.

Material and methods

The systematic literature review

The Cochrane Collaboration definition of a systematic review was applied, i.e. “A systematic

review attempts to identify, appraise and synthesize all the empirical evidence that meets pre-specified eligibility criteria to answer a given research question. Researchers conducting systematic reviews use explicit methods aimed at minimizing bias, in order to produce more reliable findings that can be used to inform decision making." (www.thecochranelibrary.com).

Literature search

The literature search included the databases PubMed, Embase, and the Cochrane Library up until April 2, 2015. The MeSH terms used were: "Down Syndrome", "Patau Syndrome", "Trisomy 18-Like Syndrome", "Chromosomes, Human, Pair 13", "Chromosomes, Human, Pair 18", and "Chromosomes, Human, Pair 21". In addition to MeSH terms, free-text words were used. For the search block regarding NIPT, only free- text terms were used, since there was no MeSH term for this concept. Detailed information about the search strategy can be found at http://www.sbu.se/upload/Publikationer/Content0/3/NIPT/Bilaga%205%20S%C3%B6kstrategier.pdf

Inclusion and exclusion criteria

Study selection was based on the following criteria using PICO (P – population, I – index test, C – control/reference test and O – outcome) (11). Population 1: Pregnant women at high risk of carrying a fetus with chromosome aberration (as determined by study authors, which could include women of varying risk level but defined as being at high risk because of different risk factors such as e.g. assessed being at high-risk on biochemical screening, first trimester combined screening (FTS), abnormal ultrasound scan or maternal age). Population 2: Pregnant women at average risk of carrying a fetus with chromosome aberration, i.e. a general pregnant population. The index test was NIPT, using cfDNA, of trisomies 13, 18, or 21. Invasive genetic testing or phenotype at birth were accepted as reference tests. Outcome measures were sensitivity, specificity and number of true positives (TP), FP, true negatives (TN) and false negatives (FN). The complementary inclusion criteria were primary study in English or Scandinavian language on the analysis of trisomy 13, 18 or 21 in singleton pregnancies, published in 1998-2015 in a peer-review journal, with reported, or data enabling calculation of, sensitivity and specificity. Exclusion criteria were RNA analysis, absence of primary data, study population < 100 women and abstract/letter/review. Formal screening of search results against inclusion and exclusion criteria and risk of bias (quality) assessment was performed according to

a pre-specified protocol, PROSPERO registration number CRD42015020076, available at http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015020076.

Study selection and data extraction

Two authors (EI, BJ) individually reviewed all abstracts and made separate decisions based on inclusion and exclusion criteria. If at least one reviewer was considering an abstract for inclusion, the full-text article was reviewed. The reviewers individually decided whether a study could be included and contained data suitable for the subsequent analysis of clinical validity, and extracted the relevant data from each selected study using a standard form. The review form was designed to capture primary data, including study type, number of samples, FN and FP results, sensitivity and specificity levels, indeterminate cases, reference test (i.e. method used to confirm NIPT results), methodology and whether the study was sponsored by a commercial company. Inter-reviewer discrepancies were resolved by discussion. Where one of the review authors was co-author of a selected study, MHA and JD replaced that author in the quality review.

Rating Quality of individual studies

The quality of each included study was rated as high, moderate or low using the QUADAS tool (12). Only studies of high or moderate quality were considered good enough for grading of scientific evidence and conclusions.

Data synthesis

All statistical analyses were conducted using the metandi and midas commands in Stata 13 (StataCorp, College Station, Texas). A bivariate random-effects model was used to estimate average sensitivity, specificity, positive and negative likelihood ratios (LR+ and LR-) and diagnostic odds ratios (DOR), with 95% confidence intervals (CI). We also constructed the summary receiver operating characteristics (SROC) curve and the corresponding area under the curve (AUC) to summarize overall test performance (13).

We used coupled forest plots and measures of variability (variances and covariance of sensitivity and specificity across studies) to assess between-study heterogeneity (14, 15). We also undertook sensitivity analysis (excluding influential studies) to verify robustness of results. Finally, we assessed publication bias by plotting DOR against the effective sample size. With no bias, the plot should have an inverted symmetrical funnel shape. The degree of asymmetry was statistically assessed by regression of the logarithm of the DOR on the inverse of the square root of the effective sample size, weighted by effective sample size (16).

Grading the scientific evidence across studies

The quality of scientific evidence of the outcomes of PICO was rated according to the four GRADE levels (17). High (⊕⊕⊕⊕) - we are very confident that the true effect lies close to that of the estimate of the effect; Moderate ⊕⊕⊕O - we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different; Low ⊕⊕OO - our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect; Very Low ⊕OOO - we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

The rated GRADE level is usually initially high, but confidence in the evidence may decrease stepwise during the analysis process for several reasons, including limitations in study design and/or quality, inconsistency or indirectness of results, imprecise estimates and probability of publication bias. Any disagreements on inclusion/exclusion criteria, rated quality of individual studies or quality of evidence of test methods were solved within the reviewer group by consensus.

Results

Selection of studies

In this systematic review, 882 abstracts met the search criteria; 453 of them were excluded based on inclusion and exclusion criteria. This resulted in 429 published full-text articles that were assessed regarding inclusion and exclusion criteria. During this step, 376 articles were excluded (Supporting Information Table S1). The scientific quality of the articles meeting the inclusion and exclusion criteria was assessed using the QUADAS tool (12). Only studies with high

or moderate quality were considered suitable to use for grading the quality of scientific evidence. Twenty-two were assessed as low-quality and excluded from analysis (Supporting Information Table S2), leaving 31 articles (32 studies) for further analysis (Figure 1) (18-48). Two other studies (49, 50) were excluded since they reported on the same patients as in Jensen et al (29).

Study characteristics

The characteristics of the included studies are summarized in Supporting Information Tables S3-5. Twenty-three studies were prospective cohort studies (21, 25-27, 31-48) and nine were case-control-studies (18-20, 22-24, 28-30). The majority of sampling had occurred during the first trimester. Study designs varied and there were also variations in planning and execution, e.g. samples may have been frozen or not and results were reported back to patients in some studies. In the majority of studies results were not reported back to patients. Generally, the number of failed analyses and need for repeat sampling are not well reported. Six of the 32 included studies did not report having a commercial partner (18, 25, 34, 39, 41, 51).

The included studies were published from January 2011 to April 2, 2015. Nine studies were from centers offering NIPT for trisomies as part of a clinical service, with the results reported back to the patient (25, 26, 31, 39, 40, 42, 46-48). The remaining 23 used biobanked samples (18-24, 27-30, 32-38, 41, 43-45). A reference test (inclusion criterion), either an invasive prenatal test or postnatal examination (phenotype), was used in all studies.

Based on QUADAS and taking into account existence of commercial partner and study design, all of the included studies were assessed as being of moderate quality (risk of bias graph shown in Supporting Information Table S6).

Five of the included studies were considered to investigate a general pregnant population (i.e. an average-risk population) (21, 33, 36, 37, 42). Two studies included both a high-risk and an average-risk population (40, 47). The remaining 25 studied a high-risk population. Thirteen of the 32 studies had a population exceeding 1,000 pregnancies (21, 26, 29, 31, 33, 35, 36, 38, 39, 42, 46-48).

In the case of the average-risk pregnant population, seven included studies reported on the performance of cfDNA analysis for T21, with a total of 156 TP T21 and 62,107 non-T21 TN singleton pregnancies (Table 1) (21, 33, 36, 37, 40, 42, 47). When it came to T18, there were a

total of 15 TP cases and 21,989 TN singleton pregnancies in the included six studies (21, 33, 36, 37, 40, 42). As for T13, there were a total of six TP cases and 14,384 TN singleton pregnancies in the included five studies (Supporting Information Tables S7 and S8) (21, 36, 37, 40, 42). In the high-risk pregnant population, there were 1,839 TP T21 cases in total and the number of included singleton pregnancies exceeded 100,000 for all three trisomies (Table 1).

The population in the study by Lau et al. (31) was classified as high-risk in our meta-analysis, since 46% of the women had an increased risk of carrying a fetus with a chromosome aberration; furthermore the median maternal age was 36. Dan et al. (26) report on 11,105 clinical NIPT analyses, but sensitivity and specificity calculations were only based on the 3,000 samples for which results of an invasive prenatal test were available. In this case, we decided to re-calculate sensitivity and specificity, in order to also include cases in which results of a postnatal clinical examination were available as a reference test. This made it possible to include 7,524 pregnancies in our meta-analysis.

Meta-analyses

A significant level of heterogeneity was observed, with greater variance in sensitivity than specificity for T21 and T13 in the high-risk population but greater variance in specifity than sensitivity for T18 in the high-risk population and T21 in the average risk population (Supporting Information Table S9). The corresponding prediction ellipses were not informative as most studies generated estimates in the upper left hand corner of the ROC plot.

The meta-analyses for T21 in the high-risk population yield a pooled sensitivity of 0.998 (95% CI 0.981 to 0.999) (Figure 2). In the average-risk population, i.e. the general pregnant population, pooled sensitivity was 0.993 (95% 0.955 to 0.999) (Figure 3). In the case of T18 in the high-risk population, pooled sensitivity was 0.977 (95% CI 0.958 to 0.987) (Figure 4) and the corresponding figure for T13 was 0.975 (95% CI 0.819 to 0.997) (Figure 5). Sensitivity and specificity for T18 and T13 could not be calculated in the average-risk population due to the low number of studies (Figure 6). For trisomies 21, 18, and 13 in the high-risk population, and for T21 in the average-risk population, pooled specificity was 0.999 (95% 0.998 to 0.999) (Figures 2-5). Individual studies clustered in the upper left-hand corner of their corresponding SROC curves, with AUC values of 1 (95% CI: 0.99 to 1.00) for trisomies 21, 18 and 13 in the high-risk population and 0.99 (95% CI: 0.98 to 1.00) for T21 in the average-risk population. Other measures of diagnostic performance (LR+ and LR- and DOR) are reported in Supporting

Information Table S10.

We found evidence of publication bias for T21 in the high-risk (p=0.001) and average-risk populations (p=0.018), as well as for T18 in the high-risk population (p=0.010). There was no evidence of publication bias for T13 in the high-risk population (p=0.085).

GRADE

GRADE was used to determine confidence in the pooled estimates of sensitivity and specificity. The quality of evidence was considered to be moderate for T21 and T18 in the high-risk population, as well as for T21 in the average-risk population. When it came to T13 in the high-risk population, the quality of evidence was limited (Table 1). No meta-analysis or grading for T13 and T18 in the average-risk population was performed due to lack of data (Figure 6).

False positive and false negative results

In this systematic review, the proportion of FP ranged between 2.7 % (T21 in the high-risk population) and 30 % (T13 in the general population) (Supporting Information Table S7). The proportion of FN was generally very low, i.e. 0.01 % at most (Supporting Information Table S8).

Discussion

This systematic review and meta-analysis show a pooled sensitivity of 0.993 (95% 0.955 to 0.999) and a pooled specificity of 0.999 (95% 0.998 to 0.999) for T21 detection with NIPT in the general pregnant population. Corresponding values for T13 and T18 could not be calculated due to the low number of studies as well as an insufficient number of trisomy cases, compared to non-trisomy cases, in the pool. The majority of studies on cfDNA analysis for aneuploidy detection have so far been performed in selected high-risk pregnant populations. The findings in this meta-analysis extend those of previous reviews by adding data from studies in general pregnant (average-risk) populations. Due to the recent publication of several additional studies, including two very large ones (36, 47), our study has enhanced power to estimate the performance of cfDNA analysis in a general pregnant population. Moreover, there is now additional data on NIPT performance in a high-risk population. Pooled sensitivities in the selected high-risk pregnant population were 0.998 (95% CI 0.981 to 0.999), 0.977 (95% CI 0.958

to 0.987) and 0.975 (95% CI 0.819 to 0.997) for T21, T18 and T13, respectively. The pooled specificity in the high-risk population for trisomies 21, 18, and 13 is 0.999 (95% 0.998 to 0.999) (Figures 2-5). In the case of T21, our meta-analysis covered a sample consisting of 169,675 singleton pregnancies and 2,004 trisomy cases, including an average-risk sample of 62,201 with 157 T21 cases, about eight times as many pregnant women and twice as many trisomy cases as in a recent review on this topic by Gil et al. (8). The review by Gil et al, with a total of 1,051 T21 cases and 21,608 unaffected fetuses, showed weighted pooled detection rates of 99.2% for T21, 96.3% for T18 and 91.0% for T13 (8).

The aim of applying GRADE is to ascertain how much confidence can be placed in a particular estimate of effect, whether the result will be sustainable and whether it is likely that new research will change the evidence. There is moderate quality of evidence, according to GRADE, underlying the pooled sensitivity of 0.993 and the pooled specificity of 0.999 for T21 detection in the general pregnant population, i.e. this result is likely to be close to the true effect. However, further improvement of cfDNA analysis may of course change these figures by enhancing method performance. GRADE assessment also demonstrates moderate quality of evidence for T21 and T18 detection in the high-risk pregnant population. Due to study quality and imprecision, the quality of evidence for T13 detection failed to reach the moderate level and was found to be low in the high-risk population. For T18 and T13, it was not possible to determine the quality of evidence in the general population due to insufficient data for these low-prevalence trisomies. Deeks' symmetry test suggested the existence of publication bias in three of the four meta-analyses (T21 in the high-risk and average-risk populations, and T18 in the high-risk population but not for T13 in the high-risk population). The cfDNA technique was developed in a commercial setting with an early introduction to the clinical market. Only six of the 32 included studies did not report having a commercial partner (18, 25, 31, 34, 39, 41). The general impression is that commercial interests may affect results, and the meta-analysis by Taylor et al confirm the presence of publication bias (9). However we have nonetheless chosen not to degrade these findings according to the GRADE protocol.

The two recently published meta-analyses that show separate data from a general pregnant population (average-risk) differ to ours in outcome (9, 10). The pooled sensitivity in the general pregnant population is higher in our meta-analysis (0.993) compared to the systematic review by Taylor et al (0.959) (9). Our reviews differ in some aspects, e.g. inclusion and exclusion criteria where we choose to exclude studies judged to be at high risk of bias (low quality studies) from the meta-analysis using QUADAS. In addition, we do not report pooled sensitivity

for T13 and T18 detection in the general pregnant population due to the low number of studies and limited data. The main difference of our data compared to Mackie et al (10) is the huge discrepancy when it comes to the number of patients included in the two different sub-populations high- and normal risk (Table 1). This might at least partly explain the different outcomes, where our meta-analysis shows a difference in sensitivity for high- compared to normal risk population (0.998 compared to 0.993), a difference that the meta-analysis by Mackie et al. do not show. In addition, our analysis shows a somewhat higher sensitivity for T21 analysis in a high-risk population 0.998 compared to Mackie et al (0.994) (10).

The majority of the included studies were not performed in a clinical context but were performed retrospectively, using frozen biobanked plasma samples, with results not reported back to the patient (18-20, 22-24, 27-30, 32-34, 37, 38, 41, 43-45). This might have affected test performance when applied clinically, e.g. the no-result rate may have been underestimated. However, several larger studies from a clinical setting have been published during 2014-2015 and the number of included patients in clinical studies thus exceeds by far the number of patients in the biobank studies (25, 26, 31, 35, 36, 39, 40, 42, 46-48). This limits the risk of such clinical consequences. The percentages of samples not generating a report back to the referring doctor is one aspect of cfDNA's clinical usefulness that is insufficiently reported in most of the studies. There can be numerous reasons for this, on different levels of the process at which sample analysis might be problematic: a sample might not fulfill the pre-analytic quality criteria (e.g. inadequate blood volume, incorrect labeling of tubes and delay in arrival at the laboratory) or analytic quality criteria (e.g. low fetal fraction or assay failure). In the studies performed in a clinical-like setting, an additional sample was required in 0.9-4.6% of cases (21, 24, 26, 31, 35, 36, 39, 42, 47, 48), in addition to the samples failing to meet the pre-analytical criteria. None of the studies included in this meta-analysis were analyzed by "intention-to-diagnose", which would have made it easier to interpret the findings from a clinical perspective. The NIPT assay used for trisomy detection is based on new sequence technology and three different approaches are in clinical use, i.e. massive parallel (shotgun) sequencing (MPSS), targeted massive parallel sequencing (t-MPS), and t-MPS with the use of single nucleotide polymorphisms (SNP), with the different respective pros and cons. The majority of the included studies used MPSS, followed by t-MPS and SNP-based analysis. In this review, we have not studied the approaches separately.

This is the first meta-analysis using GRADE that includes a sufficient sample size to permit the conclusion that NIPT performs well as a screen for fetal T21 in a general pregnant population. More data is needed concerning T18 and T13. As the data in this meta-analysis suggests nearly equally good test performance in the general population, the scenario of using cfDNA instead of FTS will be increasingly considered. This approach will have the advantage of detecting more aneuploid pregnancies, nearly eliminating false reassurance and significantly reducing the number of women requiring an invasive test for confirmation. The major limiting factor for this development is the cost of the NIPT assay, which is still at least double the cost of FTS. Moreover, although the FP rate is low compared to FTS, women should still be advised to confirm a positive NIPT result by invasive testing if termination of pregnancy is under consideration.

Acknowledgements

We are grateful for support from the Swedish Agency for Health Technology Assessment and Assessment of Social Service (SBU) and Stockholm County Council.

References

1. Nicolaides KH. Screening for fetal aneuploidies at 11 to 13 weeks. Prenat Diagn. 2011;31:7-15.

2. Akolekar R, Beta J, Picciarelli G, Ogilvie C, D'Antonio F. Procedure-related risk of miscarriage following amniocentesis and chorionic villus sampling: a systematic review and meta-analysis. Ultrasound Obstet Gynecol. 2015;45:16-26.

3. Simpson JL. Invasive procedures for prenatal diagnosis: Any future left? Best Pract Res Clin Obstet Gynaecol. 2012;26(5):625-38.

4. Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, et al. Presence of fetal DNA in maternal plasma and serum. Lancet. 1997;350:485-7.

5. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proceedings of the National Academy

of Sciences of the United States of America. 2008;105:16266-71.

6. Lo YM, Lun FM, Chan KC, Tsui NB, Chong KC, Lau TK, et al. Digital PCR for the molecular detection of fetal chromosomal aneuploidy. Proc Natl Acad Sci U S A. 2007;104(32):13116-21.

7. Bianchi DW. Pregnancy: Prepare for unexpected prenatal test results. Nature. 2015;522:29-30.

8. Gil MM, Quezada MS, Revello R, Akolekar R, Nicolaides KH. Analysis of cell-free DNA in maternal blood in screening for fetal aneuploidies: updated meta-analysis. Ultrasound Obstet Gynecol. 2015;45:249-66.

9. Taylor-Phillips S, Freeman K, Geppert J, Agbebiyi A, Uthman OA, Madan J, et al. Accuracy of non-invasive prenatal testing using cell-free DNA for detection of Down, Edwards and Patau syndromes: a systematic review and meta-analysis. BMJ open. 2016;6:e010002.

10. Mackie FL, Hemming K, Allen S, Morris RK, Kilby MD. The accuracy of cell-free fetal DNA-based non-invasive prenatal testing in singleton pregnancies: a systematic review and bivariate meta-analysis. BJOG. 2016;

11. Boudin F, Nie JY, Bartlett JC, Grad R, Pluye P, Dawes M. Combining classifiers for robust PICO element detection. BMC Med Inform Decis Mak. 2010;10:29.

12. 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.

13. Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. Ann Thorac Surg. 2005;79:16-20.

14. Naaktgeboren CA, Ochodo EA, Van Enst WA, de Groot JA, Hooft L, Leeflang MM, et al. Assessing variability in results in systematic reviews of diagnostic studies. BMC Med Res Methodol. 2016;16:6.

15. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. BMJ. 2011;342:d549.

16. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol.

2005;58:882-93.

17. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64:401-6.

18. Alberti A, Salomon LJ, Le Lorc'h M, Couloux A, Bussieres L, Goupil S, et al. Non-invasive prenatal testing for trisomy 21 based on analysis of cell-free fetal DNA circulating in the maternal plasma. Prenat Diagn. 2015;35:471-6.

19. Ashoor G, Syngelaki A, Wagner M, Birdir C, Nicolaides KH. Chromosome-selective sequencing of maternal plasma cell-free DNA for first-trimester detection of trisomy 21 and trisomy 18. American journal of obstetrics and gynecology. 2012;206:322.e1-5.

20. Ashoor G, Syngelaki A, Wang E, Struble C, Oliphant A, Song K, et al. Trisomy 13 detection in the first trimester of pregnancy using a chromosome-selective cell-free DNA analysis method. Ultrasound Obstet Gynecol. 2013;41:21-5.

21. Bianchi DW, Parker RL, Wentworth J, Madankumar R, Saffer C, Das AF, et al. DNA sequencing versus standard prenatal aneuploidy screening. N Engl J Med. 2014;370:799-808.

22. Bianchi DW, Platt LD, Goldberg JD, Abuhamad AZ, Sehnert AJ, Rava RP. Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. Obstet Gynecol. 2012;119:890-901.

23. Chen EZ, Chiu RW, Sun H, Akolekar R, Chan KC, Leung TY, et al. Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. PloS one. 2011;6:e21791.

24. Chiu RW, Akolekar R, Zheng YW, Leung TY, Sun H, Chan KC, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ (Clinical research ed). 2011;342:c7401.

25. Comas C, Echevarria M, Rodriguez MA, Prats P, Rodriguez I, Serra B. Initial experience with non-invasive prenatal testing of cell-free DNA for major chromosomal anomalies in a clinical setting. J Matern Fetal Neonatal Med. 2015;28(10):1196-201.

26. Dan S, Wang W, Ren J, Li Y, Hu H, Xu Z, et al. Clinical application of massively parallel sequencing-based prenatal noninvasive fetal trisomy test for trisomies 21 and 18 in 11,105 pregnancies with mixed risk factors. Prenat Diagn. 2012;32:1225-32.

27. Ehrich M, Deciu C, Zwiefelhofer T, Tynan JA, Cagasan L, Tim R, et al. Noninvasive

detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. Am J Obstet Gynecol. 2011;204(3):205.e1-11.

28. Guex N, Iseli C, Syngelaki A, Deluen C, Pescia G, Nicolaides KH, et al. A robust second-generation genome-wide test for fetal aneuploidy based on shotgun sequencing cell-free DNA in maternal blood. Prenat Diagn. 2013;33(7):707-10.

29. Jensen TJ, Zwiefelhofer T, Tim RC, Dzakula Z, Kim SK, Mazloom AR, et al. High-throughput massively parallel sequencing for fetal aneuploidy detection from maternal plasma. PloS one. 2013;8:e57381.

30. Jiang F, Ren J, Chen F, Zhou Y, Xie J, Dan S, et al. Noninvasive Fetal Trisomy (NIFTY) test: an advanced noninvasive prenatal diagnosis methodology for fetal autosomal and sex chromosomal aneuploidies. BMC Med Genomics. 2012;5:57.

31. Lau TK, Cheung SW, Lo PS, Pursley AN, Chan MK, Jiang F, et al. Non-invasive prenatal testing for fetal chromosomal abnormalities by low-coverage whole-genome sequencing of maternal plasma DNA: review of 1982 consecutive cases in a single center. Ultrasound Obstet Gynecol. 2014;43:254-64.

32. Liang D, Lv W, Wang H, Xu L, Liu J, Li H, et al. Non-invasive prenatal testing of fetal whole chromosome aneuploidy by massively parallel sequencing. Prenat Diagn. 2013;33:409-15.

33. Nicolaides KH, Syngelaki A, Ashoor G, Birdir C, Touzet G. Noninvasive prenatal testing for fetal trisomies in a routinely screened first-trimester population. Am J Obstet Gynecol. 2012;207:374.e1-6.

34. Nicolaides KH, Syngelaki A, Gil M, Atanasova V, Markova D. Validation of targeted sequencing of single-nucleotide polymorphisms for non-invasive prenatal detection of aneuploidy of chromosomes 13, 18, 21, X, and Y. Prenat Diagn. 2013;33:575-9.

35. Norton ME, Brar H, Weiss J, Karimi A, Laurent LC, Caughey AB, et al. Non-Invasive Chromosomal Evaluation (NICE) Study: results of a multicenter prospective cohort study for detection of fetal trisomy 21 and trisomy 18. Am J Obstet Gynecol. 2012;207:137.e1-8.

36. Norton ME, Jacobsson B, Swamy GK, Laurent LC, Ranzini AC, Brar H, et al. Cell-free DNA analysis for noninvasive examination of trisomy. N Engl J Med. 2015;372:1589-97.

37. Pergament E, Cuckle H, Zimmermann B, Banjevic M, Sigurjonsson S, Ryan A, et al. Single-nucleotide polymorphism-based noninvasive prenatal screening in a high-risk and low-risk

cohort. Obstet Gynecol. 2014;124:210-8.

38. Porreco RP, Garite TJ, Maurel K, Marusiak B, Ehrich M, Van Den Boom D, et al. Noninvasive prenatal screening for fetal trisomies 21, 18, 13 and the common sex chromosome aneuploidies from maternal blood using massively parallel genomic sequencing of DNA. Am J Obstet Gynecol. 2014;211:365.e1-.e12.

39. Quezada MS, Gil MM, Francisco C, Orosz G, Nicolaides KH. Screening for trisomies 21, 18 and 13 by cell-free DNA analysis of maternal blood at 10-11 weeks' gestation and the combined test at 11-13 weeks. Ultrasound Obstet Gynecol. 2015;45:36-41.

40. Shaw SW, Hsiao CH, Chen CY, Ren Y, Tian F, Tsai C, et al. Noninvasive prenatal testing for whole fetal chromosomal aneuploidies: a multicenter prospective cohort trial in Taiwan. Fetal Diagn Ther. 2014;35:13-7.

41. Song Y, Huang S, Zhou X, Jiang Y, Qi Q, Bian X, et al. Non-invasive prenatal testing for fetal aneuploidies in the first trimester of pregnancy. Ultrasound Obstet Gynecol. 2015;45:55-60.

42. Song Y, Liu C, Qi H, Zhang Y, Bian X, Liu J. Noninvasive prenatal testing of fetal aneuploidies by massively parallel sequencing in a prospective Chinese population. Prenat Diagn. 2013;33:700-6.

43. Sparks AB, Struble CA, Wang ET, Song K, Oliphant A. Noninvasive prenatal detection and selective analysis of cell-free DNA obtained from maternal blood: evaluation for trisomy 21 and trisomy 18. Am J Obstet Gynecol. 2012;206:319.e1-9.

44. Stumm M, Entezami M, Haug K, Blank C, Wustemann M, Schulze B, et al. Diagnostic accuracy of random massively parallel sequencing for non-invasive prenatal detection of common autosomal aneuploidies: a collaborative study in Europe. Prenat Diagn. 2014;34:185-91.

45. Verweij EJ, Jacobsson B, van Scheltema PA, de Boer MA, Hoffer MJ, Hollemon D, et al. European Non-Invasive Trisomy Evaluation (EU-NITE) study: a multicenter prospective cohort study for non-invasive fetal trisomy 21 testing. Prenat Diagn. 2013;33:996-1001.

46. Yao H, Jiang F, Hu H, Gao Y, Zhu Z, Zhang H, et al. Detection of fetal sex chromosome aneuploidy by massively parallel sequencing of maternal plasma DNA: initial experience in a Chinese hospital. Ultrasound Obstet Gynecol. 2014;44:17-24.

47. Zhang H, Gao Y, Jiang F, Fu M, Yuan Y, Guo Y, et al. Non-invasive prenatal testing for trisomies 21, 18 and 13: clinical experience from 146,958 pregnancies. Ultrasound Obstet Gynecol.

2015;45:530-8.

48. Zhou Q, Pan L, Chen S, Chen F, Hwang R, Yang X, et al. Clinical application of noninvasive prenatal testing for the detection of trisomies 21, 18, and 13: a hospital experience. Prenat Diagn. 2014;34:1061-5.

49. Palomaki GE, Deciu C, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM, et al. DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international collaborative study. Genetics in medicine : official journal of the American College of Medical Genetics. 2012;14:296-305.

50. Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM, Ehrich M, et al. DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. Genet Med. 2011;13:913-20.

51. Liau J, Romine L, Korty LA, Chao C, White K, Harmon S, et al. Simplifying the ultrasound findings of the major fetal chromosomal aneuploidies. Curr Probl Diagn Radiol. 2014;43:300-16.

52. Song K, Musci TJ, Caughey AB. Clinical utility and cost of non-invasive prenatal testing with cfDNA analysis in high-risk women based on a US population. J Matern Fetal Neonatal Med. 2013;26(12):1180-5.

Legends of Supporting Information Tables

Table S1. Full-text articles excluded according to criteria (n=376).

Table S2. Full-text articles where the quality was rated as low using the QUADAS tool (n=22).

Table S3. Characteristics of included studies reporting on non-invasive prenatal testing (NIPT) for trisomy 21 using cell-free DNA.

Table S4. Characteristics of included studies reporting on non-invasive prenatal testing (NIPT) for trisomy 18 using cell-free DNA.

Table S5. Characteristics of included studies reporting on non-invasive prenatal testing (NIPT) for trisomy 13 using cell-free DNA.

Table S6**.** Methodological quality summary of included full-text articles (n=31) based on the questions from the QUADAS tool (no 1-14) and a complementary question on commersial sponsoring (no 15).

Table S7. Proportion of false positives.

Table S8. Proportion of false negatives.

Table S9. Levels of heterogeneity in studies.

Table S10. Alternative measures of diagnostic accuracy (Positive and negative likelihood ratios [LR+ and LR-] and diagnostic odds ratio [DOR]) for trisomy 21 in high- and average-risk population, trisomy 18 in high-risk population and trisomy 13 in high-risk population.

Legends of figures

Figure 1. Flow chart summarizing selection of included studies.

Figure 2. Meta-analysis of sensitivity and specificity of non-invasive prenatal testing (NIPT) for detecting trisomy 21 in a population at high risk of carrying a fetus with chromosome aberration.

Figure 3. Meta-analysis of sensitivity and specificity of non-invasive prenatal testing (NIPT) for detecting trisomy 21 in a population at average risk of carrying a fetus with chromosome aberration.

Figure 4. Meta-analysis of sensitivity and specificity of non-invasive prenatal testing (NIPT) for detecting trisomy 18 in a population at high risk of carrying a fetus with chromosome aberration.

Figure 5. Meta-analysis of sensitivity and specificity of non-invasive prenatal testing (NIPT) for detecting trisomy 13 in a population at high risk of carrying a fetus with chromosome aberration.

Figure 6. Meta-analysis of sensitivity and specificity of non-invasive prenatal testing (NIPT) in a population at average risk of carrying a fetus with chromosome aberration: a) trisomy 18 b) trisomy 13.

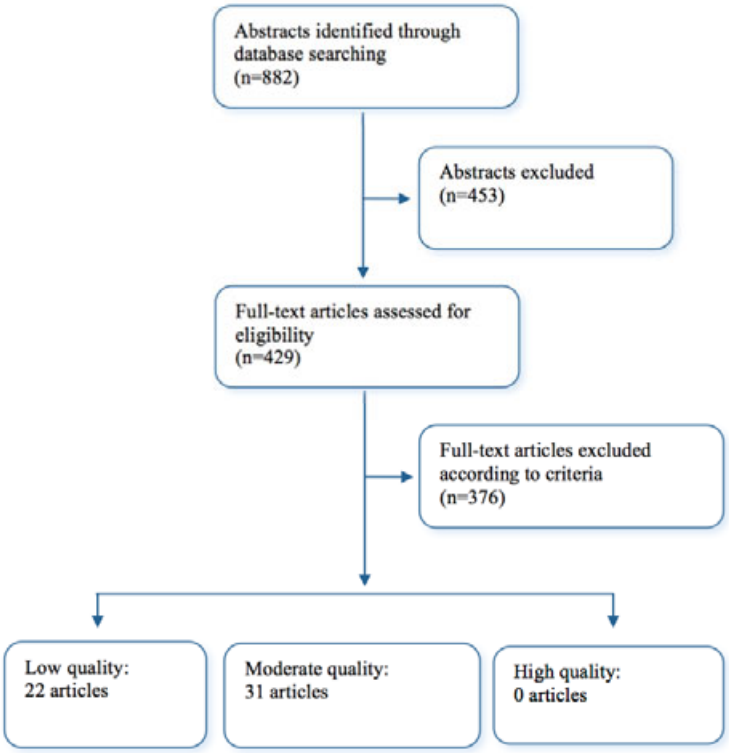
Table 1. Summary of findings and quality of evidence (GRADE).

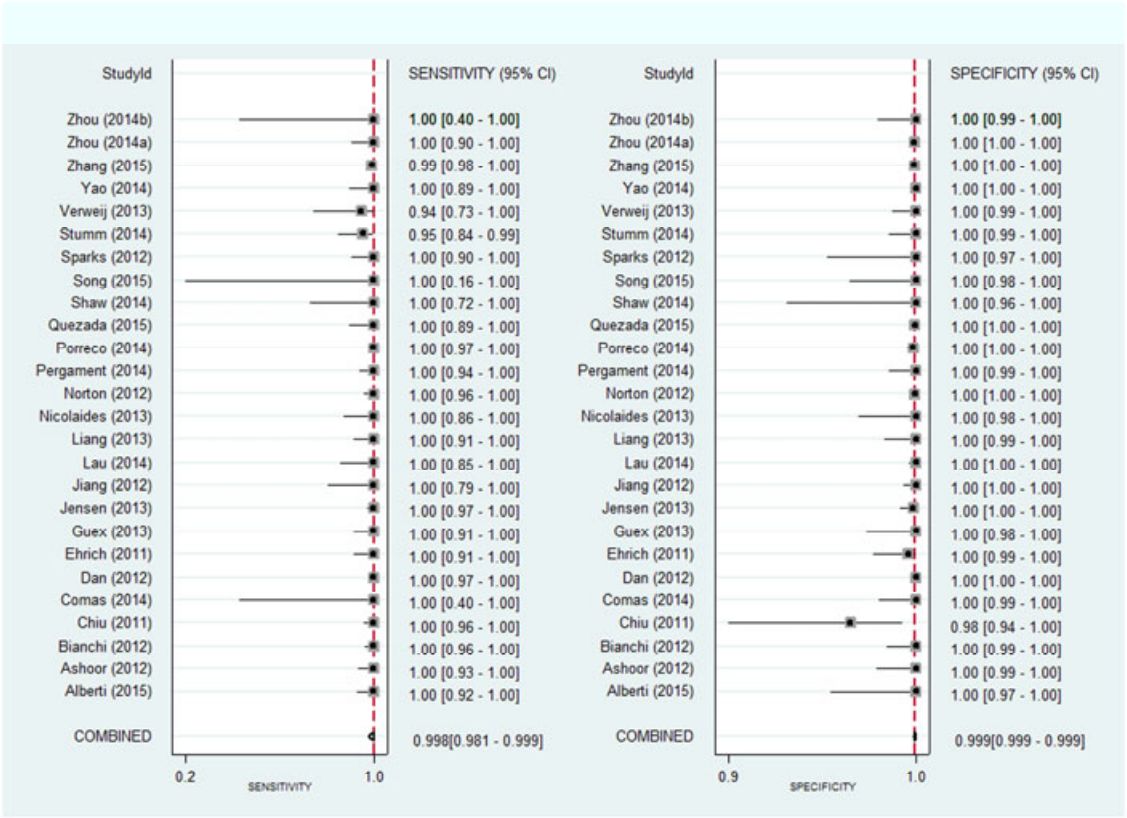
Trisomy	**Population** **Probability** **("risk")**	**Sample size (no of studies)**	**Sensitivity,** **Pooled estimates** **(95 % CI)**	**Specificity,** **Pooled estimates** **(95 % CI)**	**Quality of evidence**	**Rating items**	**True positive* (TP)** **False positive** (FP)** **False negative*** (FN)** **True Negative (TN)******
T21	High	107 474 (26)	0.998 (0.981–0.999)	0.999 (0.999-0.999)	(⊕⊕⊕O)	-1 Study design/quality	1839 TP 52 FP

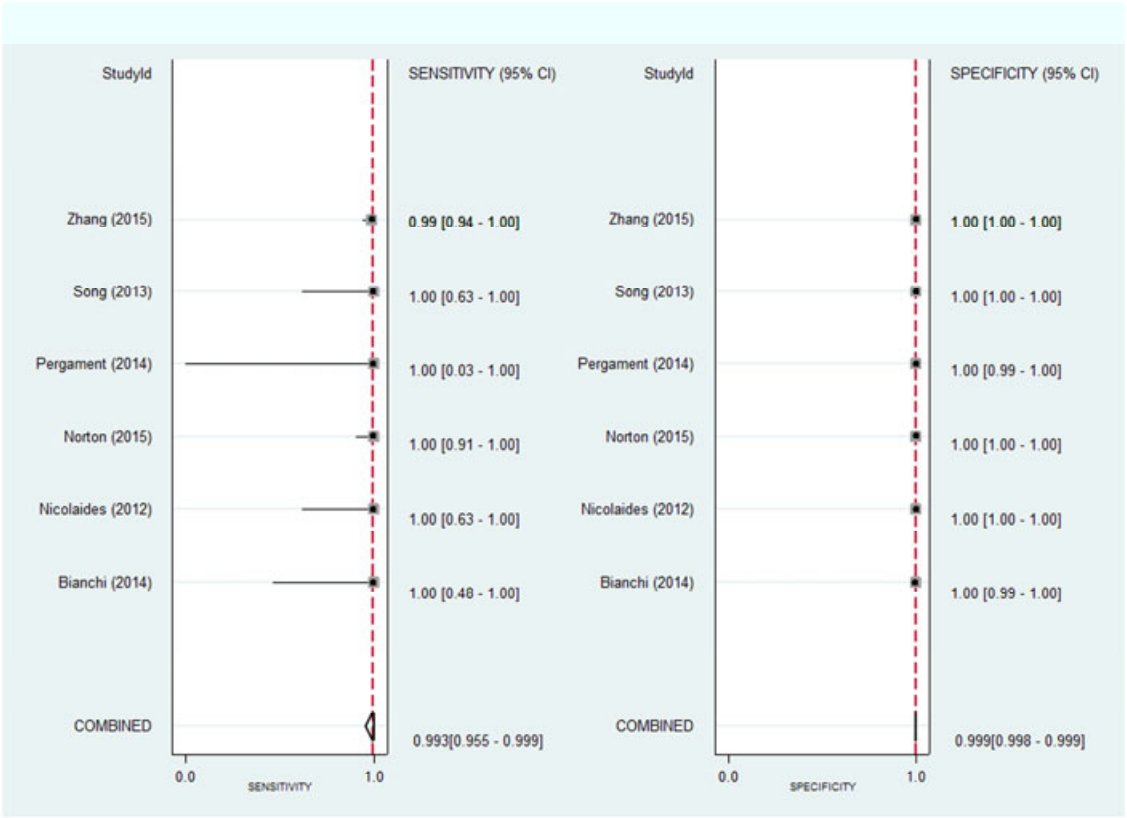
							8 FN 105 575 TN
T21	Average	62 201 (6)	0.993 (0.955–0.999)	0.999 (0.998–0.999)	(⊕⊕⊕O)	-1 Study design/quality	156 TP 37 FP 1 FN 62 107 TN
T18	High	146 465 (22)	0.977 (0.958–0.987)	0.999 (0.998–0.999)	(⊕⊕⊕O)	-1 Study design/quality	566 TP 70 FP 15 FN 146 129 TN
T13	High	137 078 (18)	0.975 (0.819–0.997)	0.999 (0.999–0.999)	(⊕⊕OO)	-1 Study design/quality -1 imprecision	134 TP 56 FP 10 FN 137 499 TN

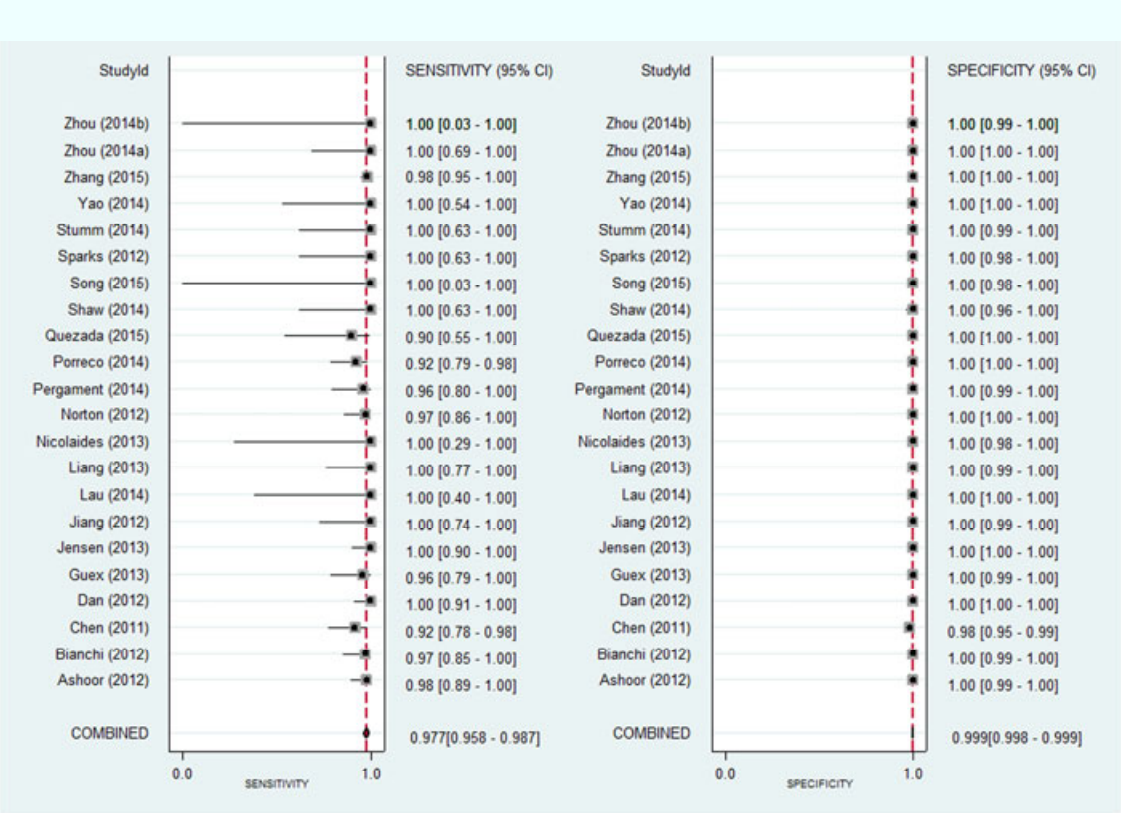
*TP= trisomy is verified; **FP= incorrectly classified as trisomy ***FN= trisomy is incorrectly classified as normal ****TN= absence of trisomy is verified.

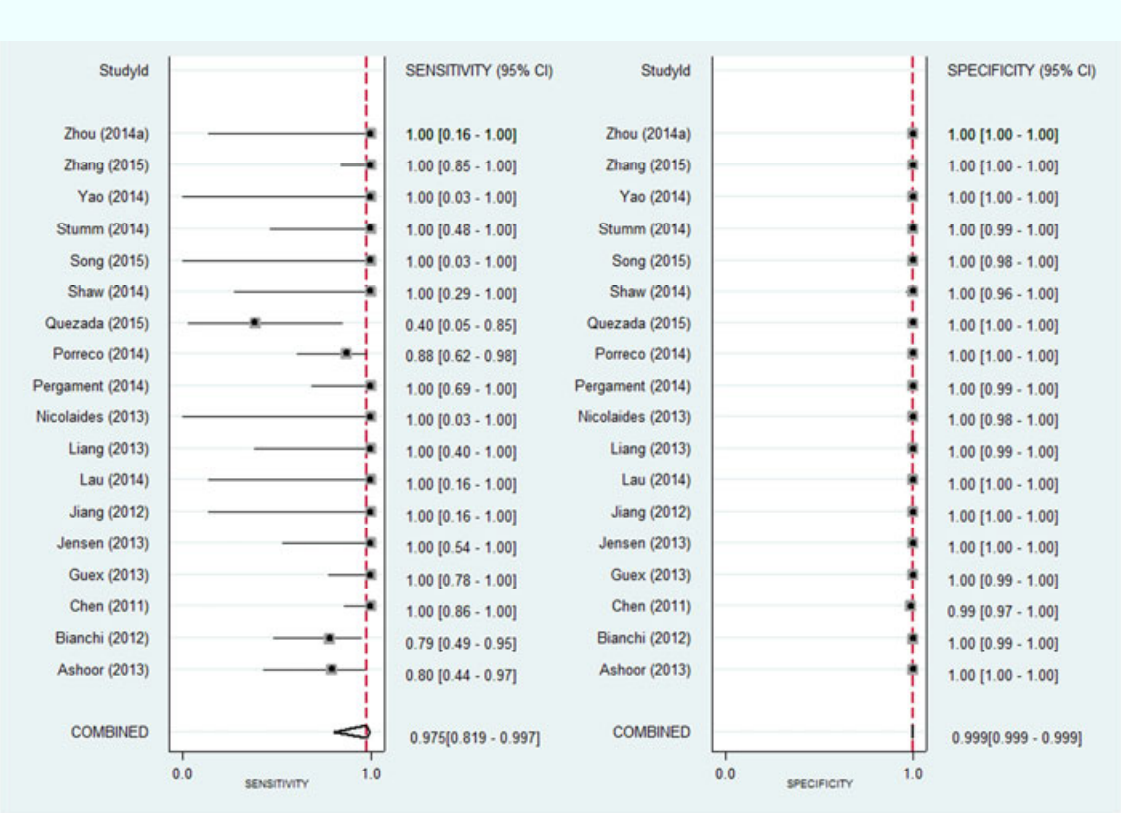
⊕⊕⊕O moderate quality of evidence, ⊕⊕OO limited quality of evidence





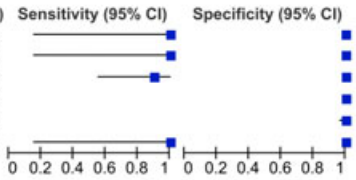






a

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bianchi 2014	2	3	0	1902	1.00 [0.16, 1.00]	1.00 [1.00, 1.00]
Nicolaides 2012	2	2	0	1945	1.00 [0.16, 1.00]	1.00 [1.00, 1.00]
Norton 2015	9	1	1	15830	0.90 [0.55, 1.00]	1.00 [1.00, 1.00]
Pergament 2014	0	0	0	474	Not estimable	1.00 [0.99, 1.00]
Shaw 2014	0	0	0	100	Not estimable	1.00 [0.96, 1.00]
Song 2013	2	1	0	1738	1.00 [0.16, 1.00]	1.00 [1.00, 1.00]



b

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bianchi 2014	1	1	0	891	1.00 [0.03, 1.00]	1.00 [0.99, 1.00]
Norton 2015	2	2	0	11181	1.00 [0.16, 1.00]	1.00 [1.00, 1.00]
Pergament 2014	2	0	0	472	1.00 [0.16, 1.00]	1.00 [0.99, 1.00]
Shaw 2014	0	0	0	100	Not estimable	1.00 [0.96, 1.00]
Song 2013	1	0	0	1740	1.00 [0.03, 1.00]	1.00 [1.00, 1.00]