

BMJ Open Accuracy of non-invasive prenatal testing using cell-free DNA for detection of Down, Edwards and Patau syndromes: a systematic review and meta-analysis

Sian Taylor-Phillips,¹ Karoline Freeman,¹ Julia Geppert,¹ Adeola Agbebiyi,¹ Olalekan A Uthman,¹ Jason Madan,¹ Angus Clarke,² Siobhan Quenby,¹ Aileen Clarke¹

To cite: Taylor-Phillips S, Freeman K, Geppert 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. doi:10.1136/bmjopen-2015-010002

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-010002>).

Received 15 September 2015
Revised 28 October 2015
Accepted 9 November 2015



CrossMark

¹Warwick Medical School, The University of Warwick, Coventry, West Midlands, UK

²Institute of Cancer & Genetics, Cardiff University School of Medicine, Cardiff, UK

Correspondence to

Professor Aileen Clarke;
aileen.clarke@warwick.ac.uk

ABSTRACT

Objective: To measure test accuracy of non-invasive prenatal testing (NIPT) for Down, Edwards and Patau syndromes using cell-free fetal DNA and identify factors affecting accuracy.

Design: Systematic review and meta-analysis of published studies.

Data sources: PubMed, Ovid Medline, Ovid Embase and the Cochrane Library published from 1997 to 9 February 2015, followed by weekly autoalerts until 1 April 2015.

Eligibility criteria for selecting studies: English language journal articles describing case-control studies with ≥ 15 trisomy cases or cohort studies with ≥ 50 pregnant women who had been given NIPT and a reference standard.

Results: 41, 37 and 30 studies of 2012 publications retrieved were included in the review for Down, Edwards and Patau syndromes. Quality appraisal identified high risk of bias in included studies, funnel plots showed evidence of publication bias. Pooled sensitivity was 99.3% (95% CI 98.9% to 99.6%) for Down, 97.4% (95.8% to 98.4%) for Edwards, and 97.4% (86.1% to 99.6%) for Patau syndrome. The pooled specificity was 99.9% (99.9% to 100%) for all three trisomies. In 100 000 pregnancies in the general obstetric population we would expect 417, 89 and 40 cases of Downs, Edwards and Patau syndromes to be detected by NIPT, with 94, 154 and 42 false positive results. Sensitivity was lower in twin than singleton pregnancies, reduced by 9% for Down, 28% for Edwards and 22% for Patau syndrome. Pooled sensitivity was also lower in the first trimester of pregnancy, in studies in the general obstetric population, and in cohort studies with consecutive enrolment.

Conclusions: NIPT using cell-free fetal DNA has very high sensitivity and specificity for Down syndrome, with slightly lower sensitivity for Edwards and Patau syndrome. However, it is not 100% accurate and should not be used as a final diagnosis for positive cases.

Trial registration number: CRD42014014947.

Strengths and limitations of this study

- This is a full systematic review with searches across multiple databases dating back to 1997, and two authors sifting all titles and abstracts.
- Two authors extracted data on prepiloted forms and appraised quality using an adapted QUADAS 2 form.
- The meta-analysis included rigorous methods of data analysis, including bivariate random-effects regression models, but required a zero-cell correction to enable model convergence which may underestimate rather than overestimate accuracy.
- The meta-analysis included a series of subgroup and sensitivity analyses to test for robustness of our pooled diagnostic accuracy estimates.
- The methods are transparent with full protocol published in PROSPERO in advance of the review.

INTRODUCTION

Non-invasive prenatal testing (NIPT) using cell-free fetal DNA (cffDNA) is a method for testing for trisomies in the fetus, using a peripheral sample of the pregnant mother's blood. It is currently marketed across 61 countries in Europe, Asia, Africa and North and South America.¹ Rapid adoption in the USA has seen increases in first trimester screening using NIPT, and concurrent decreases in the first trimester combined test and invasive testing.^{2 3} People tend to overestimate the usefulness of genetic tests, and misinterpret their meaning.³ It is possible that pregnant women will interpret a positive NIPT test as positive diagnosis, and wish to abort a pregnancy on this basis. A clear summary of test accuracy for NIPT is necessary for use by doctors and patients for use in shared and informed decision-making.

Although a previous review of NIPT test accuracy exists,⁴ it does not include two of the largest studies.^{5 6} In addition the authors use a univariate approach which is not appropriate for meta-analysis of tests since it overlooks the fact that sensitivity and specificity are usually negatively correlated across studies due to different thresholds used to define positive and negative test results. It has been shown that ignoring this correlation would be inappropriate.⁷ The weighted sums of the reported specificity are normally used to assess the value of a test, the properties of the resulting statistics depends most importantly on this correlation between the estimates, and it is exactly that is ignored in separate univariate analyses.^{8 9} Most importantly, the previous review does not provide a summary of findings which can be applied to a relevant population and used in clinician–patient shared decision-making.

The UK National Screening Committee commissioned this new review to provide a summary of the accuracy of NIPT for detection of Down, Edwards and Patau syndromes in first trimester pregnancies, to inform their decision on introduction of this test into current fetal abnormality screening in the UK.

METHODS

Identification and selection of studies

Ethical approval was granted from the University of Warwick Biomedical and Scientific Research Ethics Committee reference REGO-2015-1446. Searches were conducted in PubMed, Ovid Medline, Ovid Embase and the Cochrane Library. The search strategy used a combination of search terms for the NIPT test and trisomies, and was limited to the English language, (see online supplementary file 1). Date limits were 01.01.1997–09.02.2015. Updating autoalerts in Medline and Embase were run until 01.04.2015. Individuals and organisations were contacted for studies not freely available in the public domain. ClinicalTrials.gov, WHO International Clinical Trials Registry Platform (ICTRP) Search Portal and meeting abstracts were also searched for ongoing or recently completed trials.

Two reviewers independently screened titles and abstracts of all records obtained. Discrepancies were resolved by consensus or discussion with a third reviewer. Inclusion criteria were English language journal articles which investigated NIPT using cff DNA derived from maternal blood (serum, plasma, whole blood) in pregnant women in any trimester for the detection of Down (T21), Edwards (T18) or Patau (T13) syndromes in the fetus. The reference standard was genetic verification through amniocentesis, Chorionic Villus Sampling (CVS), cordocentesis, fetal pathological examination after abortion or postnatal phenotypic assessment. We included studies with any outcomes reported on test accuracy, or rates of test failure or indeterminate results. We excluded studies reporting the quantification of fetal cells or DNA or using elevated levels of the whole fetal

DNA or epigenetic markers. We also excluded case–control studies with fewer than 15 cases and cohort studies with fewer than 50 pregnant women as well as studies with incomplete 2×2 data or studies which reused samples from other included studies in order to prevent double counting.

Data were extracted by one reviewer and checked by a second reviewer. Disagreements were resolved by consensus or discussion with a third reviewer. Full data extraction forms are available from the authors on request.

Quality assessment

The quality of diagnostic accuracy studies was assessed using a modified QUADAS-2.¹⁰ Quality assessment was undertaken by one reviewer and checked by a second reviewer, with disagreements resolved by a third reviewer. Three modifications were made. First, an additional signalling question was added on whether the study avoided taking the sample for the index test in the 7 days after an invasive test, as fetal fraction may be elevated at this time boosting the performance of NIPT. Second, a signalling question was added to determine whether the threshold value was determined using an independent set of samples, and whether adjustment of the predefined threshold was avoided, since the threshold for testing positive is expressed as number of SDs from the mean score for a set of normal samples, rather than as an absolute threshold. Finally, the standard QUADAS-2 signalling question determining whether there was an appropriate interval between index test and reference standard was removed, as timing of an invasive test (apart from in relation to invasive testing) would not affect accuracy. Timing of the NIPT test is important as fetal fraction and therefore accuracy of NIPT increases throughout pregnancy, this was included under applicability of findings rather than as a source of bias. We also assessed the role of the sponsor in addition to QUADAS-2. This included studies that clearly declared involvement of a sponsor in the design or conduct of the study or publication, the majority of authors were employees or shareholders of companies offering NIPT or cytogenetic tests and/or other conflicts of interest (ie, patents, stock or stock options). Please see online supplementary file 2 for full information on the definition for the signalling questions of the QUADAS-2.

Statistical analysis of test accuracy studies

All eligible studies were included in a meta-analysis of performance of the NIPT test. We extracted data from the primary studies to obtain the four cell values of a diagnostic 2×2 table in order to calculate test accuracy measures. We pooled the sensitivity and specificity estimates using bivariate random-effects regression models, as recommended by the Cochrane Diagnostic Test Accuracy Working Group,¹¹ in order to take the potential trade-off between sensitivity and specificity explicitly into consideration and incorporate this negative

correlation into the analysis.⁷ We added a 0.5 cell correction to each cell where a zero was encountered. We stratified test accuracy measures according to condition (T21, T18 and T13).

Meta-analysis, subgroup and sensitivity analyses

We used sensitivity, subgroup and meta-regression analyses to explore potential sources of heterogeneity in test accuracy estimates across studies. The following variables were selected a priori as potential sources of heterogeneity: study design (cohort with consecutive sampling vs others), population risk (general, high-risk, others), population (twins vs others), first trimester (100% vs other), test type (MPSS, DANSR, single nucleotide polymorphism (SNP) technology) and publication year (2007–2013 vs 2014–2015). We conducted a series of sensitivity analyses to check the robustness of the results. We excluded all studies with zero cases of true positive and false negative results. We used Cook's distance to identify particularly influential studies and created a scatter plot of the standardised predicted random effects (standardised level 2 residuals) to check for outliers.¹² We refitted the model leaving out outliers and very influential studies.

We constructed 3×2 tables to examine the influence of the number of test failures and indeterminate results on the pooled test accuracy estimates.¹³ Test failures occur where the NIPT test has failed to produce any result, and indeterminate results where the test result is in a mid-range which is neither positive nor negative. Test failures can occur for a variety of reasons, and sometimes the cause is unknown. Test failures and indeterminate results are not included in the 2×2 tables reported, and this can lead to overestimates of sensitivity and specificity.¹⁴ We included all failures of the NIPT test, regardless of whether repeating the test on the same or a new blood sample would have given a result, but we did not include failures which could be rectified by good quality assurance procedures (such as insufficient blood or dropped samples). For the 3×2 tables we considered the following three scenarios, all non-evaluable results: (1) considered to be positive results to reflect use of the NIPT as triage for invasive testing,¹⁴ (2) considered to be negative results to reflect use of NIPT as an add-on to the combined test¹⁴ and (3) follow intention to diagnose principle to account for the first two approaches overestimating specificity and sensitivity, respectively.¹³ Intention to diagnose was defined as "including non-evaluable results either in the 'false negative' or the 'false positive' cell of a 2×2 table (worst case scenario) according to the results of the reference standard". For the intention to diagnose principle, all non-evaluable positive results were assumed to be false negative and all non-evaluable negative results were assumed to be false positive. Where the reference standard results were not reported for these cases, we assumed that they had the same prevalence of trisomy as those in the rest of the same study.

In the subgroup analyses, we computed pooled accuracy estimates in various strata to determine if accuracy is higher or lower in specific subgroups. Summary sensitivity and specificity estimates for each subgroup were generated, along with their 95% CIs. In the linear meta-regression model, studies are the units of analysis. We used the meta-regression model to generate relative diagnostic ORs.^{15 16} We used Deeks' funnel plot asymmetry test to test for publication bias, with p value<0.10 indicating significant publication bias.¹⁷ All analyses were performed using Stata V.13 for Windows including the user written commands metandi, midas, metareg and mvmeta.^{12 18–20}

RESULTS

Study selection

A total of 2012 records were identified after duplicates were removed. One-hundred and eight records remained after evaluation of title and abstract, of which 41 studies were included in the meta-analysis. Figure 1 summarises the study selection process (see online supplementary file 3 for included studies and online supplementary file 4 for reasons of exclusion for 67 full-text articles).

Characteristics of included studies

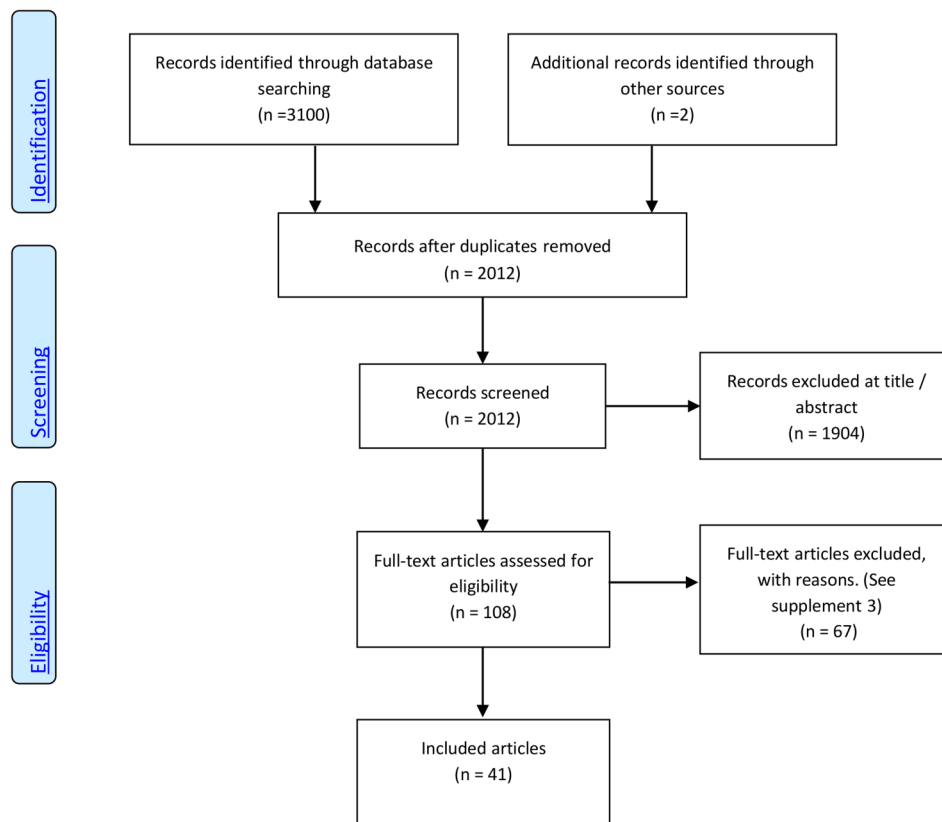
Study design, populations, reference standards

Forty-one publications, dating from 2007 to 2015, reported NIPT results for between 46 and 112 669 pregnant women for the main autosomal trisomies in relation to fetal karyotype or newborn phenotype and fulfilled our inclusion criteria (see online supplementary file 5). The majority of studies were cohort studies (n=29),^{5 6 21–47} with prospective data collection. There were 11 case-control studies^{48–58} and one of unclear design.⁵⁹ Thirty studies were undertaken in singleton pregnancies only,^{6 21 22 29–33 35–37 40–51 53–59} four studies included singleton and twin pregnancies,^{5 28 34 38} with the remainder undertaken in twin only (n=3).^{23 24 39} In four studies the reporting was unclear.^{25 26 28 52} The majority of studies (n=24) used samples from high-risk pregnant women (positive standard screening, ultrasound abnormalities, advanced maternal age, personal or family history of aneuploidies) undergoing invasive testing.^{24 26 28 30 31 33 36–38 41 44 45 47–56 58 59} Six studies were performed in the general obstetric population.^{6 21 29 35 40 43} Nine studies included pregnant women with mixed risk factors.^{5 22 27 32 34 39 42 46 57} In two the underlying risk was unclear.^{23 25} Seven studies included women in the first trimester only,^{6 23 29 30 43 47 48} while all other studies (n=34) included pregnant women with an unstated, later or broader gestational age window.^{5 21 22 24–28 31–42 44–46 49–59}

Testing strategies

Three main testing strategies were pursued by the majority of studies (see online supplementary file 6). These

Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart of included articles.



were genome-wide massively parallel shotgun sequencing (MPSS, $n=24$ studies),^{5 21 22 24–28 33–36 41 44–47 49–52 54 55 58} targeted massively parallel sequencing (DANSR, $n=9$ studies),^{6 23 29 31 37 39 43 48 56} and SNP technology ($n=5$).^{30 32 42 53 57} Two studies, performed in real clinical settings, offered more than one NIPT approach.^{38 40} Dhallan *et al*⁵⁹ used a DNA-SNP allelic ratio approach.

In 3 of the 41 studies,^{21 32 57} some of the maternal blood samples for NIPT were obtained after invasive testing and for 34 studies we concluded that tests were collected before the invasive testing.^{5 6 22–24 26–31 33–52 54 55 58}

In four studies, it was unclear if maternal blood sampling for NIPT was performed before or after an invasive procedure.^{25 53 56 59}

Forty studies reported NIPT performance for T21,^{5 6 21–49 51–59} 36 for T18,^{5 6 21–36 38–50 53–57} and 30 studies investigated non-invasive detection of T13.^{5 6 21 23 25–28 30 32–36 38–47 49 50 53–55 57} Twenty-nine studies reported test accuracy for all three main autosomal trisomies.^{5 6 21 23 25–28 30 32–36 38–47 49 53–55 57}

Methodological quality of included studies

The methodological quality of the 41 included studies, assessed by QUADAS-2¹⁰ is summarised in figures 2 and 3 and online supplementary file 7. Risk of bias was high in most studies with 25 of 41 studies considered high risk in two or more domains, and 14 studies in one domain. Two were judged as low or unclear risk of bias in all five domains. Figure 2 shows that study flow (concerned with patient follow-up) and the role of the

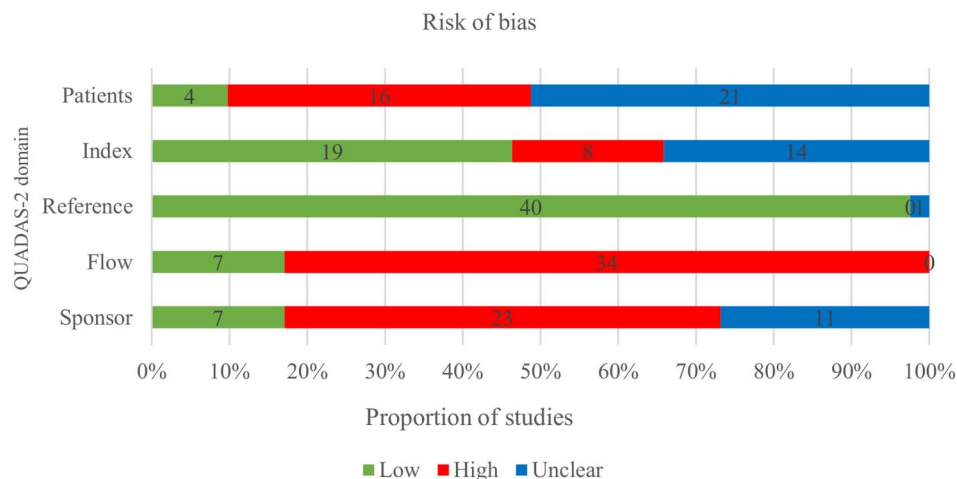
sponsor were the areas with the greatest risk of bias. Another issue was incomplete or unclear reporting, particularly of the patient selection process and the conduct of the index test, which is reflected in 21 (51.2%) and 14 (34.1%) of 41 publications scoring an unclear risk of bias in these two domains, respectively. The risk of bias regarding the reference standard was considered low in almost all studies with only one study classified as unclear.²³ Finally, risk of bias regarding the role of sponsor was deemed high in 23 studies. There were significant concerns regarding applicability of the included patient spectrum to cfDNA testing introduction in the first trimester (see figure 3), as 29 of 41 studies had significant parts (>20%) of their populations tested in the second or third trimester when fetal fraction and therefore accuracy of NIPT is higher.

Meta-analysis

There was a high likelihood of publication bias, with the slope coefficients on Deeks' funnel plot asymmetry test significant for Down syndrome ($p=0.0001$), Edwards syndrome ($p=0.0001$), and Patau syndrome ($p=0.045$) (see figure 4).

The pooled sensitivity for Down syndrome from bivariate random-effects regression of 40 studies was 99.3% (98.9% to 99.6%) and the pooled specificity was 99.9% (99.9% to 100%). For Edwards syndrome the pooled sensitivity over 33 studies was 97.4% (95.8% to 98.4%) and specificity was 99.9% (99.9% to 100%). For Patau

Figure 2 Proportion of studies with low, high or unclear risk of bias using QUADAS 2.



syndrome the pooled sensitivity over 24 studies was 97.4% (86.1% to 99.6%) and specificity was >99.9% (99.9% to 100%). Table 1 shows these pooled sensitivities and specificities applied to populations of pregnant women taking the test. In the subgroup analysis (table 2) sensitivity estimates were lower by 6.1% for Down, 10.6% for Edwards, and 12.3% for Patau syndromes for cohort studies with consecutive sampling in comparison to all other studies which are more likely to be subject to spectrum bias. Test accuracy did not appear to systematically differ between DANSR, MPSS or SNP-based test types or by publication year. Estimates of test sensitivity were higher in high-risk populations, in studies including pregnancies in the second and third trimester, and in singleton pregnancies. In high-risk populations, defined in a variety of ways, pooled sensitivity estimates were 1.4%, 6.5% and 17.8% higher than in the general obstetric population for Down, Edwards and Patau syndromes, respectively. Sensitivity estimates were 1.3%, 1.4% and 11.6% lower in studies recruiting all women in their first trimester of pregnancy in comparison to studies including women later in pregnancy. The outcomes of test accuracy of the included studies are summarised in online supplementary file 8. A forest plot of the

sensitivity and specificity from the individual studies with 95% CIs is given in figure 5.

Test failures

The rate of analytic failure (failure of the cfDNA testing) ranged from 0% to 12.7%⁵⁷ and among 5789 pregnancies with resampling, 803 (13.9%) also failed the repeat cfDNA testing. There were five papers in this review that reported indeterminate results (results in a range defined as neither positive nor negative) for trisomies 21, 18 and 13.^{21 38 49 55 60} ranging from 0% (0/2042) to 11.1% (5/45). In the study with no indeterminate results they used eight-plex testing, and where the initial score was indeterminate they repeated using one-plex which corrected any indeterminate results. There is some evidence that the rate of test failure is higher when gestational age is lower, and in trisomic pregnancies. Pergament *et al*⁸² found that failure rate at <9 weeks was 26/95 (27.4%), between 9.0 and 9.9 weeks was 6/50 (12.0%), and more than 10 weeks was 53/900 (5.9%). The same study found aneuploidy incidence was increased (20/86 (23.3%)) in samples that did not return a result when compared with the aneuploidy incidence in samples with a cfDNA testing result (105/966

Figure 3 Proportion of studies with low, high and unclear concerns regarding applicability using QUADAS 2.

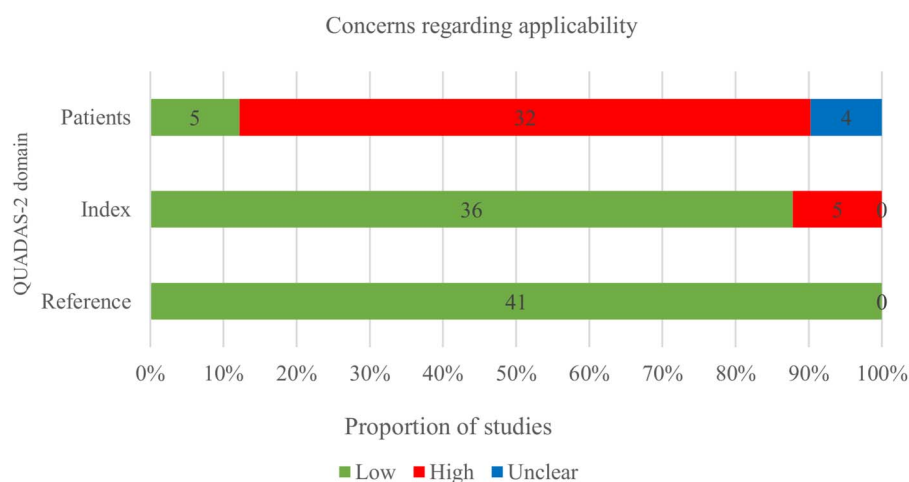
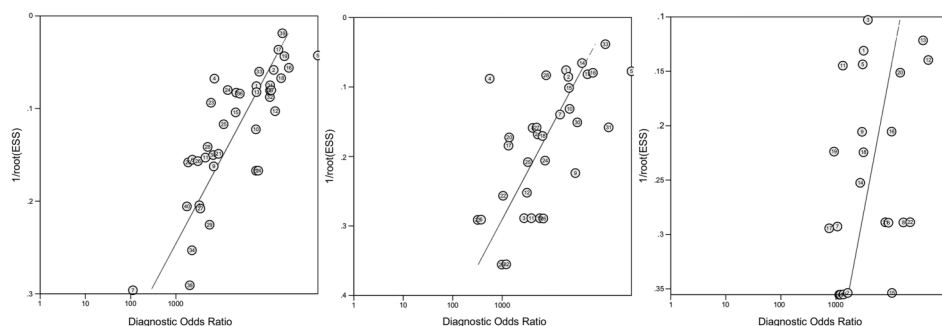


Figure 4 Deeks' funnel plot for Down (left) Edwards (centre) and Patau (right) syndromes. A vertical pattern would indicate no bias, slope is associated with publication bias.



(10.9%), $p=0.004$). Norton *et al*⁶ did not find an association between test failure and gestational age in 18 510 women between 10 and 14 weeks gestation, but found that the prevalence of aneuploidy in the group with test failure (1 in 38 (2.7%)) was higher than the prevalence of 1 in 236 (0.4%) in the overall cohort ($p<0.001$).

Including test failures in an intention to diagnose analysis in the meta-analysis decreased sensitivity estimates by 1.7% for Down, 1.6% for Edwards and 7.1% for Patau syndrome, and decreased specificity estimates by nearly 2% for all three trisomies. Excluding test failures from the calculations of test accuracy may have caused overestimation of accuracy. Similarly in the subgroup analysis sensitivity estimates were lower by 6.1% for Down, 10.6% for Edwards, and 12.3% for Patau syndromes for cohort studies with consecutive sampling in comparison to all other studies. Test accuracy did not appear to differ systematically between DANSR, MPSS or SNP technology, or by publication year. Estimates of test sensitivity were higher in high-risk populations, in studies including pregnancies in the second and third trimester, and in singleton pregnancies. In high-risk populations, defined in a variety of ways, pooled sensitivity estimates were 1.4%, 6.5% and 17.8% higher than in the general obstetric population for Down, Edwards and Patau syndromes, respectively. Sensitivity estimates were 1.3%, 1.4% and 11.6% lower in studies recruiting all women in their first trimester of pregnancy in comparison to studies including women later in pregnancy. Twin pregnancies had 8.3% lower sensitivity estimates than singletons for Down syndrome. This difference was 20.6% for Edwards syndrome, but there was only one study for Patau syndrome so we were unable to provide a pooled estimate for twins. Sensitivity and subgroup analyses are reported in [table 2](#).

DISCUSSION

In a systematic review of 2012 articles, we identified 41 articles on the test accuracy of NIPT. Quality appraisal using QUADAS-2 indicated high risk of bias, in particular due to unclear or unsystematic inclusions and exclusions of participants at study entry level as well as at the level of analysis. Applicability of findings was of concern as there is still very limited data on the screening population available. Pooled sensitivity from the meta-analysis

was 99.3% for T21, 97.4% for T18 and 97.4% for T13, with pooled specificity 99.9% (99.9% to 100%) for all three trisomies. We estimated test accuracy in a high-risk population of 10 000 pregnancies where 3.3% of fetuses have Down syndrome, 1.5% have Edwards syndrome and 0.5% have Patau syndrome. There would be 324 cases of Down syndrome detected, with 9 missed and 31 false positive results, 140 cases of Edwards syndrome detected with 11 missed and 26 false positive results, and 47 cases of Edwards syndrome detected, with 3 missed and 7 false positive results ([table 1](#)). In the general obstetric population where prevalence of trisomy is lower, there would be a lower positive predictive value. In 100 000 pregnancies in the general obstetric population we would expect 417, 89 and 40 cases of Downs, Edwards and Patau syndromes to be detected by NIPT, with 94, 154 and 42 false positive results. Therefore it is vital to follow a positive NIPT test with an invasive diagnostic test (amniocentesis or CVS) to confirm the presence of trisomy, if the woman is considering termination of pregnancy on the basis of trisomy.

The strengths of this systematic review included a comprehensive search of the literature, with quality appraisal of all included studies, with two authors sifting studies for inclusion, extracting data and appraising quality. The meta-analysis included rigorous methods of data analysis, including bivariate random-effects regression models and HSROC curve analysis. We also conducted a series of subgroup analyses and sensitivity analyses to test for robustness of our pooled diagnostic accuracy estimates. Homogeneous subgroup and sensitivity analysis summary accuracy estimates were generally similar to the overall estimates. We added predefined covariates to the model using meta-regression analyses to explain heterogeneity but considerable statistical heterogeneity remained. For some of the subgroup analyses, the relatively small number of studies available limited the generalisability of such pooled accuracy estimates. Finally we applied zero cell continuity correction of 0.5 to each cell of a study where a zero is encountered which tends to underestimate rather than overestimate test accuracy.

The findings of our review are in line with the results from previous reviews stating that NIPT has high performance in terms of sensitivity and specificity,^{61 62} that specificity is slightly higher than sensitivity,⁶¹ that the test performance is greater for T21 than for T18 and T13,⁴

Table 1 Summary of findings applied to high risk and general obstetric population

Condition	Summary accuracy	Median prevalence	Outcomes	Positive predictive value	Probability of false negative	Implications
General obstetric population (100 000 pregnancies)						
Down syndrome	Sensitivity=95.9% Specificity=99.9% (6 studies)	0.43%	TP=417 FP=94 TN=99471 FN=18	82%	1 in 5570	With prevalence of 0.4%, 435 of 100 000 pregnancies will be affected by Down syndrome. Of these 417 will be detected and 18 missed by cffDNA. Of the 99 565 who do not have Down syndrome, 94 will receive a false positive result. Therefore 82% of pregnancies which test positive will have Down syndrome
Edwards syndrome	Sensitivity=86.5% Specificity=99.8% (5 studies)	0.10%	TP=89 FP=154 TN=99744 FN=14	37%	1 in 7194	With prevalence of 0.1%, 102 of 100 000 pregnancies will be affected by Edwards syndrome. Of these 89 will be detected and 14 missed by cffDNA. Of the 99 898 who do not have Edwards syndrome, 154 will receive a false positive result. Therefore 37% of pregnancies which test positive will have Edwards syndrome
Patau syndrome	Sensitivity=77.5% Specificity=>99.9% (5 studies)	0.05%	TP=40 FP=42 TN=99906 FN=12	49%	1 in 8506	With prevalence of 0.05%, 52 of 100 000 pregnancies will be affected by Patau syndrome. Of these 40 will be detected and 12 missed by cffDNA. Of the 99 948 who do not have Patau syndrome, 42 will receive a false positive result. Therefore 49% of pregnancies which test positive will have Patau syndrome
High-risk population (10 000 pregnancies)						
Down syndrome	Sensitivity=97% Specificity=99.7% (22 studies)	3.33%	TP=324 FP=31 TN=9636 FN=9	91%	1 in 1054	With prevalence of 3.3%, 333 of 10 000 pregnancies will be affected by Down syndrome. Of these 324 will be detected and 9 missed by cffDNA. Of the 9667 who do not have Down syndrome, 31 will receive a false positive result. Therefore 91% of those who test positive will have Down syndrome
Edwards syndrome	Sensitivity=93% Specificity=99.7% (19 studies)	1.50%	TP=140 FP=26 TN=9824 FN=11	84%	1 in 930	With prevalence of 1.5%, 151 of 10 000 pregnancies will be affected by Edwards syndrome. Of these 140 will be detected and 11 missed by cffDNA. Of the 9850 who do not have Edwards syndrome, 26 will receive a false positive result. Therefore 84% of those who test positive will have Edwards syndrome
Patau syndrome	Sensitivity=95% Specificity=99.9% (11 studies)	0.50%	TP=47 FP=7 TN=9943 FN=3	87%	1 in 4265	With prevalence of 0.5%, 50 of 10 000 pregnancies will be affected by Patau syndrome. Of these 47 will be detected and 3 missed by cffDNA. Of the 9950 who do not have Patau syndrome, 7 will receive a false positive result. Therefore 87% of those who test positive will have Patau syndrome

Median prevalence determined from cohort studies included in meta-analysis for relevant populations. Estimates of sensitivity and specificity are from meta-analysis sub-groups for studies in high risk and general obstetric populations. The systematic review investigated test accuracy of non-invasive prenatal testing using cell-free DNA derived from maternal blood (serum, plasma, whole blood) in pregnant women in any trimester for the detection of Down, Edwards or Patau syndromes in the fetus. The reference standard was genetic verification through amniocentesis, CVS, cordocentesis, fetal pathological examination after abortion and postnatal phenotypic assessment. Findings should be interpreted with caution. Assessment using QUADAS-2 identified high risk of bias in included studies, particularly for selection of women and flow. Deeks' funnel plots indicated there was high risk of publication bias in included studies. Zero-cell corrections may have reduced accuracy estimates.

cffDNA, cell-free fetal DNA; CVS, Chorionic Villus Sampling; FN, false negative; FP, false positive; TN, true negative; TP, true positive.

Table 2 Accuracy estimates from sensitivity and subgroup analyses of the included studies by different study characteristics†

Variables	Down (trisomy 21)			Edwards (trisomy 18)			Patau (trisomy 13)		
	N	SN (95% CI)	SP (95% CI)	n	SN (95% CI)	SP (95% CI)	n	SN (95% CI)	SP (95% CI)
<i>All studies</i>	40	0.993 (0.989 to 0.996)	0.999 (0.999 to 1.000)	33	0.974 (0.958 to 0.984)	0.999 (0.999 to 1.000)	24	0.974 (0.861 to 0.996)	1.000 (0.999 to 1.000)
<i>Sensitivity analyses</i>									
Excluding outliers‡	37	0.993 (0.989 to 0.996)	1.000 (0.999 to 1.000)	32	0.977 (0.961 to 0.986)	0.999 (0.999 to 1.000)	22	0.977 (0.818 to 0.998)	1.000 (0.999 to 1.000)
<i>Test failures</i>									
Assuming all+ve	40	0.997 (0.990 to 0.999)	0.981 (0.972 to 0.988)	33	0.973 (0.956 to 0.983)	0.983 (0.974 to 0.990)	24	0.979 (0.873 to 0.997)	0.981 (0.966 to 0.989)
Assuming all−ve	40	0.962 (0.948 to 0.973)	1.000 (0.999 to 1.000)	33	0.942 (0.913 to 0.962)	0.999 (0.999 to 1.000)	24	0.885 (0.796 to 0.939)	1.000 (0.999 to 1.000)
Intention to diagnosis	40	0.976 (0.959 to 0.986)	0.981 (0.972 to 0.989)	33	0.958 (0.927 to 0.976)	0.983 (0.973 to 0.990)	24	0.903 (0.811 to 0.953)	0.981 (0.966 to 0.989)
Assuming all+ve	40	0.994 (0.989 to 0.997)	0.999 (0.999 to 1.000)	33	0.974 (0.958 to 0.985)	0.999 (0.999 to 1.000)	24	0.974 (0.863 to 0.996)	1.000 (0.999 to 1.000)
Assuming all−ve	40	0.993 (0.987 to 0.996)	0.999 (0.999 to 1.000)	33	0.970 (0.945 to 0.984)	0.999 (0.999 to 1.000)	24	0.976 (0.855 to 0.996)	1.000 (0.999 to 1.000)
Intention to diagnosis	40	0.993 (0.988 to 0.996)	0.999 (0.999 to 1.000)	33	0.972 (0.950 to 0.985)	0.999 (0.999 to 1.000)	24	0.976 (0.855 to 0.996)	1.000 (0.999 to 1.000)
<i>Subgroup analyses</i>									
<i>Study design</i>									
Cohort	5	0.932 (0.853 to 0.971)	0.999 (0.996 to 1.000)	4	0.868 (0.591 to 0.968)	0.998 (0.994 to 0.999)	3	0.851 (0.498 to 0.971)	0.999 (0.995 to 1.000)
Others	35	0.976 (0.963 to 0.985)	0.998 (0.997 to 0.999)	29	0.941 (0.914 to 0.960)	0.998 (0.997 to 0.999)	21	0.970 (0.852 to 0.994)	1.000 (0.999 to 1.000)
<i>Population risk</i>									
General	6	0.959 (0.874 to 0.987)	0.999 (0.998 to 1.000)	4	0.865 (0.627 to 0.961)	0.998 (0.997 to 0.999)	4	0.775 (0.135 to 0.987)§	1.000 (0.999 to 1.000)
High	22	0.973 (0.951 to 0.985)	0.997 (0.994 to 0.998)	19	0.930 (0.892 to 0.955)	0.997 (0.995 to 0.999)	11	0.953 (0.864 to 0.985)	0.999 (0.996 to 1.000)
Others	12	0.974 (0.940 to 0.989)	0.999 (0.998 to 0.999)	10	0.958 (0.907 to 0.982)	0.999 (0.999 to 1.000)	9	0.988 (0.547 to 1.000)	1.000 (0.999 to 1.000)
<i>Population</i>									
Others	36	0.977 (0.965 to 0.985)	0.998 (0.997 to 0.999)	31	0.943 (0.917 to 0.960)	0.998 (0.997 to 0.999)	23	0.974 (0.861 to 0.996)	1.000 (0.999 to 1.000)
Twins	4	0.894 (0.750 to 0.960)	0.996 (0.996 to 0.996)	2	0.737 (0.202 to 0.969)§	0.998 (0.986 to 1.000)	1*		
<i>First trimester</i>									
100%	7	0.960 (0.887 to 0.987)	0.999 (0.998 to 1.000)	5	0.925 (0.814 to 0.972)	0.998 (0.997 to 0.999)	5	0.850 (0.770 to 0.906)§	0.999 (0.998 to 0.999)
Others	33	0.973 (0.958 to 0.983)	0.998 (0.997 to 0.999)	28	0.939 (0.910 to 0.960)	0.998 (0.997 to 0.999)	19	0.966 (0.872 to 0.992)	1.000 (0.999 to 1.000)
<i>Test types</i>									
DANSR	9	0.958 (0.898 to 0.983)	0.999 (0.997 to 1.000)	6	0.948 (0.879 to 0.979)	0.998 (0.996 to 0.999)	3	0.606 (0.216 to 0.895)	1.000 (0.998 to 1.000)
MPSS	25	0.978 (0.963 to 0.987)	0.998 (0.997 to 0.999)	23	0.936 (0.899 to 0.960)	0.998 (0.997 to 0.999)	16	0.959 (0.989 to 0.991)	1.000 (0.999 to 1.000)
SNP technology	4	0.984 (0.937 to 0.996)	0.998 (0.993 to 1.000)	4	0.918 (0.751 to 0.976)	0.998 (0.994 to 1.000)	5	0.870 (0.647 to 0.960)	0.998 (0.992 to 0.999)
<i>Publication year</i>									
2007–2013	18	0.977 (0.958 to 0.988)	0.998 (0.995 to 0.999)	15	0.954 (0.919 to 0.975)	0.998 (0.995 to 0.999)	9	0.933 (0.799 to 0.980)	0.999 (0.993 to 1.000)
2014–2015	22	0.966 (0.939 to 0.981)	0.999 (0.998 to 0.999)	18	0.915 (0.853 to 0.952)	0.996 (0.998 to 0.999)	15	0.984 (0.770 to 0.999)	1.000 (0.999 to 1.000)

*Bivariate model inestimable for only one study in the subgroup.²³

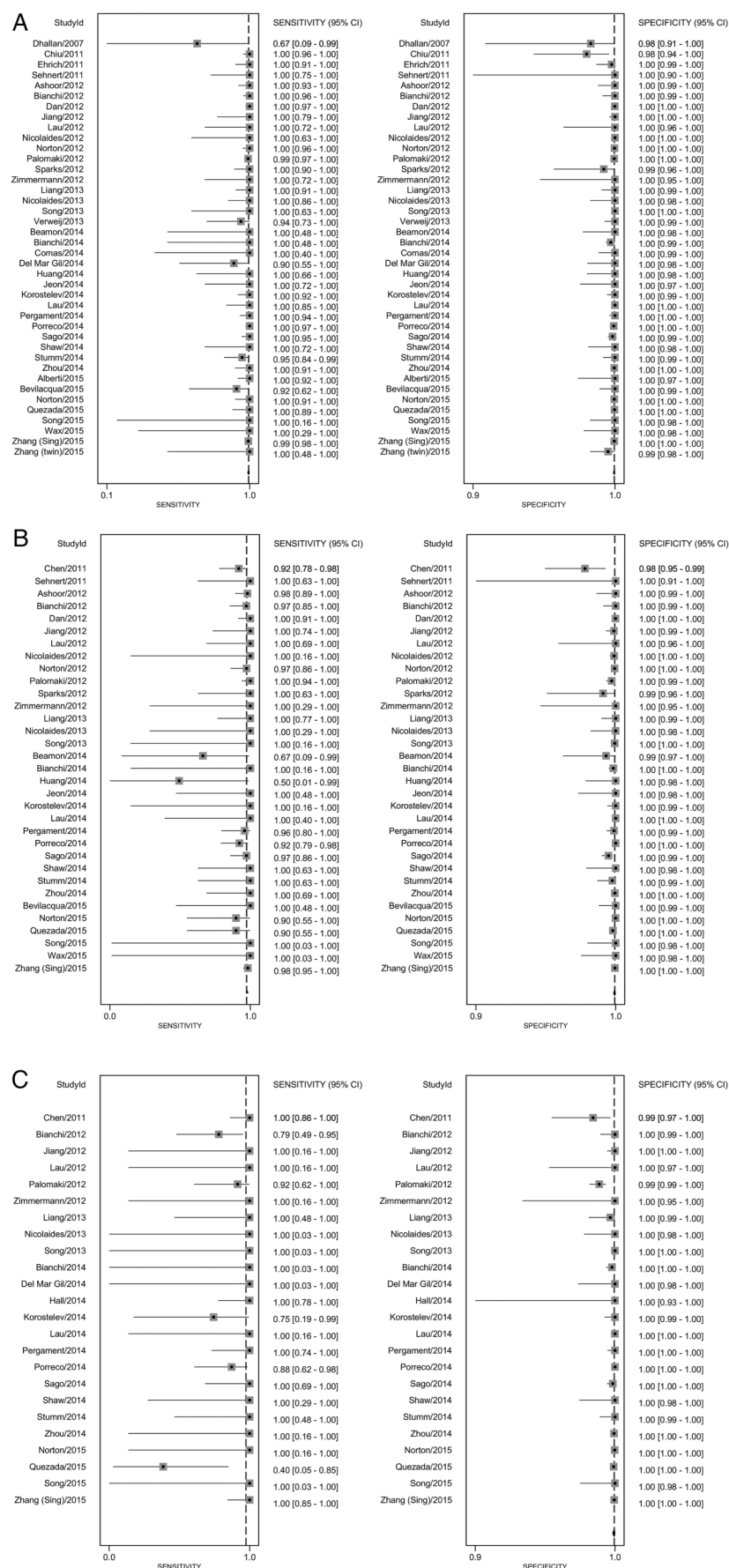
†Excluded studies with inestimable sensitivity (T21—Hall 2014; T18—Comas 2014, Hall 2014, Zhang (twins) 2015; T13—Sehnert 2011, Beamon 2014, Comas 2014, Bevilacqua 2015, Wax 2015, Zhang (twins) 2015).

‡Excluded outliers (T21—Dhallan 2007, Chiu 2011, Sparks 2012; T18—Chen 2011; T13—Chen 2011, Palomaki 2012).

§p Value for subgroup differences <0.05 (statistically significant).

SN, sensitivity; SNP, single nucleotide polymorphism; SP, specificity.

Figure 5 Individual and pooled sensitivity and specificity for non-invasive prenatal testing (NIPT) for the detection of a. Down syndrome b. Edwards syndrome and c. Patau syndrome.



and that NIPT is less successful in twin pregnancies than in singleton pregnancies.⁴ However, we found evidence of significant publication bias, converted results into a format interpretable by clinicians, and concluded that the test is not diagnostic. There are two key differences between our review and the previous publications. First, we included more studies, several of which have been published since the most recent review⁴ (including two of the largest studies with test accuracy for 128 510 women).^{5 6} Second, the two previous reviews conducted separate pooling of the diagnostic test accuracy measures using a univariate approach using standard methods for proportion^{4 62} which is not recommended for reviews of test accuracy. Berkey *et al*⁶³ show that a bivariate meta-regression is more efficient than separate univariate meta-regressions for assessing study-level covariates, due to the inclusion of correlation. We used Deeks' funnel plots and found evidence of publication bias, whereas the previous review used an Egger's bias applied to sensitivity and specificity separately and found no evidence of bias, although their method may not be appropriate for studies of test accuracy.¹⁷ Studies with a larger effective sample size tended to report higher diagnostic ORs. This may be due to publication bias in large laboratory cohort or case-control studies with a lack of systematic or consecutive sampling, or the fact that studies in the general obstetric population tend to have lower test accuracy and fewer cases. It may be partly due to our methods in that the zero-cell correction may disadvantage small studies, or simply that the test is performed to a higher standard in larger studies, perhaps due to more advanced protocols used in later large scale studies.

The implications for policymakers and clinicians are that NIPT using cfDNA has very high sensitivity and specificity, and can contribute to screening programmes for Down, Edwards and Patau syndromes. It is clear that test accuracy is very good but not perfect. This is particularly true when considering populations in terms of risk and gestational age. Our subgroup analyses showed that test performance is better in high-risk populations as well as in studies including pregnancies in the second and third trimester. Consideration of NIPT as a screening test for the general obstetric population primarily tested in the first trimester of pregnancy has to take into account the lower sensitivity of NIPT in this population. There is also some indication that higher maternal weight, and conception by in vitro fertilisation (IVF) are potential predictors of NIPT test failure³⁹ suggesting that NIPT may not work equally well in all subpopulations. We consider that for this reason cfDNA should not be regarded as a diagnostic test and that confirmation of a positive NIPT result by amniocentesis or CVS is necessary to make a diagnosis of trisomy. This is essential if parents are considering termination of pregnancy on the basis of trisomy, because in the general obstetric population as many as 20% of positive NIPT results for Down syndrome may be false positive. This proportion will be higher for Edwards

and Patau syndromes. Because the source of cfDNA is the placenta, confined placental mosaicism may explain a proportion of discordant NIPT results.⁶⁴ Furthermore, early fetal demise of an affected fetus^{53 64} and unknown chromosomal abnormality in the mother^{5 64} can lead to false positive results. Finally, in some cases discordance between NIPT and fetal karyotype results might be due to lab error.⁶⁴ The role of low fetal fraction as contributor to false positive or false negative results is unclear: Zhang *et al*⁵ reported no major influence, whereas Quezada *et al*⁴³ found lower fetal fractions in discordant than in those with concordant results.

Communicating to clinicians and patients that this genetic test is not perfect will be key for safe implementation, and pretest and post-test information provision and counselling for positive and negative NIPT results should be given careful consideration. The NIPT test may be particularly attractive to parents who are not considering termination of pregnancy, but who would like to know in advance if their pregnancy is affected by a trisomy, since NIPT gives broadly accurate results, without the slightly increased risk of miscarriage associated with invasive procedures such as amniocentesis and CVS. The final consideration for implementation is the range of test failure rates from <1% to >12%, with some evidence that presence of trisomy may be a predictor of test failure. Quality assurance to minimise test failures would minimise delays due to repeated testing, which may be a priority for pregnant women. However, if the test failure is due to insufficient fetal fraction a retest is also likely to fail.

This test is used worldwide, mostly provided directly by private providers rather than national health systems. Further research into how the test is being interpreted and understood by clinicians and pregnant women will be key to understanding the balance of benefits and harms from the provision of the test. In particular, how this understanding leads to decisions about whether to continue the pregnancy, and whether this may be influenced by how the test is presented to parents both by companies, and by clinicians. Finally if it is implemented into national screening programmes, keeping accurate records of outcomes and test failures would enable the test performance to be evaluated in practice. This may differ from the test accuracy in the included studies in this paper, due to the high risk of bias in included studies of cfDNA, and the unexplained heterogeneity illustrating the uncertainties in transferring results from research studies into everyday practice.

Acknowledgements The authors would like to thank Jill Colquitt and Emma Loveman from Effective Evidence LLP, Eastleigh, for sourcing parameters for the economic model, Paul Hewitson for his work on the search strategy and sifting, Frances Taggart for her work on test accuracy at different thresholds, and Pam Royle for her work updating the searches.

Contributors STP led the research design and implementation. JG and AA sifted and quality appraised the studies. OA conducted the meta-analysis. All authors contributed to study design, data collection and interpretation, and writing and redrafting the paper, and approved the final version of the paper.

Funding This is independent research commissioned by the UK National Screening Committee. Sian Taylor-Phillips is supported by a clinical trials fellowship from the National Institute for Health Research (NIHR). Aileen Clarke is supported by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West Midlands at University Hospitals Birmingham NHS Foundation Trust. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, the UK National Screening Committee, Public Health England or the Department of Health. Any errors are the responsibility of the authors.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>

REFERENCES

- Chandrasekharan S, Minear MA, Hung A, *et al.* Noninvasive prenatal testing goes global. *Sci Transl Med* 2014;6:231f15.
- Chetty S, Garabedian MJ, Norton ME. Uptake of noninvasive prenatal testing (NIPT) in women following positive aneuploidy screening. *Prenat Diagn* 2013;33:542–6.
- Leighton JW, Valverde K, Bernhardt BA. The general public's understanding and perception of direct-to-consumer genetic test results. *Public Health Genomics* 2012;15:11–21.
- Gil MM, Quezada MS, Revello R, *et al.* Analysis of cell-free DNA in maternal blood in screening for fetal aneuploidies: updated meta-analysis. *Ultrasound Obstet Gynecol* 2015;45:249–66.
- Zhang H, Gao Y, Jiang F, *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.
- Norton ME, Jacobsson B, Swamy GK, *et al.* Cell-free DNA analysis for noninvasive examination of trisomy. *N Engl J Med* 2015;372:1589–97.
- Reitsma JB, Glas AS, Rutjes AW, *et al.* Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
- Hand DJ. Evaluating diagnostic tests: the area under the ROC curve and the balance of errors. *Stat Med* 2010;29:1502–10.
- Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. *Stat Med* 2011;30:2481–98.
- Whiting PF, Rutjes AW, Westwood ME, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
- Leeflang MM, Deeks JJ, Gatsonis C, *et al.* Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008;149:889–97.
- Dwamena BA. *MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies*. Boston, MA: Boston College Department of Economics, 2008.
- Schuetz GM, Schlattmann P, Dewey M. Use of 3x2 tables with an intention to diagnose approach to assess clinical performance of diagnostic tests: meta-analytical evaluation of coronary CT angiography studies. *BMJ* 2012;345:e6717.
- Shinkins B, Thompson M, Mallett S, *et al.* Diagnostic accuracy studies: how to report and analyse inconclusive test results. *BMJ* 2013;346:f2778.
- Glas AS, Lijmer JG, Prins MH, *et al.* The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003;56:1129–35.
- Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002;21:1525–37.
- 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.
- Harbord RM, Higgins JPT. Meta-regression in Stata. *Stata J* 2008;8:493–519.
- Harbord RM, Whiting P. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata J* 2009;9:211–29.
- White IR. Multivariate random-effects meta-regression: updates to mvmeta. *Stata J* 2011;11:255–70.
- Bianchi DW, Parker RL, Wentworth J, *et al.* DNA sequencing versus standard prenatal aneuploidy screening. *N Engl J Med* 2014;370:799–808.
- Dan S, Ren J, Hu H, *et al.* Clinical application of massively parallel sequencing-based prenatal noninvasive fetal trisomy test for trisomies 21 and 18 in 11105 pregnancies with mixed risk factors. *Prenat Diagn* 2012;32:1225–32.
- del Mar Gil M, Quezada MS, Bregant B, *et al.* Cell-free DNA analysis for trisomy risk assessment in first-trimester twin pregnancies. *Fetal Diagn Ther* 2014;35:204–11.
- Huang X, Zheng J, Chen M, *et al.* Noninvasive prenatal testing of trisomies 21 and 18 by massively parallel sequencing of maternal plasma DNA in twin pregnancies. *Prenat Diagn* 2014;34:335–40.
- Jiang F, Ren J, Chen F, *et al.* Noninvasive Fetal Trisomy (NIFTY) test: an advanced noninvasive prenatal diagnosis methodology for fetal autosomal and sex chromosomal aneuploidies. *BMC Medical Genomics* 2012;5:57.
- Lau TK, Chen F, Pan X, *et al.* Noninvasive prenatal diagnosis of common fetal chromosomal aneuploidies by maternal plasma DNA sequencing. *J Matern Fetal Neonatal Med* 2012;25:1370–4.
- Lau TK, Cheung SW, Lo PS, *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.
- Liang D, Lv W, Wang H, *et al.* Non-invasive prenatal testing of fetal whole chromosome aneuploidy by massively parallel sequencing. *Prenat Diagn* 2013;33:409–15.
- Nicolaides KH, Syngelaki A, Ashoor G, *et al.* Noninvasive prenatal testing for fetal trisomies in a routinely screened first-trimester population. *Am J Obstet Gynecol* 2012;207:374.e1–6.
- Nicolaides KH, Syngelaki A, Gil M, *et al.* 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.
- Norton ME, Brar H, Weiss J, *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.
- Pergament E, Cuckle H, Zimmermann B, *et al.* Single-nucleotide polymorphism-based noninvasive prenatal screening in a high-risk and low-risk cohort. *Obstet Gynecol* 2014;124(2 Pt 1):210–18.
- Porreco RP, Garite TJ, Maurel K, *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.e1–12.
- Shaw SW, Hsiao CH, Chen CY, *et al.* Noninvasive prenatal testing for whole fetal chromosomal aneuploidies: a multicenter prospective cohort trial in Taiwan. *Fetal Diagn Ther* 2014;35:13–17.
- Song Y, Liu C, Qi H, *et al.* Noninvasive prenatal testing of fetal aneuploidies by massively parallel sequencing in a prospective Chinese population. *Prenat Diagn* 2013;33:700–6.
- Stumm M, Entezami M, Haug K, *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.
- Verweij EJ, Jacobsson B, van Scheltema PA, *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.
- Beamon CJ, Hardisty EE, Harris SC, *et al.* A single center's experience with noninvasive prenatal testing. *Genet Med* 2014;16:681–7.
- Bevilacqua E, Gil MM, Nicolaides KH, *et al.* Performance of screening for aneuploidies by cell-free DNA analysis of maternal blood in twin pregnancies. *Ultrasound Obstet Gynecol* 2015;45:61–6.
- Comas C, Echevarria M, Rodriguez MA, *et al.* 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;28:1196–201.
- Jeon YJ, Zhou Y, Li Y, *et al.* The feasibility study of non-invasive fetal trisomy 18 and 21 detection with semiconductor sequencing platform. *PLoS ONE* 2014;9:e110240.
- Korostelev S, Totchiev G, Kanivets I, *et al.* Association of non-invasive prenatal testing and chromosomal microarray analysis

- for prenatal diagnostics. *Gynecol Endocrinol* 2014;30(Suppl 1): 13–16.
43. Quezada MS, Gil MM, Francisco C, *et al.* 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.
 44. Sago H, Sekizawa A. Nationwide demonstration project of next-generation sequencing of cell-free DNA in maternal plasma in Japan: one-year experience. *Prenat Diagn* 2015;35:331–6.
 45. Wax JR, Cartin A, Chard R, *et al.* Noninvasive prenatal testing: impact on genetic counseling, invasive prenatal diagnosis, and trisomy 21 detection. *J Clin Ultrasound* 2015;43:1–6.
 46. Zhou Q, Pan L, Chen S, *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.
 47. Song Y, Huang S, Zhou X, *et al.* Non-invasive prenatal testing for fetal aneuploidies in the first trimester of pregnancy. *Ultrasound Obstet Gynecol* 2015;45:55–60.
 48. Ashoor G, Syngelaki A, Wagner M, *et al.* Chromosome-selective sequencing of maternal plasma cell-free DNA for first-trimester detection of trisomy 21 and trisomy 18. *Am J Obstet Gynecol* 2012;206:322.e1–5.
 49. Bianchi DW, Platt LD, Goldberg JD, *et al.* Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. *Obstet Gynecol* 2012;119:890–901.
 50. Chen EZ, Chiu RW, Sun H, *et al.* Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. *PLoS ONE* 2011;6:e21791.
 51. Chiu RW, Akolekar R, Zheng YW, *et al.* Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. *BMJ* 2011;342:c7401.
 52. Ehrich M, Deciu C, Zwiefelhofer T, *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;204:205.e1–11.
 53. Hall MP, Hill M, Zimmermann B, *et al.* Non-invasive prenatal detection of trisomy 13 using a single nucleotide polymorphism- and informatics-based approach. *PLoS ONE* 2014;9:e96677.
 54. Palomaki GE, Deciu C, Kloza EM, *et al.* DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international collaborative study. *Genet Med* 2012;14:296–305.
 55. Sehntert AJ, Rhee B, Comstock D, *et al.* Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free fetal DNA from maternal blood. *Clin Chem* 2011;57:1042–9.
 56. Sparks AB, Struble CA, Wang ET, *et al.* 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.
 57. Zimmermann B, Hill M, Gemelos G, *et al.* Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci. *Prenat Diagn* 2012;32:1233–41.
 58. Alberti A, Salomon LJ, Le Lorc'h M, *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.
 59. Dhallan R, Guo X, Emche S, *et al.* A non-invasive test for prenatal diagnosis based on fetal DNA present in maternal blood: a preliminary study. *Lancet* 2007;369:474–81.
 60. Futch T, Spinosa J, Bhatt S, *et al.* Initial clinical laboratory experience in noninvasive prenatal testing for fetal aneuploidy from maternal plasma DNA samples. *Prenat Diagn* 2013;33:569–74.
 61. Mersy E, Smits LJ, van Winden LA, *et al.* Noninvasive detection of fetal trisomy 21: systematic review and report of quality and outcomes of diagnostic accuracy studies performed between 1997 and 2012. *Hum Reprod Update* 2013;19:318–29.
 62. Verweij EJ, van den Oever JM, de Boer MA, *et al.* Diagnostic accuracy of noninvasive detection of fetal trisomy 21 in maternal blood: a systematic review. *Fetal Diagn Ther* 2012;31:81–6.
 63. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, *et al.* Meta-analysis of multiple outcomes by regression with random effects. *Stat Med* 1998;17:2537–50.
 64. McCullough RM, Almasri EA, Guan X, *et al.* Non-invasive prenatal chromosomal aneuploidy testing—clinical experience: 100,000 clinical samples. *PLoS ONE* 2014;9:e109173.

Supplement 1 Search strategy

Ovid Medline (1997 to 9th February 2015)

1. ((noninvasive or non-invasive or non invasive) adj3 (prenatal or pre?natal* or pregnanc* or diagnos* or test* or detect* or screen* or assess*)).mp.
2. (NIPD or NIPT).mp.
3. (cf?DNA or cff?DNA or ccff?DNA or cell?free?DNA).mp.
4. (DNA adj1 (cell or free or cell?free or f?etal)).mp.
5. (maternal adj1 (blood or plasma or DNA)).mp.
6. (MPS or DANSR or parental support or MaterniT21 or Verifi* or Harmony or Panorama*).mp.
7. 1 or 2 or 3 or 4 or 5 or 6
8. Trisomy/
9. trisom*.mp.
10. Aneuploidy/
11. aneuploid*.mp.
12. Down Syndrome/
13. (down* adj1 syndrom*).mp.
14. (edward* adj1 syndrom*).mp.
15. (Patau adj1 syndrom*).mp.
16. ("T21" or "T18" or "T13").mp.
17. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16
18. 7 and 17
19. limit 18 to yr="1997 -Current"
20. limit 19 to english language

Ovid Embase (1997 to 9th February 2015)

1. ((noninvasive or non-invasive or non invasive) adj3 (prenatal or pre?natal* or pregnanc* or diagnos* or test* or detect* or screen* or assess*)).mp.
2. (NIPD or NIPT).mp.
3. (cf?DNA or cff?DNA or ccff?DNA or cell?free?DNA).mp.
4. (DNA adj1 (cell or free or cell?free or f?etal)).mp.
5. (maternal adj1 (blood or plasma or DNA)).mp.

6. (MPS or DANSR or parental support or MaterniT21 or Verifi* or Harmony or Panorama*).mp.
7. 1 or 2 or 3 or 4 or 5 or 6
8. Trisomy/
9. trisom*.mp.
10. Aneuploidy/
11. aneuploid*.mp.
12. Down Syndrome/
13. (down* adj1 syndrom*).mp.
14. (edward* adj1 syndrom*).mp.
15. (Patau adj1 syndrom*).mp.
16. ("T21" or "T18" or "T13").mp.
17. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16
18. 7 and 17
19. limit 18 to yr="1997 -Current"
20. limit 19 to english language

Cochrane Library – all sections – February 2015

((noninvasive or non-invasive or non invasive) near/3 (prenatal or pre?natal* or pregnanc* or diagnos* or test* or detect* or screen* or assess*)) in Title, Abstract, Keywords or (NIPD or NIPT) in Title, Abstract, Keywords or (cfDNA or cffDNA or ccffDNA or "cell free DNA") in Title, Abstract, Keywords or (DNA near/3 (cell or free or cell?free or f?etal)) in Title, Abstract, Keywords or (maternal near/3 (blood or plasma or DNA)) in Title, Abstract, Keywords (Word variations have been searched)

Supplement 2: Modified QUADAS-2 and guidance notes for NIPT test accuracy papers

Domain 1: Patient selection

As a proportion of studies used a case-control design, the selection of study participants is of concern. This includes exclusion of hard to diagnose cases including twin pregnancies, pregnancies featuring mosaicism or translocations and homozygous fetuses in the approaches based on SNP markers.

A. Risk of bias

Guidance:

Was a consecutive or random sample of patients enrolled?

This question should only be answered with 'yes' if the study clearly states that pregnancies (rather than samples) were recruited consecutively or randomly.

Was a case-control design avoided?

For the head to head comparison question we would ideally hope for randomization to NIPT and combined test or at least a screening observational study where all participants received both tests.

For the NIPT performance question we would at least expect a prospective cohort design. Therefore, if the study is a case-control study this question should be answered with No.

Did the study avoid inappropriate exclusions?

If the study excludes >10% of participants with or without specifying reasons, the exclusions should be considered as inappropriate. This cut-off has been determined pragmatically.

B. Concerns regarding applicability

Guidance:

As the research question aims to address NIPT test performance in the first trimester and in comparison with the first trimester combined test, applicability should be regarded low if <80% of women were recruited in the first trimester.

A screening and diagnostic context should be considered separately. Low risk women without prior tests should be considered for the screening context, while high risk women should be considered for the diagnostic context (this includes add-on and triage). Both scenarios match the different research questions but the study results will be applicable only to one of the two different contexts.

The setting where samples are taken is unlikely to have an effect on the spectrum of patients. However, the setting of the study might have an impact on the applicability of the study results to general practice in terms of feasibility, if the equipment or standards of the study setting are unlikely to be met by the routine laboratory carrying out the tests in clinical practice. Some of the technologies used in the studies might not be feasible to be carried out in routine laboratories. It needs to be decided how applicable the results of these studies are to routine practice but also whether the index test is likely to be carried out in routine laboratories or in a few specialised centers. In the UK foetal testing for sex-linked disorders and RHD genotyping is carried out in a small number of specialised centres.

Domain 2: Index test

The main sources of bias introduced by conducting and interpreting the index test are blinding and defining the threshold. Furthermore, concentrating on pregnancies with increased foetal material will bias the results, therefore, sampling should be carried out before or 7 days after invasive procedures, to avoid testing when foetal DNA levels are increased due to the invasive procedure.¹ If the reference standard is carried out before the index test (e.g. in case control studies) it is important to blind personnel to the karyotype results of the fetuses.

The QUADAS 2 tool requires a threshold to be pre-specified in the methods in order to avoid adjustment of the threshold according to the test outcome. However, the testing strategies considered in this review present a further level of concern. While an explicit threshold can be reported by studies (e.g. $z\text{-score} > 3\text{ SD}$), the value of the threshold is determined by the study using either an independent set of samples or the study controls. The study threshold is therefore study specific and is dependent on the participants sampled and/or the study protocol used. This was demonstrated by one study that needed to adjust a pre-specified threshold value that a previous study had determined.² Since the population mean and standard deviation are not known, studies will have to determine their own threshold values. This review will, therefore, consider independent samples of participants to determine the threshold value as aiming to reduce bias.

A. Risk of bias

Were the index test results interpreted without knowledge of the results of the reference standard?

Due to the sequence of the tests, the studies need to report blinding clearly in order to answer this question with 'yes'. Blinding can also take place by carrying out tests at different locations.

Was the sample for the index test taken before the invasive test or 7 days after invasive testing?

If the answer to this question is 'no', the risk of bias should be considered as 'high', since the accuracy of the index test will be affected by the increased amount of foetal material in the maternal circulation following invasive procedures. Lo et al. (1999) showed that testing before and 7 days after amniocentesis did not result in different DNA levels due to rapid clearance of fetal DNA from maternal blood.¹

Was a threshold explicitly pre-specified?

For this question to be answered with 'yes' the study needs to mention what kind of threshold was to be used (e.g. $z\text{-score} > 3SD$, $\text{mean} \pm 1.96SD$) and clearly state that it was specified before the start of the study.

Was the threshold value determined using an independent set of samples?

If the study used a sample of euploid controls to define an interval/threshold, the question should be answered with 'no' and the risk of bias is 'high'. A threshold determined in this way is unlikely to be robust and would lead to poorer results in an independent sample.

Studies with blinding to reference standard, blood sampling prior invasive testing, but insufficient information on the threshold used, can be classified as low-risk of bias when a commercially available non-invasive prenatal test was used.

B. Concerns about applicability

Concerns about applicability should be classified as 'high' if the index test included paternal genetic samples for all NIPT analyses.

If the study uses different screening tests to the first trimester combined test in $>80\%$, the applicability of studies comparing NIPT to the first trimester combined test should be classed as 'high' concern about the applicability.

Domain 3: Reference standard

Due to the nature of the reference standards there is little concern about bias introduced by the choice of reference standard. We accepted prenatal or postnatal karyotyping or phenotypic newborn assessment as appropriate reference standard. They all display a detection rate of over 99% and are routine procedures in prenatal diagnosis³. If the index test is carried out before the reference standard, blinding to the results of the index tests is important.

A. Risk of bias

Is the reference standard likely to correctly classify the target condition?

Amniocentesis and CVS achieve a sensitivity and specificity of close to 100%³. Several attempts to retrieve the sample might be necessary but diagnosis is very accurate. For studies that used the stated reference standards this question should be answered with 'yes'.

Were the reference standard results interpreted without knowledge of the results of the index test?

This question should be answered with 'yes' if the routine reference standards are carried out at a different location to the index test or if the samples for the index test were stored and the index test carried out after the reference standard. However, if the question is answered with 'unclear', the risk of bias can still be regarded as low, since the laboratories carrying out the reference standards as routine tests, are unlikely to be influenced by the index test.

B. Concerns about applicability

The concern of applicability of the reference standard will be low if one of the pre-defined reference standards was used in the studies.

Domain 4: Flow and Timing

Since foetal trisomies are not progressive conditions, time intervals do not affect the performance of NIPT tests. Furthermore, all reference test have close to 100% accuracy, therefore verification bias is of little concern in studies where low risk women do not receive an invasive test but are followed up till birth. However, the exclusion of difficult to test patients and the exclusion of samples from the analysis are of great concern. These include exclusion from the study, inconclusive / intermediate results, homozygotes not testable in SNP studies, test failures and uninterpretable results.

A. Risk of bias

Did all patients receive a reference standard?

This question can be answered with 'yes' if the participants are recruited on the basis of their karyotype results.

Did all patients receive the same reference standard?

Even if this question is answered with 'no', the risk of bias can be considered as being low as long as all participants received a reference standard because all included reference standards have equally high accuracy.

Were all patients included in the analysis?

If samples were excluded due to sample issues that can be resolved by re-sampling, the risk of bias can be considered as low even if it is answered with 'no'.

However, if samples were excluded because they did not pass quality controls (e.g. amount of DNA), the risk of bias is high because this might include early pregnancies or intermediate risk pregnancies where foetal DNA levels are low.

If inconclusive or intermediate results are not included the question should be answered with 'no' and the risk of bias considered high.

Domain 5: Role of sponsor

Studies sponsored by companies are likely to be biased if the company has influence on the study design, conduct, interpretation of results and decision to publish.

A. Risk of bias

Did the funding source/sponsor play no role in design of study, interpretation of results and publication?

The risk of bias regarding the role of sponsor should be considered as 'high' if studies were funded by profit-making companies and involvement of the sponsor in the design or conduct of the study or publication was stated and/or if the majority of authors or main authors were employees or shareholders of companies offering NIPT or cytogenetic tests and/or other conflicts of interest (i.e. patents, stock or stock options) were declared.

To answer this question with 'yes', the study needs to clearly state that sponsors played no role.

References

1. Lo YM, Lau TK, Zhang J, Leung TN, Chang AM, Hjelm NM, et al. Increased fetal DNA concentrations in the plasma of pregnant women carrying fetuses with trisomy 21. *Clin Chem* 1999;45(10):1747-51.
2. 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. *Genetics in medicine : official journal of the American College of Medical Genetics* 2011;13(11):913-20.
3. Dick P. Periodic health examination, 1996 update: 1. Prenatal screening for and diagnosis of Down syndrome. Canadian Task Force on the Periodic Health Examination. *Canadian Medical Association Journal* 1996;154(4):465-79.

Supplement 3 List of included studies

1. 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.
2. 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. *Am J Obstet Gynecol*. 2012;206(4):322.e1-5.
3. Beamon CJ, Hardisty EE, Harris SC, Vora NL. A single center's experience with noninvasive prenatal testing. *Genet Med*. 2014;16(9):681-7.
4. Bevilacqua E, Gil MM, Nicolaides KH, Ordonez E, Cirigliano V, Dierickx H, et al. Performance of screening for aneuploidies by cell-free DNA analysis of maternal blood in twin pregnancies. *Ultrasound Obstet Gynecol*. 2015;45(1):61-6.
5. 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(9):799-808.
6. Bianchi DW, Platt LD, Goldberg JD, Abuhamad AZ, Sehnert AJ, Rava RP. Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. *Obstetrics and Gynecology*. 2012;119(5):890-901.
7. 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(7):e21791.
8. 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*. 2011;342:c7401.
9. 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*. 2014;Epub ahead of print:1-6.
10. Dan S, Ren J, Hu H, Xu Z, Lau TK, Zhao W, et al. Clinical application of massively parallel sequencing-based prenatal noninvasive fetal trisomy test for trisomies 21 and 18 in 11105 pregnancies with mixed risk factors. *Prenat Diagn*. 2012;32(13):1225-32.
11. del Mar Gil M, Quezada MS, Bregant B, Syngelaki A, Nicolaides KH. Cell-free DNA analysis for trisomy risk assessment in first-trimester twin pregnancies. *Fetal Diagn Ther*. 2014;35(3):204-11.
12. Dhallan R, Guo X, Emche S, Damewood M, Bayliss P, Cronin M, et al. A non-invasive test for prenatal diagnosis based on fetal DNA present in maternal blood: a preliminary study. *Lancet*. 2007;369(9560):474-81.
13. 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.
14. Hall MP, Hill M, Zimmermann B, Sigurjonsson S, Westemeyer M, Saucier J, et al. Non-invasive prenatal detection of trisomy 13 using a single nucleotide polymorphism- and informatics-based approach. *PLoS ONE*. 2014;9(5):e96677.
15. Huang X, Zheng J, Chen M, Zhao Y, Zhang C, Liu L, et al. Noninvasive prenatal testing of trisomies 21 and 18 by massively parallel sequencing of maternal plasma DNA in twin pregnancies. *Prenat Diagn*. 2014;34(4):335-40.
16. Jeon YJ, Zhou Y, Li Y, Guo Q, Chen J, Quan S, et al. The feasibility study of non-invasive fetal trisomy 18 and 21 detection with semiconductor sequencing platform. *PLoS ONE*. 2014;9(10):e110240.
17. Jiang F, Ren J, Chen F, Zhou Y, Dan S, Su Y, et al. Noninvasive Fetal Trisomy (NIFTY) test: An advanced noninvasive prenatal diagnosis methodology for fetal autosomal and sex chromosomal aneuploidies. *BMC Medical Genomics*. 2012;5(57).
18. Korostelev S, Totchiev G, Kanivets I, Gnetetskaya V. Association of non-invasive prenatal testing and chromosomal microarray analysis for prenatal diagnostics. *Gynecological Endocrinology*. 2014;30 Suppl 1:13-6.
19. Lau TK, Chen F, Pan X, Pooh RK, Jiang F, Li Y, et al. Noninvasive prenatal diagnosis of common fetal chromosomal aneuploidies by maternal plasma DNA sequencing. *J Matern Fetal Neonatal Med*. 2012;25(8):1370-4.
20. 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(3):254-64.
21. 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(5):409-15.

22. 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(5):374.e1-6.
23. 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(6):575-9.
24. 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. *American Journal of Obstetrics and Gynecology.* 2012;207(2):137.e1-.e8.
25. 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(17):1589-97.
26. 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. *Genet Med.* 2012;14(3):296-305.
27. 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(2 Pt 1):210-8.
28. 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(4):365.e1-12.
29. 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(1):36-41.
30. Sago H, Sekizawa A. Nationwide demonstration project of next-generation sequencing of cell-free DNA in maternal plasma in Japan: one-year experience. *Prenat Diagn.* 2014.
31. Sehnert AJ, Rhees B, Comstock D, de Feo E, Heilek G, Burke J, et al. Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free fetal DNA from maternal blood. *Clin Chem.* 2011;57(7):1042-9.
32. 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(1):13-7.
33. 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(1):55-60.
34. 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(7):700-6.
35. 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(4):319.e1-9.
36. 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(2):185-91.
37. 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(10):996-1001.
38. Wax JR, Cartin A, Chard R, Lucas FL, Pinette MG. Noninvasive prenatal testing: Impact on genetic counseling, invasive prenatal diagnosis, and trisomy 21 detection. *Journal of Clinical Ultrasound.* 2015;43(1):1-6.
39. Zhang H, Gao Y, Jiang F, Fu M, Yuan Y, Guo Y, et al. Noninvasive Prenatal Testing for Trisomy 21, 18 and 13 - Clinical Experience from 146,958 Pregnancies. *Ultrasound Obstet Gynecol.* 2015.
40. 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(11):1061-5.
41. Zimmermann B, Hill M, Gemelos G, Demko Z, Banjevic M, Baner J, et al. Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci. *Prenat Diagn.* 2012;32(13):1233-41.

Supplement 4 Table of excluded studies with reason

Reference	Reason for exclusion
1. Anonymous. Cell-free fetal DNA tests for trisomy show promise in women at lower risk of affected pregnancies: lower rates of false-positive returns, higher positive predictive value are associated with cfDNA tests versus standard screening panels, say experts. <i>Am J Med Genet A</i> 2014;164A(6):viii-ix.	Commentary
2. Anonymous. Trisomy 21 DNA test (MaterniT21) for detecting Down syndrome in the first trimester. <i>Manag Care</i> 2012;21(4):19-20.	Commentary
3. 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. <i>Ultrasound Obstet Gynecol</i> 2013;41(1):21-5. 269	Case control studies: <15 cases
4. Bianchi DW, Lamar Parker R, Wentworth J, Madankumar R, Saffer C, Das AF, et al. DNA sequencing versus standard prenatal aneuploidy screening. <i>Obstetrical and Gynecological Survey</i> . 2014;69(6):319-21.	Editorial
5. Canick, J.A., et al., DNA sequencing of maternal plasma to identify Down syndrome and other trisomies in multiple gestations. <i>Prenatal Diagnosis</i> , 2012. 32(8): p. 730-4.	Nested case-control study: < 15 cases
6. Chiu, R.W., et al., Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2008. 105(51): p. 20458-63.	Case-control study: < 15 cases
7. Dar P, Curnow KJ, Gross SJ, Hall MP, Stosic M, Demko Z, et al. Clinical experience and follow-up with large scale single-nucleotide polymorphism-based noninvasive prenatal aneuploidy testing. <i>Am J Obstet Gynecol</i> . 2014;211(5):527.e1-.e17.	Incomplete 2x2 table; index test results used as inclusion criteria so incomplete 2x2 table
8. Deng, Y.H., et al., Non-invasive prenatal diagnosis of trisomy 21 by reverse transcriptase multiplex ligation-dependent probe amplification. <i>Clinical Chemistry & Laboratory Medicine</i> , 2011. 49(4): p. 641-6.	Not cff DNA (cell-free fetal RNA)
9. Dugo N, Padula F, Mobili L, Brizzi C, D'Emidio L, Cignini P, et al. Six consecutive false positive cases from cell-free fetal DNA testing in a single	Case series: < 15 cases

referring centre. Journal of Prenatal Medicine. 2014;8(1-2):31-5.	
10. Faas, B.H., et al., Non-invasive prenatal diagnosis of fetal aneuploidies using massively parallel sequencing-by-ligation and evidence that cell-free fetal DNA in the maternal plasma originates from cytotrophoblastic cells. Expert Opinion on Biological Therapy, 2012. 12 Suppl 1: p. S19-26.	Case series: < 40 women
11. Fairbrother, G., et al., Clinical experience of noninvasive prenatal testing with cell-free DNA for fetal trisomies 21, 18, and 13, in a general screening population. Prenatal Diagnosis, 2013. 33(6): p. 580-583.	No reference standard results
12. Fang Y, Wang G, Wang C, Suo F, Gu M, Xia Y. The Diagnosis Pattern of Mid-Trimester Fetal Chromosomal Aneuploidy in Xuzhou and the Clinical Applications. Cell biochemistry and biophysics. 2015.	Incomplete 2x2 table; unclear reporting, incomplete follow up of cffDNA testing negative cases
13. Feenstra, H., et al., Complexity of noninvasive prenatal screening and diagnostic testing for an unbalanced translocation involving chromosomes 5 and 18. Prenatal Diagnosis, 2014. 34: p. 195-198.	Case report
14. Futch T, Spinosa J, Bhatt S, de Feo E, Rava RP, Sehnert AJ. Initial clinical laboratory experience in noninvasive prenatal testing for fetal aneuploidy from maternal plasma DNA samples. Prenat Diagn. 2013;33(6):569-74.	Incomplete 2x2 table; index test results used as inclusion criteria so incomplete 2x2 table
15. Ghanta, S., et al., Non-invasive prenatal detection of trisomy 21 using tandem single nucleotide polymorphisms. PLoS ONE [Electronic Resource], 2010. 5(10): p. e13184.	Case-control study: < 15 cases
16. Gil, M.M., et al., Implementation of maternal blood cell-free DNA testing in early screening for aneuploidies. Ultrasound in Obstetrics & Gynecology, 2013. 42(1): p. 34-40.	Cohort study: < 50 women with index and reference test result
17. Grati, F.R., et al., Fetoplacental mosaicism: potential implications for false-positive and false-negative noninvasive prenatal screening results. Genetics in Medicine, 2014. 16(8): p. 620-4.	Not cff DNA (cytogenetic material from CVS/Amnio)
18. Gromminger, S., et al., Fetal aneuploidy detection by cell-free DNA sequencing for multiple pregnancies and quality issues with vanishing twins. Journal of Clinical Medicine, 2014. 3(3): p. 679-692.	Cohort study: < 50 women
19. Guex, N., et al., A robust second-generation	Letter

genome-wide test for fetal aneuploidy based on shotgun sequencing cell-free DNA in maternal blood. Prenatal Diagnosis, 2013. 33: p. 707-710.	
20. Guo, Q., et al., Simultaneous detection of trisomies 13, 18, and 21 with multiplex ligation-dependent probe amplification-based real-time PCR. Clinical Chemistry, 2010. 56(9): p. 1451-9.	Participants not pregnant women
21. Hayes Inc., Harmony Prenatal Test (Structured abstract). Health Technology Assessment Database, 2012.	Abstract of review
22. Hayes Inc., Noninvasive Prenatal Testing (NIPT) for fetal aneuploidy (Structured abstract). Health Technology Assessment Database, 2013.	Abstract of review
23. Hill, M., et al., Evaluation of non-invasive prenatal testing (NIPT) for aneuploidy in an NHS setting: a reliable accurate prenatal non-invasive diagnosis (RAPID) protocol. BMC Pregnancy & Childbirth, 2014. 14: p. 229.	Protocol, no data presented
24. Hyett J. Non-invasive prenatal testing for down syndrome. Australian Prescriber. 2014;37(2):51-5.	Review
25. 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(3):e57381.	Re-uses some of the same samples as Palomaki et al. (2012); excluded to prevent double counting
26. Jorgez, C.J., et al., Elevated levels of total (maternal and fetal) beta-globin DNA in maternal blood from first trimester pregnancies with trisomy 21. Human Reproduction, 2007. 22(8): p. 2267-72.	Measurement of total blood DNA levels
27. Juneau K, Bogard PE, Huang S, Mohseni M, Wang ET, Ryvkin P, et al. Microarray-based cell-free DNA analysis improves noninvasive prenatal testing. Fetal Diagn Ther. 2014;36(4):282-6.	Reference standard not fetal karyotyping or postnatal phenotype
28. Kagan KO, Wright D, Nicolaides KH. First-trimester contingent screening for trisomies 21, 18 and 13 by fetal nuchal translucency and ductus venosus flow and maternal blood cell-free DNA testing. Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2014.	Modelled data
29. Lambert-Messerlian G, Kloza EM, Williams IJ, Loucky J, O'Brien B, Wilkins-Haug L, et al. Maternal plasma DNA testing for aneuploidy in pregnancies achieved by assisted reproductive technologies. Genetics in Medicine.	No additional diagnostic accuracy data to Palomaki 2011[62]

2014;16(5):419-22.	
30. Larion S, Warsof SL, Romary L, Mlynarczyk M, Peleg D, Abuhamad AZ. Uptake of noninvasive prenatal testing at a large academic referral center. <i>American Journal of Obstetrics & Gynecology</i> . 2014;211(6):651.e1-7.	No diagnostic accuracy data
31. Lee da, E., et al., Non-invasive prenatal testing of trisomy 18 by an epigenetic marker in first trimester maternal plasma. <i>PLoS ONE [Electronic Resource]</i> , 2013. 8(11): p. e78136.	Nested case-control study: < 15 cases
32. Levy B, Norwitz E. Non-invasive prenatal aneuploidy testing: technologies and clinical implication. <i>MLO Med Lab Obs</i> 2013;45(6):8, 10, 12 passim; quiz 16.	Review
33. Liao C, Yin AH, Peng CF, Fu F, Yang JX, Li R, et al. Noninvasive prenatal diagnosis of common aneuploidies by semiconductor sequencing. <i>Proc Natl Acad Sci U S A</i> . 2014;111(20):7415-20.	Incomplete 2x2 table; used cross-validation method to evaluate sensitivity and specificity so no 2 x 2 table
34. Lim, J.H., et al., Disease specific characteristics of fetal epigenetic markers for non-invasive prenatal testing of trisomy 21. <i>BMC Medical Genomics [Electronic Resource]</i> , 2014. 7: p. 1.	Method development study
35. Lim, J.H., et al., Non-invasive detection of fetal trisomy 21 using fetal epigenetic biomarkers with a high CpG density. <i>Clinical Chemistry & Laboratory Medicine</i> , 2014. 52(5): p. 641-7.	Nested case-control study: < 15 cases
36. Lim, J.H., et al., Non-invasive epigenetic detection of fetal trisomy 21 in first trimester maternal plasma. <i>PLoS ONE [Electronic Resource]</i> , 2011. 6(11): p. e27709.	Epigenetic approach
37. Lo KK, Boustred C, Chitty LS, Plagnol V. RAPIDR: an analysis package for non-invasive prenatal testing of aneuploidy. <i>Bioinformatics</i> . 2014;30(20):2965-7.	No information on population and reference standard
38. Louis-Jacques, A., et al., Effect of commercial cell-free fetal DNA tests for aneuploidy screening on rates of invasive testing. <i>Obstetrics & Gynecology</i> , 2014. 123 Suppl 1: p. 67S.	Abstract
39. Louis-Jacques, A., et al., Use of commercial tests for aneuploidy screening using cell-free fetal DNA in clinical practice. <i>Obstetrics & Gynecology</i> , 2014. 123 Suppl 1: p. 154S.	Conference abstract
40. Manegold-Brauer, G., et al., A new era in prenatal care: non-invasive prenatal testing in Switzerland. <i>Swiss Medical Weekly</i> , 2014. 144: p. w13915.	Cohort study: < 50 women

41. McCullough RM, Almasri EA, Guan X, Geis JA, Hicks SC, Mazloom AR, et al. Non-invasive prenatal chromosomal aneuploidy testing--clinical experience: 100,000 clinical samples. PLoS ONE. 2014;9(10):e109173.	Incomplete 2x2 table; no reasonable estimate for FN or FP in 2x2 table. Reliant on clinicians reporting results back to the company on an ad-hoc basis
42. Nicolaides, K.H., et al., First-trimester contingent screening for trisomies 21, 18 and 13 by biomarkers and maternal blood cell-free DNA testing. Fetal Diagnosis & Therapy, 2014. 35(3): p. 185-92.	No diagnostic accuracy data
43. Nicolaides, K.H., et al., Prenatal detection of fetal triploidy from cell-free DNA testing in maternal blood. Fetal Diagnosis & Therapy, 2014. 35(3): p. 212-7.	NIPT for triploidy
44. Norton ME, Jelliffe-Pawlowski LL, Currier RJ. Chromosome abnormalities detected by current prenatal screening and noninvasive prenatal testing. Obstetrics & Gynecology. 2014;124(5):979-86.	No diagnostic accuracy data
45. O'Brien BM, Kloza EM, Halliday JV, Lambert-Messerlian GM, Palomaki GE. Maternal plasma DNA testing: experience of women counseled at a prenatal diagnosis center. Genetic Testing & Molecular Biomarkers. 2014;18(10):665-9.	No diagnostic accuracy data
46. 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(11):913-20.	Uses the same samples as Palomaki et al. (2012); excluded to prevent double counting
47. Papageorgiou, E.A., et al., Fetal-specific DNA methylation ratio permits noninvasive prenatal diagnosis of trisomy 21. Nature Medicine, 2011. 17(4): p. 510-3.	Case-control study: < 15 cases
48. Pettit KE, Hull AD, Korty L, Jones MC, Pretorius DH. The utilization of circulating cell-free fetal DNA testing and decrease in invasive diagnostic procedures: an institutional experience. Journal of Perinatology. 2014;34(10):750-3.	No diagnostic accuracy data
49. Platt LD, Janicki MB, Prosen T, Goldberg JD, Adashek J, Figueroa R, et al. Impact of noninvasive prenatal testing in regionally dispersed medical centers in the United States. American Journal of Obstetrics & Gynecology. 2014;211(4):368.e1-7.	No diagnostic accuracy data
50. Rabinowitz, M., et al., Noninvasive aneuploidy detection by multiplexed amplification and sequencing of polymorphic Loci. Obstetrics & Gynecology, 2014. 123 Suppl 1: p. 167S.	Conference abstract

51. Shaw, S.W., C.P. Chen, and P.J. Cheng, From Down syndrome screening to noninvasive prenatal testing: 20 years' experience in Taiwan. <i>Taiwanese Journal of Obstetrics & Gynecology</i> , 2013. 52(4): p. 470-4.	Review
52. Shea JL, Diamandis EP, Hoffman B, Lo YM, Canick J, van den Boom D. A new era in prenatal diagnosis: the use of cell-free fetal DNA in maternal circulation for detection of chromosomal aneuploidies. <i>Clin Chem</i> 2013;59(8):1151-9.	Interview
53. Shi X, Zhang Z, Cram DS, Liu C. Feasibility of noninvasive prenatal testing for common fetal aneuploidies in an early gestational window. <i>Clinica Chimica Acta</i> . 2015;439:24-8.	Cohort study: < 50 women with index and reference test result
54. Skinner, J., et al., Analysis of fetal DNA in the maternal venous blood for abnormalities of chromosomes 13, 16, 18 and 21 in first-trimester spontaneous miscarriage. <i>Journal of Obstetrics & Gynaecology</i> , 2003. 23(3): p. 228-32.	Maternal plasma samples after first trimester spontaneous miscarriage vs. genetic analysis of evacuated products of the uterus
55. Sparks AB, Wang ET, Struble CA, Barrett W, Stokowski R, McBride C, et al. Selective analysis of cell-free DNA in maternal blood for evaluation of fetal trisomy. <i>Prenat Diagn</i> . 2012;32(1):3-9.	Incomplete 2x2 table; no reference standard for cffDNA testing negative cases
56. Struble CA, Syngelaki A, Oliphant A, Song K, Nicolaides KH. Fetal fraction estimate in twin pregnancies using directed cell-free DNA analysis. <i>Fetal Diagnosis & Therapy</i> . 2014;35(3):199-203.	No diagnostic accuracy data
57. Stumm, M., et al., Noninvasive prenatal detection of chromosomal aneuploidies using different next generation sequencing strategies and algorithms. <i>Prenatal Diagnosis</i> , 2012. 32(6): p. 569-77.	Method development
58. Tong, Y.K., et al., Noninvasive prenatal detection of fetal trisomy 18 by epigenetic allelic ratio analysis in maternal plasma: Theoretical and empirical considerations. <i>Clinical Chemistry</i> , 2006. 52(12): p. 2194-202.	Case series: < 50 women
59. Tong, Y.K., et al., Noninvasive prenatal detection of trisomy 21 by an epigenetic-genetic chromosome-dosage approach. <i>Clinical Chemistry</i> , 2010. 56(1): p. 90-8.	Case-control study: < 15 cases
60. Tsaliki, E., et al., MeDIP real-time qPCR of maternal peripheral blood reliably identifies trisomy 21. <i>Prenatal Diagnosis</i> , 2012. 32(10): p. 996-1001.	Epigenetic approach
61. van den Oever, J.M., et al., Single molecule	Case control: < 15 cases

sequencing of free DNA from maternal plasma for noninvasive trisomy 21 detection. Clinical Chemistry, 2012. 58(4): p. 699-706.	
62. van den Oever, J.M., et al., Successful noninvasive trisomy 18 detection using single molecule sequencing. Clinical Chemistry, 2013. 59(4): p. 705-9.	Case control: < 15 cases
63. Wang JC, Sahoo T, Schonberg S, Kopita KA, Ross L, Patek K, et al. Discordant noninvasive prenatal testing and cytogenetic results: a study of 109 consecutive cases. Genet Med. 2014.	Incomplete 2x2 table; index test results used as inclusion criteria so incomplete 2x2 table
64. Willems PJ, Dierickx H, Vandenakker E, Bekedam D, Segers N, Debouille K, et al. The first 3,000 Non-Invasive Prenatal Tests (NIPT) with the Harmony test in Belgium and the Netherlands. Facts views vis. 2014;6(1):7-12.	Incomplete 2x2 table; incomplete follow up of cffDNA testing negative cases
65. Wu, D., et al., Prenatal diagnosis of Down syndrome using cell-free fetal DNA in amniotic fluid by quantitative fluorescent polymerase chain reaction. Chinese Medical Journal, 2014. 127(10): p. 1897-901.	Not cff DNA (amniotic fluid)
66. Yu SC, Chan KC, Zheng YW, Jiang P, Liao GJ, Sun H, et al. Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. Proc Natl Acad Sci U S A. 2014;111(23):8583-8.	Re-uses the same samples as Chen et al. (2011) and Chiu et al. (2011); excluded to prevent double counting
67. Zhang, M., et al., Non-invasive prenatal diagnosis of trisomy 21 by dosage ratio of fetal chromosome-specific epigenetic markers in maternal plasma. Journal of Huazhong University of Science and Technology. Medical Sciences, 2011. 31(5): p. 687-92.	Epigenetic approach

Supplement 5 Table of study characteristics of included studies

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
Alberti 2015[56] France Study start date: March 2010	Prospective case-control (cases with abnormal karyotype matched with a balanced number of randomly selected pregnancies with euploid karyotypes) Number of centres: 3	N=976 enrolled in cohort. Women with singleton pregnancies, high-risk of foetal T21. N=225 in case-control for sequencing. Mean age (SD): 35.2 (6.7) years. Mean gestational age (SD): 14 (2) weeks. 1 st and 2 nd trimester.	N=0 from cohort. N=751 (76.9%): Not included in case-control study.	T21	All high risk for foetal T21 (>1:250) based on the combination of maternal age with ultrasound and maternal serum markers during the first or second trimester.	MPS (whole genome) performed in a cytogenetics laboratory in a university teaching hospital	CVS or amniocentesis and foetal karyotype	None	NIPT performance for T21 detection.	Accuracy of NIPT
Ashoor 2012[46] UK	Nested case-control of stored maternal	N=400 (50 T21, 50 T18, 300 euploid)	Pregnant by IVF or multiple pregnancy	T21, T18	All high risk: Combined 1st trimester screen	DANSR, FORTE	Karyotyping after CVS	None	FORTE risk score for aneuploidies, sensitivity and specificity for	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
Study start date: NR	<p>samples:</p> <p>Controls matched with T21/T18 cases for sample storage time in 3:1 ratio.</p> <p>Number of centres: 1</p>	<p>Singleton pregnancies, high-risk women.</p> <p>1st trimester 100%;</p> <p>All 11-13 weeks' gestation.</p> <p>Ethnicity:</p> <p>White 89%,</p> <p>'Afro Caribbean' 5%,</p> <p>South/ East Asian 6%,</p> <p>Mixed 0.5%.</p>	N=NR		risk >1:300	Aria Diagnostics (USA)			detection of T21 and T18	
<p>Beamon 2014[36]</p> <p>USA</p> <p>Study start</p>	<p>Prospective cohort</p> <p>Number of centres: 1</p>	<p>N=208</p> <p>High-risk pregnancies who chose NIPT as triage test, singleton or dichorionic twin gestations, ≥10</p>	<p>Multiple pregnancy</p> <p>N=NR</p>	T21, T18, T13	<p>All high-risk:</p> <p>AMA: 148 (71.2%),</p> <p>AMA alone: 121 (58.2%),</p>	<p>MPS (whole genome)</p> <p>Sequenom Center for</p>	<p>Karyotyping after amniocentesis, cordocentesis or</p>	None	Test performance for T13, T18 and T21 detection.	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
date: January 2012		weeks' gestation. Mean age (SD), range: 36 (5.5), 19-47 years. Mean gestational age (SD), range: 15.6 (4.3), 10-34 weeks. Trimester: 1 st : 111 (53.4%), 2 nd : 95 (45.7%), 3 rd : 2 (1%).			AMA + other: 27 (13.0%), Ultrasound abnormality: 26 (12.5%), Abnormal serum screen: 29 (13.9%), Combined FTS: 16 (7.7%), Quadruple: 12 (5.8%), Integrated: 1 (0.5%), Affected family member: 3 (1.4%), Other: 2 (1.0%), Twins (growth discordance): 1 (0.5%), Maternal anxiety:	Molecular Medicine (USA) (n=163, 78.4%) or Verinata Health (USA) (n=45, 21.6%).	CVS, phenotype of newborn			

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
					1 (0.5%).					
Bevilacqua 2015[37] Belgium, UK, Spain Study start date: May 2013	Prospective multicentre cohort Number of centres: NR	N=515 included. Twin pregnancies at mixed risk for aneuploidies. Median gestational age (range): 13.0 (10.0-28.0) weeks. 1 st trimester: 68.5%.	Criteria for exclusion from study NR	T21, T18, T13	Mixed risk: High risk for foetal trisomy by 1 st -trimester combined test or 2 nd -trimester triple/quadruple test or ultrasound or NIPT as primary method of screening.	DANSR, FORTE Harmony Prenatal test Ariosa Diagnostics (USA)	Karyotyping after amniocentesis, cordocentesis or CVS, or newborn phenotypic examination	None	1) Factors influencing failure rate in twin and singleton pregnancies. 2) NIPT performance for T13, T18 and T21 detection in twins.	Accuracy of NIPT
Bianchi 2012[47] USA Study start date: NR	Nested case-control Controls unmatched in 4:1 ratio (Part of MELISSA)	N=2,882 in cohort. N=534 in nested case-control study. Singleton pregnancies, high risk.	257/2,882 (8.9%) from MELISSA cohort: 85 multiple pregnancies, 45 no karyotype information, 127 ineligible	T21, T18, T13	All high risk: AMA (>38 years) only 152 (28.5%); Positive screen risk 91 (17.0%); Ultrasound abnormality	MPS (whole genome) Verinata-Illumina (USA)	Karyotyping after CVS	None	1) MPS performance (sensitivity and specificity) for T21, T18 and T13 detection.	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
	prospective cohort. Number of centres: 53 (of 60)	Mean age (SD), range: 35.2 (6.40), 18 – 46 years. Mean gestational age (SD), range: 15.1 (3.16), 10 – 23 weeks. Trimester: 1 st : 165 (30.9%), 2 nd : 369 (69.1%). Ethnicity: White 72.7%, African American 10.9%, Asian 9.9%, Native American or	blood sample.		122 (22.8%); Prior aneuploidy pregnancy 15 (2.8%); More than 1 risk 154 (28.9%).				2) Sex chromosome classification and Monosomy X detection.	

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		Alaska Native 0.9%, Multiracial 5.6%.								
Bianchi 2014[19] USA Study start date: July 2012	Prospective cohort Number of centres: 21	N=2,052 enrolled. N=2,042 eligible. Singleton pregnancies, general obstetric population. Trimester: 1 st : 759 (39.7%), 2 nd : 610 (31.9%), 3 rd : 545 (28.5%). Mean gestational age (SD), range: 20.3 (8.6), 8.0 – 39.4 weeks.	N=10 (0.5%): 7 insufficient blood volume, 1 late receipt of blood sample, 1 maternal age <18 years, 1 withdrawn consent.	T21, T18, T13	General obstetric population undergoing standard prenatal aneuploidy screening	MPS (whole genome) Verifi Verinata-Illumina (USA)	Newborn phenotype (97.0%) or Karyotyping (3.0%).	Standard prenatal aneuploidy screening produced by accredited clinical laboratories. Cutoff values as used by individual laboratories 1 st -trimester: Combined test (PAPP-A, β -hCG, NT) N=739 (38.6%).	1) Comparison of false positive rates of NIPT with conventional screening for T21 and T18. 2) Comparison of false positive rates for T13. Comparison of foetal fractions in low-risk with high-risk patients.	Comparison of NIPT with CT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		<p>Mean age (SD), range:</p> <p>29.6 (5.54),</p> <p>18.0 – 48.6 years.</p> <p>Assisted conception 66 (3.4%).</p>						<p>2nd-trimester:</p> <p>Quadruple</p> <p>(MS-AFP, β-hCG, estriol and inhibin A)</p> <p>N=439 (22.9%);</p> <p>Quadruple + combined test</p> <p>N= 53 (2.8%);</p> <p>Quadruple + 1st-trimester serum markers only</p> <p>N=164 (8.6%);</p> <p>Sequential:</p> <p>1st-trimester screen results reported before final report in 2nd trimester</p> <p>N=519 (27.1%).</p>		
Chen 2011[48] Hong Kong,	Case-control of stored samples and	N=392 (N=140 archived plasma samples with	NR	T18, T13	All high risk based on clinical indicators as per the existing	MPS (whole genome)	Karyotyping after CVS or amniocent	None	Diagnostic performance of MPS for T13 and T18	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
UK, Netherlands, China Study start date: NR	prospectively recruited women Number of centres: 10	and without aneuploidy matched for gestational age; N=252 prospectively recruited.) 344/392 samples analysed in a previous study [49], 48 cases newly recruited. Singleton pregnancy undergoing CVS/amniocentesis.			obstetric practice of each recruitment unit.	Sequenom (USA)	esis		detection.	
Chiu 2011[49] Hong Kong, UK, Netherlands, China Study start date: October	Case-control of stored samples and prospectively recruited women Number of	N=824 screened (N=248 archived T21 and non-T21 samples matched for gestational ages in 1:5 ratio and N=576 prospectively collected high-risk	N=60 (7.3%): 14 failed recruitment criteria (2 twin pregnancies, 12 without full	T21	High risk by conventional screening (>1:300): 582 (77%), Median risk for T21: 1 in 43.	MPS (whole genome) Sequenom (USA)	Full karyotyping after amniocentesis (18%) or CVS (82%).	None	Diagnostic sensitivity, specificity, PPV & NPV for T21 detection.	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
2008	centres: 10	<p>samples),</p> <p>N=764 included.</p> <p>Singleton pregnancies.</p> <p>Median age: 35.4 years.</p> <p>Median gestational age: 13+1 weeks.</p> <p>1st trimester: 74%.</p>	<p>karyotyping);</p> <p>46 compromised blood sample</p> <p>(3 samples collected after invasive obstetric procedure,</p> <p>2 delayed blood processing,</p> <p>3 with ambiguous information,</p> <p>12 haemolysed, 26 inadequate volume).</p>		<p>Intermediate risk by conventional screening (1:300-1:1000) 39 (5%),</p> <p>Median risk for T21: 1 in 502.</p> <p>Other indications (previous T21 pregnancy, ultrasound abnormalities, risk for monogenic diseases).</p>					
<p>Comas 2014[38]</p> <p>Spain</p> <p>Study start date: January</p>	<p>Prospective cohort</p> <p>Number of centres: 1</p>	<p>N=333</p> <p>Singleton pregnancies who chose to have NIPT.</p> <p>Mean maternal age</p>	<p>Multiple pregnancies, ultrasound anomalies or high risk of congenital malformation</p> <p>N=NR</p>	T21, T18, T13	<p>Routine general population in a real clinical setting.</p> <p>83.5% Low-risk by conventional</p>	<p>DANSR FORTE</p> <p>(Harmony Prenatal Test),</p> <p>Ariosa Diagnostics</p>	<p>Invasive testing and karyotyping or newborn phenotype</p>	None	<p>1) NIPT test performance for T13, T18, and T21.</p> <p>2) Comparison</p>	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
2013		(range): 37 (21-46) years. Mean gestational age (range): 14.6 (9.5-23.5) weeks. 1 st and 2 nd trimester.			screenings but unable to alleviate their anxiety. 16.5% High-risk from CT or referred for AMA with no prior screening.	(USA) (n=120, 36.0%) or SNP- and NATUS (Panorama) Natera Inc. (USA) (n=213, 64.0%)			of Harmony and Panorama tests, factors influencing foetal fraction.	
Dan 2012[63] China Study start date: 1 st quarter 2010	Prospective multicentre cohort Number of centres: 49	N=11,263 recruited. N=11,184 included. Singleton pregnancies, ≥ 18 years, gestational age of 9 - 28 weeks. Median age (range): 31 (18-49) years.	N=79 (0.7%): 55 unqualified gestational age, 14 multiple pregnancies, 10 foetal death.	T21, T18	Mixed risk factors Conventional T21 screening test: yes - positive: 4,522 (40.7%) yes - negative: 2,426 (21.8%) No – with 1 or	MPS (whole genome) BGI-Shenzen (China)	Full karyotyping 3,000 (26.6%) or birth questionnaire 4,524 (40.2%).	None	1) Sensitivity and specificity of MPS for T21 and T18 screening. 2) Workflow of MPS-based test.	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		<p>Median gestational age (range):</p> <p>20 (9-28) weeks.</p> <p>2nd trimester: >74%.</p> <p>42/49 centres offered test to high-risk pregnant women identified by a conventional T21 screening test,</p> <p>7/49 centres enrolled participants regardless of prior risk assessment.</p>			<p>more other risk factors (≥ 35 years, family history of aneuploidies, ultrasound abnormalities):</p> <p>2,770 (24.9%)</p> <p>No – without any risk factors:</p> <p>1,387 (12.5%).</p>					
<p>Del Mar Gil 2014[21]</p> <p>UK</p> <p>Study start</p>	<p>Retrospective cohort of stored samples</p> <p>Number of</p>	<p>N=207</p> <p>Twin pregnancies undergoing first-trimester screening for trisomies by combined test.</p>	<p>Singleton pregnancies</p> <p>N=NR</p>	T21, T18, T13	NR	<p>DANSR FORTE</p> <p>Harmony</p>	Known birth outcome	None	Performance of Harmony Test in twin pregnancies only	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
date: NR	centres: 1	<p>Age range: 26 – 41 years.</p> <p>Gestational age, range: 11 - 13 weeks.</p> <p>1st trimester: 100%.</p> <p>Ethnicity:</p> <p>Caucasian 70.0%,</p> <p>Afro-Caribbean 23.7%,</p> <p>South/East Asian 1.0%,</p> <p>Mixed 5.3%.</p>				Ariosa Diagnostics (USA)				
Dhallan 2007[57] USA	Prospective observational study	<p>N=60</p> <p>Women ≥ 18 years, singleton pregnancy.</p>	N=NR	T21	Mostly high risk. Definition unspecified.	SNP allelic ratio	Amniocentesis or newborn reports	None	Performance of SNP method in detecting T21	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
Study start date: January 2004	Number of centres: 10	Mean age (range): 32.8 (18-43* years, Mean gestational age (range): 19+6 (8+1 - 38+6) weeks, 1 st trimester: 8 (13%).				Ravgen Inc. (USA)				
Ehrich 2011[50] USA Study start date: May 2009	Prospective case-control (T21 matched 1:11 with euploid samples) Number of centres: NR	N=480 requested from independent 3 rd -party database. Pregnancies at increased risk for foetal aneuploidies with scheduled invasive diagnostic procedure (unclear if singleton or also multiple pregnancies). Median age (range): 37 (18 -47) years.	N=13 (2.7%): 9 sample volume <3.5 ml, 1 dropped, 2 mixed together, 1 tube broke during centrifugation.	T21	High risk: Positive serum screening 30.2%, AMA ≥ 35 years 68.3%, Ultrasound abnormality 12.9%, Positive family history 5.2%, Not specified 10.2%.	MPS (whole genome) Sequenom (USA)	Amniocentesis (81%) or CVS (19%) and karyotype (60%), FISH (3%), both (36%) or QF-PCR (1.6%)	None	Test performance for T21	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		Median gestational age (range): 16 (8-36) weeks.								
Hall 2014[51] USA Study start date: March 2012	Nested case-control (selected from a cohort of >1000 women, all T13 cases matched 1:3 on gestational age) Number of centres: NR	N=68 (17 T13, 51 euploid) High-risk pregnancy couples, women \geq 18 years, singleton pregnancy. Median gestational age (range): 16.0 (12.1-22.7) weeks, 1 st trimester: 23 (35.9%).	N=1/>1,000 (<0.1%) from cohort: 1 known foetal mosaicism.	T13	High-risk for foetal aneuploidy (positive serum screen, ultrasound abnormality or maternal age of greater than 35 years)	SNP- and NATUS Natera Inc. (USA)	CVS, amniocentesis or genetic testing of cord blood, buccal, saliva, or products of conception	None	1) Test performance for T13 detection. 2) Specificity of T18, T21 and Monosomy X detection.	Accuracy of NIPT
Huang 2014[22] China (Denmark,	Prospective, multicentre cohort	N=189 Twin pregnancies requiring invasive procedure (CVS/	N=NR Intrauterine death, without	T21, T18	All high risk Threshold and	MPS (whole genome)	Full karyotyping from CVS	None	Test performance for T18 and T21 detection in twin	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
Hong Kong) Study start date: NR	Number of centres: 7	amniocentesis) Median age (range): 31 (22-44) years. Median gestational age (range): 19 (11-36) weeks. 1 st trimester: $\geq 2.1\%$, 2 nd trimester: $\geq 74\%$	foetal karyotype		risk establishment NR	NIFTY test BGI-Shenzen (China)	(2.1%), amniocentesis (94.2%), or cordocentesis (3.7%)		pregnancies	
Jeon 2014[39] South Korea, China Study start date: March 2012	Prospective cohort Number of centres: 1	N=155 High-risk women scheduled for amniocentesis, ≥ 19 years old, singleton pregnancy with a gestational age of ≥ 12 weeks. Mean age (SD),	NR	T21, T18	High risk of foetal defects by standard aneuploidy screening with individual risk scores and interpretations produced by accredited clinical laboratories.	MPS (whole genome) Semiconductor sequencing	Amniocentesis and foetal karyotyping	None	T18 and T21 detection by semiconductor sequencer Ion Proton (PPV, NPV).	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		range: 30.73 (4.99), 19-43 years. Trimester: 1 st : <18.1%, 2 nd : >55.5%.								
Jiang 2012[23] China Study start date: June 2009	Prospective cohort Number of centres: 3	N=903 Inclusion criteria NR Age range: 20-45 years. Gestational age: 10-34 weeks (all trimesters).	Criteria NR No exclusions recorded	T21, T18 T13	Prevalence of aneuploidy suggests a general obstetric population but all women had invasive testing.	MPS (whole genome) BGI-Shenzhen (China)	Full karyotyping from amniocentesis	None	1) Aneuploidy detection. 2) GC content and sequencing bias. Relation between foetal fraction and gestational age.	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
Korostolev 2014[40] Russia Study start date: 2012	Prospective cohort Number of centres: NR (Moscow private clinics)	N=1,968 included, N=1,728 for NIPT. Women with singleton pregnancies, high risk for aneuploidies, >9 weeks' gestation. Mean age (range): 34.4 (26-45) years. Mean gestational age (range): 14 (9-33) weeks. 1 st trimester: "about 50%".	N=240 (12.2%): Ultrasound abnormality (increased NT, heart defects, malformations, foetal growth retardation) or presence of balanced chromosomal rearrangements in the parents.	T21, T18, T13	Mixed risk: High risk result of combined FTS 87%, AMA \geq 35 years only or women's will without any risk of chromosomal pathology 13%.	SNP and NATUS Panorama Natera Inc. (USA)	Invasive prenatal diagnosis with karyotyping or CMA (n=57), phenotypic newborn assessment (n=624), TOP and molecular study (n=1).	None	NIPT and/or invasive test based on CMA for chromosomal abnormalities diagnostics	Accuracy of NIPT
Lau 2012[24] Hong Kong, China, Japan	Prospective cohort	N=108 Pregnant women undergoing CVS or amniocentesis	NR	T21, T18, T13	Mostly high risk: Positive 1 st trimester screening 47.2%,	MPS (whole genome)	Conventional karyotyping from	None	Diagnostic accuracy of novel z-score method with internal	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
Study start date: NR	Number of centres: 1	(possibly singleton pregnancies but NR). Mean age (SD): 37 (4.3) years, Median gestational age (range): 12+5 (11+4 – 28+0) weeks. 1 st trimester: 97 (89.8%)			positive 1 st trimester sonographic markers 22.2%, other structural anomalies 1.5%, previous T21 0.9%, maternal anxiety 11.1%.	BGI-Shenzhen (China)	CVS (94.4%) or amniocentesis (5.6%)		reference chromosome.	
Lau 2014[25] Hong Kong, USA, China Study start date: August 2011	Prospective cohort Number of centres: 1	N=1,982 (1,929 singleton, 30 twin pregnancies, 23 internal control samples) Any pregnant women ≥12 weeks of gestation accepted for NIPT, regardless of whether they had undergone any	NR	T21, T18 T13	Prenatal diagnosis centre accepted referral of any pregnant woman for NIPT: Previous trisomy / Family history 53 (2.7%).	MPS (whole genome) NIFTY test BGI-Health (China)	Conventional karyotyping from CVS or amniocentesis, postnatal karyotyping or birth phenotype	None	Test accuracy for common autosomal trisomies, sex chromosomal abnormalities and other chromosome abnormalities.	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		<p>previous T21 screening.</p> <p>Mean age (SD), range: 36 (4.35), 20-46 years.</p> <p>Median gestational age: 14.5 weeks.</p> <p>1st trimester: 56.25%.</p> <p>Ethnicity:</p> <p>Chinese 90.91%,</p> <p>Caucasian 5.21%,</p> <p>Other 3.88%.</p>			<p>No prior screening test: 669 (34.2%).</p> <p>Prior screening test 1,290 (65.8%):</p> <p>High risk 593/1,290 (46.0%),</p> <p>Low risk 368/1,290 (28.5%),</p> <p>Result not available yet 329/1,290 (25.5%).</p>					

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
Liang 2013[26] China Study start date: March 2009	Prospective cohort Number of centres: 3	N=435 High-risk pregnant women scheduled for invasive prenatal diagnostics. Mean age (SD): 31 (5.9) years. Median gestational age (range): 21+3 (11+3 – 39+3) weeks. 1 st trimester: 1 (0.23%).	NR	T21, T18 T13	All high risk: AMA (≥ 35 years) 84 (19.3%), Positive serum screening 217 (49.9%), Ultrasound abnormality 67 (15.4%), Prior aneuploidy pregnancy 4 (0.9%), Multiple indications 63 (14.5%).	MPS (whole genome) Berry Genomics (China)	CVS (0.92%), cordocentesis (22.30%) or amniocentesis (76.78%) and full foetal karyotyping	None	Test accuracy for detection of foetal aneuploidies for all 24 chromosomes in one single sequencing event	Accuracy of NIPT
Nicolaides 2012[27] UK	Retrospective cohort of stored samples	N=2,230 original cohort, N=2,049 eligible	N=181 (8.1%): 74 no foetal karyotype,	T21, T18	General obstetric population undergoing first-trimester screening for	DANSR FORTE	86 (4.2%) CVS or amniocentesis and foetal	First-trimester CT (free β -hCG, PAPP-A, NT) with or without additional	1) Performance of screening by NIPT for	Comparison of NIPT with CT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
Study start date: October 2010	Number of centres: 1	<p>cases.</p> <p>Women with singleton pregnancies attending for first-trimester combined screening for aneuploidies and ultrasound (general obstetric population).</p> <p>Median age (IQR): 31.8 (27.7 – 35.4) years,</p> <p>Gestational age, range: 11+0 – 13+6 weeks,</p> <p>1st trimester: 100%.</p> <p>Ethnicity:</p> <p>Caucasian 69.8%,</p> <p>African 20.6%,</p>	<p>7 abnormal karyotype other than T21 or T18,</p> <p>29 inadequate sample volume,</p> <p>1 wrongly labelled</p> <p>70 lab mixed samples together.</p>		<p>aneuploidies as part of their routine antenatal care.</p> <p>All had 1st-trimester combined test:</p> <p>Median estimated T21 risk (range)</p> <p>1:8,469</p> <p>(1:2–1:23,527),</p> <p>Median estimated T18 risk (range)</p> <p>1:14,894</p> <p>(1:2-1:47,472).</p>	<p>Harmony Prenatal Test</p> <p>Ariosa Diagnostics (USA)</p>	<p>karyotyping.</p> <p>1963 (95.8%) phenotypic newborn examination.</p>	<p>ultrasound markers (nasal bone, tricuspid regurgitation, reversed a-wave in ductus venosus).</p> <p>Risk threshold $\geq 1:150$ (0.67%) for T21 and T18.</p>	<p>trisomies 21 and 18.</p> <p>2)</p> <p>Comparison of NIPT with detection rate and false positive rate of 1st-trimester CT with or without additional ultrasound markers.</p>	

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		South Asian 4.0%, East Asian 2.8%, Mixed 2.8%.								
Nicolaides 2013[28] UK Study start date: NR	Prospective cohort Number of centres: 1	N=242 Women with singleton pregnancies undergoing CVS at 11-13 weeks' gestation, ≥ 18 years, ≥ 10 weeks gestation. Mean age (range): 35.7 (18.5- 46.5) years. Median gestational age (range): 13.1 (11.3 – 13.9) weeks. 1 st trimester: 100%.	NR	T21, T18, T13	High risk for aneuploidies or sickle cell disease: 1 st -trimester CT $>1:300$ 227 (93.8%), AMA 5 (2.1%), Previous aneuploidy pregnancy 6 (2.5%), Sickle cell testing 4 (1.7%). Median estimated risk for T21, T18	SNP- and NATUS Natera Inc. (USA)	CVS and karyotyping	None	Performance of NIPT to detect T21, T18, T13, SCA and triploidy.	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
					or T13 by CT (range): 1:75 (1:2–1:12,433).					
Norton 2012[29] USA, Netherlands, Sweden Study start date: August 2010	Prospective, multicentre cohort study (NICE study) Number of centres: 48	N=4,002 enrolled, N=3,228 eligible: Women ≥ 18 years, gestational age ≥ 10 weeks, with singleton pregnancy, scheduled for invasive testing for any indication. Mean age (SD), range: 34.3 (6.4), 18-50 years. Mean gestational age (SD), range: 16.9 (4.1), 10-38.7 weeks.	Exclusion criteria: Multiple pregnancies, known maternal aneuploidy, active malignancy or history of metastatic cancer, already undergone CVS or amniocentesis. N=774 (19.3%): 433 samples used for assay development. 237 failed I/E	T21, T18	Undergoing invasive testing for any indication (primarily high risk women)	DANSR, FORTE Harmony Prenatal Test Ariosa Diagnostics (USA)	Karyotyping, FISH or QF-PCR from amniocentesis (74.7%) or CVS (25.3%)	None	1) Harmony Test performance for T21 and T18 at 1% risk cutoff. 2) Foetal fraction. Test performance at different risk cutoff values.	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		<p>Ethnicity:</p> <p>Caucasian 49.6%,</p> <p>African American 6.4%,</p> <p>Asian 13.4% ,</p> <p>Hispanic 22.7%,</p> <p>Other 7.9%.</p>	<p>criteria,</p> <p>84 insufficient sample volume,</p> <p>20 incorrect sample labelling.</p>							
<p>Norton 2015[6] USA, Sweden</p> <p>Study start date: March 2012</p>	<p>Prospective multicentre cohort (NEXT study)</p> <p>Number of centres: 35</p>	<p>N=18,955 enrolled.</p> <p>N=18,510 met I/E criteria.</p> <p>Women with singleton pregnancies, ≥ 18 years of age, presenting for aneuploidy screening at 10-14 weeks of gestation (NIPT and 1st-trimester CT).</p> <p>Mean age (range): 31</p>	<p>N=450 (2.4%):</p> <p>229 did not meet inclusion criteria or met exclusion criteria,</p> <p>31 had twins discovered on NT testing,</p> <p>121 had unknown ovum-donor status,</p> <p>64 withdrew or were withdrawn</p>	T21, T18, T13	General obstetric population (unselected)	<p>DANSR, FORTE</p> <p>Harmony Prenatal Test</p> <p>Ariosa Diagnostics (USA)</p>	<p>Invasive prenatal testing (135 CVS, 422 amniocentesis), 52 postnatal genetic testing,</p> <p>16 testing on products of conception, all other examinations</p>	<p>First-trimester CT (cut-off ≥1:270 for T21, ≥1:150 for T18 and T13)</p>	<p>1) Area under ROC curve for T21 screening with NIPT versus standard screening.</p> <p>2) Evaluation of NIPT and standard screening to assess the risk for T18 and T13.</p>	Comparison of NIPT with CT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		(18-48) years. Mean gestational age (range): 12.5 (10.0-14.3) weeks. 100% 1 st trimester.	by investigator.				on of the newborn.		Performance of NIPT in low-risk patients.	
Palomaki 2012[52] USA Study start date: Trial submission 6th April 2009	Nested case-control in a cohort (Part of an international clinical validation study, NCT00877292). Each pregnancy with T18 and T13 matched	N=4,664 in cohort, N=293 case-control study (62 T18, 12 T13, 219 euploid) plus 212 T21 and 1,483 matched controls reported earlier [62]. N=1,988 for NIPT. Singleton pregnancies at high risk for T21.	N=279/4,664 (6.0%) from cohort: 116 sample not adequate, 112 multiple gestation / foetal death, 51 no karyotype /outcome available. N=2,397/4,385 (54.7%):	T21, T18, T13	High risk for T21: 1 st -trimester screening positive: 7.2%, 2 nd -trimester screening positive: 4.4%, Integrated test positive: 10.2%, Ultrasound anomaly: 19.5%, AMA ≥ 38 years: 41.6%, 2 or more	MPS (whole genome) Sequenom Inc. (USA)	Amniocentesis (48.5%) or CVS (51.5%) and karyotyping	None	Correct identification of T21, T18 & T13	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
	with 3 controls based on the gestational age, enrolment site, race, and time in freezer (within 1 month). Number of centres: 27	Mean age (SD): 37.2 (5.0)* years. Median gestational age (range): 14.6 (9-22) weeks*. 1 st trimester: 52%, 2 nd trimester: 48%. Ethnicity: Caucasian 84.7%, Black 4%, Asian 5.4%, Unknown 5.4%.	Not selected for case-control study.		indications: 12.6%, Family history of aneuploidy: 3.4%, Other /unknown: 1.0%.					
Pergament 2014[30]	Prospective international multicentre cohort	N=1,064 enrolled, N=1,051 for testing (926 euploid, 67 T21,	N=13 (1.2%): 6 triploidy,	T21, T18, T13	543 (51.0%) High risk: abnormal serum screen, ultrasound	SNP- and NATUS	Amniocentesis/CVS (44.1%) and	None	Performance of single-nucleotide polymorphism	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
USA Study start date: NR	Number of centres: 36	32 T18, 14 T13, 12 Monosomy X). Singleton pregnancies of at least 7 weeks of gestation. Mean age (SD), range: 30.3 (7.4), 18-47 years. Mean gestational age (SD), range: 17.0 (4.1), 7.6-40.6 weeks.	3 foetal mosaic, 2 47,XXY, 1 47,XXX, 1 47,XXY.		abnormality, maternal age ≥ 35 years. 521 (49.0%) Low risk: maternal age < 35 years and lacking any reported high-risk indications.	Natera Inc. (USA)	karyotyping/FISH; genetic testing of cord blood, buccal sample or saliva (13.2%) or products of conception (42.8%).		-based test on both high- and low-risk pregnant women.	
Porreco 2014[31] USA Study start date:	Prospective multicentre cohort (NCT00847990) Number of	N=4,170 enrolled, N=3,430 for testing. Singleton pregnancies, high risk for foetal aneuploidy	N=740 (17.7%): 320 insufficient sample volume, 120 outside 6h lab processing window,	T21, T18, T13	High risk for foetal aneuploidy: Abnormal NT 104 (3%), Abnormal Triple/quad screen 289	MPS (whole genome) MaterniT21® PLUS	Amniocentesis (75.5%) or CVS (24.5%) and karyotype	None	Clinical performance of MPS to test for T21, T18, T13, foetal sex and SCA.	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
September 2009	centres: 31	undergoing invasive procedure. Mean age (SD), range: 35.1 (5.6), 18-50 years. Mean gestational age (SD), range: 16.3 (3.5), 9.0-37.0 weeks. Ethnicity: White 60.1%, Asian 18.7%, Hispanic or Latino 9.9%, Black 4.5%, Multiple 5.5%.	270 used as lab quality control set, 24 incomplete case report forms, 6 no amniocentesis / CVS.		(8.4%), Abnormal ultrasound 492 (14.3%), AMA \geq 35 years 1,417 (41.3%), Multiple indications 929 (27.1%), Previous or family history of aneuploidies 98 (2.9%).	Sequenom, Inc. (USA)				
Quezada	Prospective	N=2,905	N=NR	T21, T18,	No prior screening, general	DANSR,	CVS or amniocent	First-trimester CT for T21	1) Numbers and	Comparison of

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
2015[41] UK Study start date: October 2012	cohort Number of centres: 1	<p>Women with singleton pregnancies undergoing routine first-trimester screening for the major trisomies by NIPT and by the combined test.</p> <p>Mean age (range): 36.9 (20.4–51.9) years.</p> <p>Median gestational age (range): 10+4 (10+0 -11+6) weeks.</p> <p>1st trimester: 100%.</p> <p>Ethnicity:</p> <p>Caucasian 2,570 (88.5%),</p> <p>South Asian 173</p>		T13	<p>obstetric population,</p> <p>AMA \geq 35 years 1,958 (67.4%).</p>	<p>FORTE</p> <p>Harmony</p> <p>Ariosa Diagnostics (USA)</p>	<p>esis and foetal karyotyping,</p> <p>post-mortem examination and karyotyping,</p> <p>newborn phenotype</p>	<p>(PAPP-A, free β-hCG, nuchal translucency)</p> <p>Risk threshold \geq 1/100 for T21.</p>	<p>concordance of results of NIPT and 1st-trimester combined screen.</p> <p>2) Discordant results between NIPT and foetal karyotype.</p>	NIPT with CT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		(6.0%), East Asian 96 (3.3%), Afro-Caribbean 21 (0.7%), Mixed 45 (1.5%).								
Sago 2014[42] Japan Study start date: April 2013	Prospective multicentre cohort Number of centres: 15 in April 2013, 37 by March 2014	N=7,740 Women with singleton pregnancies, 10 to 18 weeks' gestation, high-risk for aneuploidy, requesting NIPT. Mean age (range): 38.3 (21-48) years. Mean gestational age (range): 13.3 (10.0-19.9) weeks.	Multiple Pregnancy N=NR	T21, T18, T13	All high-risk: Maternal age ≥ 35 years 7387 (95.4%), Prior history 226 (2.9%), Ultrasound abnormality 108 (1.4%), Serum marker 16 (0.2%), Balanced Robertsonian translocation 3 (0.04%).	MPS (whole genome) MaterniT21 PLUS Sequenom Inc. (USA)	CVS or amniocentesis and foetal karyotyping, foetal death and karyotyping or birth phenotype	None	PPV for T21, T18 and T13.	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		5.2% multi-ethnic Selected for training set: 71/435, Selected for validation set: 48/575.								
Shaw 2014[32] Taiwan, China Study start date: June 2012	Prospective cohort Number of centres: 11	N=201 Pregnant women > 12 weeks' gestation. <u>High risk</u> (n=100): Mean age (SD): 35.1 (3.2) years. Mean gestational age (SD) 17.3 (2.1) weeks. 98 singleton, 2 twin pregnancies.	N=1 (0.5%): 1 due to early gestational age (<12 weeks)	T21, T18, T13	Very high risk (T21 risk >1:30 or NT >3.0mm): N=100 Average screening risk: 1:22.8. Low risk (T21 risk <1:1,500): N=100 Average screening risk:	MPS (whole genome) Berry Genomics (China)	Amniocentesis and karyotyping or birth outcome	None	Test performance for detection of all foetal autosomal and sex chromosome aneuploidies	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		<p><u>Low risk</u> (n=100):</p> <p>Mean age (SD):</p> <p>34.6 (2.6) years.</p> <p>Mean gestational age (SD) 16.1 (3.0) weeks.</p> <p>98 singleton, 2 twin pregnancies.</p>			1:3,179.					
<p>Song 2013[33]</p> <p>China</p> <p>Study start date: April 2011</p>	<p>Prospective cohort</p> <p>Number of centres: 2</p>	<p>N=1,916</p> <p>Singleton pregnancies, women <35 years undergoing routine antenatal screening.</p> <p>Mean age (SD), range: 29.03 (2.7), 20 - 34 years.</p>	N=NR	T21, T18 T13	<p>General obstetric population < 35 years.</p> <p>High risk</p> <p>275/1,741 (15.8%):</p> <p>Positive serum screening >1:270: 249 (14.3%),</p>	<p>MPS (whole genome)</p> <p>Berry Genomics (China)</p>	<p>CVS, amniocentesis or cordocentesis and karyotyping or birth phenotype</p>	<p>2nd trimester triple serum screening</p> <p>(α-fetoprotein, free β-hcg, unconjugated estriol)</p> <p>Cutoff \geq 1:270 for T21 and T18.</p>	<p>NIPT test performance for detection of T21, T18, T13 and SCA.</p> <p>Comparison of NIPT and serum screening performance.</p>	Comparison of NIPT with CT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		<p>Mean gestational age (SD), range:</p> <p>16.57 (1.56),</p> <p>11 - 21+6 weeks.</p> <p>1st trimester: 3.4%,</p> <p>2nd trimester: 96.6%.</p> <p>Assisted conception 14 (0.8%).</p>			<p>Increased NT:</p> <p>10 (0.6%),</p> <p>Other indications</p> <p>16 (0.9%).</p> <p>Low risk</p> <p>1,466/1,741 (84.2%).</p>					
<p>Song 2015[45]</p> <p>China</p> <p>Study start date: May 2012</p>	<p>Prospective cohort</p> <p>Number of centres: 1</p>	<p>N=213</p> <p>Women with singleton pregnancies, ≥ 35 years, 8+0 – 12+6 weeks' gestation, high-risk of foetal aneuploidies, presenting for NIPT.</p> <p>Mean age (range):</p>	<p>N=1 (0.5%):</p> <p>1 with quality control failure (haemolysis)</p>	T21, T18, T13	All high-risk for foetal aneuploidies due to advanced maternal age ≥ 35 years.	<p>MPS (whole genome)</p> <p>Berry Genomics (China)</p>	<p>CVS or amniocentesis and karyotyping (n=178) or newborn phenotypic examination (n=34).</p>	None	<p>1) Clinical performance of NIPT in the first trimester.</p> <p>2) Relationship between foetal DNA fraction and early gestational</p>	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		37.25 (35-45) years. Mean gestational age (range): 9+6 (8+0 – 12+6) weeks. 100% 1 st trimester.							age.	
Sparks 2012[54] USA Study start date: NR	Prospective case-control Number of centres: NR	Number enrolled unclear. Singleton pregnancies, women \geq 18 years, \geq 10 weeks' gestation, high risk for foetal trisomies undergoing invasive testing. Subset of N=338 (250 euploid, 72 T21, 16 T18) randomised into <u>Validation set</u>	NR	T21, T18	High risk for foetal trisomy	DANSR and z statistic or FORTE Aria Diagnostics (USA)	Invasive testing with FISH and/or karyotype analysis	None	Detecting foetal aneuploidy using DANSR and z statistic or FORTE	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		<p>(n=167)</p> <p>(36 T21, 8 T18, 123 euploid):</p> <p>Mean age (SD), range: 33.5 (7.1), 18-51 years.</p> <p>Mean gestational age (SD), range: 18.6 (4.0), 11.0-36.1 weeks.</p> <p><u>Training set</u> (n=171)</p> <p>(36 T21, 8 T18, 127 euploid):</p> <p>Mean age (SD), range: 34.5 (6.3), 18-44 years.</p> <p>Mean gestational age (SD), range: 17.6 (4.4), 10.3-33.0</p>								

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		weeks.								
Stumm 2014[34] Germany , Switzerland Study start date: NR	Prospective cohort Number of centres: 5	N=522 recruited, N=504 for testing. Women with singleton pregnancy, ≥18 years, high risk for aneuploidies, with foetal karyotype. Mean age (range): 36.0 (19-47) years. Mean gestational age (range): 15.6 (11+0 – 32+1) weeks.	N=18 (3.4%): 9 no consent, 8 no karyotype, 1 sample previously tested.	T21, T18 T13	All high risk for chromosomal aberrations: AMA >35 years 69.5%, Positive serum markers 11.1%, Ultrasound abnormality 39.3%, Family history 2.1%, Parental chromosome abnormality 0.4%, Other 14.9% (more than 1 risk factor in 179/522)	MPS (whole genome) LifeCodexx (Germany)	Amniocentesis, CVS, cordocentesis and foetal karyotyping	None	1) Diagnostic accuracy for foetal T21 detection (using DAP.21). 2) Diagnostic accuracy for foetal T13 and T18 detection (using DAP.plus) and comparison of algorithms for T21.	Accuracy of NIPT
Verweij	Multicentre	N=595 enrolled,	N=75 (12.6%):	T21	91.2% increased risk for T21 based	DANSR	CVS (54%) or	None	Test performance	Accuracy

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
2013[35] Netherlands, Norway, Sweden, USA Study start date: May 2011	international prospective cohort (EU-NITE study) Number of centres: 6 (4 Dutch, 2 Swedish)	N=520 eligible. Women undergoing invasive testing, singleton pregnancy, ≥ 10 weeks' gestation. Mean age (SD), range: 36.4 (4.6), 20-47 years. Mean gestational age (SD), range: 14.0 (2.1), 10-28 weeks. Ethnicity: Caucasian 84.8%, Mediterranean 6.0%, Asian 3.3%,	21 failed I/E criteria (non-invasive procedure performed, twin pregnancy, no blood sample); 19 insufficient plasma volume; 11 logistical problems - shipping difficulties; 24 chromosome abnormalities other than T21.		on 1 st trimester screening (serum screening, NT and/or maternal age), detection of foetal anomalies on ultrasound, previous affected pregnancy or family history. 8.8% other indications (psychosocial or anxiety reasons).	FORTE Harmony Ariosa Diagnostics (USA)	amniocentesis (46%) and karyotyping or quantitative fluorescent PCR		for T21 detection by shipping whole blood samples from Europe to a laboratory in the USA.	of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		Black 1.3%, Other 4.6%.								
Wax 2015[43] USA Study start date: June 2012	Retrospective review of prospective cohort Number of centres: 1	N=1,046 eligible for NIPT, N=166 high-risk pregnant women with singleton pregnancies opted for NIPT. Mean age (SD): 34.6 (5.5) years. Gestational age: range 10+0 – 21+6 weeks. 1 st and 2 nd trimester.	Multiple pregnancy N=NR; N=880 (84.1%) chose not to have NIPT.	T21, T18, T13	All high-risk: AMA \geq 35 years 742 (70.9%), Ultrasound abnormality 280 (26.8%), Positive screen 115 (11.0%), Prior trisomy 15 (1.4%), Parental translocation 1 (0.1%).	MPS (whole genome) Manufacturer: NR	Amniocentesis (n=56) or CVS (n=50) and karyotyping, postnatal karyotyping of neonatal blood, birth phenotype from records	None	Difference in genetic counselling utilisation, invasive procedures and T21 detection before and after NIPT implementation.	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
Zhang 2015[5] China, Hong Kong (Denmark) Study start date: January 2012	Prospective multicentre cohort Number of centres: 508	N=147,314 samples received for NIPT. N=147,103 appropriate samples. Women with singleton or twin pregnancy, ≥ 9 weeks of gestation, ≥ 18 years old. Mean age (range): 30.9 (18-56) years. Mean gestational age (range): 18.7 (9-37) weeks. Trimester: 1 st (9-13 wks): 4.21%, 2 nd (14-27 wks): 94.13%,	N=211 (0.14%): 211 samples rejected due to inadequate volume, contamination, <9 gestational weeks, or improper labelling.	T21, T18, T13	Mixed (high-risk, low-risk or no prior screening): Positive T21 screening 37.83%, Negative T21 screening 21.43%, No prior screening 40.73%. AMA 23.04%, Family history of aneuploidies 0.01%, Sonographic markers of chromosomal abnormality 1.61%.	MPS (whole genome) NIFTY test BGI-Health (China)	Karyotyping or clinical follow-up results.	None	1) Clinical performance of NIPT in detecting T21, T18, and T13. 2) NIPT performance in twin pregnancies. NIPT performance for T21 detection in high-risk and low-risk subjects. Factors contributing to NIPT false-positive and false-negative results.	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
		3 rd (≥ 28 wks): 1.47%, Unknown: 0.18%. 99.45% singletons, 0.55% twins.								
Zhou 2014[44] China Study start date: November 2011	Prospective cohort Number of centres: 1	N=7,705 Women with singleton pregnancies, 12-24 weeks' gestation, high-risk or no prior T21 screening. Gestational age: 12-24 weeks. 1 st and 2 nd trimester.	Multiple pregnancy N=NR	T21, T18, T13	Mixed risk: AMA ≥ 35 years: 40.4%, High risk T21 screening: 32.1%, Low risk T21 screening: 11.3%, No prior T21 screening: 56.6%.	MPS (whole genome) NIFTY test BGI-Shenzhen, China	Amniocentesis and karyotyping (n=54), postnatal karyotype (n=2) or birth outcome (n=3,894).	None	1) NIPT performance for detection of trisomies 13, 18, and 21. 2) Confirming care flow path	Accuracy of NIPT
Zimmermann 2012[55]	Prospective case-control	N=166 (11 T21, 3 T18, 2	NR	T21, T18 T13	Mixed: Aneuploidy	SNP-based, Parental Support (PS)	Invasive testing and FISH	None	Detection of foetal aneuploidies	Accuracy of NIPT

Reference	Study design	Participants	Exclusions from study	Trisomies investigated (T21, T18, T13)	Prior risk	Type of test	Reference method	Comparator if applicable	Primary / secondary outcome	Accuracy of NIPT or comparison of NIPT with CT
USA Study start date: NR	Unblinded proof-of-principle study Number of centres: NR	T13, 2 45X, 2 47XXY, 146 putatively euploid) Singleton pregnancies, women \geq 18 years, \geq 9 weeks' gestation. Median gestational age: 17.0 and 17.5 weeks for euploid and aneuploid samples, respectively.			samples from pregnant women with invasive prenatal testing. Putative euploid samples from average-risk women without known risk indicators.	algorithm Natera Inc. (USA)	and/or karyotype in aneuploid samples, 62/146 putative euploid samples confirmed by karyotyping of post-birth child tissue.		at chromosomes 13, 18, 21, X, and Y.	

AMA, advanced maternal age; β -hCG, β -fragment of human chorionic gonadotropin; CMA, chromosomal microarray; CT, first-trimester combined test; CVS, chorionic villus sampling; DANSR, digital analysis of selected regions; DNA, deoxyribonucleic acid; FISH, fluorescence in situ hybridisation; FORTE, Foetal fraction Optimized Risk of Trisomy Evaluation; FTS, first-trimester combined test; ICD, international classification of diseases; I/E criteria, inclusion or exclusion criteria; IQR, interquartile range; IVF, in vitro fertilisation; MPS, massively parallel sequencing; MS-AFP, maternal serum alpha-fetoprotein; NATUS, Next Generation Aneuploidy Test Using SNPs; NIFTY, Non-invasive Fetal Trisomy Test; NIPT, non-invasive prenatal testing; NPV, negative predictive value; NR, not reported; NT, nuchal translucency; PAPP-A, pregnancy-associated plasma protein; PCR, polymerase chain reaction; PPV, positive predictive value; QF-PCR, quantitative fluorescent polymerase chain reaction; ROC, receiver-operating-characteristic curve; SCA, sex chromosome anomalies; SD, standard deviation; SNP, single-nucleotide polymorphism; TOP, termination of pregnancy. * Reviewer calculation from published data.

Supplement 6 Test characteristics by type of test

Test characteristics – MPSS (whole genome)

Reference	Blood sampling (volume, time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Denominator	GC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
Alberti 2015[56] France	10 ml / Before invasive testing	All fragments mapping to Chr21 (no markers)	Illumina HiSeq2000	NR (10 libraries prepared at the same time)	z-score > 3 for T21, used 23 euploid pregnancies as reference set.	Total count of unique sequences mapped in the control-sequencing run.	No / no	NR	SOAP2 / 0 mismatch
Bianchi 2012[47] USA	17 ml / Before invasive test	All fragments mapping to Chr13, Chr18 or Chr21 (no markers)	Illumina HiSeq 2000	6-plex	NCV > 4.0 aneuploid, NCV < 2.5 euploid, $2.5 \leq \text{NCV} \leq 4.0$ unclassified; Used 110 independent unaffected samples	Normalizing chromosome denominators not specified	Normalising chr denominators / NR	hg18 (UCSC)	Bowtie short read aligner (version 0.12.5) / ≤ 2 mismatches
Bianchi 2014[19] USA	10 ml / Before or > 2 weeks after invasive test	All fragments mapping to Chr13, Chr18 or Chr21 (no markers)	Illumina HiSeq 2000	8-plex	$\text{NCV} \geq 4.0$ affected, $\text{NCV} \leq 3.0$ unaffected, $3.0 < \text{NCV} < 4.0$: resequenced in 1-plex	Normalising chromosome denominators not specified	Normalising chr denominators / NR	hg18 (UCSC)	Bowtie short read aligner (version 0.12.5) / ≤ 2 mismatches
Chen 2011[48] Hong Kong,	5-10 ml / Before	All fragments mapping to Chr13 or	Genome Analyzer IIX	2-plex	z-score > 3 for T13 and T18;	Total GC-corrected read counts from a	GC correction (LOESS regression)	Hg18 NCBI.36	Short Oligonucleotide Alignment

Reference	Blood sampling (volume, time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Denominator	GC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
Netherlands, UK, China	invasive test	Chr18 (no markers)	(Illumina)		103 independent male euploid samples as controls	sample	/ non-repeat masked		Programme 2 (SOAP2); no mismatch
Chiu 2011[49] Hong Kong, Netherlands, UK, China	5-10 ml / Before invasive test	All fragments mapping to Chr21 (no markers)	Genome Analyzer IIx (Illumina) for 2-plex; Genome Analyzer II (Illumina) for 8-plex	2-plex or 8-plex	z-score > 3 for T21; used 82 and 96 independent male euploid samples as controls for 2-plex and 8-plex, respectively	Total reads sequenced from a sample	no / repeat-masked	NCBI Build 36, version 48	ELAND, version 1.0 for Genome Analyzer II and version 1.4 for Genome Analyzer IIx / NR
Dan 2012[20] China, Hong Kong	5 ml / Before invasive test	All fragments mapping to chr18 and chr21 (no markers)	Illumina GAIIx or Illumina HiSeq 2000	4-plex or 12-plex	Binary hypothesis t-test and logarithmic LR between the two t-tests (NIFTY): t > 2.5 and L > 1: test positive, t > 2.5 or L > 1: test positive, t < 2.5 and L < 1: test negative.	Total number of unique reads. Then normalisation by average <i>k</i> -mer coverage of the 22 autosomes	GC correction (Losses regression) / NR	hg18, NCBI build 36	NR / 0 mismatch
Ehrich	10 ml /	All fragments aligned to	Genome Analyzer IIx	4-plex	z-score > 2.5 for T21;	All sequence reads excluding	no / non-repeat	UCSC hg19 human	CASAVA version 1.6 / up

Reference	Blood sampling (volume, time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Denominator	GC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
2011[50] USA	Before invasive test	Chr21 (no markers)	(Illumina)	(1-plex for resequencing when foetal fraction \leq 3.9%)	used 24 independent euploid reference samples; iterative censoring to adjust for biased control group	chr X and Y	masked	reference genome	to 1 mismatch
Huang 2014[22] China, Denmark, Hong Kong	5 ml / Before invasive test	All fragments mapping to chr18 and chr21 (no markers)	Illumina GAIIX or Illumina HiSeq 2000 (from [20])	4-plex or 12-plex (from [20])	Binary hypothesis t-test and logarithmic LR between the two t-tests (NIFTY): $t > 2.5$ and $L > 1$: test positive, $t > 2.5$ or $L > 1$: test positive (or test repeated), $t < 2.5$ and $L < 1$: test negative.	For k-mer coverage: Total number of unique reads. Then normalisation by average k -mer coverage of the 22 autosomes	GC correction (Losses regression) / NR	hg18, NCBI build 36	NR / 0 mismatch
Jeon 2014[39] South Korea, China	10 ml / Before invasive testing	All fragments mapping to Chr18 or Chr21 (no markers)	Ion Proton TM System (Life Technologies, Grand Island, NY, USA)	10-plex	z-score, all 139 euploid samples from this study used as reference group. Interactive threshold.	Mapped reads without denominator used for z-score calculation	Filtered by GC contents (35%-45%) / non-repeat masked	Unmasked Human reference genome sequence (hg19)	BWA / NR
Jiang 2012[23]	5 ml /	All fragments mapping to chr13, chr18	Illumina GAIIX and Illumina	multiplex	Binary hypothesis t-test and logarithmic LR between the two t-	For k-mer coverage: total number of	GC correction (Losses regression)	hg18, NCBI build 36	NR / 0 mismatch

Reference	Blood sampling (volume, time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Denominator	GC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
China	NR	and chr21 (no markers)	HiSeq 2000		tests (NIFTY): $ t_{ij;first} > 3$ and $ t_{ij;second} < 3$ as warning criteria. Autosomal aneuploidy if $L_{ij} > 1$.	unique reads. Then normalisation by average k-mer coverage for the 22 autosomes	/ NR		
Lau 2012[24] Hong Kong, China, Japan	5 ml / Before invasive test	All fragments mapping to chr13, chr18 and chr21 (no markers)	Illumina HiSeq 2000	12-plex	z-score (with internal reference chr) ≥ 3 for trisomy; used 400 independent euploid samples as reference set.	Total number of unique reads	GC correction (internal reference chromosome: Chr4 for T13, Chr8 for T18, Chr14 for T21) / repeat-masked	NCBI build 36.1	ELAND / 0 mismatch
Lau 2014[25] Hong Kong, USA, China	5 ml / Before invasive test	All fragments mapping to chr13, chr18 and chr21 (no markers)	Illumina GAIIX and Illumina HiSeq 2000 (from [23])	Multiplex (from [23])	Binary hypothesis t-test and logarithmic LR between the two t-tests (NIFTY): $t > 2.5$ and $L > 1$: test positive, $t > 2.5$ or $L > 1$: test positive (or test repeated), $t < 2.5$ and $L < 1$: test	For k-mer coverage: total number of unique reads. Then normalisation by average k-mer coverage for the 22 autosomes (from [23])	GC correction (Losses regression) / NR	Hg18, NCBI build 36	NR / 0 mismatch

Reference	Blood sampling (volume, time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Denominator	GC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
					negative. Threshold t-value NR				
Liang 2013[26] China	5 ml / Before invasive test	All fragments mapping to chr13, chr18 and chr21 (no markers)	Illumina HiSeq 2000	8-plex or 12-plex	z-score > 3 for T21, z-score > 5.91 for T18, z-score > 5.72 for T13; reference set of 50 independent female euploid samples	Total count of sequences uniquely mapped to all autosomal chromosomes	GC correction (slope of simple linear regression) / non-repeat masked	Unmasked human reference genome (hg19)	SOAP2 / NR
Palomaki 2012[52] USA	20-50 ml / Before invasive test	All fragments mapping to Chr13, Chr18 or Chr21 (no markers)	Illumina HiSeq 2000	4-plex	FC-robust z-scores \geq 3 for T21, T18 and T13. Euploid pregnancies considered to be controls for each chromosome.	Counts for all 22 autosomes (from [62])	GC correction / non-repeat masked for T13 and T18, repeat-masked for T21 test and post hoc for T13 and T18 analysis	UCSC hg19 human reference genome (from [62])	CASAVA version 1.6 / 0 mismatches (from [62])
Porreco 2014[31] USA	20-30 ml / Before invasive test	All fragments mapping to chr13, chr18, chr21, X and Y (no markers)	Illumina HiSeq 2000	12-plex	FC-robust z-score \geq 3 for T21, Flow cell-robust z-score \geq 3.95 for T18 and T13.	Counts for all 22 autosomes (from [64])	GC correction / repeat-masked (from [64])	UCSC hg19	Bowtie version 2 / 0 mismatch (from [64])

Reference	Blood sampling (volume, time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Denominator	GC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
Sago 2014 [42] Japan	20 ml / Before invasive testing	NR (MaterniT21 Plus, Sequenom: All fragments mapping to chr13, chr18 and chr21 (no markers))	NR (Illumina HiSeq 2000[64])	NR (12-plex[64])	NR (Robust z-scores $z > 3$ for chromosome 21 and $z > 3.95$ for chromosomes 18 and 13[64]).	NR (Counts for all 22 autosomes (from [64]))	GC correction / repeat-masked (from [64])	NR (UCSC hg19 (from [64]))	NR / NR (Bowtie2 / Perfect matches within the seed sequence (from [64]))
Sehnert 2011[53] USA	20 ml / Before invasive test	All fragments mapping to chr13, chr18 and chr21 (no markers)	Genome Analyzer IIx (Illumina)	Monoplex	NCV > 4.0 aneuploid, NCV < 2.5 euploid, $2.5 \leq \text{NCV} \leq 4.0$ unclassified; Used independent euploid samples from training set.	Chr9 for Chr21, Chr8 for Chr18, Sum of Chr(2-6) for Chr13.	Normalising chr denominators / NR	hg18 (UCSC)	Bowtie short read aligner (version 0.12.5) / ≤ 2 mismatches
Shaw 2014[32] Taiwan, China	5 ml / Before invasive test	All fragments mapping to chr13, chr18 and chr21 (no markers)	Illumina HiSeq 2000	12-plex	z-score > 3 for trisomy; Used 50 independent female euploid samples as reference set	Total count of sequences uniquely mapped to all autosomes (from [26])	GC correction (slope of simple linear regression[26]) / non-repeat masked	hg19	SOAP2 / NR

Reference	Blood sampling (volume, time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Denominator	GC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
Song 2013[33] China	5 ml / Before invasive test	All fragments mapping to chr13, chr18 and chr21 (no markers)	Illumina HiSeq 2000	12-plex	z-score ≥ 3 for trisomy; Used 50 independent female euploid samples as reference set	Total count of sequences uniquely mapped to all autosomes (from [26])	GC correction (slope of simple linear regression[26]) / non-repeat masked	hg19	BWA / NR
Song 2015[45] China	NR / Before invasive testing	All fragments mapping to Chr13, Chr18 or Chr21 (no markers)	Illumina HiSeq 2000	12-plex (from [33])	z-score ≥ 3 for trisomy; Used 50 independent female euploid samples as reference set (from [33])	Total count of sequences uniquely mapped to all the autosomal chromosomes (from [26])	GC correction (slope of simple linear regression[26]) / non-repeat masked(from [33])	hg19	BWA / NR (from [33])
Stumm 2014[34] Germany, Switzerland	7-10 ml / Before invasive procedure	All fragments mapping to chr13, chr18 and chr21 (no markers)	Illumina HiSeq 2000	12-plex	MAD-based z-score ≥ 3 for T21, ≥ 3.9 for T13 and ≥ 3.2 for T18. 1%, 2%, 4%, 10%, 20% or 40% T21 DNA control samples in each FC.	Total counts of all autosomes, X and Y	DAP.21 for T21: no / repeat-masked (after unblinding DAP.plus for T13, T18 and T21 with GC correction (LOWESS))	DAP.21: hg18, DAP.plus: hg19	ELAND / 0 mismatch
Wax 2015[43] USA	NR / Before invasive testing	NR (Single commercial laboratory using MPSS)	NR	NR	NR	NR	NR / NR	NR	NR / NR

Reference	Blood sampling (volume, time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Denominator	GC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
Zhang 2015[5] China, Hong Kong, (Denmark)	5 ml / Before invasive testing	All fragments mapping to chr13, chr18 and chr21 (no markers)	Illumina HiSeq2000	24-plex	A binary hypothesis t-test and logarithmic likelihood ratio L-score between the two t-tests (NIFTY) (from [20 23]). Threshold NR	Total number of unique reads. Then normalisation by average k-mer coverage of the 22 autosomes (from [20 23])	GC correction (Losses regression) / NR (from [20 23])	hg18, NCBI build 36	NR / 0 mismatch (from [20 23])
Zhou 2014[44] China	NR / Before invasive testing	All fragments mapping to chr13, chr18 and chr21 (no markers) (from [20])	Illumina GAIIx or Illumina HiSeq 2000 (from [20])	NR	Binary hypothesis t-test and logarithmic LR between the two t-tests (NIFTY): $t > 2.5$ and $L > 1$: test positive, $t > 2.5$ or $L > 1$: test positive, $t < 2.5$ and $L < 1$: test negative (from [20])	Total number of unique reads. Then normalisation by average k-mer coverage of the 22 autosomes (from [20])	GC correction (Losses regression) / NR (from [20])	hg18, NCBI build 36 (from [20])	NR / 0 mismatch (from [20])

BWA, Burrows–Wheeler Aligner; Chr, chromosome; DNA, deoxyribonucleic acid; FC, flow cell; GC, guanine cytosine; LOESS / LOWESS, locally weighted scatterplot smoothing regression; LR, likelihood ratio; MAD, median absolute deviation; MPSS, massively parallel signature sequencing; NCBI,

National Centre for Biotechnology Information; NCV, normalised chromosome value; NIFTY, Non-Invasive Fetal Trisomy test; NR, not reported; SOAP, Short Oligonucleotide Alignment Program; UCSC, University of California, Santa Cruz.

Test characteristics - DANSR (targeted sequencing)

Reference	Blood sampling (volume / time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Denominator	GpC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
Ashoor 2012[46] UK	10 ml / Before invasive test	576 nonpolymorphic loci on each chr18 and chr21	Illumina HiSeq 2000	96-plex	FORTE risk score, threshold NR	Sum of mean cfDNA counts of the loci for chr18 and chr21	Median polish on log-transformed counts / NA	Expected locus sequences	NR / <3 mismatches
Bevilacqua 2015[37] Belgium, UK, Spain	20 ml / Before invasive testing	576 nonpolymorphic loci on each chr13, chr18 and chr21 (from [21 54])	Illumina HiSeq 2000 (from [21 54])	96-plex (from [21 54])	FORTE risk score (threshold NR, HarmonyTM Prenatal Test usually uses FORTE risk score of 1% as cutoff)	Sum of mean cfDNA counts of the loci for chr13, chr18 and chr21 (from [21 54])	Median polish on log-transformed counts / NA (from [21 54])	Expected locus sequences (from [21 54])	NR / <3 mismatches (from [21 54])
Del Mar Gil 2014[21] UK	2 ml stored plasma / NR	576 nonpolymorphic loci on each chr13, chr18 and chr21	Illumina HiSeq 2000	96-plex	FORTE risk score, threshold NR	Sum of mean cfDNA counts of the loci for chr13, chr18 and chr21	Median polish on log-transformed counts /	Expected locus sequences	NR / <3 mismatches

Reference	Blood sampling (volume / time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Denominator	GpC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
							NA		
Nicolaides 2012[27] UK	2 ml stored plasma / Before invasive test	576 nonpolymorphic loci on each chr18 and chr21	HiSeq 2000	96-plex	FORTE risk score > 1%: High risk for T18 or T21	Sum of mean cfDNA counts of the loci for chr18 and chr21	Median polish on log-transformed counts / NA	Expected locus sequences	NR / <3 mismatches
Norton 2012[29] USA, Sweden, Netherlands	20 ml / Before invasive test	576 nonpolymorphic loci on each chr18 and chr21	Illumina HiSeq 2000	96-plex	FORTE risk score > 1%: High risk for T18 or T21	Sum of mean cfDNA counts of the loci for chr18 and chr21	Median polish on log-transformed counts / NA	Expected locus sequences	NR / <3 mismatches
Norton 2015[6] USA, Sweden	NR / Before invasive testing	Harmony™ Prenatal test: 576 nonpolymorphic loci on each chr13, chr18 and chr21 for chromosome proportion.	Illumina HiSeq 2000 (from [54])	96-plex (from [54])	FORTE risk score > 1%: High risk for T13, T18 or T21, respectively.	Sum of mean cfDNA counts of the loci for chr13, chr18 and chr21 (from [54])	Median polish on log-transformed counts / NA (from [54])	Genome Reference Consortium human build 37	NR / <3 mismatches (from [54])
Quezada 2015[41] UK	20 ml / Before invasive testing	Harmony™ Prenatal test: 576 nonpolymorphic loci on each	Illumina HiSeq 2000 (from [46 54])	96-plex (from [46 54])	FORTE risk score (threshold NR, usually 1% cutoff).	Sum of mean cfDNA counts of the loci for chr13, chr18 and chr21 (from [46 54])	Median polish on log-transformed counts / NA (from [46	Expected locus sequences (from [46 54])	NR / <3 mismatches (from [46 54])

Reference	Blood sampling (volume / time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Denominator	GpC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
		chr13, chr18 and chr21 for chromosome proportion.					54))		
Sparks 2012[54] USA	8 ml / NR	576 nonpolymorphic loci on each chr18 and chr21	Illumina HiSeq 2000	96-plex	Training set: Standard Z-test of proportions; iterative censoring on each lane of 96 samples; z-score > 3. Validation set: FORTE risk score, threshold 1:100-1:300	Sum of mean cfDNA counts of the loci for chr18 and chr21	Median polish on log-transformed counts / NA	Expected locus sequences	NR / <3 mismatches
Verweij 2013[35] Netherlands, Sweden, USA	20 ml / Before invasive test	576 nonpolymorphic loci on each chr18 and chr21	Illumina HiSeq 2000	96-plex	FORTE risk score > 1%: High risk	Sum of mean cfDNA counts of the loci for chr18 and chr21	Median polish on log-transformed counts / NA	Expected locus sequence	NR / <3 mismatches

cfDNA, cell-free deoxyribonucleic acid; Chr, chromosome; DANSR, digital analysis of selected regions; FORTE, Fetal-fraction Optimized Risk of Trisomy Evaluation; NA, not applicable; NR, not reported.

Test characteristics – Single-nucleotide polymorphism-based NIPT (with PS or NATUS algorithm)

Reference	Blood sampling (volume / time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Paternal genetic sample	GpC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
Hall 2014[51] USA	NR / NR	11,000 or 19,488 SNPs on chromosomes 21, 18, 13, X, and Y	Illumina GAIIx or HiSeq sequencer	11,000-plex or 19,488-plex targeted PCR	NATUS: calls foetal genotype and foetal fraction with maximum likelihood, calculates copy number call accuracy, threshold NR	yes	NA / NR	NR	Proprietary algorithm adapted from Novoalign (Novocraft, Selangor, Malaysia) / NR (from [55])
Korostelev 2014[40] Russia	NR / Before invasive testing	>19,000 polymorphic loci covering chromosomes 21, 13, 18, X, and Y.	NR (Illumina GAIIx or HiSeq sequencer (from [55]))	NR (19,488-plex targeted PCR (from [30]))	Maximum likelihood estimate generated by the NATUS algorithm combined with maternal and gestational age prior risks. Threshold NR.	NR	NA / NR	NR	NR / NR (Proprietary algorithm adapted from Novoalign (Novocraft, Selangor, Malaysia) / NR (from [55]))
Nicolaides, 2013[28] UK	20 ml / Before invasive test	19,488 SNPs on chromosomes 21, 13, 18, X, and Y	Illumina GAIIx or HiSeq sequencer (from [55])	19,488-plex targeted PCR	NATUS: calls foetal genotype and foetal fraction with maximum likelihood, calculates copy number call accuracy, threshold NR	no	NA / NR	NR	Proprietary algorithm adapted from Novoalign (Novocraft, Selangor, Malaysia) / NR (from [55])
Pergament 2014[30]	NR / 93% before invasive test,	19,488 SNPs on chromosomes 21, 13, 18, X,	Illumina GAIIx or HiSeq sequencer	19,488-plex targeted PCR	NATUS: calls foetal genotype and foetal fraction with maximum likelihood,	yes for 48.1% of samples	NA / NR	NR	Proprietary algorithm adapted from Novoalign (Novocraft,

USA	7% at least 4 days after	and Y	(from [55])		calculates copy number call accuracy, threshold NR				Selangor, Malaysia) / NR (from [55])
Zimmermann 2012[55] USA	20-40 ml / Putative euploid samples before, most aneuploidy samples after invasive test	11,000 SNPs on chromosomes 21, 18, 13, X, and Y	Illumina GAIIX or HiSeq sequencer	11,000-plex targeted PCR	PS: calls foetal genotype and foetal fraction with maximum likelihood, calculates copy number call accuracy, threshold NR	yes	NA / NR	NR	Proprietary algorithm adapted from Novoalign (Novocraft, Selangor, Malaysia) / NR

NA, not applicable; NATUS, Next-generation Aneuploidy Test Using SNPs; NIPT, non-invasive prenatal testing; NR, not reported; PCR, polymerase chain reaction; PS, Parental Support™ algorithm; SNP, single-nucleotide polymorphism.

Test characteristics – other approaches

Reference	Blood sampling (volume / time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Paternal genetic sample	Denominator (reference chromosome)	Human reference genome	Alignment algorithm / mismatches allowed
Dhallan 2007[57] USA	25-50 ml / NR	549 SNPs on chr 13; 570 SNPs on chr 21	NA (Allelic SNP ratio: PCR followed by quantification of bands on sequencing gels)	NA	Mean log ratio of foetal DNA between chr 13 and chr 21 significantly different (two-tailed Student's t-test allowing for unequal variances, significance level <0.05)	yes	Chr 13	NA	NA / NA

Chr, chromosome; DNA, deoxyribonucleic acid; NA, not applicable; NR, not reported; PCR, polymerase chain reaction; SNP, single-nucleotide polymorphism.

Test characteristics – more than one approach

Reference	Blood sampling (volume, time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Denominator	GC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
Beamon 2014[36] USA	NR / Before invasive testing	NR (Verinata and Sequenom commercial tests: All fragments mapping to Chr13, Chr18 or Chr21 (no markers))	NR	NR	Verinata: Normalised chromosome value (NCV) >4.0 for autosomal aneuploidy and <2.5 for unaffected foetuses. NCV between 2.5 and 4.0 as “unclassified”. Sequenom: NR (Robust z-scores, cutoff NR)	NR	NR / NR	NR	NR / NR
Comas 2014[38] Spain	≤ 20 ml / Before invasive testing Panorama: Paternal genetic	NR (Harmony test: 576 nonpolymorphic loci on each chr13, chr18 and chr21. Panorama test: 19,488 polymorphic loci covering chromosomes 21, 13, 18, X,	NR	NR	NR (Harmony: FORTE risk score (usually 1% cutoff). Panorama: Maximum likelihood estimate generated by the NATUS algorithm combined with maternal and gestational age prior risks.	NR (Harmony: Sum of mean cfDNA counts of the loci for chr13, chr18 and chr21 (from [46 54]) Panorama: NA)	NR / NR	NR	NR / NR

Reference	Blood sampling (volume, time of sampling)	Type and number of markers used	Sequencing platform	Multiplexing	Threshold	Denominator	GC correction / repeat masked	Human reference genome	Alignment algorithm / mismatches allowed
	sample in 51% of samples.	and Y.)			Threshold NR.)				

cfDNA, cell-free deoxyribonucleic acid; Chr, chromosome; NA, not applicable; NCV, normalised chromosome value; NR, not reported.

Supplement 7 Study quality according to QUADAS-2[7]

Study	Risk of bias					Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Role and impact of sponsor	Patient selection	Index test	Reference standard
Alberti 2015[56]	High	High	Low	High	Low	Unclear	Low	Low
Ashoor 2012[46]	High	Unclear	Low	High	Unclear	Low	Low	Low
Beamon 2014[36]	High	Low	Low	High	Low	High	Low	Low
Bevilacqua 2015[37]	Unclear	Low	Low	High	High	High	Low	Low
Bianchi 2012[47]	High	Low	Low	High	High	High	Low	Low
*Bianchi 2014[19]	Unclear	Unclear	Low	High	High	High	High	Low
Chen 2011[48]	High	Unclear	Low	Low	High	Unclear	Low	Low
Chiu 2011[49]	High	Low	Low	High	High	High	Low	Low
Comas 2014[38]	High	Low	Low	High	High	High	Low	Low
Dan 2012[63]	Unclear	Unclear	Low	High	High	High	Low	Low
Del Mar Gil 2013 [21]	Unclear	Unclear	Unclear	High	Unclear	Low	Low	Low
Dhallan 2007[57]	High	Unclear	Low	Low	High	High	High	Low
Ehrich 2011[50]	High	High	Low	High	High	High	Low	Low
Hall 2014[51]	High	Unclear	Low	High	High	High	High	Low
Huang 2014[22]	Unclear	Unclear	Low	Low	High	High	Low	Low
Jeon 2014[39]	Unclear	High	Low	Low	Low	High	Low	Low
Jiang 2012[23]	Unclear	Unclear	Low	Low	High	High	Low	Low

Study	Risk of bias					Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Role and impact of sponsor	Patient selection	Index test	Reference standard
Korostolev 2014[40]	Unclear	Low	Low	High	Low	High	Low	Low
Lau 2012[24]	Unclear	Low	Low	Low	Unclear	Low	Low	Low
Lau 2014[25]	Low	Unclear	Low	High	Unclear	High	Low	Low
Liang 2013[26]	Unclear	Low	Low	High	Low	High	Low	Low
*Nicolaidis 2012[27]	Unclear	Low	Low	High	Unclear	High	Low	Low
Nicolaidis 2013[28]	Unclear	Unclear	Low	High	Unclear	Low	Low	Low
Norton 2012[29]	Unclear	Low	Low	High	High	High	Low	Low
*Norton 2015[6]	Unclear	Low	Low	High	High	High	Low	Low
Palomaki 2012[52]	High	High	Low	High	High	High	Low	Low
Pergament 2014[30]	Unclear	High	Low	High	High	High	Low	Low
Porreco 2014[31]	High	Low	Low	High	High	High	Low	Low
*Quezada 2015[41]	Unclear	Low / High\$	Low	High	Unclear	High	Low	Low
Sago 2014[42]	High	Low	Low	High	Unclear	Unclear	Low	Low
Sehnert 2011[53]	High	Low	Low	High	High	High	Low	Low
Shaw 2014[32]	Unclear	Low	Low	Low	Unclear	High	Low	Low
*Song 2013[33]	Unclear	Low	Low	High	Low	High	High	Low

Study	Risk of bias					Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Role and impact of sponsor	Patient selection	Index test	Reference standard
Song 2015[45]	Unclear	Low	Low	High	Unclear	Low	Low	Low
Sparks 2012[54]	High	High	Low	High	High	High	Low	Low
Stumm 2014[34]	Low	Low for DAP.21 High for DAP.plus**	Low	High	High	High	Low	Low
Verweij 2013[35]	Low	Low	Low	High	High	High	Low	Low
Wax 2015[43]	Low	Unclear	Low	High	Low	Unclear	Low	Low
Zhang 2015[5]	Unclear	Unclear	Low	High	High	High	Low	Low
Zhou 2014[44]	Unclear	Unclear	Low	High	Unclear	High	Low	Low
Zimmermann 2012[55]	High	High	Low	High	High	High	High	Low

* Studies comparing NIPT with conventional screening tests for T21, T18 and T13 (addressing Research question 2)

** A second algorithm was used for T18 and T13 during the study which was unblinded.

\$ In this study the combined test (as comparator) was also assessed.

Supplement 8 Outcomes of test accuracy

Reference	Foetal Fraction, Median (IQR)	2x2 table					Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis
		TP	TN	FP	FN							
Alberti 2015[56] France	20.11 (mean among 43 male euploid foetuses) 16.86 (mean among 23 T21 foeuses)	T21	47	136	0	0	100 (90.6-100)	100 (96.6-100)	100 (90.6-100)	100 (96.6-100)	NR	11 test failures / 0 inconclusive results / 8 used for pretesting phase, 23 used as reference set. → 42 (18.7%) excluded.
Ashoor 2012[46] UK	NR	T21	50	297	0	0	100 (91.1-100)	100 (98.4-100)	100 (91.1-100)	100 (98.4-100)	NR	3 test failures / 0 inconclusive results /
		T18	49	297	0	1	98 (88.0-99.9)	100 (98.4-100)	100 (90.9-100)	99.7 (97.8-99.98)	NR	50 T18 cases excluded from T21 performance analysis and vice versa. → 53 (13.3%) excluded.
Beamon 2014[36] USA	NR	T21	5	157	0	0	100 (46.3-100)	100 (97.0-100)	100 (46.3-100)	100 (97.0-100)	NR	3 test failures / 2 unclassified for T21,
		T18	2	160	1	1	66.7	99.4	66.7	99.4	NR	1 unclassified for T13 /

Reference	Foetal Fraction, Median (IQR)	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis	
		TP	TN	FP	FN							
			T13	0	162	1	0	(12.5-98.2) NA	(96.1-99.97) 99.4 (96.1-99.97)	(12.5-98.2) NA		(96.1-99.97) 100 (97.1-100)
	All	7	155	1	1	87.5 (46.7-99.3)	99.4 (95.9-99.97)	87.5 (46.7-99.3)	99.4 (95.9-99.97)	NR	→ 46 (22.1%), 44 (21.2%), and 45 (21.6%) excluded from T21, T18, and T13 analysis, respectively.	
Bevilacqua 2015[37] Belgium, UK, Spain	8.7 (Range 4.1-30.0)	T21	11	328	0	1	91.7 (59.8-99.6)	100 (98.6-100)	100 (67.9-100)	99.7 (98.0-99.98)	NR	16 test failures / 0 inconclusive results /
		T18	5	335	0	0	100 (46.3-100)	100 (98.6-100)	100 (46.3-100)	100 (98.6-100)	NR	7 miscarriage or stillbirth without karyotype,
		T13	0	340	0	0	NA (98.6-100)	100 (98.6-100)	NA (98.6-100)	100 (98.6-100)	NR	19 pregnancies still continuing, 138 lost to follow-up. (Overlap of 5 samples with test failure and no reference standard). → 175 (34%) excluded.
Bianchi	NR	T21	89	404	0	0	100	100	100	100	NR	2 pre-analytic failures,

Reference	Foetal Fraction, Median (IQR)	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis	
		TP	TN	FP	FN							
2012[47]												
USA		T18	35	460	0	1	(95.9-100)	(99.1-100)	(94.8-100)	(98.8-100)	NR	16 test failures /
							97.2	100	100	99.8		7 inconclusive for T21,
		T13	11	485	0	3	(85.5-99.9)	(99.2-100)	(87.7-100)	(98.6-99.99)	NR	5 inconclusive for T18,
							78.6	100	100	99.4		2 inconclusive for T13 /
												Censored complex karyotype:
												19 for T21,
												18 for T18,
												18 for T13.
												(Overlap of 3 censored and test failures.)
												→ 41 (7.7%) for T21, 38 (7.1%) for T18 and 35 (6.6%) for T13 excluded.
Bianchi 2014	NR	T21	5	1941	6	0	100	99.7	45.5	100	FP rate, %:	18 test failures /
[19]							(47.8-100)	(99.3-99.9)	(16.7-76.6)	(99.8-100)	0.3	0 inconclusive results /
USA.		T18	2	1947	3	0	100	99.8	40.0	100	0.2	48 lost to follow-up,
NIPT							(15.8-100)	(99.6-100)	(5.3-85.3)	(99.8-100)		24 no live birth and no

Reference	Foetal Fraction, Median (IQR)	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis
		TP	TN	FP	FN						
		(Range 4.2-27.9)	T18	0	312	0	0	NA	100 (98.5-100)	NA	
		T13	0	312	0	0	NA	100 (98.5-100)	NA	100 (98.5-100)	NR
Dan 2012[20]	NR	T21	139	7384	1	0	100 (96.6-100)	99.99 (99.9-100)	99.3 (95.5-99.96)	100 (99.9-100)	NR
China, Hong Kong		T18	41	7482	1	0	100 (89.3-100)	99.99 (99.9-100)	97.6 (85.9-99.9)	100 (99.9-100)	NR
Del Mar Gil 2014[21]	9.8	T21	9	182	0	1	90.0 (54.1-99.5)	100 (97.4-100)	100 (62.9-100)	99.5 (96.5-99.97)	NR
UK	(7.4-12.1) in 193 euploid pregnanci es	T18	0	192	0	0	NA (97.6-100)	100 (97.6-100)	NA (97.6-100)	100 (97.6-100)	NR
		T13	1	191	0	0	100 (5.5-100)	100 (97.5-100)	100 (5.5-100)	100 (97.5-100)	NR
Dhallan 2007[57]	32.5	T21	2	56	1	1	66.7 (12.5–98.2)	98.2 (89.4–99.9)	66.7 (12.5–98.2)	98.2 (89.4–99.9)	NR
USA	(range 17.0-93.8)										

Reference	Foetal Fraction, Median (IQR)	2x2 table					Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis
		TP	TN	FP	FN							
												0 other exclusions. → 0 (0%) excluded.
Ehrich 2011[50] USA	NR	T21	39	409	1	0	100 (89-100)	99.7 (98.5-99.9)	97.5 (85.3-99.9)	100 (98.8-100)	NR	13 pre-analytic failures, 18 test failures / 0 inconclusive results / 0 other exclusions. → 31 (6.5%) excluded.
Hall 2014[51] USA	11.1 (range 2.2-30.4)	T21	0	64	0	0	NA	100 (94.4-100)	NA	100 (92.9-100)	NR	4 test failures / 0 inconclusive results /
		T18	0	64	0	0	NA	100 (94.4-100)	NA	100 (92.9-100)	NR	0 other exclusions. → 4 (5.9%) excluded.
		T13	15	49	0	0	100 (78.2-100)	100 (98.2-100)	100 (74.7-100)	100 (90.9-100)	NR	
Huang 2014[22] China, Denmark, Hong Kong	NR	T21	9	180	0	0	100 (62.9-100)	100 (97.4-100)	100 (62.9-100)	100 (97.4-100)	NR	0 test failures / 0 inconclusive results /
		T18	1	187	0	1	50	100	100	99.5	NR	0 exclusions.

Reference	Foetal Fraction, Median (IQR)	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis	
		TP	TN	FP	FN							
				T13	3	678	0	1	<i>75.0</i> <i>(21.9-98.7)</i>	<i>100</i> <i>(99.3-100)</i>	<i>100</i> <i>(31.0-100)</i>	<i>99.85</i> <i>(99.0-99.99)</i>
Lau 2012[24] Hong Kong, China, Japan	NR	T21	11	97	0	0	<i>100</i> <i>(67.9-100)</i>	<i>100</i> <i>(95.3-100)</i>	<i>100</i> <i>(67.9-100)</i>	<i>100</i> <i>(95.3-100)</i>	NR	0 test failures / 0 inconclusive results / 0 other exclusions. → 0 (0%) excluded.
		T18	10	98	0	0	<i>100</i> <i>(65.5-100)</i>	<i>100</i> <i>(95.3-100)</i>	<i>100</i> <i>(65.5-100)</i>	<i>100</i> <i>(95.3-100)</i>	NR	
		T13	2	106	0	0	<i>100</i> <i>(19.8-100)</i>	<i>100</i> <i>(95.6-100)</i>	<i>100</i> <i>(19.8-100)</i>	<i>100</i> <i>(95.6-100)</i>	NR	
Lau 2014[25] Hong Kong, USA, China	NR	T21	23	1659	0	0	<i>100</i> <i>(82.2-100)</i>	<i>100</i> <i>(99.7-100)</i>	<i>100</i> <i>(82.2-100)</i>	<i>100</i> <i>(99.7-100)</i>	NR	0 test failures / 1 inconclusive result /
		T18	4	1678	0	0	<i>100</i> <i>(39.6-100)</i>	<i>100</i> <i>(99.7-100)</i>	<i>100</i> <i>(39.6-100)</i>	<i>100</i> <i>(99.7-100)</i>	NR	299 without reference standard. → 300 (15.1%) excluded.
		T13	2	1680	0	0	<i>100</i> <i>(19.8-100)</i>	<i>100</i> <i>(99.7-100)</i>	<i>100</i> <i>(19.8-100)</i>	<i>100</i> <i>(99.7-100)</i>	NR	
Liang 2013[26]	NR	T21	40	372	0	0	100	100	100	100	NR	12 test failures /

Reference	Foetal Fraction, Median (IQR)	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis		
		TP	TN	FP	FN								
China		T18	14	398	0	0	(89.1-100) 100 (73.2-100)	(98.7-100) 100 (98.8-100)	(89.1-100) 100 (73.2-100)	(98.7-100) 100 (98.8-100)	NR	0 inconclusive results / 11 failed karyotyping. → 33 (7.6%) excluded.	
		T13	5	407	1	0	100 (46.3-100)	99.75 (98.4-99.99)	83.3 (36.5-99.1)	100 (98.8-100)	NR		
Nicolaides 2012[27] UK. NIPT	10.0 (7.8-13.0)	T21	8	1941	0	0	100 (59.8-100)	100 (99.8-100)	100 (59.8-100)	100 (99.8-100)	NR		100 test failures (not included in either test) / 0 inconclusive results / 0 other exclusions. → 100 (4.9%) excluded.
		T18	2	1945	2	0	100 (19.8-100)	99.9 (99.6-99.98)	50 (9.2-90.8)	100 (99.8-100)	NR		
		All	10	1937	2	0	100 (65.5-100)	99.9 (99.6-99.98)	83.3 (50.9-97.1)	100 (99.8-100)	FP rate, %: 0.1		
Combined FTS (≥1:150 for T18 and T21)	NA	T21	8	NR	NR	NR	NR	NR	NR	NR	NR		
		T18	2	NR	NR	NR	NR	NR	NR	NR	NR		
		All	10	1852	87	0	100 (65.5-100)	95.5 (94.5-96.4)	10.3 (5.3-18.6)	100 (99.7-100)	FP rate, %: 4.5		
Nicolaides	≥3.95	T21	25	204	0	0	100	100	100	100	NR	13 test failures /	

Reference	Foetal Fraction, Median (IQR)	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis	
		TP	TN	FP	FN							
2013[28] UK		T18	3	226	0	0	(86.3-100) 100 (31.0-100)	(98.2- 100) 100 (97.9-100)	(83.4-100) 100 (31.0-100)	(97.7-100) 100 (97.9-100)	NR	0 inconclusive results / 0 other exclusions. → 13 (5.4%) excluded.
		T13	1	228	0	0	100 (5.5-100)	100 (97.9-100)	100 (5.5-100)	100 (97.9-100)	NR	
Norton 2012[29] USA, Sweden, Netherlands	Mean 11 SD 4.5 (range 4.2-51.3)	T21	81	2887	1	0	100 (95.5-100)	99.97 (99.8-99.99)	98.8 (92.5-99.9)	100 (99.8-100)	NR	148 test failures / 0 inconclusive results /
		T18	37	2886	2	1	97.4 (86.5-99.9)	99.93 (99.75- 99.98)	94.9 (81.4-99.1)	99.96 (99.8-100)	NR	73 other chromosomal abnormalities excluded; 38 T18 cases excluded for T21 test performance; 81 T21 cases excluded for T18 test performance. → 259 (8.0%) for T21 and 302 (9.4%) for T18 excluded.
Norton 2015[6] USA, Sweden. NIPT	NR	T21	38	15794	9	0	100 (90.7-100)	99.9 (99.9-100)	80.9 (66.7-90.9)	100 (99.9-100)	LR+: 1755.9 LR-: 0 FP rate, %:	384 pre-analytic failures, 488 NIPT failures, 308 no standard-screening

Reference	Foetal Fraction, Median (IQR)	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis			
		TP	TN	FP	FN									
Combined FTS (≥1:270 for T21, ≥1:150 for T13 and T18)		T18	9	15830	1	1	90.0 (55.5-99.7)	100 (99.9-100)	90.0 (55.5-99.7)	100 (99.9-100)	0.06 (0.03-0.11) AUC: 0.999 FP rate, %: 0.01 (0-0.04)	result / 0 inconclusive results / 1,489 lost to follow-up. → 2,669 (14.4%) excluded for T21 and T18 from either test. For T13, another 4,656 patients enrolled before September 2012 were excluded. → 7,325 (39.5%) excluded for T13 from either test.		
		T13	2	11181	2	0	100 (15.8-100)	100 (99.9-100)	50.0 (6.8-93.2)	100 (99.9-100)	FP rate, %: 0.02 (0-0.06)			
		NA	T21	30	14949	854	8	78.9 (62.7-90.4)	94.6 (94.2-94.9)	3.4 (2.3-4.8)	99.9 (99.9-100)		LR+: 14.6 LR-: 0.22 FP rate, %: 5.4 (5.1-5.8) AUC: 0.958	
			T18	8	15782	49	2	80.0 (44.4-97.5)	99.7 (99.6-99.8)	14.0 (6.2-25.8)	100 (99.9-100)		FP rate, %: 0.31 (0.23-0.41)	
		T13	1	11155	28	1	50.0 (1.2-98.7)	99.7 (99.6-99.8)	3.4 (0.1-17.8)	100 (99.9-100)	FP rate, %: 0.25 (0.17-0.36)			
	Palomaki	4-50%	T21	210	1758	1	2	99.1	99.9	99.5	99.9		FP rate, %:	17 test failures /

Reference	Foetal Fraction, Median (IQR)	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis	
		TP	TN	FP	FN							
2012[52] USA	accepted	T18	59	1907	5	0	(96.3-99.8) 100 (92.4-100)	(99.6-100) 99.7 (99.4-99.9)	(97.0-99.98) 92.2 (82.0-97.1)	(99.5-99.98) 100 (99.7-100)	0.1 (<0.1-0.3) 0.3 (0.1-0.7)	0 inconclusive results / 0 other exclusions. → 17 (0.9%) excluded.
		T13	11	1943	16	1	91.7 (59.8-99.6)	99.2 (98.6-99.5)	40.7 (23.0-61.0)	99.9 (99.7-100)	0.9 (0.5-1.5)	[FP rate = FP / 1688 euploid samples.]
Pergament 2014[30] USA	NR	T21	58	905	0	0	100 (93.8-100)	100 (99.6-100)	100 (92.2-100)	100 (99.5-100)	NR	85 test failures, 8 test failures for 1/5 chromosomes (includes 2 no-calls for Monosomy X) /
		T18	24	938	1	1	96.0 (79.7-99.9)	99.9 (99.4-100)	96.0 (77.7-99.8)	99.9 (99.3-99.99)	NR	0 inconclusive results /
		T13	12	953	0	0	100 (73.5 -100)	100 (99.6-100)	100 (69.9-100)	100 (99.5-100)	NR	0 other exclusions. → 88 (8.4%) for T21, 87 (8.3%) for T13, 86 (8.2%) for T13 excluded.
Porreco 2014[31] USA	4-50% accepted	T21	137	3182	3	0	100 (97.34 -100)	99.92 (99.7-99.98)	97.9 (93.9-99.56)	100 (99.88-100)	FP rate, %: 0.1	54 test failures / 0 inconclusive results /
		T18	36	3283	0	3	92.3 (79.1-98.38)	100 (99.89-100)	100 (90.26-100)	99.9 (99.7-99.98)	0.0	56 complex karyotypes. (Overlap of 2 with test

Reference	Foetal Fraction, Median (IQR)	2x2 table					Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis
		TP	TN	FP	FN							
		T13	14	3306	0	2						
							87.5 (61.65- 98.45)	100 (99.89-100)	100 (76.84-100)	99.9 (99.8-99.99)	0.0	failure and complex karyotype). → 108 (3.1%) excluded.
Quezada 2015[41] UK. NIPT	11% (Range 4- 40%)	T21	32	2752	1	0	100 (86.7-100)	99.96 (99.8-100)	97.0 (82.5-99.8)	100 (99.8-100)	FP rate, %: 0.04	54 test failures / 0 inconclusive results /
		T18	9	2770	5	1	90.0 (54.1-99.5)	99.8 (99.6-99.9)	64.3 (35.6-86.0)	99.96 (99.8-100)	FP rate, %: 0.19	48 miscarriages or stillbirths with unknown karyotype; 21 lost to follow up.
		T13	2	2778	2	3	40.0 (7.3-83.0)	99.9 (99.7-99.99)	50.0 (9.2-90.8)	99.9 (99.7-99.97)	FP rate, %: 0.07	(Overlap of 3 patients without NIPT and reference standard result.)
		All	43	2730	8	4	91.5 (78.7-97.2)	99.7 (99.4-99.9)	84.3 (70.9-92.5)	99.9 (99.6-99.95)	FP rate, %: 0.3	→ 120 (4.1%) excluded.
Combined FTS (≥1:100 for T21)	NA	T21	34	2663	139	0	100 (87.4-100)	95.0 (94.2-95.8)	19.7 (14.2-26.5)	100 (99.8-100)	FP rate, %: 5.0	12 without FTS result / 48 miscarriages or stillbirths with unknown karyotype;
		All	49	2663	124	0	100 (90.9-100)	95.6 (94.7-96.3)	28.3 (21.9-35.8)	100 (99.8-100)	FP rate, %: 4.4	21 lost to follow up. (Overlap of 12 without

Reference	Foetal Fraction, Median (IQR)	2x2 table					Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis
		TP	TN	FP	FN							
												combined FTS and reference standard result.) → 69 (2.4%) excluded.
Sago 2014[42] Japan	NR	T21	71	1694	3	0	100 (93.6-100)	99.8 (99.4-99.95)	95.9 (87.8-98.9)	100 (99.7-100)	NR	4 test failures /
		T18	36	1723	8	1	97.3 (84.2-99.9)	99.5 (99.1-99.8)	81.8 (66.8-91.3)	99.9 (99.6-100)	NR	0 inconclusive results /
		T13	10	1756	2	0	100 (65.5-100)	99.9 (99.5-99.98)	83.3 (50.9-97.1)	100 (99.7-100)	NR	3 TOP without karyotype;
		All	NR	NR	NR	NR	NR	NR	NR	NR	FN rate, %: <0.1	9 foetal deaths without karyotype; 5,956 women without birth outcome. → 5,972 (77%) excluded.
Sehnert 2011[53] USA. Test set	NR	T21	13	34	0	0	100 (71.7-100)	100 (87.4-100)	100 (71.7-100)	100 (87.4-100)	NR	0 test failures /
		T18	8	39	0	0	100 (59.8-100)	100 (88.8-100)	100 (59.8-100)	100 (88.8-100)	NR	1 inconclusive for T13 /
		T13	0	46	0	0	NA	100 (90.4-100)	NA	100 (90.4-100)	NR	1 twin sample removed. → 1 (2.1%) for T21 and T18 excluded, 2 (4.2%) for T13 excluded.

Reference	Foetal Fraction, Median (IQR)	2x2 table					Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis
		TP	TN	FP	FN							
Shaw 2014[32] Taiwan, China	NR	T21	11	189	0	0	100 (67.9-100)	100 (97.5-100)	100 (67.9-100)	100 (97.5-100)	FP rate 0%	0 test failures / 0 inconclusive results / 1 case excluded due to early gestational age (10 weeks). → 1 (0.5%) excluded.
		T18	8	192	0	0	100 (59.8-100)	100 (97.6-100)	100 (59.8-100)	100 (97.6-100)	FN rate 0%	
											FP rate 0%	
											FN rate 0%	
		T13	3	197	0	0	100 (31.0-100)	100 (97.6-100)	100 (31.0-100)	100 (97.6-100)	FP rate 0%	
											FN rate 0%	
Song 2013[33] China. NIPT	NR	T21	8	1733	0	0	100 (59.77-100)	100 (99.72 -100)	100 (59.8-100)	100 (99.7-100)	FP rate, %: 0.00	73 test failures / 0 inconclusive results / 111 no birth outcome. (Overlap of 9 without NIPT and reference standard result) → 175 (9.1%) excluded for either test.
											FN rate, %: 0.00	
		T18	2	1738	1	0	100 (19.79-100)	99.94 (99.6-99.99)	66.67 (12.5-98.2)	100 (99.7-100)	FP rate 0.06%	
											FN rate 0.00%	
		T13	1	1740	0	0	100 (5.46-100)	100 (99.73- 100)	100 (5.5-100)	100 (99.7-100)	FP rate 0.00%	
											FN rate 0.00%	
		All	11	1729	1	0	100 (67.86-100)	99.94 (99.6-99.99)	91.67 (59.8-99.6)	100 (99.7-100)	FP rate 0.06%	
											FN rate 0.00%	

Reference	Foetal Fraction, Median (IQR)	2x2 table					Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis
		TP	TN	FP	FN							
Serum screening ($\geq 1:270$ for T18 and T21)	NA	All	6	1487	243	5	54.55 (24.6-81.7)	85.95 (84.2-87.5)	2.41 (0.98-5.4)	99.7 (99.2-99.9)	FP rate 14.05% FN rate 45.45%	
Song 2015[45] China	8.54 (range 2.69- 18.75) (n=100 male foetuses)	T21	2	202	0	0	100 (19.8-100)	100 (97.7-100)	100 (19.8-100)	100 (97.7-100)	NR	1 pre-analytic failure / 0 inconclusive results / 2 IUFD without karyotype, 1 TOP without karyotype, 5 spontaneous miscarriages without karyotype. → 9 (4.2%) excluded.
		T18	1	201	0	0	100 (5.5-100)	100 (97.7-100)	100 (5.5-100)	100 (97.7-100)		
		T13	1	201	0	0	100 (5.5-100)	100 (97.7-100)	100 (5.5-100)	100 (97.7-100)		
Sparks 2012[54] USA. Training set Validation set	NR	T21	35	120	1	0	100 (87.7-100)	99.2 (94.8-99.96)	97.2 (83.8-99.9)	100 (96.1-100)	NR	8 test failures in training set, 0 test failures in validation set / 0 inconclusive results / For both sets:
		T18	7	121	0	0	100 (56.1-100)	100 (96.2-100)	100 (56.1-100)	100 (96.2-100)	NR	
	NR	T21	36	122	1	0	100 (88.0-100)	99.2 (94.9-99.96)	97.3 (84.2-99.9)	100 (96.2-100)	NR	T18 cases excluded from T21 test performance and vice versa. → 15 (8.8%) for T21 and 43 (25.1%) for T18 excluded
		T18	8	122	1	0	100	99.2	88.9	100	NR	

Reference	Foetal Fraction, Median (IQR)	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis	
		TP	TN	FP	FN							
												from training set. 8 (4.8%) for T21 and 36 (21.6%) for T18 excluded from validation set.
Stumm 2014[34] Germany, Switzerland	NR	T21	40	430	0	2	95.2 (82.6-99.2)	100 (98.9-100)	100 (89.1-100)	99.5 (98.2-99.9)	NR	32 test failures / 0 inconclusive results / 0 other exclusions. → 32 (6.3%) excluded.
		T18	8	463	1	0	100 (59.8-100)	99.8 (98.6-99.99)	88.9 (50.7-99.4)	100 (99.0-100)	NR	
		T13	5	467	0	0	100 (46.3-100)	100 (99.0-100)	100 (46.3-100)	100 (99.0-100)	NR	
Verweij 2013[35] Netherlands, Sweden, USA	Mean 11.1, SD 4.1 (range 4- 30)	T21	17	486	0	1	94.4 (72.7 -99.9)	100 (99.4-100)	100 (77.1-100)	99.8 (98.7-99.99)	NR	30 pre-analytic failures, 16 test failures / 0 inconclusive results / 24 other chromosomal abnormalities besides T21. → 70 (12.2%) excluded.
Wax 2015[43] USA	NR	T21	3	161	0	0	100 (31.0-100)	100 (97.1-100)	100 (31.0-100)	100 (97.1-100)	NR	0 test failures / 0 inconclusive results /

Reference	Foetal Fraction, Median (IQR)	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis
		TP	TN	FP	FN						
		T18	1	163	0	0					
		T13	0	164	0	0	NA	100 (97.1-100)	100 (97.1-100)	NR	1 miscarriage without karyotype, 1 IUFD without karyotype. → 2 (1.2%) excluded
Zhang 2015[5] China, Hong Kong (Denmark). Overall performance (n=112,669)	NR	T21	720	111882	61	6	99.17 (98.52-99.83)	99.95 (99.93-99.96)	92.19 (90.31-94.07)	99.99 (99.99-100)	211 pre-analytic failures, 145 test failures / 0 inconclusive results / 34,289 without karyotyping or clinical follow-up. → 34,645 (23.5%) excluded.
		T18	167	112448	51	3	98.24 (94.93-99.63)	99.95 (99.94-99.97)	76.61 (70.99-82.23)	100 (99.99-100)	
		T13	22	112602	45	0	100 (84.56-100)	99.96 (99.95-99.97)	32.84 (21.59-44.08)	100 (99.99-100)	
		All	909	111594	157	9	99.02 (98.38-99.66)	99.86 (99.84-99.88)	85.27 (83.14-87.40)	99.99 (99.99-100)	
	Twins only (n=404)	T21	5	397	2	0	100 (47.82-100)	99.50 (98.20-99.94)	71.43 (29.04-96.33)	100 (99.08-100)	
Zhou 2014[44] China. NIPT	NR	T21	38	3910	2	0	100 (88.6-100)	99.9 (99.8-99.99)	95.0 (81.8-99.1)	100 (99.9-100)	4 test failures / 0 inconclusive results /
		T18	10	3938	2	0	100	99.9	83.3	100	5 TOP without karyotype,

Reference	Foetal Fraction, Median (IQR)	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Other	Test failures / inconclusive results / exclusions from analysis	
		TP	TN	FP	FN							
	implementation study		T13	2	3946	2	0	(65.5-100) 100 (19.8-100)	(99.8-99.99) 99.9 (99.8-99.99)	(50.9-97.1) 50.0 (9.2-90.8)		(99.9-100) 100 (99.9-100)
Zimmermann 2012[55] USA	Mean 12.0 Range 2.0-30.8	T21	11	66	0	0	100 (67.9-100)	100 (93.1-100)	100 (67.9-100)	100 (93.1-100)	NR	21 test failures / 0 inconclusive results /
		T18	3	74	0	0	100 (31.0-100)	100 (93.9-100)	100 (31.0-100)	100 (93.9-100)	NR	68 putative euploid samples without reference standard. → 89 (53.6%) excluded.
		T13	2	75	0	0	100 (19.8-100)	100 (93.9-100)	100 (19.8-100)	100 (93.9-100)	NR	

AUC, area under the receiver-operating-characteristic curve; cfDNA, cell-free deoxyribonucleic acid; chr, chromosome; CI, confidence interval; DNA, deoxyribonucleic acid; FP, false positive; FP rate = FP / (FP+TN) = 1 – Specificity; FN, false negative; FN rate = FN / (FN+TP) = 1 – Sensitivity; FTS, first-trimester screening; IQR, interquartile range; IUFD, intrauterine foetal death; LR+, positive likelihood ratio; LR-, negative likelihood ratio; MX, Monosomy X; NA, not applicable; NR, not reported; NPV, negative predictive value; PPV, positive predictive value; SCA, sex chromosome abnormalities; SD, standard deviation; TOP, termination of pregnancy; TN, true negative; TP, true positive. Note: Numbers in italics were calculated based on information given in the paper. Confidence intervals in italics were calculated using the Wilson score interval with continuity correction. Numbers and confidence intervals not in italics were extracted directly from the papers