Systematic review and cost-consequence assessment of cell-free DNA testing for T21, T18 and T13 in the UK – Final report

Review group:	Sian Taylor-Phillips
	Karoline Freeman
	Julia Geppert
	Jason Madan
	Olalekan Uthman
	Adeola Agbebiyi
	Angus Clarke
	Siobhan Quenby
	Aileen Clarke
Correspondence to:	Dr Sian Taylor-Phillips
	Populations, Evidence and Technologies
	Division of Health Sciences
	Warwick Medical School
	University of Warwick
	Coventry CV4 7AL
Tel:	02476 575882
Email:	S.Taylor-Phillips@warwick.ac.uk

Date completed: 4th July 2015

Plain English Summary

We investigated how well a new blood test (called cell-free DNA testing – cfDNA for short) for pregnant women works for detecting Down syndrome (trisomy 21), Edwards syndrome (trisomy 18) and Patau syndrome (trisomy 13) in the fetus. We systematically searched for published studies. We found high risk of bias in the research studies, meaning that test performance might be lower in real life than the studies suggest. We combined 41 different research studies to get an overall estimate of test accuracy. We found that test accuracy is very good but not 100%, so the test should not be used to give a final diagnosis. We estimated how well cfDNA would work if it was used 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. We predict that 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. One large study in the general pregnant population estimated that 19 in 100 pregnancies testing positive for Down syndrome did not actually have a baby with the condition. Because of the possibility of the test giving an inaccurate result cfDNA testing should not be considered as a diagnostic test for trisomies. Pregnant women with positive results should be offered an invasive diagnostic test (such as amniocentesis or chorionic villus sampling [CVS], which carry a small risk of miscarriage) to give a conclusive diagnosis.

We made an economic model to compare three options for the NHS. The first option is keeping the current NHS screening programme using the combined test (a combination of a blood test and ultrasound) with pregnant women given a screening risk of having a baby with Down's, Edwards' or Patau's syndrome of greater than 1/150 offered an invasive diagnostic test. The second option was using the combined test with women given a risk greater than 1/150 offered the new cfDNA test, and if they tested positive offered an invasive diagnostic test. This option resulted in similar numbers of trisomies detected, 43 fewer miscarriages of healthy pregnancies because of many fewer women choosing to have invasive tests than currently, and may cost approximately the same as currently. The third option is to use the new cfDNA test as the first test offered instead of the combined test. This option would cost an extra £105 million to the NHS, and would result in more invasive tests than the second option.

In summary, the new cfDNA test is very accurate, but does not give a definite answer. Offering the new cfDNA test to pregnant women who test positive using the current combined test could reduce the number of invasive tests, and therefore the number of miscarriages of unaffected fetuses caused by invasive testing. Because the cfDNA test cannot give a definitive answer as to whether the baby has a trisomy, a CVS or amniocentesis would be recommended before parents considered termination of pregnancy.

Funding Acknowledgement

This research was 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. The authors have no conflicts of interest.

Acknowledgements

We 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, Dave Wright for his assistance with the economic model, and Pam Royle for her work updating the searches.

Contents

List of Figures

List of Tables

List of abbreviations

AUC	Area under the hierarchical summary receiver-operating characteristics curve
β-hCG	beta-subunit of human chorion gonadotropin
BMI	Body mass index
cf	Cell-free
cfDNA	Cell-free deoxyribonucleic acid
CI	Confidence interval
CT	First-trimester combined test
CVS	Chorionic villus sampling
DANSR	Digital analysis of selected regions
DNA	Deoxyribonucleic acid
DOR	Diagnostic odds ratio
FASP	Fetal Anomaly Screening Programme
FN	False negative
FORTE	Fetal-fraction Optimized Risk of Trisomy Evaluation
FP	False positive
GC	Guanine-cytosine
hCG	Human chorionic gonadotropin
HSROC	Hierarchical summary receiver-operating characteristic
ICTRP	International Clinical Trials Registry Platform
MAD	Median absolute deviation
MPSS	Massively parallel shotgun sequencing
NA	Not applicable
NATUS	Next-generation Aneuploidy Test Using SNPs
NCV	Normalised chromosome value
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NIPT	Non-invasive prenatal testing
NLR	Negative likelihood ratio
NR	Not reported
NSC	National Screening Committee
NT	Nuchal translucency
PAPP-A	Pregnancy-associated plasma protein A
PCR	Polymerase chain reaction
PLR	Positive likelihood ratio
PPV	Positive predictive value
PS	Parental Support algorithm
PSSRU	Personal Social Services Research Unit
RDOR	Relative diagnostic odds ratio
RNA	Ribonucleic acid
ROC	Receiver operating characteristic curve
SD	Standard deviation
SN	Sensitivity

SNP	Single-nucleotide polymorphism
SP	Specificity
T13	Trisomy 13, Patau syndrome
T18	Trisomy 18, Edwards Syndrome
T21	Trisomy 21, Down Syndrome
TN	True negative
TOP	Termination of pregnancy
TP	True positive

1. Scientific summary

Introduction

Currently screening for trisomy in the UK depends on a combination of biochemical and ultrasound tests. If the combined test result is reported as high risk, genetic verification via an invasive diagnostic test is offered, either through chorionic villus sampling (CVS) or amniocentesis. Cell-free DNA (cfDNA) testing, sometimes known as Non-Invasive Prenatal Testing or NIPT measures the underlying genetic pathology of trisomy 21, trisomy 18 and trisomy 13 directly, by analysing fetal genetic material present in the maternal circulation. Several testing strategies have been developed and are commercially available (including massively parallel shotgun sequencing (MPSS), digital analysis of selected regions (DANSR) and targeted sequencing of single nucleotide polymorphisms (SNPs). In various studies cfDNA testing has been shown to have a high sensitivity and specificity for detection of trisomy. The aim of this report is therefore to undertake a systematic review and cost-consequence assessment of cfDNA testing in the first trimester for T21, T18 and T13 to contribute advice to the UK National Screening Committee (UKNSC) on its potential implementation in the UK. Specific questions for the review are shown below:

1a) What is the accuracy of cfDNA testing in predicting T21, T18 and T13 in pre-defined high risk (1:150) pregnant women following a combined test?

1b) How does changing the threshold for defining high risk following a combined test affect the accuracy of cfDNA testing?

2) What is the most accurate primary prenatal screening tool for T21, T18 and T13 in the first trimester when cfDNA testing and the combined test are compared in a general obstetric population?
3) What diagnostic accuracy is achievable by integrating cfDNA testing into the combined test?

4) What is the rate of cfDNA testing failure (number of inconclusive and excluded samples / total number of samples)?

5) What are the costs and consequences (cases detected, test-related miscarriages avoided) for the current NHS screening programme when cfDNA testing is used

- In sequence with the combined test
- As a replacement for the combined test
- In combination with (i.e. alongside) the combined test

Methods

Search strategies were designed to include pregnant women in the general population (or those with an increased risk for question 1), tested using cfDNA testing on maternal blood in the first trimester. Outcomes investigated included diagnostic accuracy, detection rate, sensitivity and specificity, predictive values, likelihood ratios, diagnostic odds ratios, receiver operating characteristic (ROC) curves, numbers of true positive, false positive, true negative and false negative results and data on inconclusive, indeterminate and excluded samples to determine test failure rates.

Searches were conducted in PubMed, Ovid Medline, Ovid Embase and the Cochrane Library. The search strategy used a combination of search terms for cfDNA testing and trisomies, with a limit to English language. The date limits were 1997 to 9th February 2015.

We included studies of pregnant women at high risk or in the general obstetric population, who had been given non-invasive prenatal testing using cell-free (cf) DNA derived from maternal blood

(serum, plasma, whole blood), and a reference standard of either genetic verification through amniocentesis, CVS, cordocentesis and fetal pathologic examination after abortion or postnatal phenotypic assessment. We included studies with and without a comparator of any "conventional" screening test. We also included studies with any diagnostic accuracy information, but excluded from the meta-analysis those from which a full 2x2 table could not be adequately constructed.

We excluded studies reporting the quantification of fetal cells or DNA and using elevated levels of the whole fetal DNA or epigenetic markers as a screening/diagnostic tool, case-control studies with <15 cases and cohort studies with <50 pregnant women, non-English studies, letters, reviews, editorials, grey literature, conference abstracts and communications containing insufficient information on methods and no numerical outcomes data.

Two reviewers independently screened all records identified with discrepancies resolved through a third reviewer. Study quality was assessed using QUADAS-2. We analysed reported performance of tests in the first trimester general obstetric population. Where possible quantitative analysis (meta-analysis) was used to provide pooled summary estimates for test accuracy across studies stratified by trisomy type. Where this was not possible a narrative review was undertaken. We used bivariate random-effects regression models and generated hierarchical summary receiver-operating characteristic (HSROC) curves. We investigated heterogeneity, publication bias, the effect of test failures and indeterminate results and undertook sensitivity analyses. We investigated the effects of different characteristics using meta-regression to generate relative diagnostic odds ratios (RDOR).

We constructed a decision tree to estimate the impact of alternative antenatal screening strategies on the annual performance of the England FASP. Costs used for the combined test was £27, for cfDNA testing was £232, amniocentesis £383 and CVS £319. The model was used to evaluate four strategies:

i) Combined test, followed by invasive testing offered to all those whose risk of trisomy is 1/150 or higher

ii) Combined test, followed by cfDNA testing offered to all whose risk of trisomy is 1/150 or higher, followed by invasive test offered to all those with a positive cfDNA testing result

iii) Combined test, followed by cfDNA testing offered to all whose risk of trisomy is 1/1000 or higher, followed by invasive test offered to all those with a positive cfDNA testing result

iv) cfDNA testing offered to all, followed by invasive test offered to all those with a positive cfDNA testing result

Results

We identified 2012 unique records of which 52 articles met the inclusion criteria. Risk of bias was high with 35 of 52 studies (67.3%) considered high risk in two or more domains of the modified QUADAS-2.

Meta-Analysis of test accuracy across all studies

A total of 41 studies were included in the meta-analysis for T21, 37 for T18 and 30 for T13. We found that pooled sensitivity was 97.1% (CI 95.5% to 98.1%) for T21, 93.1% (CI 90.0% to 95.3%) for T18, and 82.7% (CI 74.7% to 88.5%) for T13. The pooled specificity was 99.8% (CI 99.7% to 99.9%) for each of the three trisomies. These may be underestimates due to use of a zero cell correction to enable model convergence. Without the zero cell correction for TP and FN we found that pooled sensitivity was 99.4% (CI 98.9% to 99.6%) for T21, 97.4% (CI 95.8% to 98.4%) for T18, and 97.4% (CI 86.2% to 99.6%) for T13. The pooled specificity was 99.9% (CI 99.9% to 100%) for all three trisomies. These may be overestimates due to publication bias, risk of bias in included studies, and inclusion of second and third trimester pregnancies where there is a higher fetal fraction enhancing cfDNA test performance. We applied estimates of sensitivity and specificity without zero cell corrections to a theoretical cohort of 10,000 pregnant women from a high risk population. Population prevalence

was determined as the median prevalence for the studies included in high risk groups. In a high risk population with prevalence 3% for T21, 1.5% for T18 and 0.5% for T13 the positive predictive value was 91% for T21, 84% for T18 and 87% for T13, indicating that even when using the most favourable meta-analysis results which may be overestimates, cfDNA testing must not be considered a diagnostic test.

We found that sensitivity was significantly lower in twin than singleton pregnancies. Sensitivity was reduced by 9% for T21, 28% for T18 and 22% for T13. Sensitivity was also lower by 1% for T21, 3% for T18 and 22% for T13 when studies which included only first trimester pregnancies were compared to all other studies with pregnancies across all trimesters, but this was not statistically significant. Pooled sensitivity was also lower by 2% for T21, 9% for T18, and 26% for T13 in studies in the general obstetric population, in comparison to studies in high risk populations, but this difference was only statistically significant for T13. Although some of these results are not statistically significant, they may be potentially clinically significant and therefore merit further investigation.

Using Deeks' funnel plots we found publication bias in the included studies so test accuracy may be overestimated, because publications with positive results may have been more likely to have been published than those with less promising results. In an analysis including only study designs which were less likely to result in overestimation due to spectrum bias (cohort studies with consecutive sampling) we found sensitivity estimates were lower by 4% for T21, 6% for T18 and 2% for T13, but these differences were not statistically significant.

Overall test accuracy estimates from the meta-analysis may have been deflated by use of a zero cell correction, but also inflated by including studies in the second and third trimester of pregnancy when fetal fraction is higher, inflated by publication bias, and inflated by bias in the included studies.

Research question 1a: What is the accuracy of cfDNA testing in predicting T21, T18 and T13 in pre-defined high risk (1:150) pregnant women following a combined test?

There were no studies reporting the performance of cfDNA testing after the UK combined test at threshold 1:150, but the meta-analysis showed that in high risk populations defined in a range of ways, pooled sensitivity (without zero cell corrections for TP and FN) was 97.2% (CI 95.1% to 98.4%) for T21, 92.9% (CI 89.2% to 95.4%) for T18 and 95.3% (CI 86.3% to 98.5%) for T13. Pooled specificity was >99.7% for all three trisomies. Applied to a high risk population with prevalence 3.3%, 1.5% and 0.5% for T21, T18 and T13 respectively, would give positive predictive values of 91%, 84% and 87% respectively. Therefore whilst it is a very good test, even using our highest estimates of accuracy it must not be considered a diagnostic test.

Research question 1b: How does changing the threshold for defining high risk following a combined test affect the accuracy of cfDNA testing?

Ten studies reported risk thresholds from screening tests prior to cfDNA testing. There was no difference between high and low risk pregnancies in the number of false negatives (FN) and false positives (FP) for T21, T18 or T13 in two studies that investigated two different risk thresholds. However, the evidence was weak and was based on small studies that did not report performance separately by risk group. One large study reported similar sensitivities and specificities in the study population of mainly second trimester pregnancies divided into two risk groups, however, the risk was assessed using a number of methods and the study used two different risk thresholds from previous screening tests. The remaining 7 studies did not contribute to the decision question. We are therefore unable to present cfDNA testing performance at different risk cut-offs ranging from very high to low risk or present an optimal risk cut-off to maximise cfDNA testing performance in clinical practice.

The meta-analysis showed a reduction in sensitivity of cfDNA testing in the general obstetric population in comparison to high risk groups, but this was only statistically significant for T13. In the general obstetric population sensitivity was 95% (CI 87% to 99%) for T21, (decreased from 97% (CI 95% to 98%) over all studies) and 84% (CI 60% to 95%) for T18 (decreased from 93% (CI 89% to 95%) for all studies, and 60% (CI 30% to 84%) for T13 (decreased from 83% (CI 75% to 89%) for all studies.

Research Question 2: What is the most accurate primary prenatal screening tool for T21, T18 and T13 in the first trimester when cfDNA testing and the combined test are compared in a general obstetric population?

It was not possible to include a comparison of test performance of cfDNA testing and the combined test in the meta-analysis, due to a lack of studies making the comparison, and heterogeneity in the definition of the combined test. Individual studies have provided evidence that the specificity of the combined test is considerably worse than that of cfDNA testing, which is to be expected as the tests have different thresholds. The threshold for positive results on the combined test is designed to be at a risk of ≥1/150 or ≥1/270, so the chosen threshold will by its nature include a large number of false positives. One large study in the US found that sensitivity for T21 is also better for cfDNA testing (detected 36/36 cases) than the combined test at threshold ≥1/270 (detected 30/36 cases). For T18 and T13 it is more difficult to make comparisons due to the lower prevalence.

Research Question 3: What diagnostic accuracy is achievable by integrating cfDNA testing into the combined test?

One option for implementation of cfDNA testing is the integration of cfDNA testing into the current first trimester combined screening test to provide one risk score integrating all screening information. This may be of advantage as cfDNA testing is not 100% accurate and the combined screening test may provide additional information therefore achieving a potentially higher test performance with an integrated test result. There was one included study and one theoretical study discussing the potential of integrating cfDNA testing and the combined test but no studies which demonstrated test accuracy after implementing this approach.

Research Question 4: What is the rate of cfDNA testing failure (number of inconclusive and excluded samples / total number of samples)?

The rate of initial analytic failure (failure of the initial cfDNA testing) ranged from 0% to 12.7% and among 5,789 pregnancies with resampling, 803 (13.9%) also failed repeat cfDNA testing. Including analytic failures as an intention to diagnose in the meta-analysis (assuming they are false positives and false negatives) reduced the sensitivity by 1-3% and the specificity by 2-3%, but this was not statistically significant. There were 5 papers in this review that reported indeterminate results (results in a range defined as neither positive nor negative) for trisomies 21, 18 and 13 ranging from 0% (0/2042) to 11.1% (5/45). In the study with no indeterminate results they used 8-plex testing, and where the initial score was indeterminate they repeated using 1-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. (2014) 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 ten 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 [10.9%], p=0.004). Norton et al. (2015)[46] 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).

Two studies have shown that test failure may be associated with a small increased risk of aneuploidy, but other studies found no association, and pregnancies should not be considered high

risk on the basis of a failed test alone. Many other factors may cause test failures, including testing earlier in pregnancy, and testing methods such as multiplexing.

Research question 5: Economic Evaluation

In comparison to the current UK screening programme, if cfDNA testing is offered initially to those identified as at high risk (>1/150) of trisomy, the model predicts that 9,912 tests would be carried out, and 350 retests would be required. As a consequence, the number of invasive tests required would fall from 7,910 to 1,434. The majority of invasive tests avoided are in trisomy-free pregnancies, so that the number of test-related miscarriages of healthy pregnancies falls from 46 to 3 per year. If cfDNA testing were to cost £232 per test then the additional cost to the screening programme would be £120,000. This includes savings due to a reduction in the number of invasive tests. However, using conservative estimates for test performance 13 fewer trisomies would be detected by screening because cfDNA testing is not a perfect test (sensitivity < 100%) resulting in false negative test results when cfDNA testing is scheduled after the combined screening test. Using more optimistic estimates for test performance means that an estimated 24 more trisomies would be detected. If estimates of the risks associated with the combined test are taken from Danish registry data rather than Australian Medicare data then the estimate of the reduction in the number of test-related miscarriages from adding the cfDNA testing increases from 46 to 112.

If the combined test threshold were to be relaxed to 1/1000 and followed by cfDNA testing then 93 more trisomies would be detected per year in comparison to the current programme, and there would be 36 fewer test-related miscarriages of healthy pregnancies, but this would cost £8million extra per year.

If cfDNA testing were to replace the combined test the additional cost to the NHS would be £105 million, (including 12million cost of continuing the combined test) and there would be 38 fewer test-related miscarriages of healthy pregnancies, and 117 extra trisomies detected.

It was not possible to model the implications of cfDNA testing in combination with the combined test, as there was no primary research demonstrating test accuracy with which to populate the model.

Conclusions and research recommendations

Pooled test sensitivity (without zero cell corrections for TP and FN) was 99.7% (CI 96.3% to 99.9%) for T21, 96.5% (CI 93.7% to 98.1%) for T18 and 95.3% (CI 86.3% to 98.5%) for T13. Pooled specificity was 99.9% for all three trisomies. Applied to a high risk population with prevalence 3.3%, 1.5% and 0.5% for T21, T18 and T13 respectively, would give positive predictive values of 91%, 84% and 87% respectively. Therefore whilst it is a very good test, because of the increased risk of termination of unaffected pregnancies it would be entirely inappropriate to use cfDNA testing as a diagnostic test. Test sensitivity was significantly lower for twins than singleton pregnancies for T21 and T18, and in the general obstetric population for T13. Estimates of sensitivity were lower in subgroups of studies which only included women in the first trimester, which only included women in the general obstetric population (rather than high risk groups), and only including cohort studies with consecutive sampling (these are less prone to spectrum bias), but this was not statistically significant. Deeks' funnel plots showed publication bias so test performance may be overestimated. The failure rate of the cfDNA testing was very variable, it ranged from 0% to 12.7% on the initial test and was 13.9% overall for repeat testing. Including analytic failures as an intention to diagnose in the meta-analysis (assuming they are false positives and false negatives) reduced the sensitivity by 1-3% and the specificity by 2-3%, but this was not statistically significant. There was limited evidence in the UK, and generalisability of findings to the UK should be carefully considered.

Our economic model predicts that the cost (£232/test) of implementing cfDNA testing would be partially offset by the reduced number of invasive tests if implemented after the current combined test at threshold 1/150, leading to an estimated annual cost increase of £120,000. Implementing cfDNA testing in this way would reduce the number of test-related miscarriages of healthy pregnancies by 43, detection of similar numbers of trisomies, and potentially delays to the diagnosis of 927 trisomies due to adding the extra test to the pathway. Lowering the threshold of the combined test reduces the number of missed trisomies but increases the costs. Implementing cfDNA testing as the primary screen for T21, T18 and T13 would cost an extra £105 million.

2. Introduction

Assessment for autosomal trisomy, a type of chromosomal abnormality, is the most common reason why women choose to undergo invasive prenatal diagnosis. However, the current indication for invasive diagnostic testing in the NHS is based on screening tests with a false positive rate (defined as 1-specificity) of about 5%.[1] Reducing the number of invasive procedures performed is desirable since invasive tests carry an increased risk of fetal loss. Cell-free DNA (cfDNA) testing, sometimes known as Non-Invasive Prenatal Testing or NIPT for trisomies, based on the analysis of fragments of fetal DNA in maternal blood, may have the potential to meet this aspiration. A number of clinical studies have been published, and a review to evaluate the performance of such tests is needed before implementation into UK clinical practice can be considered.

2.1 Antenatal screening for trisomies

The Fetal Anomaly Screening Programme (FASP) offers tests to pregnant women to assess risk of fetal chromosomal anomalies such as Down syndrome, Edwards syndrome and Patau syndrome.

These conditions are typically caused by an additional (third) chromosome in an otherwise normal (euploid) karyotype of 23 pairs of chromosomes. In other words, affected individuals carry an aneuploidy karypotype characterised by:

- 3 chromosomes 21 in Down syndrome (Trisomy 21 [T21])
- 3 chromosomes 18 in Edwards syndrome (Trisomy 18 [T18])
- 3 chromosomes 13 in Patau syndrome (Trisomy 13 [T13])

This underlying genetic imbalance causes an increase in gene dosage which in turn results in phenotypic characteristics and the symptoms of trisomy. Addition of only part of a third chromosome (translocation) or inheritance of the third chromosome in only a proportion of cells (mosaicism) are sometimes associated with milder symptoms than complete trisomy. These special cases also need to be considered in antenatal screening programmes.

2.2 Symptoms and prognosis

Down syndrome

The symptoms of Down syndrome include congenital cardiac defects and other malformations, mental impairment, learning difficulty, shortened life expectancy, and increased risk of leukaemia, Alzheimer's disease, thyroid disorders and diabetes. The severity of symptoms varies among Down syndrome individuals and people with Down syndrome now have a life expectancy of over 50 years.[2]

Edwards syndrome

The symptoms of Edwards syndrome include physical abnormalities, heart and kidney problems, growth problems, inability to walk and talk, breathing and feeding problems, bone abnormalities and severe learning disabilities.[3] About 75% of fetuses with Edwards syndrome are lost due to spontaneous miscarriage or termination.[4] Of those that are born alive, 5-10% survive beyond one year. The median life expectancy of babies with Edwards syndrome is 4 days and the mean survival is 48 days.[5]

Patau syndrome

Patau syndrome is associated with very severe physical and mental impairment. Physical signs and symptoms of the condition are numerous and include a low birth weight, abnormalities of the brain (failure to develop into two hemispheres), heart, kidneys and gastrointestinal tract, facial malformations, abnormalities of the hands and feet leading to the inability to walk, and severe learning disabilities.[6] 70-90% of fetuses with Patau syndrome are either miscarried spontaneously or

terminated.[4] Mortality is high with about 50% of live births surviving beyond one week, with a median survival of 2.5 days and a mean survival of 10 days; 5-10% of infants live beyond one year.[7]

2.3 Epidemiology

Down syndrome

There are currently just under 700,000 live births in England and Wales annually.[8] Down syndrome is the most common chromosomal abnormality (1,982 diagnoses in 2012 in England and Wales) and affects about 2.7 in 1,000 pregnancies in England and Wales (just under 2,000 annually) with a live birth prevalence of 1.1 per 1,000 live births.[4] It is prevalent in all ethnicities.[9] The risk of a T21 pregnancy, however, is dependent on maternal age and rises from 1:1300 in 25 year olds to 1:380 in 35 year olds and further to 1:28 in women aged 45.[10] The number of T21 pregnancies increased between 1989 and 2003 due to an increase in maternal age. However, improvement in prenatal screening methods towards 2003 and an increase in prenatal diagnosis, especially in younger women, followed by pregnancy termination actually resulted in an overall drop in T21 births between 1989 and 2003.[11]

Edwards syndrome

In 2012 there were 526 diagnoses of Edwards syndrome in England and Wales. The prevalence of Edwards syndrome is 0.7 per 1,000 births of which an estimated 68 were live births (life birth prevalence of 0.09 per 1,000 live births).[4] Prevalence increased significantly with maternal age. Mothers of affected births had a mean age of 36.3 years compared to 29.8 years of all mothers.[4]

Patau syndrome

Patau syndrome is slightly less common with 229 diagnoses in 2012 in England and Wales. The prevalence of Patau syndrome is 0.3 per 1,000 births and affects about 0.03 per 1,000 live births in England and Wales. This equated to about 22 live births in 2012.[4] Prevalence increased significantly with maternal age. Mothers of affected births had a mean age of 34.4 years compared to 29.8 years of all mothers.[4]

Due to the rate of spontaneous miscarriage and termination following prenatal diagnosis, the prevalence at the time of first trimester screening is higher for all three trisomies but particularly for T18 and T13.

2.4 Current service provision

Historically, prenatal screening for Down syndrome was based solely on maternal age (>35 years), identifying about 30% of pregnancies with Down syndrome.[12] With the availability of biochemical and ultrasound tests in the 1980s,[13] the specificity and sensitivity of Down syndrome screening increased dramatically. Biochemical serum screening tests detect protein levels in maternal blood that are down or up regulated in Down syndrome pregnancies, while ultrasound looks for physical features associated with Down syndrome fetuses, namely increased nuchal translucency (NT) thickness. The UK government announced the introduction of a screening programme for T21 offered to all pregnant women independent of age, in 2001 following the recommendations of the UKNSC.[13]

In order to achieve high detection rates, multiple screening tests, which are only meaningful in a narrow gestational window need to be offered. The current screening programme, therefore, involves a range of first and second trimester biochemical and ultrasound tests for women booking at different gestational ages. The National Institute for Health and Care Excellence (NICE) identified the combined test, which combines a maternal serum test of two pregnancy related proteins, and a nuchal translucency scan from 10 weeks, as the most cost effective for women booking in the first trimester (see Figure 1 for the screening pathway for women booking in the first trimester).[14] For

women booking later in pregnancy, a blood test is available which is not covered in this review. As a result of first trimester combined screening and the fetal anomaly scan 64% of all Down syndrome cases were detected prenatally in 2012.[4] Additionally, about 90% of cases of Edwards and Patau syndrome were detected prenatally with nearly 50% of these before 15 weeks.[4] Review of the available evidence encouraged the NSC to officially roll out the first trimester combined test for Edwards and Patau syndrome in 2015. The association between the different screening markers of the combined test and the trisomies is summarised in Table 1.

The results of any screening test for the trisomies are reported in the form of a ratio which expresses the individual's personal risk of having a baby with trisomy for the pregnancy tested. Since differential screening for Edwards and Patau syndrome is unsuccessful, the risk for Edwards or Patau syndrome is reported jointly. If the risk of any of the three trisomies exceeds 1 in 150 the result is regarded as positive and an invasive diagnostic test is offered.[15] The current gold standard diagnostic test for fetal trisomy is genetic verification either through chorionic villus sampling (CVS) or amniocentesis.[16] These invasive tests take fetal material from the placenta and amniotic fluid, respectively and trisomy is confirmed by either karyotyping or rapid testing[17] with an accuracy close to 100%.[18] The concern over the invasive tests is that they carry an increased yet unclear (possibly up to 2%) risk of miscarriage.[16] Hence there is a balance between the benefit of detecting fetal anomaly and the risk of causing the miscarriage of a normal fetus. This balance determines the level of risk at which invasive procedures are offered.

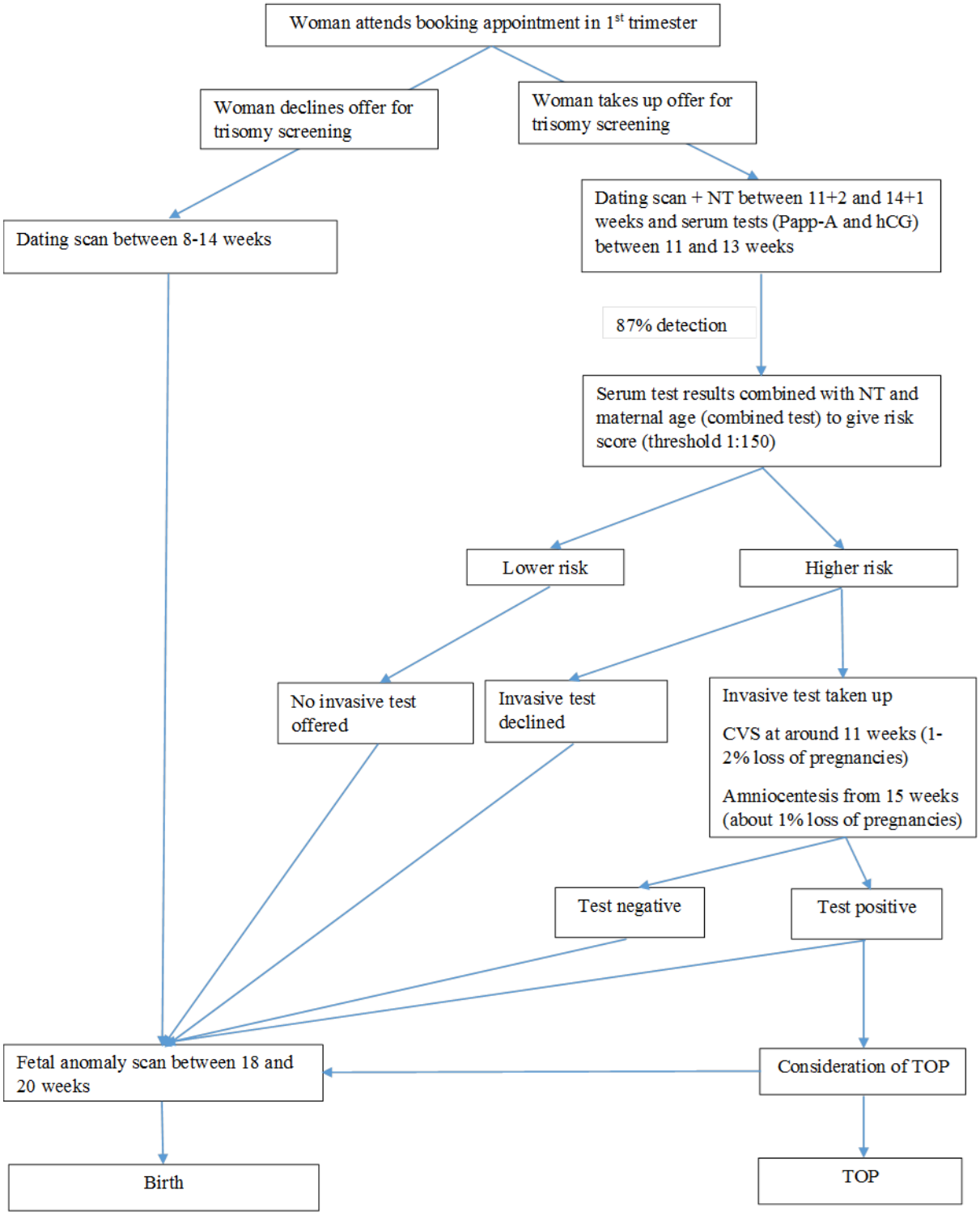


Figure 1 Screening pathway for trisomy women booking in the first trimester
NT – nuchal translucency; Papp-A – Pregnancy-associated plasma protein A; hCG – human chorionic gonadotropin; CVS – chorionic villus sampling; TOP – termination of pregnancy

Table 1 Association of first trimester combined serum screening markers with trisomies 21, 18 and 13

Trisomy	Marker of combined screening test	Levels associated with trisomy
T21	Maternal age NT PAPP-A Free β-hCG	increased increased decreased increased
T18 or T13	Maternal age NT PAPP-A Free β-hCG	increased increased decreased decreased

T21 – trisomy 21, T18 – trisomy 18; T13 – trisomy 13; NT – nuchal translucency; PAPP-A - Pregnancy associated plasma protein-A; β-hCG - β-human chorionic gonadotropin

2.5 Non-invasive prenatal testing (cfDNA testing) – technology under assessment

While current screening tests for trisomies target indirect, phenotypic (observable or measurable) characteristics that are associated with a trisomy pregnancy, i.e. pregnancy related protein levels and physical features of the fetus, cfDNA testing measures the underlying genetic pathology of trisomy 21, trisomy 18 and trisomy 13 directly by investigating fetal genetic material.

In the mid-1990s, detection of fetal cells[19] and free fetal nucleic acids (DNA and RNA)[20] in the maternal circulation were recognised as a potential route for non-invasive prenatal diagnosis for a number of conditions, promising earlier diagnosis, improved patient care and a reduction in invasive testing.[16] The assessment of trisomy presents a quantitative challenge of detecting the additional genetic information contributed by the third chromosome in an overwhelming background of maternal cells and DNA. cfDNA testing for trisomy therefore aims to quantify information from chromosome 21, 18 and 13 (which is present in the mother and the fetus) in order to identify the slight overrepresentation of the chromosome in question in affected pregnancies contributed by a fetus carrying three copies of chromosome 21, 18 or 13.

Several testing strategies have been developed to address the challenge and some tests are now commercially available offered by companies including Sequenom, Verinata Health, Ariosa, Natera,[21] as well as LifeCodexx,[22] Berry Genomics[23] and BGI.[24] These employ three different testing strategies, namely massively parallel shotgun sequencing (MPSS), digital analysis of selected regions (DANSR) and targeted sequencing of SNPs, which are described in turn below. Other strategies are under development but are currently not commercially available and were therefore not included in the review. These include for instance the epigenetic approach.

Massively parallel shotgun sequencing (MPSS)

The general idea of cfDNA testing using MPSS is to count random maternal and fetal DNA fragments in a maternal blood sample. This is done by sequencing the DNA fragments (determining the identity of the individual building blocks) and mapping them against a human reference genome of known sequence using a computer algorithm. This sequence and mapping process provides information on the chromosomal origin of the DNA fragments. A ratio is determined of the number of fragments that map against the chromosomes of interest (chromosomes 21, 18 and 13) to the number of fragments mapping to one (or more) chromosome that is not involved in trisomy (e.g. chromosome 1). This ratio is subsequently compared to the ratio derived from a set of normal pregnancies for the

same chromosomes. In order to distinguish affected from unaffected pregnancies this method relies on a minimum amount of fetal DNA in the maternal circulation, currently about 4%, which is termed the fetal fraction. In order to increase turnover and decrease cost, several samples can be sequenced in parallel in one reaction, which is termed multiplexing. This will result in fewer DNA molecules to be sequenced per sample relative to the total amount of DNA assessable in one reaction.

Digital analysis of selected regions (DANSR)

DANSR represents a variation to MPSS, in which a selected number of nonpolymorphic regions on the chromosomes of interest and a reference chromosome are considered. By sequencing only the fragments aligning to these specific regions, the sequencing load decreases from about 25 million fragments in MPSS to about 1 million in DANSR.[25]

Targeted sequencing of SNPs

While in DANSR the targeted regions are nonpolymorphic, the targeted sequencing of SNPs approach targets single nucleotide polymorphisms (SNPs), that is fetal DNA regions that differ in one DNA building block depending on whether it was inherited from the mother or the father. By investigating about 2000 SNPs per chromosome of interest this method allows the number of chromosomes 21, 18 and 13 to be determined.[21] This method does not require a reference chromosome and retains its accuracy at low fetal fractions.[26]

Once the sequencing data is available using any of the three strategies, different algorithms are used for data analysis in order to classify a sample as trisomy positive or negative.

Z-score approach

MPSS typically uses a z-score approach to classify trisomy positive and negative samples. An externally-referenced z-score is defined as the difference in the proportion of chromosome 21, 18 or 13 fragments in a tested sample and the mean proportion of fragments in the reference population, divided by its standard deviation (SD). For each sample the z-score is determined and compared typically with a threshold of 3 SD. A z-score of <3 is classified as trisomy negative and a z-score of >3 is classified as trisomy positive. This approach can be further optimised in a number of ways to improve the accuracy for the specific chromosomes tested and by using information on chromosomal variations within the sample set.

FORTE algorithm

In the DANSR method the z-score approach can be replaced by the Fetal-fraction Optimized Risk of Trisomy Evaluation (FORTE) algorithm which combines information on the fetal fraction and maternal age with the information from the z-score approach to report an individualised risk score.[25] It is the first approach to incorporate different risk factors with the outcome of cfDNA testing. Furthermore it does not require the information from and testing against external reference samples.

NATUS algorithm

The sequencing data obtained from the SNP approach is analysed using the Next-generation Aneuploidy Test Using SNPs (NATUS) algorithm. This algorithm uses Bayesian statistics and Maximum Likelihood Estimation to first create hypotheses for various fetal genotypes and then test these expected genotypes against actual observed results from the sequencing data. It reports the actual disease state with calculated accuracy for each test result and fetal fraction in the sample.[26]

3. Methods

3.1 Aim of Review

The aim of the evidence review is to undertake a systematic review and cost-consequence assessment of cfDNA testing in the first trimester for T21, T18 and T13 in the UK. The decision questions for this project are shown in the box below:

1a) What is the accuracy of NIPT in predicting T21, T18 and T13 in pre-defined high risk (1:150) pregnant women following a combined test?

1b) How does changing the threshold for defining high risk following a combined test affect the accuracy of NIPT?

2 What is the most accurate primary prenatal screening tool for T21, T18 and T13 in the first trimester when NIPT and the combined test are compared in a general obstetric population?

3 What diagnostic accuracy is achievable by integrating NIPT into the combined test?

4 What is the rate of NIPT failure (number of inconclusive and excluded samples / total number of samples)?

5 What are the costs and consequences (cases detected, test-related miscarriages avoided) for the current NHS screening programme when NIPT is used

In sequence with the combined test (Question 1);
As a replacement for the combined test as the primary screen (Question 2);
In combination with (i.e. alongside) the combined test (Question 3)?

3.2 Identification and selection of studies

3.2.1 Search strategies for clinical effectiveness

Searches were conducted in PubMed, Ovid Medline, Ovid Embase and the Cochrane Library. The search strategy used a combination of search terms for cfDNA testing and trisomies, restricted to the English language. The date limits were 1997 to 9th February 2015. Thereafter, weekly auto-alerts in Medline and Embase were run until 1st April 2015 to check for any new articles.

Individuals and organisations were contacted for studies that were not freely available in the public domain. Also, ClinicalTrials.gov, WHO International Clinical Trials Registry Platform (ICTRP) Search Portal and meeting abstracts were searched for ongoing or recently completed trials.

The search strategy in Appendix 1 was developed for Medline and was adapted as appropriate for other databases. Two reviewers independently screened the titles and abstracts of all records identified by the searches using the inclusion criteria detailed below. Discrepancies were resolved by consensus or discussion with a third reviewer.

3.2.2 Inclusion and exclusion of relevant studies

Inclusion criteria

Studies which satisfied the following criteria were included:

Population	*For Questions 1a and 3*: Pregnant women with increased risk of one (or more) of T21, T18 or T13 according to a combined test outcome

	For Question 1b: Pregnant women with a risk score ranging from 1:3000 to 1:1 according to the combined test *For question 2*: Pregnant women in the general obstetric population
Target condition	Trisomies 21, 18 and 13 in the fetus, and also including translocation or mosaicism
Intervention	Non-invasive prenatal testing using cell-free (cf) DNA derived from maternal blood (serum, plasma, whole blood)
Reference standard	1) Genetic verification through amniocentesis, CVS, cordocentesis and fetal pathologic examination after abortion and 2) postnatal phenotypic assessment
Comparator	*For question 2:* Any "conventional" screening test *For questions 1, 3 and 4:* No comparator (no further management)
Outcome	Any type of diagnostic performance as an outcome measure including outcomes reported as: accuracy, detection rate, sensitivity and specificity, predictive values, likelihood ratios, diagnostic odds ratios, receiver operating characteristic (ROC) curves and numbers of true positive, false positive, true negative and false negative results. Data on inconclusive, indeterminate and excluded samples to determine test failure rates
Study type	All study types

Exclusion criteria

Studies were excluded for the following reasons:

Intervention	Studies reporting the quantification of fetal cells or DNA and using elevated levels of the whole fetal DNA (does not apply to T18 and elevated DNA levels are also associated with other conditions) or epigenetic markers (in development and not available as commercial test) as a screening/diagnostic tool
Study type	• Case-control studies with <15 cases and cohort studies with <50 pregnant women • Non-English studies • Letters, reviews, editorials and communications containing insufficient information on methods and no numerical outcomes data • Grey literature and conference abstracts

3.3 Review strategy

Two reviewers independently screened the titles and abstracts of all records identified by the searches for inclusion or exclusion and discrepancies in decisions were resolved through consensus or discussion with a third reviewer. Full copies of all studies deemed potentially relevant, were obtained and two reviewers independently assessed these for inclusion; any disagreements were resolved by consensus or discussion with a third reviewer. Records rejected at full text stage and reasons for exclusion were documented.

Data extraction strategy

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

3.4 Quality assessment strategy

The quality of diagnostic accuracy studies were assessed using a modified QUADAS-2.[27] The results of the quality assessment provided an overall description of the quality of the included studies. Quality assessment was undertaken by one reviewer and checked by a second reviewer, any disagreements were resolved by a third reviewer through discussion.

Modifications to the QUADAS-2 tool:

1. Addition of one signalling question to domain 2: Was the sample for the index test taken before the invasive test or at least 7 days after?

Rationale: Invasive testing increases the amount of fetal material in the maternal circulation which will affect the performance of cfDNA testing.

2. Addition of another signalling question to domain 2: Was the threshold value determined using an independent set of samples or was adjustment of the predefined threshold value avoided?

Rationale: While an explicit threshold can be reported by studies (e.g. z- score>3 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.

3. Removal of one signalling question from domain 4: Was there an appropriate interval between index test(s) and reference standard.

Rationale: T21, T18 and T13 are not progressive conditions; therefore, the time interval does not affect the performance of cfDNA testing. The timing of cfDNA testing in the pregnancy may affect its performance, but this is addressed separately and is unrelated to the timing of the reference standard.

4. Addition of one domain: Role of sponsor: Did the funding source/sponsor play no role in design of study, interpretation of results and publication?

Rationale: 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.

As recommended by the QUADAS-2 group, an overall quality score was not determined.[27] The results of each quality item will be presented in table and graph form.

3.5 Methods of analysis/synthesis

3.5.1 Overall approach

Depending on the available evidence, analyses were stratified according to condition (T21, T18 and T13). Study, population, test and outcome characteristics were summarised and compared in text and tables.

3.5.2 Statistical analysis of test accuracy studies

A meta-analysis of performance of cfDNA testing was performed. Papers were included in the narrative synthesis and not in the quantitative analysis (meta-analysis) for two reasons. Firstly, if it was clear from the methods section that more than one included paper used the same samples as one another to test for the same trisomies, then only one was included to prevent double counting. Secondly if the methods used in the study resulted in an inability to produce a complete 2x2 table.

For example, using results of either the reference standard or the index test as inclusion criteria, or if there was a lack of follow-up to establish trisomy status.

We included studies which provided a 2x2 table in the meta-analysis, even if they did not sample consecutive women, therefore the estimates from the meta-analysis may be subject to spectrum bias with atypical cases and controls selected. To determine whether spectrum bias was affecting results we included a sensitivity analysis including only cohort studies which enrolled a consecutive series of women or randomly sampled from a consecutive series of women.

We extracted data from the primary studies to obtain the four cell values of a diagnostic 2x2 table in order to calculate test accuracy measures: sensitivity (SN), specificity (SP), positive likelihood ratio (PLR), negative likelihood ratio (NLR) and diagnostic odds ratio (DOR) and corresponding 95% CI. We added a 0.5 cell correction to each cell where a zero was encountered, which was necessary for model convergence. We stratified test accuracy measures according to condition (T21, T18 and T13).

Sensitivity was defined as the proportion of positive test results among those with the target disease: specificity as the proportion of negative test results among those without the disease. In a clinical setting, likelihood ratios are considered useful. The diagnostic odds ratio (DOR), or the odds of a positive result in diseased individuals compared to the odds of a positive result in non-diseased individuals, combines both likelihood ratios and is a global measure of test performance.[28] A value of 1 would indicate that the test cannot discriminate between people with and without disease.

We pooled the sensitivity and specificity estimates using bivariate random-effects regression models, as recommended by the Cochrane Diagnostic Test Accuracy Working Group.[29] The bivariate model takes into consideration the potential trade-off between sensitivity and specificity by explicitly incorporating this negative correlation in the analysis.[30] We then used the results of the bivariate model to draw hierarchical summary receiver-operating characteristic (HSROC) curves.[31] The closer the curve is to the upper left-hand corner of the HSROC curve plot (i.e. sensitivity and specificity are both 100%), the better the overall accuracy of the test. We also calculated the area under the hierarchical summary receiver-operating characteristics curve (AUC).[32-34] An AUC between 0.9 and 1.0 indicates that cfDNA testing is highly accurate, while an AUC of 0.5 indicates poor diagnostic accuracy.[32-34]

Meta-analysis sub-groups 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 versus others), population risk (general, high-risk, others), population (twins versus others), first trimester (100% versus other) and publication year (2007-2013 versus 2014-2015). Cohort studies with consecutive sampling will be subject to less spectrum bias than other designs, and this was included to investigate whether spectrum bias was affecting test accuracy estimates. 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 to explore the effects of the zero cell correction. 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.[35] We then refitted the model leaving out any outliers and very influential studies.

We also constructed 3x2 tables to examine the influence of the number of test failures and indeterminate results on the pooled test accuracy estimates.[36] Test failures occur where cfDNA testing 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 2x2

tables reported, and this can lead to overestimates of test sensitivity and specificity.[37] We included all failures of cfDNA testing, 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 clearly be rectified by good quality assurance procedures (such as insufficient blood or dropped samples). For the 3x2 tables we considered the following three scenarios, all non-evaluable results: (1) considered to be positive results to reflect use of the cfDNA testing as triage for invasive testing,[37] (2) considered to be negative results to reflect use of cfDNA testing as an add-on to the combined test,[37] and (3) follow intention to diagnose principle to account for the first two approaches overestimating specificity and sensitivity respectively.[36] 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. A P value below 0.05 for sensitivity or specificity was used to determine whether there was a statistically significant difference in sensitivity, specificity, or both among the levels of a particular subgroup. In the linear meta-regression model, studies are the units of analysis. The DOR was the outcome (dependent) variable. The independent variables are the covariates that might be associated with the variability in the DOR (listed above). We used the meta-regression model to generate relative diagnostic odds ratios (RDOR) as the output.[28, 38] An RDOR is a ratio of two DORs. An RDOR of 1.0 indicates that a particular covariate does not affect the overall DOR. An RDOR >1.0 indicates that studies with a particular characteristic have a higher DOR than studies without this characteristic. For a RDOR <1.0, the reverse holds. We used Deeks' funnel plot asymmetry test to test for publication bias, with p-value <0.10 indicating significant publication bias.[39] All analyses were performed using Stata version 13 for Windows including the user written commands metandi, midas, metareg and mvmeta.[35, 40-42]

3.5.3 Test accuracy of cfDNA testing in different populations (Questions 1 and 2)

Where possible quantitative analysis (meta-analysis) was used to provide estimates for test accuracy across studies. Where this was not possible a narrative review was undertaken. Question 1a investigates the test accuracy of cfDNA testing if it were to be offered to people testing positive at the current first-trimester combined test in the UK at threshold 1:150. Therefore for question 1a the optimal study design was considered a test accuracy review using the UK combined test as inclusion criteria. Question 1b investigates the scenario that cfDNA testing is introduced following the combined test, but the threshold for the combined test is relaxed as further testing in this case will be non-invasive testing rather than invasive diagnostic testing. The question of the performance of cfDNA testing at various levels of risk as determined by the combined test is therefore examined here. The ideal study to answer this question is a study which includes women with a combined test result and subsequently performs cfDNA testing using a range of cut-offs to investigate cfDNA testing performance dependent on risk threshold. Alternatively, studies utilising different thresholds from the combined test as an inclusion criteria and reporting cfDNA testing performance at one specific set cut-off could be used to establish a profile of cfDNA testing performance over a range of risk thresholds. This approach is less desirable because of the heterogeneity of studies that will also impact on test performance. Question 2 investigates the possibility of implementing cfDNA testing as an alternative to the UK combined test. The ideal study design here compares the performance of the two tests in the same population. We aimed to synthesise such studies in a meta-analysis to compare the performance across a range of studies.

3.5.4 Combination of cfDNA testing and a combined test result (Question 3)

A narrative is provided of the available evidence to explore whether the combined screening test can improve the cfDNA test result when it was based on a low fetal fraction and whether a combined risk score could achieve diagnostic potential.

3.5.5 Analysis of the implications of cfDNA testing failure (Question 4)

Failure rates at different gestational ages were investigated in tables and a narrative summary and addressed the questions below:

- What is the failure rate of cfDNA testing at:
  - 8 weeks
  - 10 weeks
  - 12 weeks
  - 14 weeks
- What are the common causes of cfDNA testing failure?
- Are there risk factors that can predict cfDNA testing failure?
- How many repeat tests are successful?

3.6 Economic Model Methods

3.6.1 Model structure and assumptions

We constructed a decision tree to estimate the impact of alternative antenatal screening strategies on the annual performance of the England and Wales FASP. The decision tree is included in Appendix 3. The model was used to evaluate four strategies:

i) Combined test, followed by invasive testing offered to all those whose risk of trisomy is 1/150 or higher

ii) Combined test, followed by cfDNA testing offered to all whose risk of trisomy is 1/150 or higher, followed by invasive test offered to all those with a positive cfDNA testing result

iii) Combined test, followed by cfDNA testing offered to all whose risk of trisomy is 1/1000 or higher, followed by invasive test offered to all those with a positive cfDNA testing result

iv) cfDNA testing offered to all, followed by invasive test offered to all those with a positive cfDNA testing result

The parameter estimates used in the economic model were taken from our systematic review and meta-analysis, from the NHS fetal anomaly screening programme, and from targeted searches of the published literature.

For each strategy, we calculated the total annual numbers performed of each test, the total cost of screening, the numbers of trisomies detected by screening, the numbers of trisomies present at the time of screening but not detected by the programme, and the number of healthy pregnancies lost as a result of undergoing invasive testing. Probabilistic sensitivity analyses were carried out to derive uncertainty around predictions. The assumptions and parameter values/distributions used in the analysis are described below, and detailed in Table 2.

We assume that, based on the 2012 FASP, 448,676 women take up the offer of a combined test out of 723,913 annual pregnancies. To calculate this we used FASP data from 2012/13 for the number of women who take up the offer of the combined test. The number of women who are offered the combined test is not collected in the routine data, so we used data from the National Maternity Survey of recent mothers[43] where 89% reported being offered screening and 69% of these reported taking up the offer. The prevalence of T13, T18 and T21 in that population was assumed to be 32, 73 and 274 per 100,000 pregnancies respectively, based on diagnosed cases of each trisomy as

reported by Morris et al. (2014).[4] The probability of being identified as 'high-risk' by the combined test, as a function of trisomy status, for any given risk threshold, was taken from an analysis of FASP data (personal communication David Wright). For a risk threshold of 1/150, these were assumed to be 2.1% for non-trisomy pregnancies, and 72.1%, 83.6% and 86.3% for T13, T18 and T21 pregnancies respectively. With a risk threshold of 1/1,000, these probabilities become 9.6%, 86.6%, 94.3% and 94.6% respectively.

We assume 72.4% acceptance of the offer of an invasive test after a 'high-risk' combined test result, irrespective of the threshold used. We assume that the proportion of invasive tests involving amniocentesis is 47.8% and CVS is 52.2%. Both of these assumptions are derived from FASP 2012/13 data. We assume that the probability of test-induced pregnancy loss is 0.6% for amniocentesis and 0.7% for CVS, based on Australian Medicare data.[44] If offered cfDNA testing following a 'high-risk' combined test result, we assume that 90.7% will accept, 2.9% will choose an invasive test, and 6.4% will opt for no further testing. This is derived from antenatal clinic survey data.[45] The failure rate for cfDNA testing in the model is 10% if a woman has a trisomy and 3% if not. Estimates were taken from the largest study in the systematic review to report these.[46] All women in the model who experience cfDNA testing failure are offered a retest. We assume that the proportion of women accepting the retest, choosing an invasive test, or opting for no further testing are the same as for the choice made following the initial offer of cfDNA testing. We assume that the success rate of the retest is 46%.[47] The proportion of women who accept invasive testing after a positive cfDNA testing result is assumed to be 83%.[48] We assume this is true whether the positive result came from the initial test or a retest. For those women who proceed to invasive testing, we assume the proportion receiving amniocentesis is the same as when cfDNA testing is not included in the strategy. We assume that the cost of cfDNA testing is £232 from Hill et al. (2011),[49] using the calculation of laboratory based costs only and excluding the cost of giving results. For all other tests, we base our costs on the bottom-up costing estimates provided in the NHS FASP Decision Planning Tool,[50] which were inflated to 2014 prices using the hospital and community health services pay & prices index,[51] see Table 2.

Table 2 Parameter estimates used in the economic model

Description	Source	Estimate/ Value	Sensitivity analysis
Number of women who are offered the combined test per year in UK	NHS Fetal Anomaly Screening Programme (FASP) 2012/13 Rowe et al. (2008)[43]	723,913	NA
Number of women who accept the combined test per year in UK	FASP 2012/2013	448,676	NA
Prevalence of trisomy	Morris et al. (2014) [4]	T21: 274/100,000 T18: 73/100,000 T13: 32/100,000	NA
Proportion identified by combined test as having risk of trisomy >1/150	FASP (Dave Wright personal communication)	Non-trisomy pregnancies: 2.1% T21 pregnancies: 86.3% T18 pregnancies: 83.6% T13 pregnancies: 72.1%,	NA

Description	Source	Estimate/ Value	Sensitivity analysis
Proportion identified by combined test as having risk of trisomy >1/200	FASP (Dave Wright personal communication)	Non-trisomy pregnancies: 2.7% T21 pregnancies: 87.8% T18 pregnancies: 85.4% T13 pregnancies: 74.4%,	NA
Proportion identified by combined test as having risk of trisomy >1/1000	FASP (Dave Wright personal communication)	Non-trisomy pregnancies: 9.6% T21 pregnancies: 94.6% T18 pregnancies: 94.3% T13 pregnancies: 86.6%,	NA
Proportion identified by combined test as having risk of trisomy >1/2000	FASP (Dave Wright personal communication)	Non-trisomy pregnancies: 15.8% T21 pregnancies: 96.6% T18 pregnancies: 96.4% T13 pregnancies: 91.1%,	NA
Strategy i (combined test only) Proportion of those offered invasive test following combined test who accept	FASP 2012/13	72.4%	NA
Strategy i (combined test only) Proportion of invasive tests following combined test that are amniocentesis	FASP 2012/13	47.8%	NA
Probability of loss of pregnancy from amniocentesis	O'Leary et al. (2013)[44]	0.6% (0.5-0.7%)	1. Tabor et al. (2009)[52] 457/32852=1.4% (CI 1.3 to 1.5) 2. Akolekar et al. (2015)[53] 0.11% (CI –0.04 to 0.26)
Probability of loss of pregnancy from CVS.	O'Leary et al. (2013)[44]	0.7% (0.3-0.14%)	1. Tabor et al. (2009)[52] 589/31355=1.9% (CI 1.7 to 2.0)

Description	Source	Estimate/ Value	Sensitivity analysis
			2. Akolekar et al. (2015)[53] CVS 0.22 (CI –0.71 to 1.16)
Proportion of those above combined test risk threshold who accept cfDNA testing	Lewis et al. (2014)[45]	381/420=90.7%	Gil et al. (2015)[48] 147/260=57.3%
Proportion of those with risk of trisomy >1/150 who opt for invasive test rather than cfDNA testing	Lewis et al. (2014)[45]	12/420=2.9%	Gil et al. (2015)[48] 104/260=40%
Proportion of those with risk of trisomy >1/150 who opt for no further testing	Lewis et al. (2014)[45]	27/420=6.4%	Gil et al. (2015)[48] 7/260=2.7%
Success rate of cfDNA testing in women with trisomy	Norton et al. (2015)[46] (largest study that reports failure by trisomy)	54/60=90%	NA
Success rate of cfDNA testing in women without trisomy	Norton et al. (2015)[46] (largest study that reports failure by trisomy)	15787/16269= 97.0%	NA
Success rate of cfDNA testing retest	Willems et al. (2014)[47] (most relevant to UK population from our review – Netherlands and Belgium)	23/50=46%	NA
Proportion of invasive tests following cfDNA testing that are amniocentesis	FASP 2012/13 – assume same as after combined test	47.8%	NA
Proportion of those with positive cfDNA testing who accept invasive test	Gil et al. (2015)[48]	5/6=83.3%	NA
Sensitivity	Our Meta-analysis pooled estimates	T21 *97.1% (CI 95.5%-98.1%)* T18 *93.1% (CI 90.0%-95.3%)* T13 *82.7% (CI 74.7%-88.5%)*	Estimates using consecutively enrolled cohorts only: T21: *93.2% (CI 85.3%-97.1%)* T18: *86.8% (CI 59.1%-96.8%)* T13: *81.1% (CI 46.4%-95.5%)* Estimates using all

Description	Source	Estimate/ Value	Sensitivity analysis
			first trimester screened only: T21: 96.0% (CI 88.6%-98.7% T18: 90.7% (CI 74.0%-97.1%) T13: 62.7% (CI 33.5%-84.9%) Estimates excluding studies requiring zero cell corrections for TP and FN: T21: 99.4% (CI 98.9%-99.6%) T18: 97.4% (CI 95.8%-98.4%) T13: 97.4% (CI 86.2%-99.6%)
Specificity	Our Meta-analysis pooled estimates over 41 studies	T21 *99.8% (CI 99.7%-99.9%)* T18 *99.8% (CI 99.7%-99.9%)* T13 *99.8% (CI 99.7%-99.9%)*	Estimates using consecutively enrolled cohorts only: T21: 99.9% (CI 99.6%-99.9%) T18: 99.8% (CI 99.4%-99.9%) T13: 99.9% (CI 99.5%-99.9%) Estimates using all first trimester screened only: T21: 99.9% (CI 99.8%-99.9%) T18: 99.8% (CI 99.7%-99.9%) T13: 99.9% (CI 99.8%-100%) Estimates excluding studies requiring zero cell corrections for TP and FN: T21: 99.9% (CI 99.9%-100%) T18: 99.9% (CI 99.9%-100%)

Description	Source	Estimate/ Value	Sensitivity analysis
			T13: 99.9% (CI 99.9%-100%)
Total cost of combined test	NHS FASP decision planning tool 2011[50] (£26.10) inflated to 2014 prices	£27.11	NA
Cost of Amniocentesis	NHS FASP decision planning tool 2011[50] (£368.93) inflated to 2014 prices	£383.31	£515 Department of health reference costs for 2013/14
Cost of CVS	NHS FASP decision planning tool 2011[50] (£306.93) inflated to 2014 prices	£318.90	£515 Department of health reference costs for 2013/14
Cost of cfDNA testing	Hill et al. (2011)[49] lab costs UK	£232	£100-£500

Test prices have been inflated using the Hospital & community health services pay & prices index reported in PSSRU Unit Costs of Health and Social Care 2014, NA, not applicable; TP, true positive; FN, false negative; CVS, chorionic villus sampling; cfDNA, cell-free DNA

3.6.2 Sensitivity analysis

Our meta-analysis of cfDNA testing performance found evidence of publication bias. We therefore estimated the impact of basing test sensitivity on studies with consecutively enrolled cohorts only, to ameliorate this bias. We also explored the impact of basing the risk of test-related early miscarriage of pregnancy on published data from a Danish registry,[52] which reported higher risk than our reference case (1.4% and 1.9%), and from a recent meta-analysis[53] which reported lower risks (0.1% and 0.2%). Finally, we explored the impact of varying the cost per cfDNA testing.

4. Results

First we present the results for the full systematic review in three parts: a description of the included studies (Section 4.1 and 4.2), an assessment of the quality and risk of bias in those studies (Section 4.3), and a meta-analysis synthesising test performance (Section 4.4). The meta-analysis also examines how the test accuracy measured is affected by the study design, whether test failures are included, different test types, publication year, and the population screened (including twins vs singleton pregnancies, first trimester vs later in pregnancy, and high risk vs general obstetric population). Then we present a summary of results relevant to each of the five research questions. In Section 4.5.1 we describe the evidence from individual studies and from the meta-analysis combining studies regarding the accuracy of cfDNA testing in high risk groups (question 1a), and in Section 4.5.2 we extend this to evidence about the variation in accuracy of cfDNA testing in populations with different risk (question 1b). In Section 4.6 we present results comparing cfDNA testing performance in comparison to the combined test (question 2). In Section 4.7 results regarding integrating cfDNA testing into the combined test are presented (Question 3). Section 4.8 covers test failure rates, including analytic failures and indeterminate results (Question 4). Finally in Section 5 we present the health economic model (question 5).

Ethical approvals for both the clinical and cost effectiveness parts of this project were given by the University of Warwick Biomedical and Scientific Research Ethics Committee reference REGO-2015-1446.

4.1 Systematic review of included studies

Figure 2 provides the PRISMA flow diagram for the cfDNA testing clinical effectiveness review. Our searches identified 2012 unique records of which 108 full text articles were assessed. Of these, 56 articles were subsequently excluded using the pre-defined inclusion / exclusion criteria (see Appendix 2 for excluded studies with reason). This left 52 articles that met the inclusion criteria and were included in the narrative synthesis. Forty-one of these articles were also included in the quantitative synthesis.

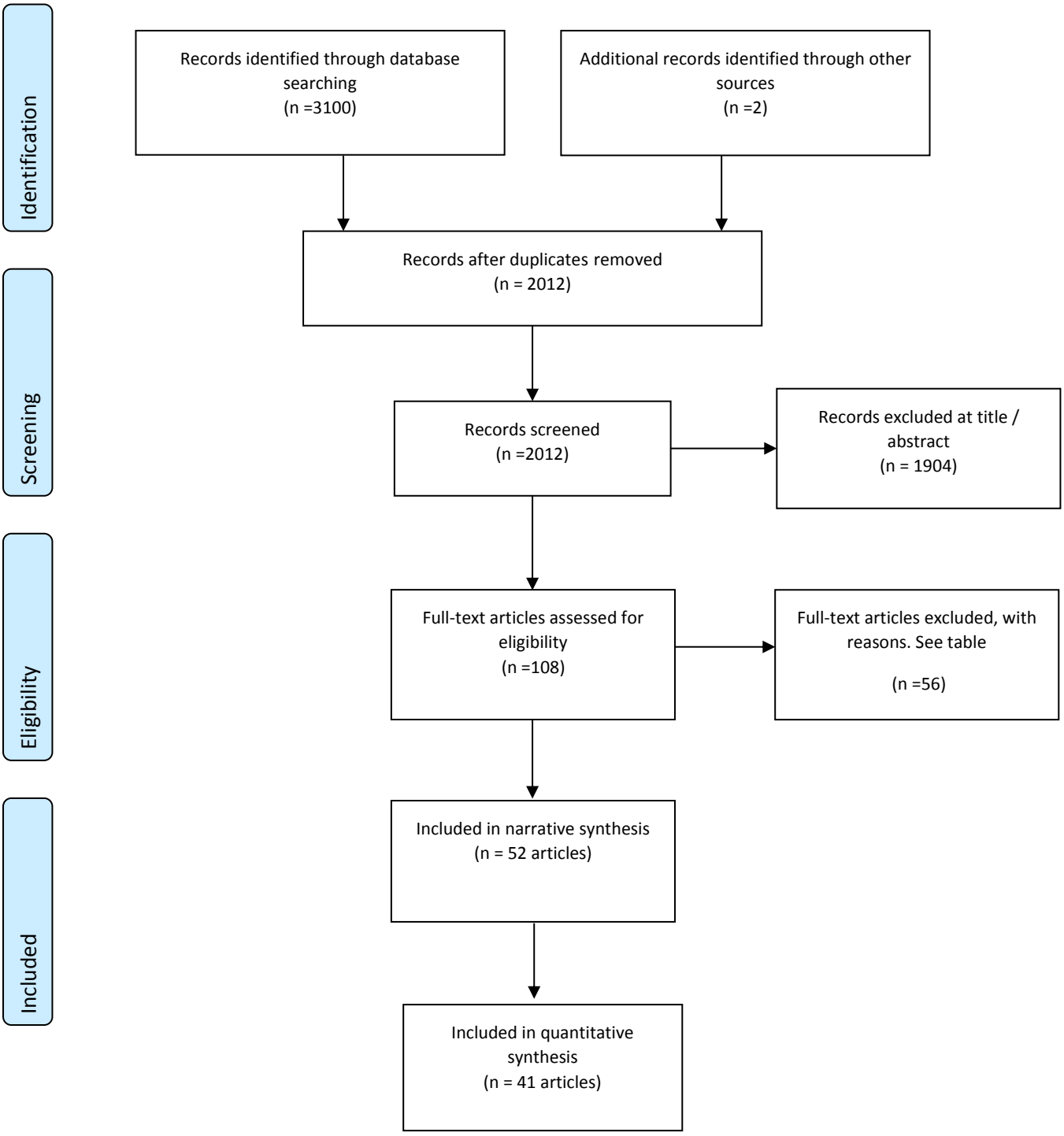


Figure 2 PRISMA Flow Diagram: cfDNA testing Clinical Effectiveness

4.2 Characteristics of included studies

4.2.1 Study design

Fifty-two publications, dating from 2007 to 2015, reported cfDNA testing results for the main autosomal trisomies in relation to fetal karyotype or newborn phenotype and fulfilled our inclusion criteria (see Table 12 and Figure 3).

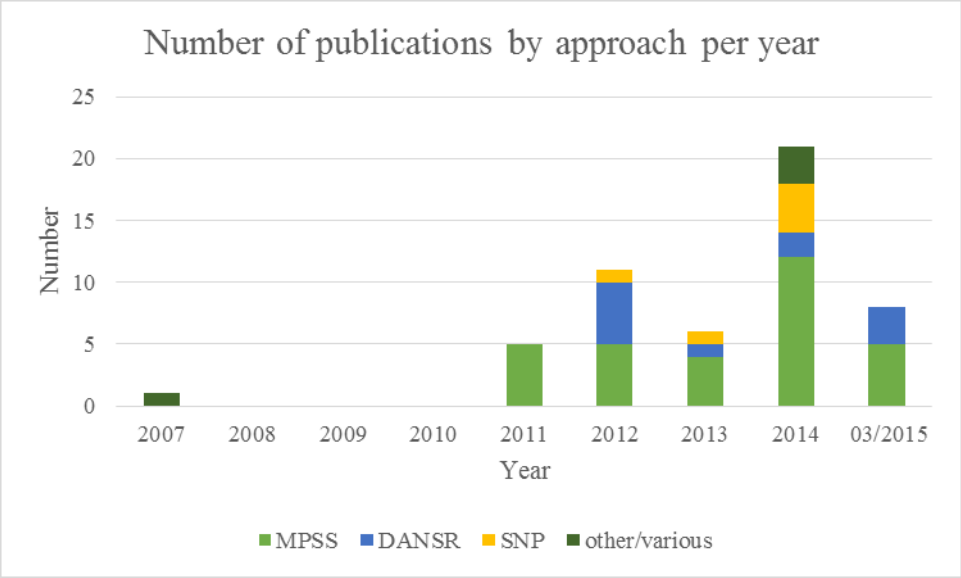


Figure 3 Number of publications by approach per year (up to March 2015)
MPS, massively parallel sequencing (whole genome); DANSR, digital analysis of selected regions (targeted sequencing); SNP, single-nucleotide polymorphism-based approach

Fifteen studies were case-controls with ≥ 15 combined trisomy cases,[25, 26, 54-66] 35 were cohort studies with ≥ 50 participants,[22-24, 46, 47, 67-96] and two studies had an unclear observational design.[97, 98]

Three publications reported on samples of two nested case-control studies[61, 62] and from one study of unclear design[98] selected from the same cohort. Another three studies had an overlap of samples analysed.[56, 57, 65]

Nine included studies used a retrospective study design of stored samples from pregnancies with known outcome[54, 60, 65, 69, 76] or performed a retrospective database review of prospectively collected and analysed clinical samples.[86, 89, 92, 93] Two studies used both archived and prospectively collected maternal plasma samples,[56, 57] while the remaining 41 studies used prospectively collected maternal samples only.[22-26, 46, 47, 55, 58, 59, 61-64, 66-68, 70-75, 77-85, 87, 88, 90, 91, 94-98]

The majority of studies (n=47) only addressed the accuracy of cfDNA testing by comparing cfDNA testing outcomes to a reference standard (fetal karyotype or newborn phenotype),[22, 24-26, 47, 54-66, 68-75, 77-89, 91-98] whereas five studies compared the performance of cfDNA testing for the main autosomal trisomies with that of standard screening methods (serum markers with or without ultrasound markers) in the general population.[23, 46, 67, 76, 90] Only three of them[46, 76, 90] compared cfDNA testing performance head-to-head to the first-trimester combined test.

4.2.2 Population

Thirty-six studies were performed in women with singleton pregnancies only,[22, 23, 25, 26, 46, 54-57, 59, 61-68, 70, 76-80, 82, 85-88, 90, 91, 93, 94, 96-98] five studies included women with both singleton and twin pregnancies[24, 75, 81, 83, 89], three studies enrolled pregnant women with twin gestations only,[69, 71, 84] and in the remaining eight studies it was unclear if participating women were pregnant with one or more fetuses.[47, 58, 60, 72, 73, 75, 92, 95]

The majority of studies used samples from primarily high-risk pregnant women with a range of different indications for invasive testing (n=31),[22, 25, 54, 55,371, 57-63, 65, 66, 70, 71, 73, 75, 77, 78, 80, 82, 83, 87, 89, 91, 93, 95-98] six studies were performed in the general obstetric population,[23, 46, 67, 76, 85, 90] 12 studies included pregnant women with mixed risk factors,[24, 26, 47, 64, 68, 74, 79, 81, 84, 86, 88, 94] and in three studies, the prior aneuploidy risk was unclear.[69, 72, 92]

Seven studies included pregnant women in the first trimester only,[46, 54, 69, 76, 77, 90, 96] one study included 90% first-trimester pregnancies,[73] and all other studies (n=44) included pregnant women with an unstated, later or broader gestational age window.[22-26, 47, 55-68, 70-72, 74, 75, 78-89, 91-95, 97, 98]

4.2.3 Testing strategies

Three main testing strategies were pursued by the majority of studies (see Figure 1). These were genome-wide massively parallel shotgun sequencing (MPSS, n=31 studies),[22-24, 55-58, 60-63, 65-68, 70-75, 80, 81, 87, 89, 91, 93-96, 98] targeted massively parallel sequencing (DANSR, n=11 studies),[25, 46, 47, 54, 64, 69, 76, 78, 82, 84, 90] and single-nucleotide polymorphism (SNP)-based methods (n=6).[26, 59, 77, 79, 86, 88] Three studies, which were performed in real clinical settings, offered more than one cfDNA testing approach to their patients.[83, 85, 92] The study by Dhallan et al.(2007)[97] explored a methodology that did not fall into any of the three approaches.

In three of the 52 studies, some of the maternal blood samples for cfDNA testing were obtained after invasive testing.[26, 67, 79] In 44 studies it was explicitly reported or strongly assumed from the study design and setting that samples for cfDNA testing were collected before the invasive testing.[22-24, 46, 47, 54-58, 60-63, 65, 66, 68-71, 73-78, 80-96, 98] In five studies, it was unclear if maternal blood sampling for cfDNA testing was performed before or after the invasive procedure.[25, 59, 64, 72, 97]

The majority of studies reported cfDNA testing performance for T21 (n=51) [22-26, 46, 47, 54, 55, 57-98] and T18 detection (n=46);[22-26, 46, 47, 54-56, 59, 60, 62-65, 67-81, 83-96, 98] 38 studies investigated non-invasive detection of T13.[22-24, 26, 46, 47, 55, 56, 59, 60, 62, 63, 65, 67, 69, 70, 72-75, 77, 79-81, 83-94, 96, 98] Thirty-five studies investigated all three trisomies (see Table 12 for details).

4.2.4 Testing strategies of included studies

Random MPSS

Random MPSS-based counting approaches differed in many points between the 31 studies: the level of multiplexing (monoplex up to 24-plex), the sequencing platform used (Illumina Genome Analyzer GAIIx, Illumina HiSeq, or semiconductor sequencing with Ion Proton Sequencer), the denominator to derive a chromosome ratio of the chromosome of interest to a reference, the human reference genome used for alignment, the alignment algorithm and number of mismatches allowed, whether or not a guanine-cytosine (GC)-content adjustment of the raw data was undertaken, and the threshold used for autosomal aneuploidy detection (see Table 13).

Four studies used a normalised chromosome value (NCV) as threshold,[55, 63, 67, 70] twelve studies a mean and standard deviation (SD)-based z-score using a euploid reference group,[23, 56-58, 60, 65, 66, 73, 75, 81, 87, 96] and seven studies a median and median absolute deviation (MAD)-based z-score (also called robust z-score) computed on the basis of the chromosomal variations observed in the sample set without using an external reference set.[22, 61, 62, 80, 89, 91, 98] Six studies used a binary hypothesis t-test

and logarithmic likelihood ratio between the two t-tests for autosomal aneuploidy detection.[24, 68, 71, 72, 74, 94] Two studies using the MPSS counting approach did not report details on the methodology used.[93, 95] Yu et al. (2014)[65] analysed paired-end sequencing data also by DNA size using a size-based z-score.

Targeted MPSS

The digital analysis of selected regions (DANSR) approach was performed in all 11 studies by one single laboratory and differed in the number of non-polymorphic loci on chromosomes 13, 18, and 21 which were amplified and sequenced (384 loci[64] versus 576 loci[25, 46, 47, 54, 69, 76, 78, 82, 84, 90] on each chromosome) as well as in the threshold used to distinguish between euploid and aneuploid samples (see Table 14). A z-score of chromosome proportions with external reference set was used in the initial exploration of the DANSR assay,[25, 64] whereas later studies applied the Fetal-fraction Optimized Risk of Trisomy Evaluation (FORTE) algorithm, which incorporates fetal fraction and the prior risk of aneuploidy associated with the subject's maternal and gestational age, and needs no external reference data set.[25, 46, 47, 54, 76, 78, 82, 90] For assessment of risk for trisomies in twin pregnancies, the lower fetal fraction contribution of the two fetuses was used by the FORTE algorithm.[69, 84]

Single-nucleotide polymorphism (SNP)-based approach

This approach was carried out in all six studies by one single laboratory and differed in respect to whether a paternal genetic sample was available, how many SNPs covering chromosomes 13, 18, 21, X, and Y were amplified and sequenced (11,000[26, 59] or 19,488 SNPs[59, 77, 79, 85, 86, 88]), and whether the Parental Support (PS)[26] or Next-generation Aneuploidy Test Using SNPs (NATUS) algorithm[59, 77, 79, 85, 86, 88] was used for copy number classification (Table 15). PS and the advanced NATUS version of the algorithm both use Bayesian statistics to determine the relative likelihood of each possible monosomic, disomic, and trisomic fetal genotype at measured loci (each considered as a separate hypothesis). The hypothesis with the maximum likelihood is selected as the copy number and fetal fraction.

Other

Dhallan et al. (2007)[97] amplified multiple DNA-SNPs on chromosomes 21 and 13 isolated from the maternal plasma, maternal buffy coat, and paternal buffy coat by PCR and quantified only SNPs showing a unique fetal allele in the maternal plasma (Table 16). The ratio of the unique fetal allele signal to the combined maternal and fetal allele signal was calculated and mean log ratios of SNPs on chromosomes 13 and 21 were compared by a two-tailed Student's t-test.

More than one approach

Three studies (Table 17), which were performed in real clinical settings, offered more than one cfDNA testing approach to their patients[83, 85] or performed confirmatory cytogenetic testing on clinical samples with prior cfDNA testing.[92] cfDNA testing used were commercially available and based on genome-wide MPSS, targeted sequencing, or SNPs, respectively. The cfDNA testing methodology was not further described in these three publications.

4.3 Quality Appraisal

4.3.1 Methodological quality of included studies

The methodological quality of the 52 included studies, assessed by QUADAS-2[27] is summarised in Figure 4, Figure 5 and Table 18. These illustrate the risk of bias regarding the five assessed domains (patient selection, index test, reference standard, flow and timing, and the role of sponsor). Concerns regarding applicability of the studies in terms of study participants, index test and reference standard were assessed separately for diagnostic and screening context.

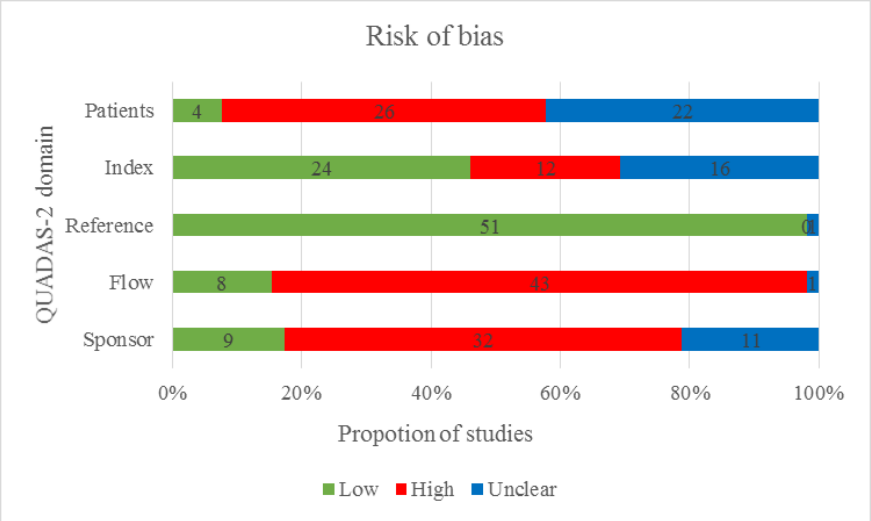


Figure 4 Proportion of studies with low, high or unclear risk of bias

4.3.2 Risk of bias

A study was considered to be at low-risk of bias regarding *patient selection* if a consecutive or random sample of patients was enrolled, a case-control design was avoided, and the exclusions from the study were described and appropriate (< 10%).

The risk of selection bias (Figure 4) was judged to be low in only four studies.[22, 74, 82, 93] Twenty-two cohort studies were classified as unclear risk of bias because it was not explicitly stated that patients (rather than samples) were recruited randomly or consecutively, and exclusions from the study were not further described.[23, 24, 46, 67-69, 71-73, 75-79, 81, 84, 87, 88, 90, 94-96] Twenty-six studies were classified as being high-risk of bias because a case-control design was used,[25, 26, 54-66] patients were not recruited randomly or consecutively[47, 70, 83, 85, 86, 89, 91, 92, 97] or study exclusions were higher than 10%.[80, 98]

A study was considered to be at low-risk of bias regarding the *index test* if laboratory personnel were blinded to reference standard results, if the blood sample for the index test was taken before or at least seven days after invasive testing, and the threshold was explicitly pre-specified and (if appropriate) determined using an independent set of samples.

Risk of bias was judged as low in 17 studies.[23, 46, 55, 57, 61, 63, 70, 73, 75, 76, 78, 80-82, 86, 89, 96] An additional seven studies with blinding to reference standard, blood sampling prior invasive testing, but insufficient information on the threshold used, were classified as low-risk of bias when commercial non-invasive prenatal tests were used.[47, 83-85, 88, 90, 91] Risk of bias was judged as unclear for 16 studies with inadequately reported methodology, mostly missing information on the threshold and/or reference sample set used.[24, 54, 56, 59, 67-69, 71, 72, 74, 77, 92-95, 97] Twelve studies were classified as high-risk of bias because lab personnel were not blinded to reference standard results,[22, 26, 64, 65, 87, 98] blood was taken

less than seven days after invasive testing,[79] threshold was adjusted or not explicitly pre-specified [22, 25, 26, 58, 62, 64-66, 87] or samples used for the reference set were not independent.[60, 87]

In one study by Quezada et al. (2015)[90], the conduct of the comparator test (first-trimester combined test) was considered to be at high-risk of bias as ultrasound measurement of nuchal translucency was performed unblinded to cfDNA testing results.

The risk of bias regarding the *reference standard* was considered to be low if the reference standard was likely to correctly classify trisomies 21, 18 and 13. We accepted prenatal or postnatal karyotyping or phenotypic newborn assessment as appropriate reference standard. Only one study[69] was classified as unclear risk of bias as the reference standard used was not reported; in all other studies there was little concern about bias.[22-26, 46, 47, 54-68, 70-98]

In the fourth domain, relating to *flow and timing*, a study was considered to be at low-risk of bias if all patients in the study received a result from both cfDNA testing and reference standard and all patients were included in the analysis. This was true for only eight of 52 studies.[56, 71-73, 81, 87, 97, 98] In one study,[60] test failures and exclusions from analysis were not reported and it was therefore classified as unclear risk of bias. In all other studies (n=43), the risk of bias was judged as high because cfDNA testing was not performed or failed to provide a result and/or exclusion of inconclusive results and/or no complete follow-up of birth outcomes.[22-26, 46, 47, 54, 55, 57-59, 61-70, 74-80, 82-86, 88-96] Fourteen of the 52 included studies (26.9%) reported cfDNA testing and reference standard results (fetal karyotype or newborn phenotype) for less than 85% of the original study population.[24, 26, 46, 64, 66, 68, 70, 83, 84, 86, 88, 89, 91, 94]

The risk of bias regarding the *role of sponsor* was 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 cfDNA testing or cytogenetic tests and/or other conflicts of interest (i.e. patents, stock or stock options) were declared.

In only nine of the 52 included studies, the risk of bias regarding the role of sponsor was judged as low.[23, 60, 66, 75, 83, 87, 88, 93, 95] Risk of bias was classified as unclear in eleven studies [54, 69, 73, 74, 76, 77, 81, 90, 91, 94, 96] and as high in all other studies (n=32).[22, 24-26, 46, 47, 55-59, 61-65, 67, 68, 70-72, 78-80, 82, 84-86, 89, 92, 97, 98]

In summary, risk of bias was high in most studies with 35 of 52 studies (67.3%) considered high risk in two or more domains, 15 studies (28.8%) in one domain, and only two studies (3.8%) were judged as low or unclear risk of bias in all five domains. No study scored low risk of bias in all domains. Figure 4 shows that the study flow (exclusions from analysis) and the role of the sponsor presented areas with the greatest risk of bias introduced into the studies. Another issue was incomplete or unclear reporting, particularly of the patient selection process and the conduct of the index test, which is reflected in 22 (42.3%) and 16 (30.8%) of 52 publications scoring an unclear risk of bias in these two domains, respectively.

4.3.3 Concerns regarding applicability

As the research questions aim to address cfDNA testing performance in the first trimester and in comparison with the first-trimester combined test, applicability of *included patients* should be regarded as low if <80% of women were recruited in the first trimester. In the context of screening,

cfDNA testing should have been carried out in pregnant women without prior aneuploidy screening (general obstetric population) and also include multiple gestations.

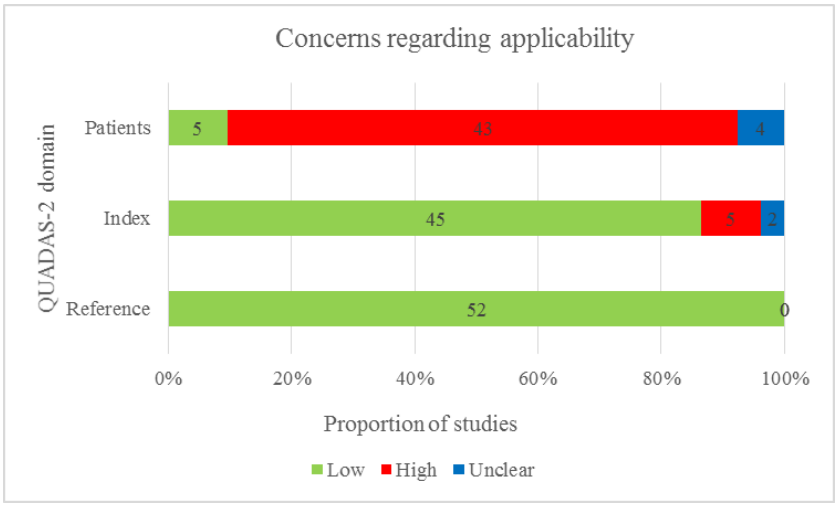


Figure 5 Proportion of studies with low, high and unclear concerns regarding applicability

Only five studies included at least 80% of pregnant women in their first trimester and were classified as giving rise to low levels of concern regarding applicability of patient spectrum (Figure 5); all investigated cfDNA testing accuracy only.[54, 69, 73, 77, 96] Another three studies included first-trimester pregnant women without prior screening for head-to-head comparison of cfDNA testing and first-trimester combined test performance, but were judged as high concerns regarding patient applicability as all three excluded multiple pregnancies.[46, 76, 90] In the two other studies addressing cfDNA testing performance in comparison to standard screening, concerns regarding patient applicability were high as <80% of participants were in their first trimester of pregnancy and only women with singleton pregnancies were included.[23, 67] In four studies, concerns regarding patient applicability were unclear as patient characteristics, particularly percentage of first-trimester pregnancies, were not reported.[56, 66, 91, 93] In all other studies (n=38) we had high levels of concern regarding applicability to cfDNA testing introduction in the first trimester as <80% of participants were less than 14 weeks of gestation.[22, 24-26, 47, 55, 57-65, 68, 70-72, 74, 75, 78-89, 92, 94, 95, 97, 98]

In terms of the *index test*, applicability of studies comparing cfDNA testing to a standard screening test was classed as high concern in two studies as different screening tests to the first-trimester combined test were used in >20% of cases.[23, 67]

Concerns regarding applicability of studies evaluating cfDNA testing accuracy only was classified as high in three studies[26, 59, 97] as the index test included paternal genetic samples for all cfDNA testing analyses. In two studies,[92, 95] information about the index test methodology was not available or incomplete and was judged as unclear concerns regarding applicability. In all other studies evaluating cfDNA testing accuracy only (n=45), the concerns about index test applicability were low.[22, 24, 25, 46, 47, 54-58, 60-66, 68-91, 93, 94, 96, 98]

Applicability concerns regarding the *reference standard* were low in all studies.

In summary, there were significant concerns regarding applicability of the included patient spectrum to cfDNA testing introduction in the first trimester, as most studies had significant parts of their

populations tested in the second or third trimester. Another concern for the head-to-head comparison of cfDNA testing performance with the first-trimester combined test was that all studies addressing this review question excluded multiple pregnancies and two of five studies used different screening tests to the combined test in 61% and 100% of participants, respectively.

4.4 Meta-analysis of cfDNA testing accuracy studies

4.4.1 Inclusions and Exclusions from Meta-analysis

In total eight studies were excluded from the meta-analysis,[47, 60, 64, 70, 86, 89, 92, 95] with an additional 3 papers excluded because they used the same samples as included studies, see Table 3.[61, 65, 98] The 8 excluded studies that could not give a complete 2x2 table for the meta-analysis were made up of 5 studies where index test results were used as study inclusion criteria or incomplete follow up of cfDNA testing negative cases,[47, 70, 86, 92, 95] one study where there was no reference standard for cfDNA testing negative cases,[64] one study which used cross-validation to estimate sensitivity and specificity so did not produce a 2x2 table,[60] and one study which only had reference standard data for cases where it had been reported back to the company by physicians on an ad-hoc basis.[89]

Table 3 Articles included in the narrative analysis but excluded from the quantitative synthesis

Paper	Reason	Detail of reason
Dar et al. (2014)[86]	Incomplete 2x2 table	Index test results used as inclusion criteria so incomplete 2x2 table
Fang et al. (2015)[95]	Incomplete 2x2 table	Unclear reporting, incomplete follow up of cfDNA testing negative cases
Futch et al. (2013)[70]	Incomplete 2x2 table	Index test results used as inclusion criteria so incomplete 2x2 table
Jensen et al. (2013)[98]	Same samples as other study so removed to prevent double counting	Re-uses some of the same samples as Palomaki et al. (2012)[62] Excluded to prevent double counting
Liao et al. (2014)[60]	Incomplete 2x2 table	Used cross-validation method to evaluate sensitivity and specificity so no 2 x 2 table
McCullough et al. (2014)[89]	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
Palomaki et al. (2011)[61]	Same samples as other study so removed to prevent double counting	Uses the same samples as Palomaki et al. (2012)[62] Excluded to prevent double counting
Sparks et al. (2012)[64]	Incomplete 2x2 table	No reference standard for cfDNA testing negative cases
Wang et al. (2014)[92]	Incomplete 2x2 table	Index test results used as inclusion criteria so incomplete 2x2 table
Willems et al. (2014)[47]	Incomplete 2x2 table	Incomplete follow up of cfDNA testing negative cases
Yu et al. (2014)[65]	Same samples as other study so removed to prevent double counting	Re-uses the same samples as Chen et al. (2011)[56] and Chiu et al. (2011)[57] Excluded to prevent double counting

4.4.2 Meta-analysis results for Trisomy 21 (Down syndrome)

Overall accuracy of cfDNA testing

Figure 6 shows a forest plot of the sensitivity and specificity with 95% confidence intervals for each study included in the bivariate meta-analysis. The summary receiver operating characteristics plot (Figure 7) shows the summary sensitivity and specificity and the 95% confidence and prediction regions. As shown in Figure 6 and Figure 7, specificity seemed to be more consistent across studies than sensitivity. Overall, for all cfDNA testing studies for the detection of trisomy 21 (n = 41), the pooled sensitivity from bivariate random-effects regression was 97.1% (CI 95.5% to 98.1%) and the pooled specificity was 99.8% (CI 99.7% to 99.9%). This corresponds to a positive likelihood ratio of 501 (CI 302 to 831) and a negative likelihood ratio of 0.029 (CI 0.019 to 0.045). A positive likelihood ratio value of 501 suggests that a positive cfDNA testing increases the odds of a woman having a T21 pregnancy by 501 times. The negative likelihood ratio of 0.029 suggests that a negative cfDNA testing result reduces the odds of a woman having a T21 pregnancy by 34 times. The HSROC curve was positioned near the desirable upper left corner and the AUC was 0.992 (CI 0.989 to 0.994), which indicated a high level of overall accuracy. The results of Deeks' funnel plot asymmetry test showed that the slope coefficient was associated with p value of 0.0001, suggesting a high likelihood of publication bias (Figure 8).

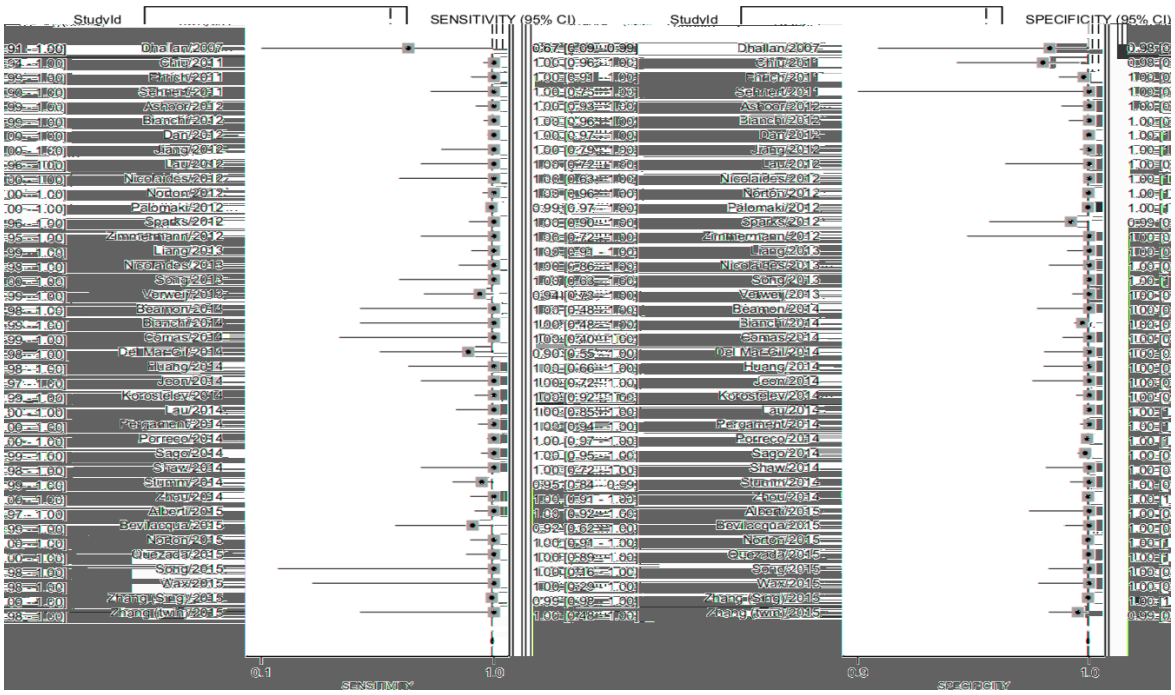


Figure 6 Individual and pooled sensitivity and specificity for cfDNA testing for the detection of trisomy 21 for included studies

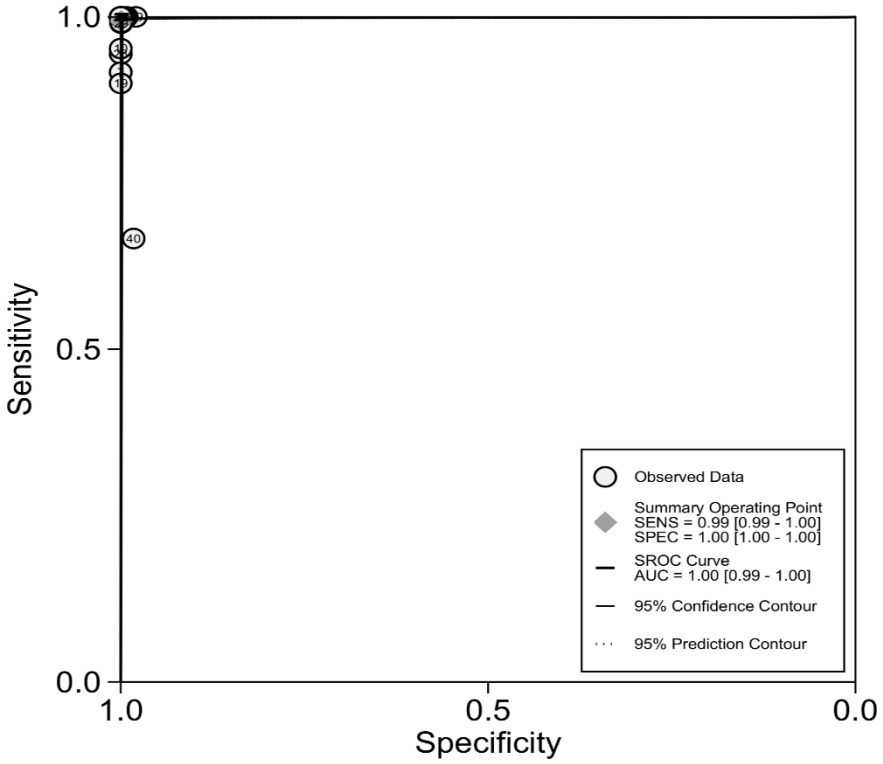


Figure 7 Hierarchical summary receiver-operating characteristics curve plot of cfDNA testing for the detection of trisomy 21

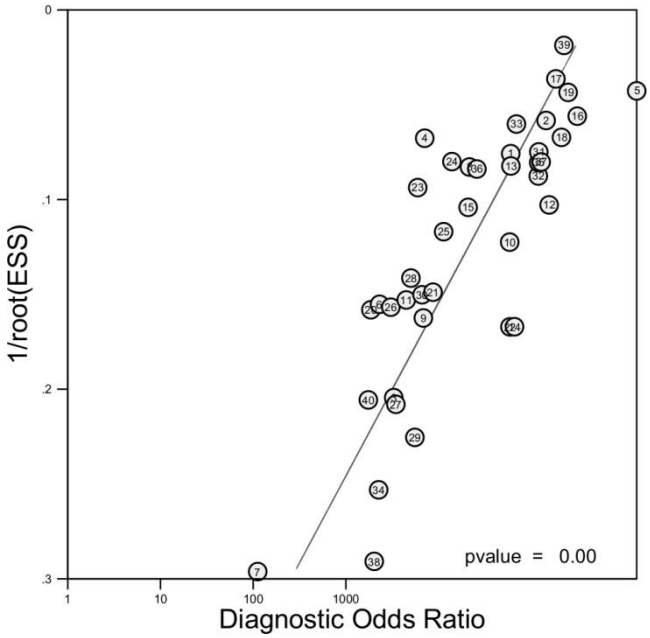


Figure 8 Deeks' funnel plot for asymmetry test for cfDNA testing for the detection of trisomy 21 for included studies

Table 4 Accuracy estimates from sensitivity and subgroup analyses of the included studies by different study characteristics

	Trisomy 21				Trisomy 18				Trisomy 13		
Variables	N	Sensitivity (95% CI)	Specificity (95% CI)		N	Sensitivity (95% CI)	Specificity (95% CI)		n	Sensitivity (95% CI)	Specificity (95% CI)
All studies[*]	41	0.971 (0.955-0.981)	0.998 (0.997-0.999)		37	0.931 (0.900-0.953)	0.998 (0.997-0.999)		30	0.827 (0.747-0.885)	0.998 (0.997-0.999)
Sensitivity analyses											
Excluding tp/fn = 0[¥]	40	0.994 (0.989-0.996)	0.999 (0.999-1.000)		33	0.974 (0.958-0.984)	0.999 (0.999-1.000)		24	0.974 (0.862-0.996)	0.999 (0.999-1.000)
Excluding outliers[‡]	38	0.971 (0.955-0.981)	0.999 (0.998-0.999)		36	0.927 (0.892-0.952)	0.998 (0.997-0.999)		28	0.802 (0.711-0.870)	0.999 (0.998-0.999)
Test failures											
Assuming all +ve	41	0.976 (0.962-0.984)	0.999 (0.998-0.999)		37	0.935 (0.905-0.956)	0.976 (0.966-0.983)		30	0.842 (0.768-0.897)	0.975 (0.963-0.983)
Assuming all -ve	41	0.938 (0.920-0.953)	0.998 (0.997-0.999)		37	0.894 (0.853-0.925)	0.998 (0.997-0.999)		30	0.785 (0.709-0.846)	0.998 (0.997-0.999)
Intention to diagnosis	41	0.956 (0.942-0.966)	0.970 (0.961-0.978)		37	0.906 (0.868-0.934)	0.975 (0.965-0.982)		30	0.802 (0.726-0.861)	0.975 (0.963-0.983)
Indeterminate results											
Assuming all +ve	41	0.971 (0.956-0.981)	0.998 (0.997-0.999)		37	0.932 (0.904-0.953)	0.998 (0.997-0.999)		30	0.835 (0.760-0.890)	0.998 (0.997-0.999)
Assuming all -ve	41	0.969 (0.954-0.980)	0.998 (0.997-0.999)		37	0.921 (0.886-0.946)	0.998 (0.997-0.999)		30	0.808 (0.719-0.874)	0.998 (0.997-0.999)
Intention to diagnosis	41	0.971 (0.957-0.981)	0.998 (0.997-0.999)		37	0.924 (0.893-0.948)	0.998 (0.997-0.999)		30	0.811 (0.724-0.875)	0.998 (0.997-0.999)
Subgroup analyses											
Study design											
Cohort	5	0.932 (0.853-0.971)	0.999 (0.996-0.999)		4	0.868 (0.591-0.968)	0.998 (0.994-0.999)		4	0.811 (0.464-0.955)	0.999 (0.995-0.999)
Others	36	0.975 (0.960-0.984)	0.998 (0.997-0.999)		33	0.934 (0.903-0.956)	0.998 (0.997-0.999)		26	0.929 (0.745-0.889)	0.998 (0.997-0.999)
Population risk											
General	6	0.959 (0.874-0.987)	0.999 (0.998-1.000)		5	0.841 (0.597-0.949)	0.998 (0.997-0.999)		5	0.597 (0.298-0.838)[§]	0.999 (0.998-0.999)
High	23	0.970 (0.948-0.983)	0.998 (0.994-0.998)		20	0.927 (0.889-0.953)	0.997 (0.995-0.999)		14	0.864 (0.763-0.926)	0.996 (0.992-0.998)
Others	12	0.974 (0.940-0.989)	0.999 (0.998-0.999)		12	0.943 (0.870-0.976)	0.999 (0.998-0.999)		11	0.839 (0.673-0.929)	0.999 (0.998-1.000)
Population											
Others	37	0.976 (0.962-0.984)	0.998 (0.997-0.999)		33	0.939 (0.913-0.958)	0.998 (0.997-0.999)		27	0.835 (0.755-0.893)	0.998 (0.997-0.999)
Twins	4	0.894 (0.749-0.960)[§]	0.996 (0.989-0.998)		4	0.656 (0.298-0.895)[§]	0.998 (0.992-1.000)		3	0.616 (0.165-0.929)	0.998 (0.992-0.999)
First trimester											
100%	7	0.960 (0.886-0.987)	0.999 (0.998-0.999)		6	0.907 (0.740-0.971)	0.998 (0.997-0.999)		5	0.627 (0.335-0.849)	0.999 (0.998-1.000`)
Others	34	0.972 (0.955-0.982)	0.997 (0.996-0.999)		31	0.934 (0.901-0.956)	0.998 (0.997-0.999)		25	0.851 (0.772-0.905)	0.998 (0.997-0.999)
Test types											
DANSR	9	0.958 (0.898-0.983)	0.999 (0.997-0.999)		7	0.940 (0.860-0.976)	0.998 (0.996-0.999)		4	0.579 (0.241-0.856)	0.999 (0.998-1.000)
MPSS	25	0.978 (0.963-0.987)	0.998 (0.997-0.999)		24	0.933 (0.895-0.958)	0.998 (0.997-0.999)		20	0.851 (0.763-0.909)	0.998 (0.997-0.999)
SNP	5	0.972 (0.880-0.994)	0.998 (0.991-0.999)		5	0.894 (0.684-0.971)	0.998 (0.993-0.999)		5	0.870 (0.647-0.960)	0.998 (0.992-0.999)
Publication year											
2007-2013	18	0.977 (0.958-0.988)	0.998 (0.995-0.999)		15	0.954 (0.919-0.975)	0.998 (0.996-0.999)		10	0.848 (0.704-0.929)	0.996 (0.990-0.998)
2014-2015	23	0.963 (0.934-0.980)	0.998 (0.998-0.999)		22	0.896 (0.823-0.942)	0.998 (0.997-0.999)		20	0.805 (0.687-0.929)	0.999 (0.998-0.999)

CI – confidence interval

*-with zero cell corrections;

¥excluded studies sensitivity not estimable (T21 – Hall et al. (2014);[59] T18 – Comas et al. (2014),[85] Hall et al. (2014);[59] Zhang (twins) et al. (2015);[24] T13 – Sehnert et al. (2011),[63] Beamon et al. (2014),[83] Comas et al. (2014),[85] Bevilacqua et al. (2015),[84] Wax et al. (2015),[93] Zhang (twins) et al. (2015);[24]

‡excluded outliers (T21 – Dhallan et al. (2007),[97] Chen et al. (2011),[56] Sparks et al. (2012);[25] T18 – Chen et al. (2011);[56] T13 – Chen et al. (2011), [56] Palomaki et al. (2012)[62]

§p-value for subgroup differences < 0.05 (statistically significant)

Table 5 Results of univariable meta-regression analysis

	Trisomy 21				**Trisomy 18**				**Trisomy 13**		
Variables	RDOR (95% CI)	p-value	Explained variance (%)		RDOR (95% CI)	p-value	Explained variance (%)		RDOR (95% CI)	p-value	Explained variance (%)
Study design			0.0				0.0				0.0
Cohort vs others	2.28 (0.25-21.01)	0.457			1.91 (0.15-24.52)	0.610			0.98 (0.09-10.19)	0.989	
Population risk											0.0
General	1 (reference)		5.5		1 (reference)		22.1		1 (reference)		
High	0.43 (0.05-3.35)	0.408			1.49 (0.19-11.64)	0.697			1.01 (0.19-6.10)	0.990	
Others	1.32 (0.15-12.04)	0.799			5.59 (0.61-50.88)	0.122			2.83 (0.41-19.69)	0.281	
Population			17.3				7.4				0.0
Others vs twins	0.12 (0.01-1.12)	0.062			0.12 (0.01-1.65)	0.108			0.24 (0.014-4.11)	0.310	
First trimester			0.0				0.0				0.0
100% vs others	0.79 (0.12-5.42)	0.806			2.20 (0.330-16.26)	0.431			1.97 (0.35-1.098)	0.427	
Test types			3.9				0.0				0.0
MPS vs others	1.96 (0.46-8.32)	0.350			1.91 (0.41-8.91)	0.398			1.68 (0.39-7.14)	0.469	
Publication year			0.0				0.0				0.0
2007-2013 vs 2014-2015	0.86 (0.20-3.61)	0.830			0.83 (0.19-3.60)	0.793			2.30 (0.57-9.35)	0.232	

RDOR – Ratio of Diagnostic Odds Ratio; CI – confidence interval

Investigation of heterogeneity

Sensitivity and subgroup analyses were conducted to investigate heterogeneity in sensitivity, and to a lesser degree, in specificity (Table 4). A priori sensitivity analysis excluding studies with zero cells for true positive and false negative results resulted in similar but slightly higher estimates of test accuracy. Including test failures yielded slightly lower pooled test accuracy estimates. On the basis of the Cook's distance, we found the following studies to be the most influential in the meta-analysis (in descending order): Chiu et al. (2011), Stumm et al. (2014), Sparks et al. (2012), Dhallan et al. (2007), Dan et al. (2012), and Zhang et al. (2015, singleton pregnancies) (Figure 9a).[22, 24, 25, 57, 68, 97] Of these, Chiu et al. (2011), Sparks et al. (2012), and Dhallan et al. (2007) were identified as outliers having the highest standardised residuals for specificity (Figure 9b).[25, 57, 97] After refitting the model and leaving these studies out, we found no significant change in sensitivity and specificity. The result of subgroup analysis showed statistically significant lower pooled sensitivity among twins than singleton pregnancies (89.4% versus 97.6%, p=0.016). Table 5 shows the results of the univariable meta-regression analyses. None of the study-level covariates included in the meta-regression analyses were found to be a statistically significant source of heterogeneity (all p value >0.05). However, population, population risk and test type explained 17%, 6% and 4% of the variability in the pooled test accuracy estimates between studies respectively.

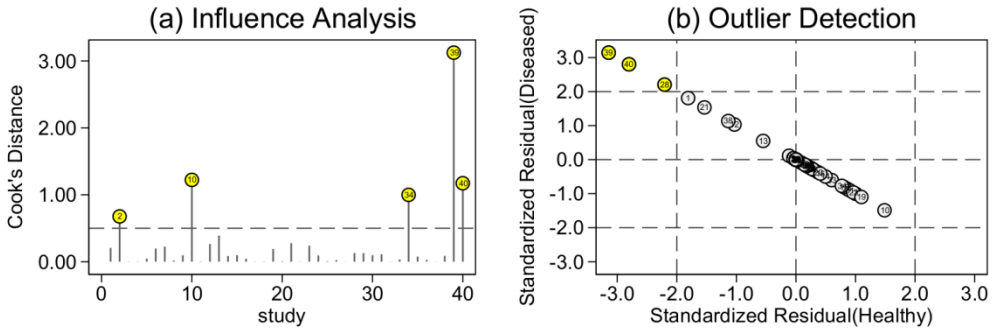


Figure 9 Goodness-of-fit (a) influence analysis (b) outlier detection for cfDNA testing for the detection of trisomy 21 for included studies

Note: (2) Zhang (Singleton pregnancies) 2015, (10) Stumm 2014, (28) Sparks 2012, (34) Dan 2012, (39) Chiu 2011, and (40) Dhallan 2007.[22, 24, 25, 57, 68, 97]

Sensitivity and specificity may be underestimated due to use of a zero cell correction. Estimates of sensitivity and specificity without the zero cell correction for TP and FN may be overestimates, due to high risk of bias in included studies and evidence of publication bias.

4.4.3 Meta-analysis results for Trisomy 18 (Edwards syndrome)

Overall accuracy of cfDNA testing

Figure 10 shows a forest plot of the sensitivity and specificity with 95% confidence intervals for each study included in the bivariate meta-analysis. The summary receiver operating characteristics plot (Figure 11) shows the summary sensitivity and specificity and the 95% confidence and prediction regions. As shown in Figure 10 and Figure 11, specificity seemed to be more consistent across studies than sensitivity. Overall, for all cfDNA testing studies for the detection of trisomy 18 (n = 37), the pooled sensitivity from bivariate random-effects regression was 93.1% (CI 90.0% to 95.3%) and the pooled specificity was 99.8% (CI 99.7% to 99.9%). This corresponds to a positive likelihood ratio of 514 (CI 316 to 835) and a negative likelihood ratio of 0.070 (CI 0.047 to 0.100). A positive likelihood ratio value of 514 suggests that a positive cfDNA test increases the odds of a woman having a T18 pregnancy by 514 times. The negative likelihood ratio of 0.07 suggests that a negative

cfDNA testing result reduces the odds that the woman has a T18 pregnancy by 14 times. The HSROC curve was positioned near the desirable upper left corner and the AUC was 0.989 (CI 0.984 to 0.992), which indicated a high level of overall accuracy. The results of Deeks' funnel plot asymmetry test showed that the slope coefficient was associated with a p value of 0.0001, suggesting a high likelihood of publication bias (Figure 12).

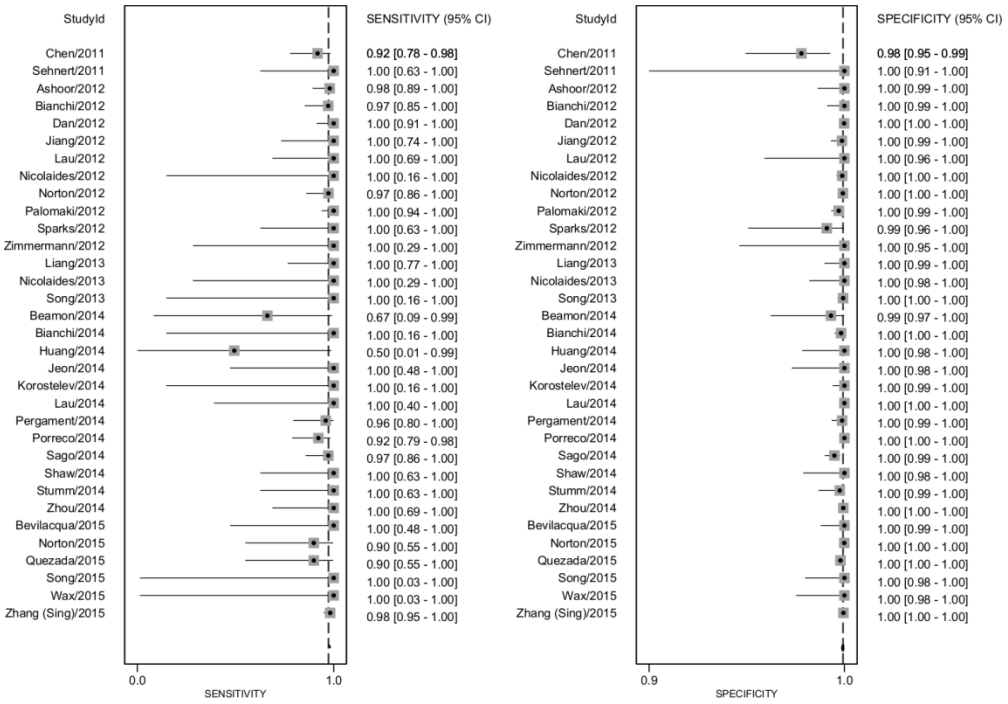


Figure 10 Individual and pooled sensitivity and specificity for cfDNA testing for the detection of trisomy 18 for included studies

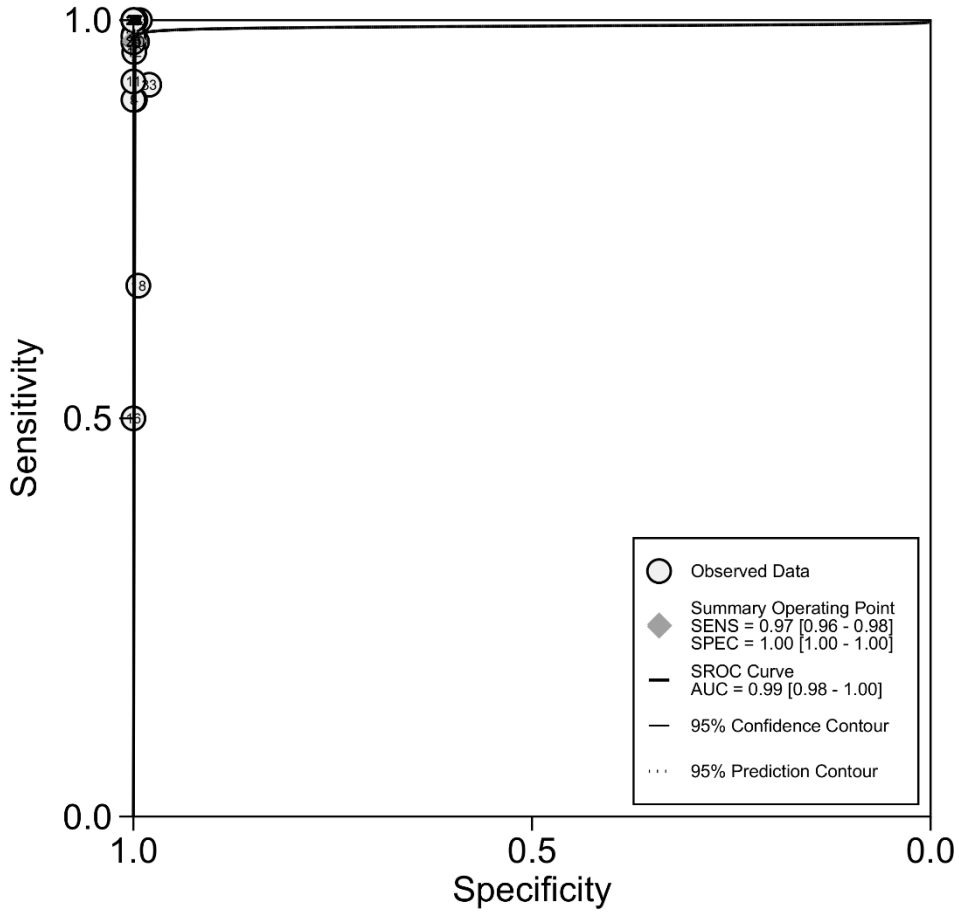


Figure 11 Hierarchical summary receiver-operating characteristics curve plot of cfDNA testing for the detection of trisomy 18

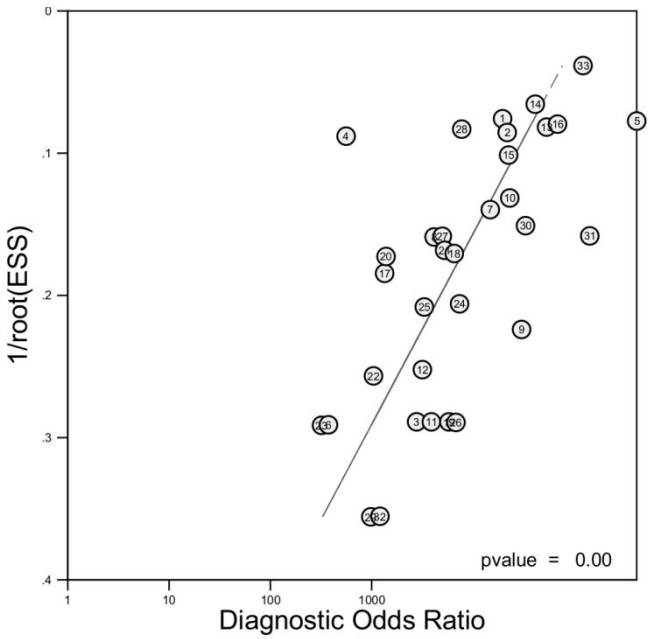


Figure 12 Deeks' funnel plot for asymmetry test for cfDNA testing for the detection of trisomy 18 for included studies

Investigation of heterogeneity

Sensitivity and subgroup analyses were conducted to investigate heterogeneity in sensitivity, and to a lesser degree, in specificity (Table 4). A priori sensitivity analysis excluding studies with zero cells for true positive and false negative results resulted in similar but slightly higher estimates of test accuracy. Including test failures yielded slightly lower pooled test accuracy estimates. On the basis of the Cook's distance, we found the following studies to be the most influential in the meta-analysis (in descending order): Chen et al. (2011), Porreco et al. (2014), Norton et al. (2015) and Palomaki et al. (2012) (Figure 13a).[46, 56, 62, 80] Of these, only Chen et al. (2011)[56] was identified as an outlier having the highest standardised residuals for specificity (Figure 13b). After refitting the model and leaving these studies out, we found no significant change in sensitivity and specificity. The pooled sensitivity tended to be lower among general population, compared with populations at high risk (84.1% versus 92.7%), however, this difference did not reach a statistically significant level (p=0. 066). The result of subgroup analysis showed statistically significant lower pooled sensitivity among twins than singleton pregnancies (65.6% versus 93.9%, p=0.009). Similarly, the pooled sensitivity analysis tended to be lower among recent studies published between 2014 and 2015 compared with studies published between 2007 and 2013 (89.6% versus 95.4%), however this difference did not reach a statistically significant level (p=0.056). As shown in Table 5, none of the study-level covariates included in the meta-regression analyses were found to be a statistically significant source of heterogeneity (all p value >0.05). However, population risk and population explained 22% and 7% of the variability in the pooled test accuracy estimates between studies respectively.

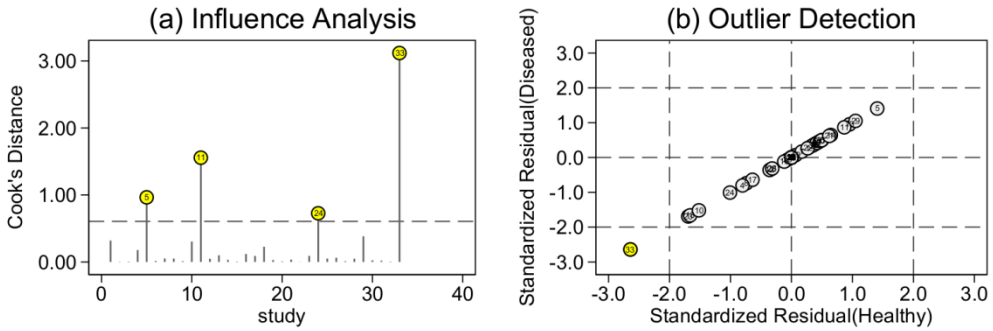


Figure 13 Goodness-of-fit (a) influence analysis (b) outlier detection for cfDNA testing for the detection of trisomy 18 for included studies

Note: (5) Norton et al. (2015), (11) Porreco et al. (2014), (24) Palomaki et al. (2012), and (33) Chen et al. (2011). [46, 56, 62, 80]

Again, sensitivity and specificity may be underestimated due to use of a zero cell correction. Estimates of sensitivity and specificity without the zero cell correction may be overestimates, due to high risk of bias in included studies and evidence of publication bias.

4.4.4 Meta-analysis results for Trisomy 13 (Patau syndrome)

Overall accuracy of cfDNA testing

Figure 14 shows a forest plot of the sensitivity and specificity with 95% confidence intervals for each study included in the bivariate meta-analysis. The summary receiver operating characteristics plot (Figure 15) shows the summary sensitivity and specificity and the 95% confidence and prediction regions. As shown in Figure 14 and Figure 15, specificity seemed to be more consistent across studies than sensitivity. Overall, for all cfDNA testing studies for the detection of trisomy 13 (n = 30), the pooled sensitivity from bivariate random-effects regression was 82.7% (CI 74.7% to 88.5%) and the pooled specificity was 99.8% (CI 99.7% to 99.9%). This corresponds to a positive likelihood ratio

of 512 (CI 301 to 870) and a negative likelihood ratio of 0.174 (CI 0.117 to 0.258). A positive likelihood ratio value of 512 suggests that a positive cfDNA test increases the odds of a mother having a T13 pregnancy by 512 times. The negative likelihood ratio of 0.174 suggests that a negative cfDNA testing result reduces the odds that the mother has a T13 pregnancy by 6 times. The HSROC curve was positioned near the desirable upper left corner and the AUC was 0.982 (CI 0.975 to 0.987), which indicated a high level of overall accuracy. The results of Deeks' funnel plot asymmetry test showed that the slope coefficient was associated with a p value of 0.045, suggesting a high likelihood of publication bias (Figure 16).

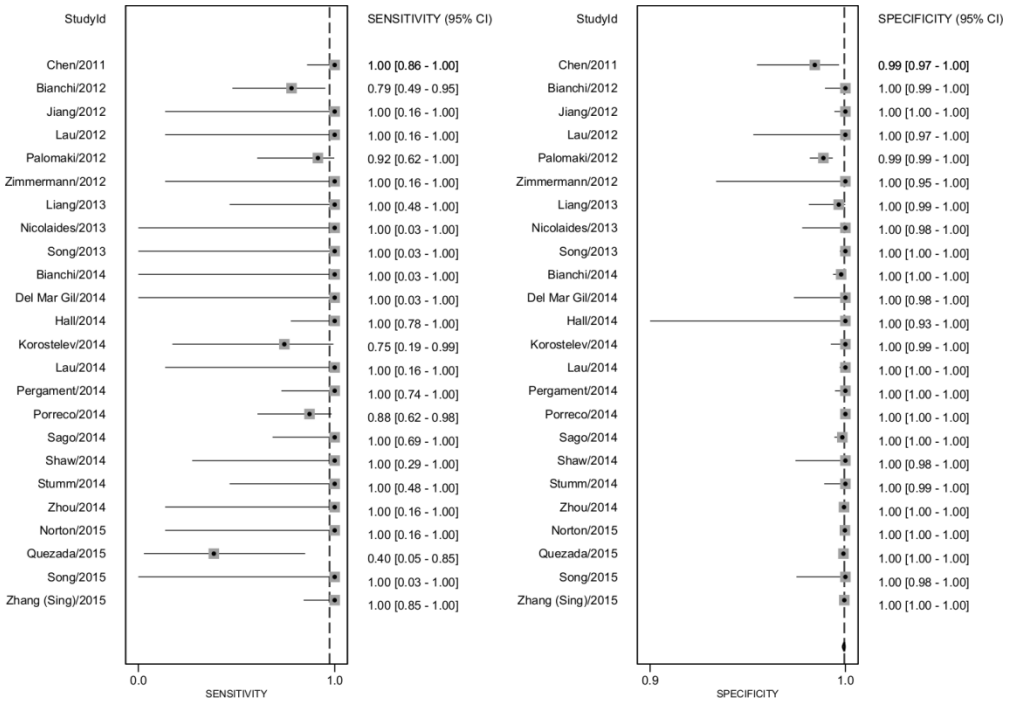


Figure 14 Individual and pooled sensitivity and specificity for cfDNA testing for the detection of trisomy 13 for included studies

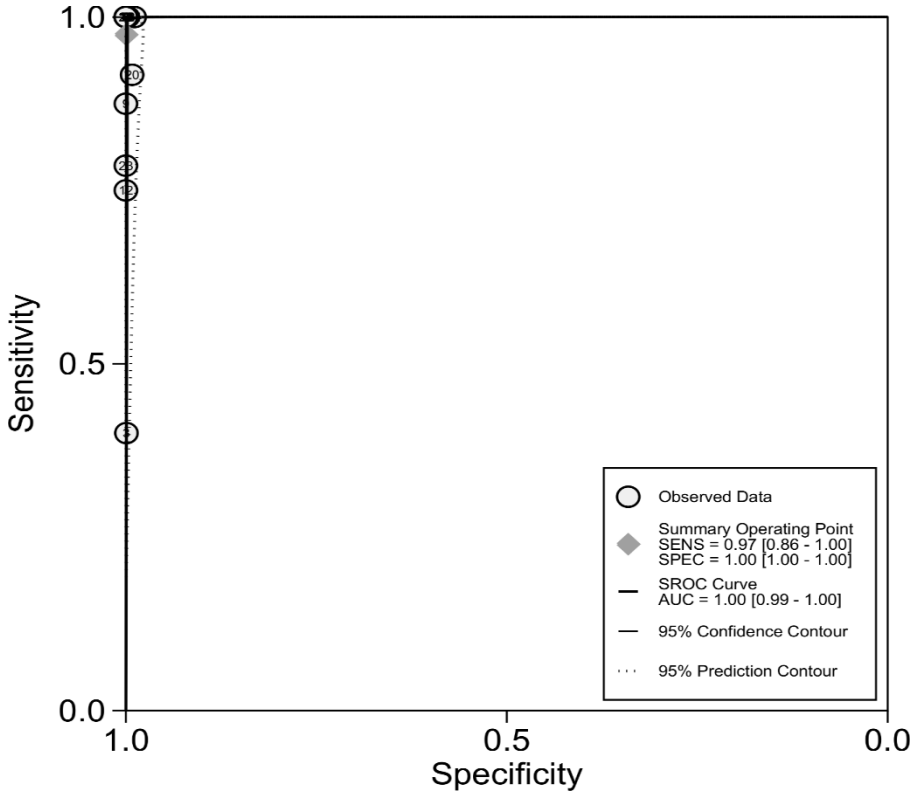


Figure 15 Hierarchical summary receiver-operating characteristics curve plot of cfDNA testing for the detection of trisomy 13

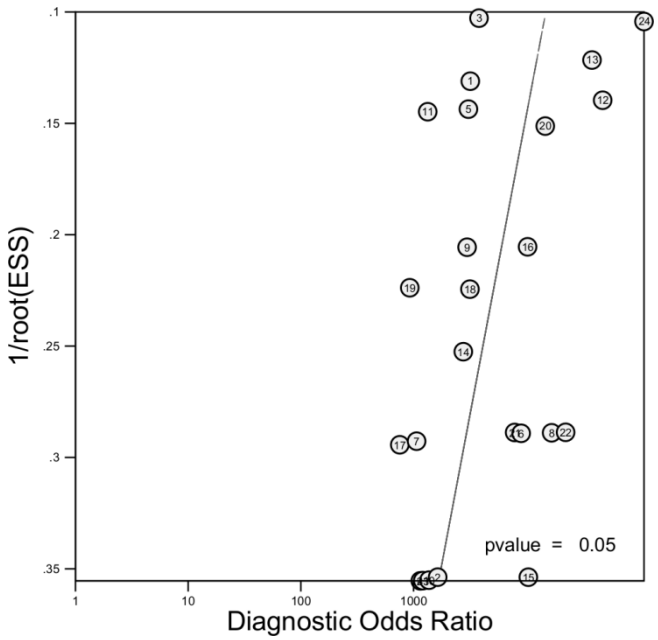


Figure 16 Deeks' funnel plot for asymmetry test for cfDNA testing for the detection of trisomy 13 for included studies

Investigation of heterogeneity

Sensitivity and subgroup analyses were conducted to investigate heterogeneity in sensitivity, and to a lesser degree, in specificity (Table 4). A priori sensitivity analysis excluding studies with zero cells for true positive and false negative results resulted in similar but slightly higher estimates of test accuracy. Including test failures yielded slightly lower pooled test accuracy estimates. On the basis of the Cook's distance, we found the following studies to be the most influential in the meta-analysis (in descending order): Palomaki et al. (2012), Quezada et al. (2015), Chen et al. (2011), and Porreco et al. (2014) (Figure 17a).[56, 62, 80, 90] Of these, only Palomaki et al. (2012)[62] and Chen et al. (2011)[56] were identified as an outlier having the highest standardised residuals for specificity (Figure 17b). After refitting the model and leaving these studies out, we found no significant change in sensitivity and specificity. The pooled sensitivity was statistically significantly lower among the general population, compared with populations at high risk (59.7%% versus 86.4%, p=0. 044). Similarly, the pooled sensitivity analysis tended to be lower among recent studies that recruited pregnant women in the first trimester compared with other trimesters (62.7% versus 85.1%), however this difference did not reach a statistically significant level (p=0. 066). As shown in Table 5, none of the study-level covariates included in the meta-regression analyses were found to be a statistically significant source of heterogeneity (all p value >0.05).

Figure 14:

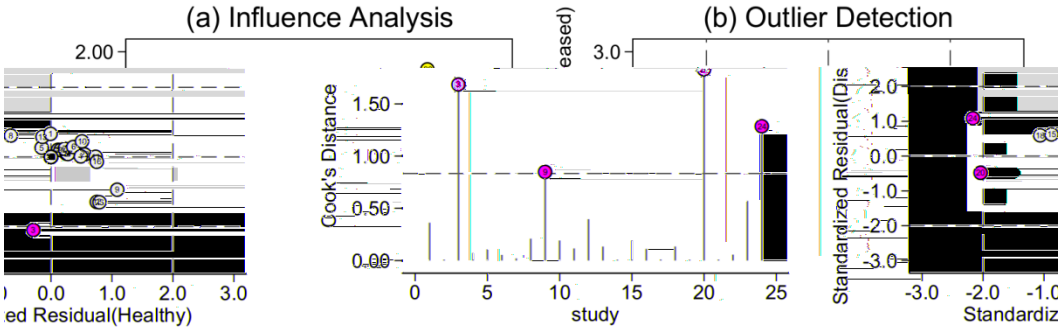


Figure 17 Goodness-of-fit (a) influence analysis (b) outlier detection for cfDNA testing for the detection of trisomy 13 for included studies

Note: (3) Quezada 2015, (9) Porreco 2014, (20) Palomaki 2012, and (24) Chen 2011.[56, 62, 80, 90]

As for the other two trisomies, sensitivity and specificity may be underestimated due to use of a zero cell correction. Estimates of sensitivity and specificity without the zero cell correction may be overestimates, due to high risk of bias in included studies and evidence of publication bias.

4.4.5 Interpreting Meta-Analysis results in different populations

The test accuracy estimates from the meta-analysis may be underestimates due to use of the zero cell correction to enable model convergence. Estimates without the zero cell correction may be overestimates due to risk of bias of included studies and publication bias. We applied estimates of sensitivity and specificity without zero cell corrections for TP and FN to a theoretical cohort of 10,000 pregnant women from a high risk population as described in Table 6. Population prevalence was determined as the median prevalence for the studies included in high risk groups. In a high risk population with prevalence 3% for T21, 1.5% for T18 and 0.5% for T13 the positive predictive value was 91% for T21, 84% for T18 and 87% for T13, indicating that even when using the most favourable meta-analysis results which may be overestimates, cfDNA testing must not be considered a diagnostic test. Positive predictive values for all three trisomies would be significantly lower in a general obstetric population in comparison to a high risk population, and cfDNA must not be

considered a diagnostic test. In the general obstetric population using cfDNA as a diagnostic test for T18 would result in an incorrect diagnosis more than 50% of the time.

Table 6 Summary of findings applied to high risk populations

Condition	Population	Summary accuracy % (96% CI)	Median prevalence	Outcomes	Positive predictive value	Probability of test negative when actually positive	Implications	Quality and comments
Down syndrome	High Risk (10,000 pregnancies)	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 out of 10,000 pregnancies will be affected by Down syndrome. Of these 324 will be detected and 9 missed by cfDNA. 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.	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.
Edwards syndrome	High Risk (10,000 pregnancies)	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 out of 10,000 pregnancies will be affected by Edwards syndrome. Of these 140 will be detected and 11 missed by cfDNA. 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	High Risk (10,000 pregnancies)	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 out of 10,000 pregnancies will be affected by Patau syndrome. Of these 47 will be detected and 3 missed by cfDNA. 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 high risk population. Estimates of sensitivity and specificity are from the meta-analysis sub-groups for studies in general obstetric populations without zero cell correction for TP and FN. Prevalence here refers to prevalence in the population tested, and not risk scores from another test which would be affected by both pre-test probabilities and accuracy of that test. TP, True Positive; FP, False Positive; TN, True Negative; FN, False Negative

Summary

A total of 41 studies were included in the meta-analysis for T21, 37 for T18 and 30 for T13. We found that pooled sensitivity was 97.1% (CI 95.5% to 98.1%) for T21, 93.1% (CI 90.0% to 95.3%) for T18, and 82.7% (CI 74.7% to 88.5%) for T13. The pooled specificity was 99.8% (CI 99.7% to 99.9%) for all three trisomies. These may be underestimates due to use of a zero cell correction to enable model convergence. Without the zero cell correction for TP and FN we found that pooled sensitivity was 99.4% (CI 98.9% to 99.6%) for T21, 97.4% (CI 95.8% to 98.4%) for T18, and 97.4% (CI 86.2% to 99.6%) for T13. The pooled specificity was 99.9% (CI 99.9% to 100%) for all three trisomies. These may be overestimates due to publication bias, risk of bias in included studies, and inclusion of second and third trimester pregnancies where there is a higher fetal fraction enhancing cfDNA test performance. We applied estimates of sensitivity and specificity without zero cell corrections to a theoretical cohort of 10,000 pregnant women from a high risk population. Population prevalence was determined as the median prevalence for the studies included in high risk groups. In a high risk population with prevalence 3% for T21, 1.5% for T18 and 0.5% for T13 the positive predictive value was 91% for T21, 84% for T18 and 87% for T13, indicating that even when using the most favourable meta-analysis results which may be overestimates, cfDNA testing must not be considered a diagnostic test.

We found that sensitivity was significantly lower in twin than singleton pregnancies. Sensitivity was reduced by 9% for T21, 28% for T18 and 22% for T13. Sensitivity was also lower by 1% for T21, 3% for T18 and 22% for T13 when studies which included only first trimester pregnancies were compared to all other studies with pregnancies across all trimesters, but this was not statistically significant. Pooled sensitivity was also lower by 2% for T21, 9% for T18, and 26% for T13 in studies in the general obstetric population, in comparison to studies in high risk populations, but these estimates were not statistically significant either.

Using Deeks' funnel plots we found publication bias in the included studies so test accuracy may be overestimated, because publications with positive results may have been more likely to have been published than those with less promising results. In an analysis including only study designs which were less likely to result in overestimation due to spectrum bias (cohort studies with consecutive sampling) we found sensitivity estimates were lower by 4% for T21, 6% for T18 and 2% for T13, but these differences were not statistically significant.

4.5 Accuracy of cfDNA testing in high risk groups and in different risk groups (Question 1a and 1b)

The accuracy of cfDNA testing in every included study is shown in Table 19. This includes numbers of true positive, false positive, true negative and false negative results, where reported. Sensitivity, specificity, positive predictive value and negative predictive value are included as reported in the papers, or calculated using information provided in the papers. Positive and negative predictive values are dependent on population prevalence and so are only applicable to the prevalence of trisomies in the individual study.

4.5.1 Accuracy of cfDNA testing in high risk groups (Question 1a)

Research question 1a is: What is the accuracy of cfDNA testing in predicting T21, T18 and T13 in pre-defined high risk (1:150) pregnant women following a combined test? No studies were identified that carried out cfDNA testing in pregnant women with a high risk threshold typically used in the UK screening programme (1:150) as estimated by the first trimester combined screening test. The majority of studies (n=31) used samples from primarily high-risk pregnant women with a range of different indications which included cut-offs either different to 1:150 or unreported. Twelve studies included pregnant women with mixed (high and low) risk levels and three studies did not clearly report the risk of trisomy of the included population. Six studies were undertaken in populations

resembling the general obstetric population. Analysis of the performance variation in groups at different levels of risk is reported in Section 4.6.

As no study directly answered research question 1a we included all studies in the meta-analysis (Section 4.4) to estimate the overall performance of cfDNA testing in the populations with varying risk indicators and level of risk. Then we undertook sub-group analyses to investigate the impact of population risk on cfDNA testing performance by classifying the studies according to the risk level of the included populations (general obstetric population, high-risk, and others (mixed, unknown risk)).

When only the studies which recruited high risk populations were included the test accuracy estimates were very similar to those for all studies. In high risk populations defined in a variety of ways pooled sensitivity (without zero cell corrections for TP and FN) was 99.7% (CI 96.3% to 99.9%) for T21, 96.5% (CI 93.7% to 98.1%) for T18 and 95.3% (CI 86.3% to 98.5%) for T13. Pooled specificity was 99.9% for all three trisomies

Summary

For objective 1a, there were no studies reporting the performance of cfDNA testing after the UK combined test at threshold 1:150, but the meta-analysis showed that in high risk populations defined in a range of ways, the pooled sensitivity was 97.2% (CI 95.1% to 98.4%) for T21, 92.9% (CI 89.2% to 95.4%) for T18 and 95.3% (CI 86.3% to 98.5%) for T13. Pooled specificity was >99.7% for all three trisomies. Applied to a high risk population with prevalence 3.3%, 1.5% and 0.5% for T21, T18 and T13 respectively, would give positive predictive values of 91%, 84% and 87% respectively. Therefore whilst it is a very good test, even using our highest estimates of accuracy it must not be considered a diagnostic test.

4.5.2 Accuracy of cfDNA testing in populations with different risk (Question 1b)

Research question 1b is How does changing the threshold for defining high risk following a combined test affect the accuracy of cfDNA testing? For objective 1b no studies reported a comparison of cfDNA test accuracy following different thresholds of the combined test. There were also insufficient studies in populations at different defined risk thresholds to make comparisons of performance between different studies, with a range of different thresholds of the combined test. However there were studies in both high risk populations (defined in a variety of ways) and studies in the general obstetric population, so we were able to compare cfDNA testing performance in the general obstetric population to that in high risk groups in the meta-analysis.

Studies comparing cfDNA testing accuracy in populations with different risk

Generally, studies reported cfDNA testing accuracy for women at high risk with a number of indications or not specifying the cut-off or for women from the general obstetric population. Ten studies reported a risk cut-off from previous screening tests.[23, 47, 54, 57, 66, 76, 77, 81, 95] The accuracy of cfDNA testing at different risk threshold was the study objective in two of the included articles.[81] Shaw et al. (2014)[81] recruited 100 high risk women and 100 women with a low risk from 11 centres in Taiwan. 201 pregnant women at >12 weeks gestation were recruited in total. The high risk group included pregnancies with a >1:30 risk of having a trisomy fetus or an NT of >3.0 cm. The low risk group included pregnancies with a <1:1,500 risk. Prior risk was assessed by the first trimester combined screening test and blood samples were analysed using MPSS technology. Population characteristics were similar in both groups but the results of test performance were not reported separately for the two risk groups. However, no FPs or FNs were recorded for the detection of T21, T18 or T13, and all trisomy cases (11 T21, 8 T18 and 3 T13) were detected in the high risk group. The authors therefore concluded that the test performs with 100% sensitivity and 100% specificity in the two different risk groups for the three most common autosomal trisomies. The study therefore provides limited information to inform the research question. Furthermore, the study population was small and highly selected to achieve such a high prevalence.

Zhang et al. (2015)[24] assessed the clinical performance of MPSS in detecting T21, T18 and T13 in over 140,000 clinical samples compared test performance in low-risk and high-risk pregnancies. Pregnancies with a cfDNA testing outcome were divided into high risk (maternal age > 35 years, positive conventional T21 screening test at cut-off 1/270 or 1/300, abnormal ultrasound markers, family history of aneuploidy or a previous pregnancy with trisomic fetus, n = 72,382) and low risk none of the high-risk factors, n = 40,287) and reported no significant difference (sensitivity 99.21% vs. 98.97%, p=0.82; specificity 99.95% vs. 99.95%, p=0.98) of the performance for T21 detection between between both risk groups, except the decreased PPV in the low-risk group (81.36% vs. 94.12%, p < 0.00001) which was expected due to lower prevalence of T21. Due to the low numbers of T13 and T18 cases in the low risk group, the difference in test performance was not investigated in these two trisomies.

One further study reported two different risk groups of the included study population. Chiu et al. (2011)[57] recruited pregnant women referred for invasive testing due to a >1:300 risk of fetal T21 (high risk group, n=582), an intermediate risk (between 1:300 and 1:1,000; n=39) both estimated by conventional prenatal screening (91.8% combined screening) or other risk indications (n=132). Plasma samples for cfDNA testing were assessed using MPSS. The accuracy of cfDNA testing was not reported separately for the different risk groups but the study reported that the number of FPs and FNs was not significantly different for different risk levels.

These three studies give an indication that cfDNA testing performance at different cut-offs might be comparable, but no firm conclusions of the effect of different risk thresholds on the accuracy of cfDNA testing can be inferred from these three studies that included and reported different risk levels of included participants.

Studies reporting cfDNA testing accuracy after a combined test at a single threshold

Three studies reported a single defined risk threshold for the inclusion of the population under investigation.[54, 66, 77] The risk threshold was indicative for invasive testing. Nicolaides et al. (2013)[77] recruited UK women with high risk trisomy of which 94% (227/242) had a combined screening outcome of >1:300. Alberti et al. (2015)[66] included 976 French pregnant women with a T21 risk of >1:250 estimated by the first or second trimester screening test. Ashoor et al. (2012)[54] recruited 400 pregnant UK women with a risk >1:300 established by the first trimester combined test. cfDNA testing was undertaken by the targeted sequencing of SNPs method,[77] MPSS,[66] and DANSR.[54] Table 7 gives an overview of the consistently high cfDNA testing performance in terms of sensitivity and specificity for the three studies.

Table 7 cfDNA testing performance of three studies defining a single risk threshold for study inclusion

Reference	Trisomy	Sensitivity % (95%CI)	Specificity % (95%CI)
Nicolaides et al. (2013)[77]	T21 T18 T13	100 (86.3-100) 100 (31.0-100) 100 (5.5–100)	100 (98.2-100) 100 (97.9-100) 100 (97.9-100
Alberti et al. (2015)[66]	T21	100 (90.6-100)	100 (96.6-100)
Ashoor et al. (2012)[54]	T21 T18	100 (91.1-100) 98 (88.0-99.9)	100 (98.4-100) 100 (98.4-100)

The three studies used similar thresholds and did not cover a big enough range of risk cut-offs to enable an assessment of cfDNA testing performance by risk threshold. Furthermore, the three different studies used the three different main testing strategies and reported equally high sensitivities and specificities with CI that reflect the different prevalence of T21, T18 and T13 cases.

The final four studies reporting a risk cut-off from previous screening tests included studies where either only a subgroup of included participants were included on the basis of a high risk estimated by a screening test and a set risk threshold[23, 47] or which reported the median risk of the included study population where the risk threshold was not reported as the inclusion criteria.[76, 95]

Willems et al. (2014)[47] included women that chose to undergo cfDNA testing. Of the included women 22% had an indication of >1:200 (patients from the Netherlands) or >1:300 (patients from Belgium) following a first trimester screening test. The remaining women had various other indications or no indication for cfDNA testing. In the study by Song et al. (2013)[23] 14.3% of the included participants from the general obstetric population were included on the basis of a >1:270 risk of a positive serum screening test.

Fang et al. (2015)[95] reported the median estimated risk for 65% of the included high risk population who received second trimester dual serological screening to be 1:270 for T21 and 1:350 for T18. Nicolaides et al. (2012)[76] included pregnant women undergoing first trimester combined screening. The reported median T21 risk was 1:8,469 with a range from 1:2 to 1:23,527. Equivalent numbers for the median and range for T18 were reported (1:14,894, 1:2 to 47,472).

Both sets of studies do not lend themselves to the assessment of cfDNA testing performance by risk threshold.

Meta-analysis comparing cfDNA testing accuracy in high-risk and general obstetric population

The meta-analysis would not converge to give estimates in the general obstetric population without a zero cell correction for T21 and T18. Estimates of sensitivity to detect all three trisomies was lower in the general obstetric population, and this was statistically significant for T13.

Summary

Ten studies reported risk thresholds from screening tests prior to cfDNA testing. There was no difference between high and low risk pregnancies in number of FNs and FPs for T21, T18 or T13 in two studies that investigated two different risk thresholds. However, the evidence was weak and was based on small studies that did not report performance separately by risk group. One study reported similar sensitivities and specificities in the study population divided into high and low risk of having a T21 pregnancy, where the high risk was determined by various methods. The remaining 7 studies did not contribute to the decision question. We are therefore unable to present cfDNA testing performance at different risk cut-offs ranging from very high to low risk or present an optimal risk cut-off to maximise cfDNA testing performance in clinical practice. The meta-analysis showed a reduction in sensitivity of cfDNA testing in the general obstetric population in comparison to high risk groups, but this was only statistically significant for T13.

4.6 cfDNA test performance in comparison to the combined test (Question 2)

It was not possible to conduct a meta-analysis of the performance of the combined test due to lack of relevant studies reporting trisomies separately, and heterogeneity in the test. We only included studies which reported performance of the combined test in comparison to or in parallel with cfDNA testing, and did not investigate the combined test performance alone. Therefore a narrative summary of these papers is included.

Five of the studies included in the meta-analysis of cfDNA test accuracy[23, 46, 67, 76, 90] also reported accuracy of a combined test on the same or some of the same pregnancies. Two of these were using the UK combined test[76, 90] but the type of trisomy (T21, T18, or T13) is not reported. There was also

one large population based screening study in six countries reported by Norton and colleagues[46] which separately reports performance for T18, T21 and T13 using first trimester screening at different thresholds to the UK (≥1:270 for T21, ≥1:150 for T18 and T13), and a smaller study from the US[67] which does not report thresholds used. There is one study from China,[23] using the same thresholds as the large study by Norton et al. (2015),[46] but this also does not report performance of the three trisomies separately.

Norton et al. (2015)[46] compared cfDNA testing to the first trimester combined test in 15,841 singleton pregnancies across six countries using DANSR (FORTE). This study found that cfDNA testing detected all 38 cases of T21 with only 9 false positive results, whereas the combined test detected only 30 cases of T21, with 8 false negatives and 854 false positives. For T18 cfDNA testing detected 9 out of 10 cases, with 1 false negative and 1 false positive, whereas the combined test detected 8 out of 10, with 2 false negatives and 49 false positives. For T13 there were only 2 cases, cfDNA testing detected both cases with only 2 false positives, and the combined test detected 1 case with 1 false negative and 28 false positive results. This study is in a relevant population given first trimester screening. The threshold for the combined test was ≥1/270 for T21, and ≥1/150 for T13 and T18, therefore we would expect specificity for T21 to be worse than this in the UK as the threshold is at ≥1/150.

There are two studies from the UK.[76, 90] Quezada et al. (2015)[90] included 2,785 pregnancies which had cfDNA testing and a combined test and were followed up to pregnancy outcome. All 32 cases of T21 were detected by both cfDNA testing and the combined test, with 1 false positive on cfDNA testing and 139 for the combined test. For trisomies 18 and 13 together, cfDNA testing detected 11 out of 15, with the combined test detecting all 15, but there were just 7 false positives on cfDNA testing, with 158 on the combined test. However the cfDNA testing results were available to clinicians at the time of measuring nuchal translucency so there may have been overestimation of the performance of the combined test. The second study from the same group[76] shows that 8 cases of T21 and 2 cases of T18 were detected on both cfDNA testing and the combined test, but no further comparisons are made between the two tests. Both studies used the DANSR (FORTE) testing strategy.

Bianchi et al. (2014)[67] compared the performance of a combined test to cfDNA testing (MPSS) in a screening population in the US. The study was powered to measure differences in specificity but not sensitivity. In 1,909 women who had both tests for T21 there were 69 false positives for the combined test, in comparison to 6 false positive results for cfDNA testing. In 1,905 women who had both tests for T18, there were 11 false positives for the combined test and 3 for cfDNA testing. There were only five cases of T21 and 2 cases of T18 and they were all detected by both tests. And finally in 899 women who had both tests for T13 there were 6 false positives for the combined test and just one for cfDNA testing. This result is unsurprising as both tests have very different thresholds for recall. These results may be limited in their generalisability to the UK context as the 'conventional' standard screening test they used was serum biochemical assays with or without nuchal translucency measurement in the first or second trimester with unspecified threshold. Furthermore the women included were from the first, second and third trimester of pregnancy, so many will be from later in pregnancy when the performance of cfDNA testing is improved due to greater fetal fraction.

Finally a study from China[23] with just 11 cases of trisomy (8 T21, 2 T18 and 1 T13) found that cfDNA testing using MPSS detected all eight with only one false positive, and serum screening with a threshold of ≥1/270 only detected 6 cases, with 5 false negatives and 243 false positives.

Summary

The specificity of the combined test is considerably worse than that of cfDNA testing, which is to be expected as they have different thresholds. The threshold for positive results on the combined test is designed to be at a risk of ≥1/150 or ≥1/270, so the chosen threshold will by its nature include a large number of false positives. One large study in the US found that sensitivity for T21 is also better for cfDNA testing (detected 36/36 cases) than the combined test at threshold ≥1/270 (detected 30/36 cases). For T18 and T13 it is more difficult to make comparisons due to the lower prevalence.

4.7 Integrating cfDNA testing into the combined test (Question 3)

To date, no study or paper reports original data about the diagnostic accuracy that could be achieved when integrating cfDNA testing into the combined test. Only one of the included studies by Quezada et al. (2015)[90] state in their discussion that most false-positive and false-negative results from cell-free DNA testing could be avoided if the a priori risk from the combined test is taken into account in the interpretation of individual risk. A recent meta-analysis reports positive likelihood ratios for T21, T18, and T13 of 1238, 645, and 461, respectively.[99] The negative likelihood ratios for T21, T18, and T13 are 0.01, 0.03 and 0.08, respectively. Therefore, with a negative cfDNA result for these trisomies, there is a 100-fold, 31-fold and 13-fold reduction in the a priori risk.[100]

Wright et al. (2015)[101] performed a theoretical study using Bayes' theorem to combine measures from cfDNA testing with maternal age and first-trimester biomarkers in screening for fetal trisomies. They found that in screening for fetal trisomies by cfDNA analysis of maternal blood, the detection rate (=sensitivity) at a given false-positive rate (=1-specificity) and the positive and negative likelihood ratios depend on the fetal fraction and the precision of its measurement. According to this study, combining all available data from the cfDNA test with those of any prior method of screening will provide more accurate patient-specific risks and improve performance of screening by cfDNA testing. Even when the fetal fraction is < 4% and cfDNA testing is reported to have failed, data from cfDNA test could potentially be used to improve the performance of screening by the combined test. In a population with the maternal-age distribution of pregnancies in England and Wales in 2011, screening for T21 by the first-trimester combined test (risk cut-off 1:100) and cfDNA testing with known fetal fractions had a theoretical sensitivity of 99.9% and a false-positive rate of 0.02% (corresponding to a specificity of 99.98%) for all fetal fractions combined as well as fetal fractions ≥ 4% only.[101] Screening by the first-trimester combined test (risk cut-off 1:100) and cfDNA testing with estimated fetal fractions gave an overall sensitivity and false-positive rate for the whole population of 99.8% and 0.05%, respectively, while for the subgroup with a fetal fraction of ≥ 4%, the overall detection rate and false-positive rate were 99.8% and 0.04%, respectively.

Summary

There was one study and a further theoretical study which discuss the potential of integrating cfDNA testing and the combined test[90, 101] but no studies which demonstrated test accuracy after implementing this approach. We are therefore unable to determine if combining the conventional screening tests with the cfDNA test would offer an increased accuracy.

4.8 Test failure of cfDNA testing (Question 4)

The test failure rate, reasons for failure and whether repeat testing after failure was successful is detailed for included studies in Table 20. There are broadly three types of cfDNA testing failure. Firstly errors in taking, storing or transporting the sample such as insufficient blood volume, labelling errors, late receipt or haemolysis. Data relating to pre-analytic failures were provided by 16 studies [24, 46, 54-56, 58, 61, 62, 67, 70, 76, 78, 80-82, 86, 96, 98] and ranged from 0.14%[24] to 10.6%.[80] We do not give much further consideration to this first type as they can be reduced or eradicated by effective quality assurance procedures. Secondly, in cases of analytic failures where cfDNA testing failed to produce a result, these may be resolved by retesting the same blood, or may require a fresh blood sample, or

may not provide a result even after retesting. Reasons for this are low fetal fraction (usually below 4%) or assay failure for a variety of reasons, such as failed DNA extraction, amplification and sequencing. The third type is indeterminate results, meaning that the test result is in a mid-range of neither positive nor negative. Test failures and inconclusive results are detailed by karyotype in Table 21.

Including test failures and inconclusive results in the meta-analysis did not make a statistically significant difference to sensitivity and specificity. Including inconclusive results resulted in a very small (if any) change in estimates because so few studies define an intermediate range. However including test failures in an intention to diagnose analysis reduced the sensitivity for T21 from 97% (CI 96% to 98%) to 96% (CI 94%-97%), for T18 from 93% (CI 90% to 95%) to 91% (CI 87% to 93%), and for T13 from 83% (CI 74% to 89%) to 80% (CI 73% to 86%). Similarly, including test failures in an intention to diagnose analysis reduced the specificity from 99.8% (CI 99.7% to 99.9%) for all three trisomies to 97% (CI 96% to 98%) for T21, 98% (CI 97% to 98%) for T18, and 98% (CI 96% to 98%) for T13. These reductions in sensitivity could be an overestimation because intention to diagnose is the worst case scenario, and because of use of a zero cell correction, but they may also be an underestimation as some studies did not report numbers of test failures so these were assumed to be zero.

4.8.1 Analytic Failures

Thirty-five studies [22-26, 46, 47, 54, 55, 57-59, 61, 62, 67-70, 74-80, 82-86, 88-91, 94] reported analytic failures, with 16 [22, 25, 26, 46, 54, 58, 59, 61, 62, 69, 74, 76, 77, 79, 85, 90] reporting the reference standard results for the failed cases. Of the remaining 17 studies, 15 [56, 63-65, 71-73, 81, 87, 88, 92, 93, 96-98] reported no test failures and two[60, 95] did not report whether there were test failures.

In a study of 147,103 women in China with singleton or twin pregnancy of nine or more weeks of duration using MPSS, there were 3,213 (2.2%) initial test failures, reduced to 145 (0.1%) after a redraw of blood.[24] Similarly, in 1,982 women primarily of Chinese ethnicity using MPSS there were 23 (1.2%) initial test failures and 1 (<0.1%) failure upon retest,[74] and in 7,705 women with singleton pregnancies from China using MPSS there were 141 (1.8%) test failures of which 4 (<0.1%) failed repeat testing.[94] In 3,000 women from the Netherlands and Belgium using DANSR, there were 55 (1.8%) cfDNA testing failures, with 27 (0.9%) remaining unsuccessful after retest.[47] Of 1,988 predominantly Caucasian women from the US using MPSS,[62] 110 (5.5%) initially failed cfDNA testing, with 17 (0.9%) test failures after retesting. In 520 women from the Netherlands and Sweden using DANSR there were 51 (9.8%) initial test failures of which 16 (3.1%) failed repeat testing (7 due to low fetal fraction and 9 due to assay failure).[82] In 30,705 samples from single pregnancies in the US submitted for commercial cfDNA testing using SNP-approach there were 1,966 (6.4%) test failures; 85% of which due to low fetal fraction.[86] A redraw analysis in 21,292 cases found 317 (35.7%) of 888 remaining unsuccessful after retest. In 6,017 eligible samples submitted for commercial MPSS testing in the US, there were 43 (0.7%) technical cancellations with no details about retesting.[70] Similarly in 16,329 women with singleton pregnancies in the US using DANSR there were 488 (3.0%) test failures (192 [39%] had a low fetal DNA <4%, 83 [17%] had fetal fraction that could not be measured, and 213 [44%] had high assay variance or an assay failure), with no details about retesting.[46] In 100,000 samples from a commercial laboratory in the US offering MPSS testing, there were 1,928 (1.9%) initial test failures (842 [44%] due to insufficient fetal DNA and 1,086 samples [56%] failing other laboratory quality metric including library and sequencing passing criteria or generally technical or mechanical failures) of which 1,330 (1.3%) were not reportable after redrawing blood.[89]

Overall, among 363,572 initial cfDNA tests performed, there were 8,968 (2.5%) analytical failures due to insufficient fetal DNA or assay failure for a variety of reasons. The initial analytical failure rate provided by 50 studies ranged from 0.0% in 15 studies [56, 63-65, 71-73, 81, 87, 88, 92, 93, 96-98] to 12.7%.[26] In 16

studies, further details were given for the reason of the test failure: low fetal fraction rates in the initial blood sample ranged from 0.0% [54, 75] to 6.1%.[79] The failure rate after resampling was reported in 14 studies and ranged from 2.8% (4/141)[94] to 100% (1/1).[83] Overall, among 5,789 pregnancies with resampling, 803 (13.9%) failed also the repeat cfDNA test.

4.8.2 Indeterminate Results

Indeterminate results occur when a mid-range of test results are classified as indeterminate, rather than positive or negative. There were 5 papers in this review that reported indeterminate results for trisomies 21, 18 and 13.[55, 63, 70, 83,185] Bianchi et al. (2012)[55] in a study of 534 high risk singleton pregnancies across the first and second trimester found 7 indeterminate results for T21 (of which one was affected by T21), 5 indeterminate results for T18 (of which 2 had T18) and 2 indeterminate results for T13 (both had T13). This study used 6-plex MPSS using the Illumina HiSeq 2000, with Normalised Chromosome Value (NCV) > 4.0 classified as aneuploid, NCV < 2.5 euploid, and 2.5 ≤ NCV ≤ 4.0 indeterminate. A later study by the same group[67] used the same testing technique, but increased to 8-plex testing but with results in the range 3.0 ≤ NCV ≤ 4.0 resequenced in 1-plex. This second study reported no indeterminate results after this repeat testing, in a general obstetric population of 2,042 singleton pregnancies. Futch et al. (2013)[70] reviewed 6,123 samples submitted for commercial cfDNA testing, using the 6-plex MPSS using the Illumina HiSeq 2000 with the same thresholds as the first study by Bianchi et al. (2012).[55] They reported 173 indeterminate results, of which 3 were later confirmed as T18, 46 were euploid, 1 had monosomy X and the trisomy status of the remainder was unknown. Sehnert et al. (2011)[63] in a test validation cohort of 575 singleton pregnancies found 1 indeterminate result for T13 (NCV=3) which was a T13 pregnancy. They used MPSS monoplex sequencing using a Genome Analyzer IIx (Illumina) with NCV > 4.0 classified as aneuploid, NCV < 2.5 euploid, 2.5 ≤ NCV ≤ 4.0 indeterminate. Beamon et al. (2014)[83] reviewed their institution's experience with the use of commercial cfDNA testing in a cohort of 208 high risk pregnancies. 45 samples were tested using MPSS testing with Verinata commercial test, which classified NCV > 4.0 as aneuploid, NCV < 2.5 as euploid, and 2.5 ≤ NCV ≤ 4.0 as indeterminate. They reported 5 (11.1%) indeterminate results of which 1 was T18, 2 were euploid, and 2 were second-trimester fetal demises without post-mortem genetic testing.

Other studies have used similar sequencing techniques using the Illumina HiSeq 2000 but with different thresholds which do not create indeterminate results, for example McCulloch et al. (2014)[89] used robust z-scores z > 3 for chromosome 21 and z > 3.95 for chromosomes 18 and 13.

4.8.3 Predictors of Test Failure

Only one study reported failure rate by gestational age. Pergament et al. (2014)[79] found in a cohort of 1,051 singleton pregnancies of at least 7 weeks gestation 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 ten weeks was 53/900 (5.9%). Norton et al. (2015)[46] did not find an association between test failure and gestational age in 18,510 women between 10 and 14 weeks gestation.

Sixteen studies included[22, 25, 26, 46, 54, 58, 59, 61, 62, 69, 74, 76, 77, 79, 85, 90] give details both of the number of test failures, and the karyotype of those babies, see Table 21. One study with 488 test failures found that of these 3 had T21, 1 had T18, and 2 had T13, with 475 euploid pregnancies and 7 with other abnormalities.[46] 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). Another study with 85 test failures had among them 8 with T21, 7 with T18, 2 with T13, 66 euploid and 2 with other chromosomal abnormalities.[79] 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 [10.9%], p=0.004). In contrast, Palomaki et al. (2012)[62] found a similar aneuploidy risk in the 17 women with repeat test failures (17.6%) to the 15% occurring in the

population with successful testing. Another study with 148 test failures (57 due to low fetal fraction, 91 due to assay failure),[78] did not report rates of trisomy in test failures, but reported that assay failure was similar in normal vs trisomy cases. Furthermore they could not find an association between assay failure and gestational age, maternal age or race/ethnicity. However, one included study with 29 initial test failures (5.6%) in 515 twin gestations and 32 initial test failures (1.7%) in 1,847 singleton pregnancies reported using univariable regression that twin pregnancy, higher maternal weight, and conception by IVF were all significant predictors of failure.[84] However the univariable regression approach would result in a high type 1 error rate so these conclusions should be treated with caution. This same study did not find that gestational age at test, cigarette smoking or origin of oocyte were significant predictors. The three studies that included only twins have 15/208 (7.2%) test failures using DANSR,[69] 0/189 (0%) using MPSS,[71] and 16/515 (3.1%) using DANSR.[78] It is unclear whether these rates are higher than for singleton pregnancies.

Norton et al. (2015)[46] found a higher median maternal weight in women with a low fetal fraction as compared with women with successful cfDNA testing (93.7 kg vs 65.8 kg, p<0.001). Beamon et al. (2014)[83] reported that all 3 observed test failures because of insufficient fetal fraction (1.4%) occurred in obese women with body mass indices (BMI) of 38.2, 44.6, and 47.3 kg/m^2 at 11, 10, and 12 weeks' gestation, respectively. Palomaki et al. (2011)[61] found that 0.35% of maternal plasma samples had insufficient (<4%) fetal fraction, and a strong negative association of fetal fraction with maternal weight was observed in cases and controls, with weights of 100, 150, and 250 pounds associated with predicted fetal fractions of 17.8%, 13.2%, and 7.3%, respectively. No association was found for gestational age, maternal race, or indication for testing. Bianchi et al. (2012)[55] found no distinguishing clinical features for 16 samples (3.0%) with no fetal DNA detected. McCullough et al. (2014)[89] found that whilst overall there was insufficient fetal fraction for just 0.9% of samples, this increased with increasing maternal BMI. For BMI between 10 and 30 kg/m^2 there was insufficient fetal fraction for <1%, at 30-40kg/m^2 this increased to 3%, at 40-50kg/m^2 to 9% and 50-60 kg/m^2 to around 18%. At BMI >60kg/m^2 there was insufficient fetal fraction for 18.3% of patients (reported as around 10% on the graph in the publication). The authors argue that whilst cfDNA testing is more likely to fail for more obese patients, it still produces results for around 80%, and they may have fewer other testing options.

Summary

The rate of analytic failure (failure of the cfDNA testing) ranged from 0% to 12.7%[26] and among 5,789 pregnancies with resampling, 803 (13.9%) also failed the repeat cfDNA testing. There were 5 papers in this review that reported indeterminate results (results in a range defined as neither positive nor negative) for trisomies 21, 18 and 13.[55, 63, 70, 83, 185] ranging from 0% (0/2042) to 11.1% (5/45). In the study with no indeterminate results they used 8-plex testing, and where the initial score was indeterminate they repeated using 1-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. (2014)[79] 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 ten 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 [10.9%], p=0.004). Norton et al. (2015)[46] did not find an association between test failure and gestational age in 18,510 women between 10 and 14 weeks gestation, but found that [46] 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).

5. Results of Economic Model (Question 5)

5.1 Summary of Economic Model

Table 8 presents the predicted impact of each strategy under the reference case, with simplified results illustrated in Figure 18 to Figure 21. If cfDNA testing is offered initially to those identified as at high risk (>1/150) of trisomy, the model predicts that 9,912 tests would be carried out, and 350 retests would be required. As a consequence, the number of invasive tests required would fall from 7,910 to 1,434. The majority of invasive tests avoided are in trisomy-free pregnancies, so that the number of test-related miscarriages of healthy pregnancies falls from 46 to fewer than 3 per year. The reduction in invasive tests, given the costs listed in Table 2, partially offsets total cfDNA testing costs. The projected annual cost of testing increases by £120,000 when cfDNA testing is offered to all those above the current first trimester combined test threshold. The number of trisomies detected also falls as a result, from 1032 to 1019 per year. As a result, the cost per trisomy detected increases from £14,472 to £14,764. The number of trisomies present but not detected in the programme, increases from 1705 to 1717. This does, however, include 1037 trisomies in women who do not accept combined testing, so that the number of women who accept combined testing with a trisomy that is not detected through the programme increases from 668 to 680. Of the 1019 trisomies detected, 927 would have had cfDNA testing and therefore a potentially delayed diagnosis.

This model uses the more conservative estimates of sensitivity and specificity from the meta-analysis, which may be underestimates as they use a zero cell correction. Section 5.2 uses the same model with the more optimistic estimates of sensitivity and specificity, which do not include a zero cell correction for TP and FN, but may be overestimates due to high risk of bias and publication bias in included studies, and including women in the second and third trimester of pregnancy.

Table 8 also illustrates the impact of relaxing the threshold at which cfDNA testing is offered. The introduction of cfDNA testing leads an increase in the number of trisomies detected if the combined test threshold is relaxed to 1/200 or beyond. As the threshold is relaxed further, the number of trisomies detected continues to increase. However, the number of tests required (cfDNA testing and invasive tests) also increases, so that the total cost and the cost-per-trisomy also increase. If cfDNA testing were to be offered to all women regardless of their combined test-defined risk, the total cost of the screening programme would increase to £109m, giving a cost per trisomy detected of £84,709. This estimate assumes that cfDNA testing would remove the need for combined testing. If combined testing was still required, costs would increase by a further £12.2m.

Outcome	Number
Test-related miscarriage	46
Downs cases identified	764
Downs cases undetected	1219
Edwards cases identified	196
Edwards cases undetected	331
Patau cases identified	74
Patau cases undetected	155
Invasive tests/trisomy detected	7.4
Cost per case detected	£14,472
Total cost (£millions)	14.9
cffDNA initial test failures	N/A

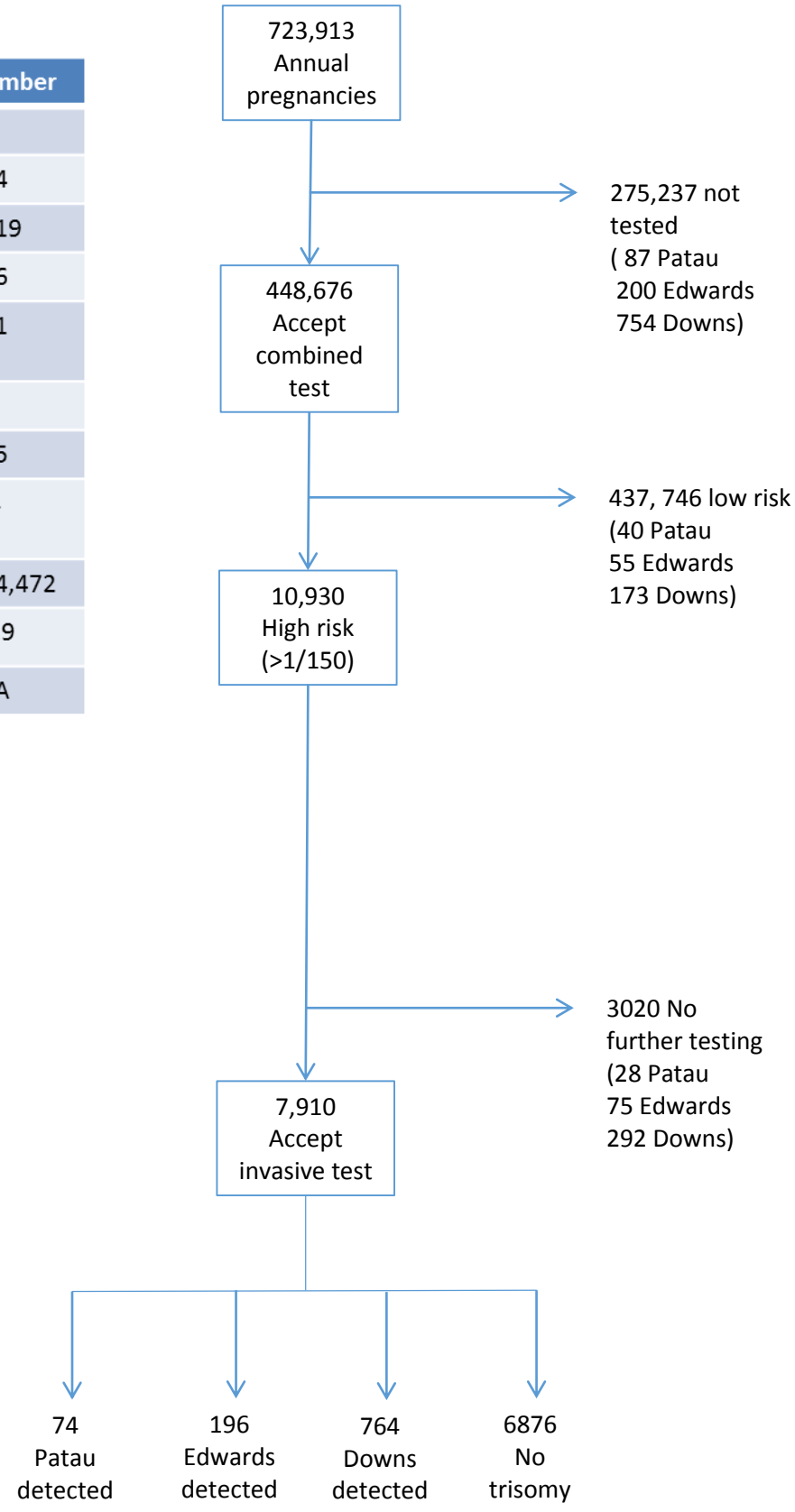


Figure 18 Simplified flow chart illustrating the economic model of a screening programme with the combined test only at threshold 1/150 (no cfDNA)

Outcome	Number
Test-related miscarriage	3
Downs cases identified	767
Downs cases undetected	1215
Edwards cases identified	189
Edwards cases undetected	337
Patau cases identified	64
Patau cases undetected	165
Invasive tests/trisomy detected	1.4
Cost per case detected	£14,764
Total cost (£millions)	15
cffDNA initial test failures	385

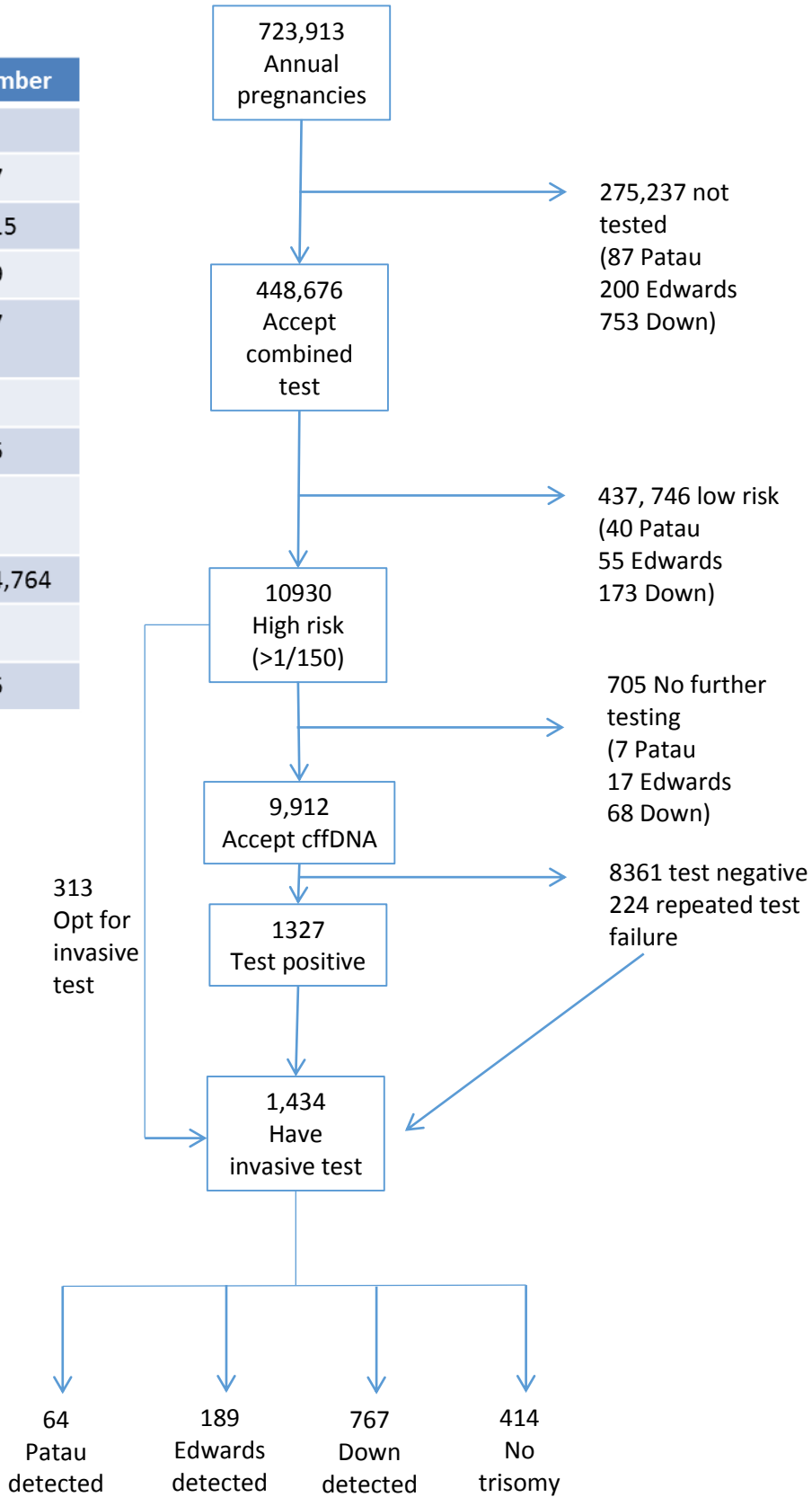


Figure 19 Simplified flow chart illustrating the economic model of combined test at threshold 1/150 followed by cfDNA testing

Outcome	Number
Test-related miscarriage	10
Downs cases identified	839
Downs cases undetected	1144
Edwards cases identified	214
Edwards cases undetected (missed)	312
Patau cases identified	77
Patau cases undetected	152
Invasive tests/trisomy detected	2.3
Cost per case detected	£20,094
Total cost (£millions)	22.7
cffDNA initial test failures	1,296

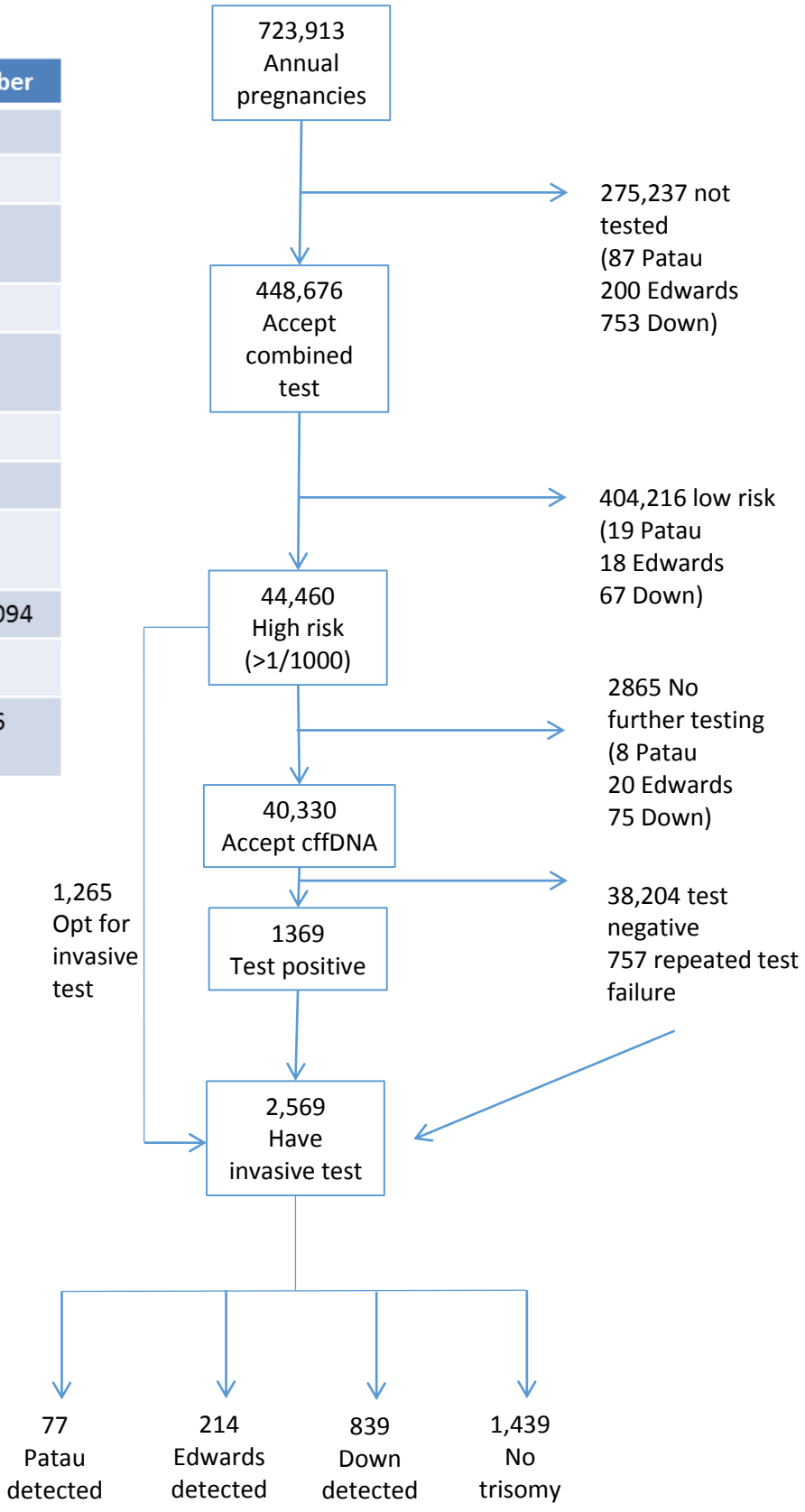


Figure 20 Simplified flow chart illustrating the economic model of combined test at threshold 1/1000 followed by cfDNA testing

Outcome	Number
Test-related miscarriage	8
Downs cases identified	941
Downs cases undetected (missed)	1041
Edwards cases identified	240
Edwards cases undetected (missed)	286
Patau cases identified	93
Patau cases undetected (missed)	136
Invasive tests/trisomy detected	2.0
Cost per case detected	£84,709
Total cost (£millions)	108
cffDNA initial test failures	13,410

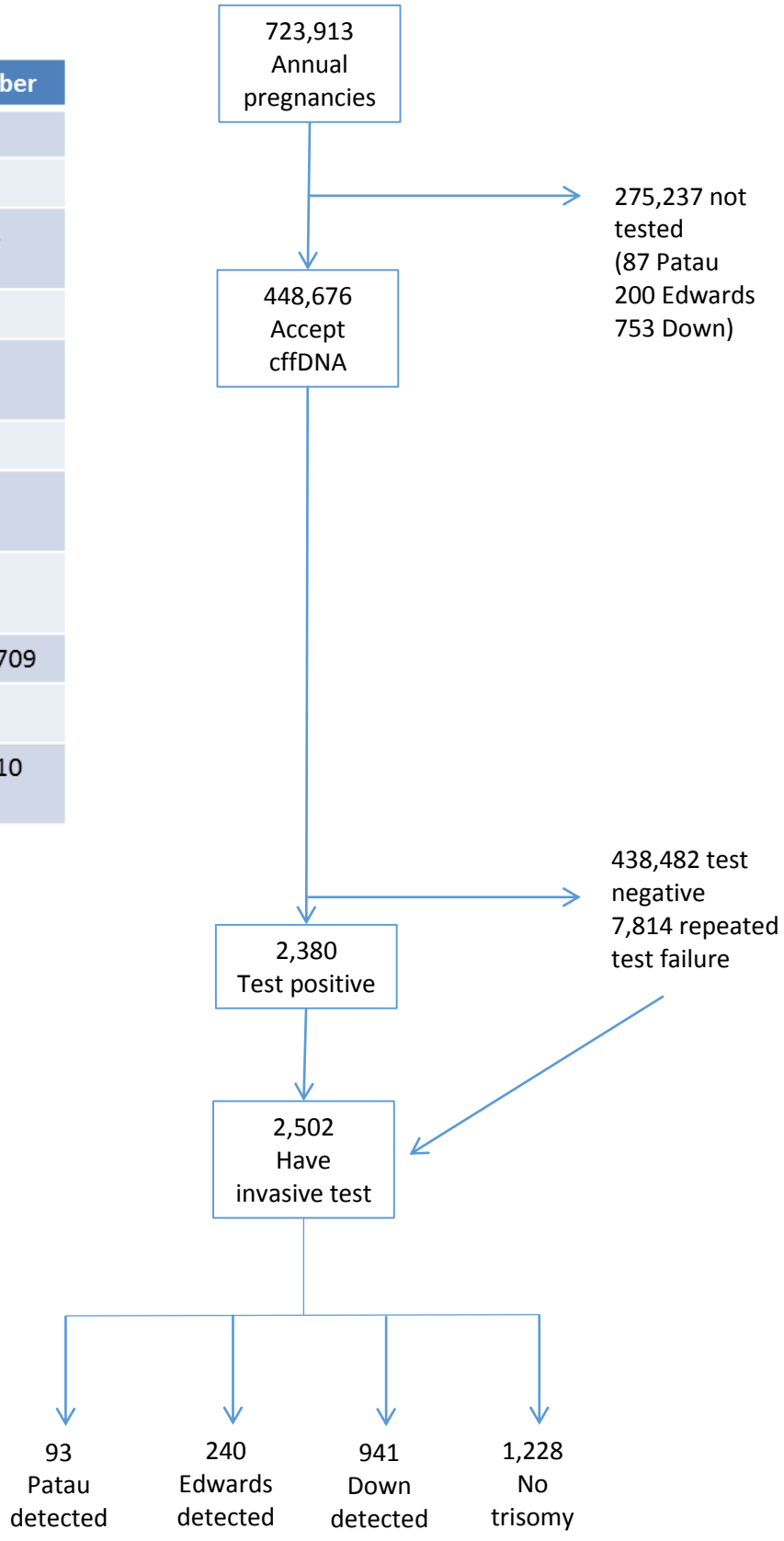


Figure 21 Simplified flow chart illustrating the economic model of cfDNA as the primary screening test

Table 8 Reference case predictions for annual FASP performance in England and Wales. 95% credible intervals for mean estimates are provided in brackets

	Combined test alone	cfDNA testing if combined test result >1/150	cfDNA testing if combined test result >1/200	cfDNA testing if combined test result >1/1000	cfDNA testing if combined test result >1/2000	cfDNA testing alone
Number of combined tests performed	448676	448676	448676	448676	448676	0
Number of initial cfDNA tests performed	0	9912 (9513, 10300)	12150 (11670, 12600)	40330 (38940, 41600)	65660 (63470, 67650)	448676
Number of cfDNA test failures	0	385.3 (300.3,503.1)	453.1 (364.1,574.6)	1296 (1153,1464)	2050 (1848,2268)	13410 (12240,14610)
Number of cfDNA retests performed	0	349.7 (271,458.3)	411.2 (328.4,523.7)	1176 (1033,1339)	1860 (1651,2084)	12160 (11040,13330)
Number of amniocenteses performed	3781 (3683, 3879)	685.5 (457.2,885.2)	730.6 (488.7,944.4)	1228 (842.4,1684)	1649 (1089,2339)	1196 (753.3,1556)
Number of CVS performed	4129 (4022, 4236)	748.6 (499.3,966.6)	797.9 (533.7,1031)	1341 (920,1839)	1800 (1189,2554)	1306 (822.7,1699)
Total cost (£m)	14.93 (14.9, 15)	15.05 (14.86,15.22)	15.61 (15.42,15.79)	22.69 (22.33,23.03)	29.03 (28.5,29.53)	107.8 (107.4,108.2)
T13 cases detected by testing	73.7 (64.4, 83.5)	63.6 (36.97,84.23)	65.62 (38.09,86.6)	76.59 (44.7,99.95)	79.65 (45.49,103.9)	92.5 (52,121)
T18 cases detected by testing	196.4 (180,214)	189.1 (111.3,236.7)	193.2 (113,241.9)	213.9 (126.7,267)	217.6 (126.4,272.5)	239.6 (137.3,299.7)
T21 cases detected by testing	763.9 (730,798)	766.7 (452.2,935.2)	779.4 (457.4,950.5)	838.7 (499.2,1022)	854.9 (499.4,1044)	940.5 (536.7,1148)
Total trisomies detected	1031.65 (964, 1103)	1019.4 (600, 1256)	1038.22 (608.49,1279)	1129.19 (671,1389)	1152.15 (671,1420)	1272.59 (726,1568)
Undetected T13 cases after testing	155 (135,178)	165.3 (138.3,199.9)	163.3 (136.4,198.1)	152.4 (124.5,189.9)	149.4 (121.9,188.8)	136.3 (108,182)
Undetected T18 cases after testing	330.5 (302,362)	337 (285.2,419.6)	333 (280.9,418)	312.3 (258.3,404.2)	308.6 (253.8,401.5)	286.2 (229,393)
Undetected T21 cases after testing	1219 (1163, 1277)	1215 (1044,1532)	1203 (1029,1530)	1144 (960,1488)	1127 (938.9,1485)	1041 (837.3,1446)
Test-related terminations of healthy pregnancy	46.1 (30,69)	2.8 (1.2,5.7)	3.3 (1.4,6.6)	9.7 (4.5,17.7)	15.4 (7.2,28.2)	8.2 (4.1,14.4)
Cost per trisomy detected through testing (£/trisomy)	14472 (13605, 15414)	14764 (12117, 24747)	15035 (12346, 25341)	20094 (16581, 33299)	25196 (20790, 42456)	84709 (68992, 148003)

5.2 Economic Model Sensitivity analysis

5.2.1 Test accuracy estimates

The first sensitivity analysis uses the more optimistic estimates of test accuracy. These do not include a zero cell correction for TP and FN, but may be overestimates due to the inclusion of women later in pregnancy in the studies, and high risk of bias and publication bias. Results are described in Table 9, and in Figure 22 to Figure 25.

Table 9 Predictions for annual FASP performance in England and Wales when test performance is estimated excluding studies with zero cell correction for TP and FN 95% credible intervals for mean estimates are provided in brackets

	Combined test alone	**cfDNA testing if combined test result >1/150**	**cfDNA testing if combined test result >1/200**	**cfDNA testing if combined test result >1/1000**	**cfDNA testing if combined test result >1/2000**	**cfDNA testing alone**
Number of combined tests performed	448676	448676	448676	448676	448676	0
Number of initial cfDNA tests performed	0	9912 (9513,10300)	12140 (11650,12600)	40330 (38940,41600)	65660 (63470,67650)	448676
Number of cfDNA test failures	0	385.3 (300.3,503.1)	452.8 (364.9,574.6)	1296 (1153,1464)	2049 (1846,2268)	13410 (12240,14610)
Number of cfDNA retests performed	0	349.7 (271,458.3)	410.8 (328.5,524.2)	1176 (1033,1339)	1859 (1652,2078)	12160 (11040,13330)
Number of amniocenteses performed	3781 (3683,3879)	698.7 (466.2,896)	743.1 (499.1,957)	1238 (844.8,1698)	1646 (1098,2333)	1051 (677.5,1319)
Number of CVS performed	4129 (4022,4236)	763 (509.1,978.5)	811.5 (545.1,1045)	1351 (922.6,1854)	1798 (1200,2547)	1148 (739.8,1441)
Total cost (£m)	14.93 (14.86,15)	15.06 (14.87,15.22)	15.62 (15.42,15.8)	22.7 (22.34,23.04)	29.03 (28.49,29.52)	107.7 (107.3,108)
T13 cases detected by testing	73.7 (64.4,83.5)	74.45 (42.71,98.36)	76.81 (44.29,101.1)	89.76 (52.74,116.5)	93.41 (54.52,121)	109 (61.33,141)
T18 cases detected by testing	196.4 (180.0, 213.6)	197.2 (117.3,247)	201.7 (118.5,251.8)	223.6 (133.5,278.6)	227.7 (134.2,283.8)	250.3 (144.5,312.2)
T21 cases detected by testing	763.9 (730.4,797.7)	784.1 (466.2,954.7)	798 (469.8,974)	859.2 (514.6,1048)	877.1 (518.5,1067)	961.9 (556.6,1173)
Total trisomies detected	1031.65 (964.08,1102.51)	1055.75 (626.21,1300.06)	1076.51 (632.59,1326.9)	1172.56 (700.84,1443.1)	1198.21 (707.22,1471.8)	1321.2 (762.43,1626.2)
T13 cases undetected after testing	155 (135.1,177.6)	154.7 (126.7,191)	152 (124.7,190.5)	139.2 (111.2,181.6)	135.5 (107.5,178.3)	120.1 (90.62,170.2)
T18 cases undetected after testing	330.5 (301.5,361.5)	328.9 (276.4,414.4)	324.4 (270.6,412.9)	302.8 (247.4,395.9)	298.5 (243,395.4)	275.8 (217.1,387.2)
T21 cases undetected after testing	1219 (1163,1277)	1198 (1022,1522)	1184 (1007,1519)	1123 (937.1,1474)	1105 (915.8,1466)	1019 (812.1,1433)
Test-related miscarriage of healthy pregnancy	46.05 (29.98,68.85)	2.717 (1.135,5.614)	3.196 (1.366,6.445)	9.49 (4.273,17.66)	15.05 (6.895,27.87)	5.884 (3.116,10.3)
Cost per trisomy detected through testing (£/trisomy)	14472 (13605,15414)	14265 (11707,23746)	14510 (11907,24376)	19359 (15966,31876)	24228 (20057,40284)	81517 (66412,140734)

Outcome	Number
Test-related miscarriage	3
Downs cases identified	784
Downs cases undetected	1198
Edwards cases identified	197
Edwards cases undetected	329
Patau cases identified	74
Patau cases undetected	155
Invasive tests/trisomy detected	1.4
Cost per case detected	£14,265
Total cost (£millions)	15
cffDNA initial test failures	385

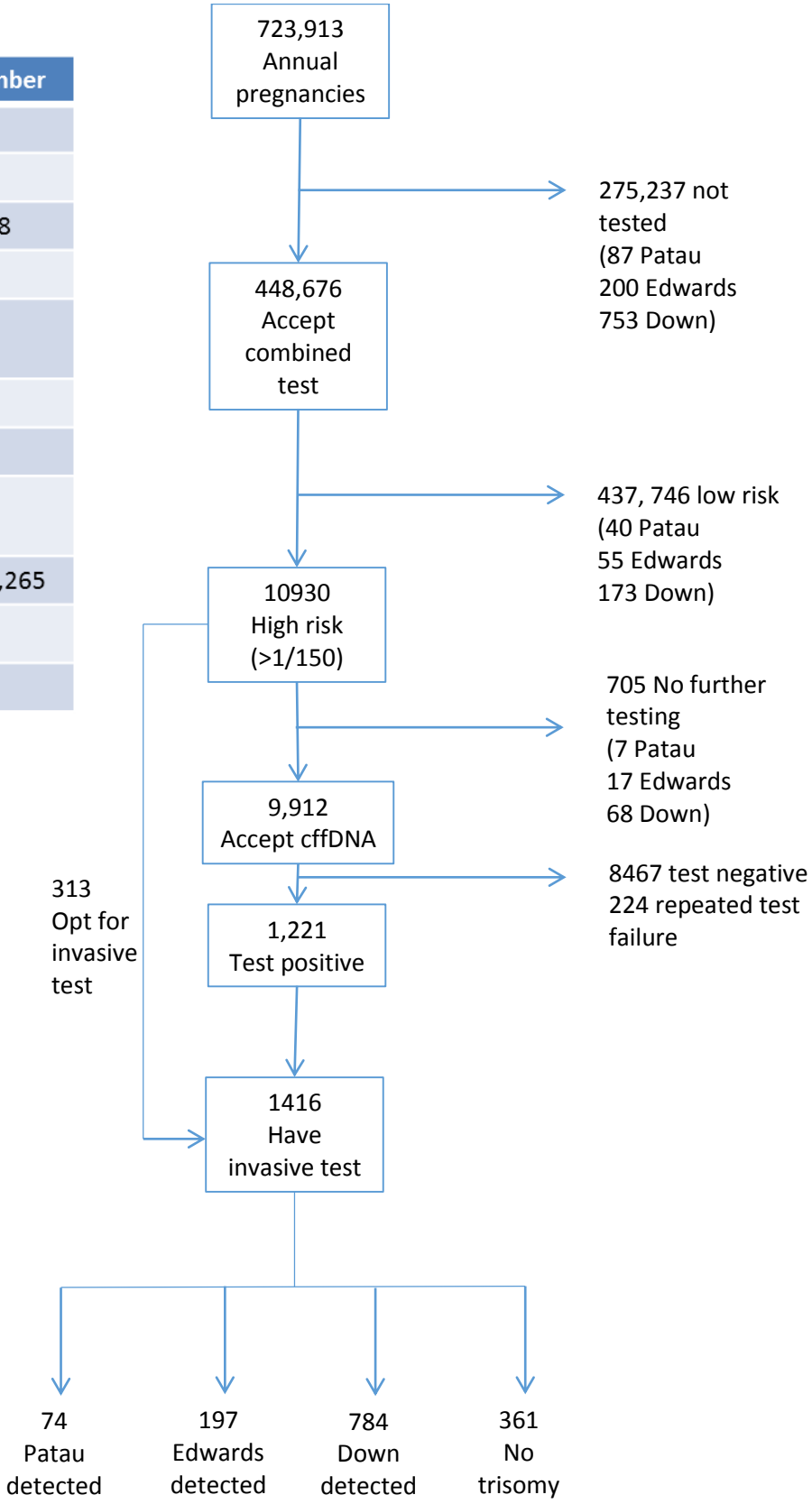


Figure 22 Simplified flow chart illustrating the economic model of combined test at threshold 1/150 followed by cfDNA testing, using higher estimates of test accuracy without zero cell correction for TP and FN

Outcome	Number
Test-related miscarriage	9
Downs cases identified	859
Downs cases undetected	1123
Edwards cases identified	224
Edwards cases undetected (missed)	303
Patau cases identified	90
Patau cases undetected	139
Invasive tests/trisomy detected	2.2
Cost per case detected	£19,359
Total cost (£millions)	22.7
cffDNA initial test failures	1,296

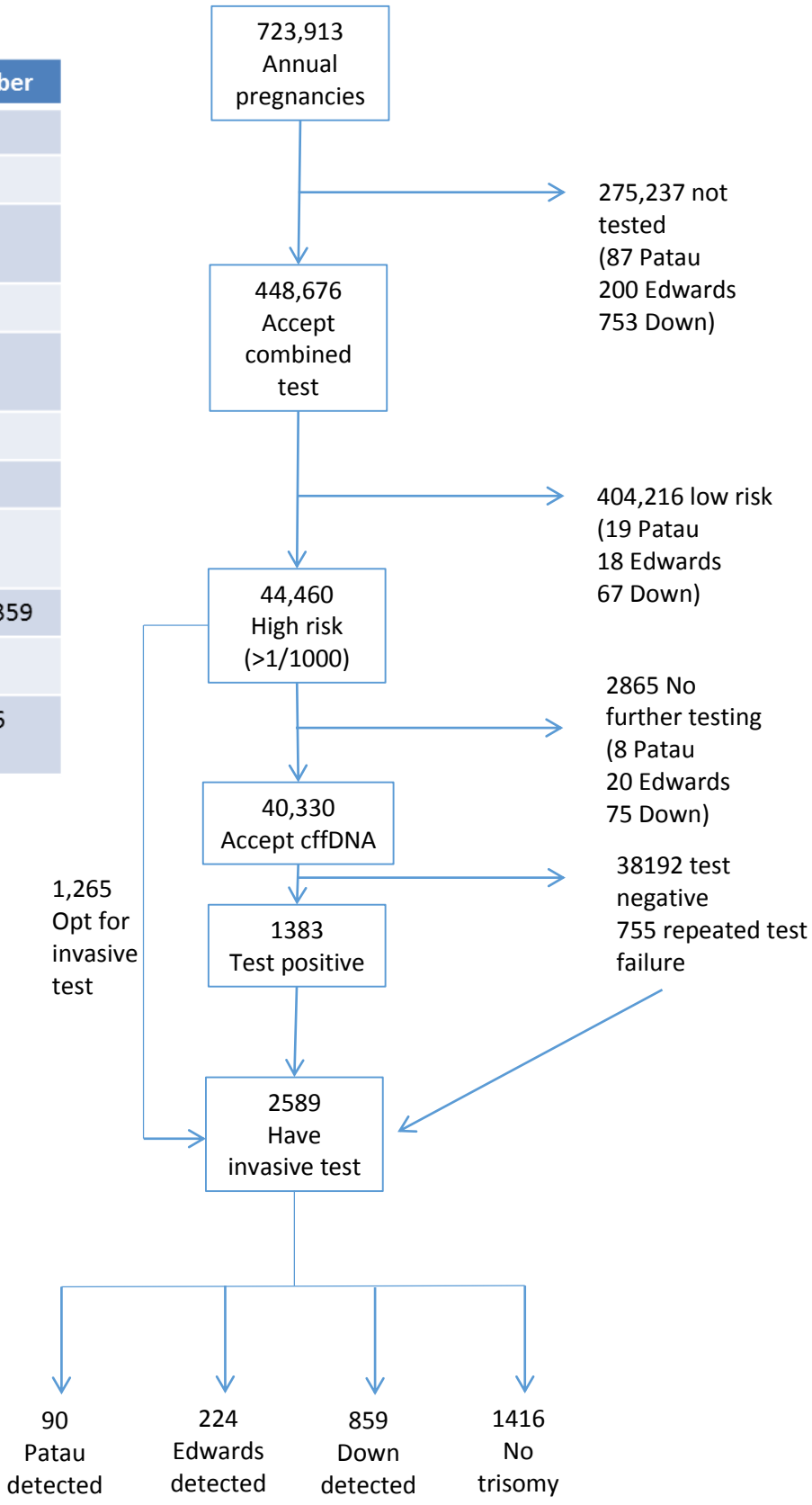


Figure 23 Simplified flow chart illustrating the economic model of combined test at threshold 1/1000 followed by cfDNA testing, using higher estimates of test accuracy without zero cell correction for TP and FN

Outcome	Number
Test-related miscarriage	15
Downs cases identified	877
Downs cases undetected	1105
Edwards cases identified	228
Edwards cases undetected (missed)	299
Patau cases identified	93
Patau cases undetected	136
Invasive tests/trisomy detected	2.9
Cost per case detected	£24,228
Total cost (£millions)	29
cffDNA initial test failures	2049

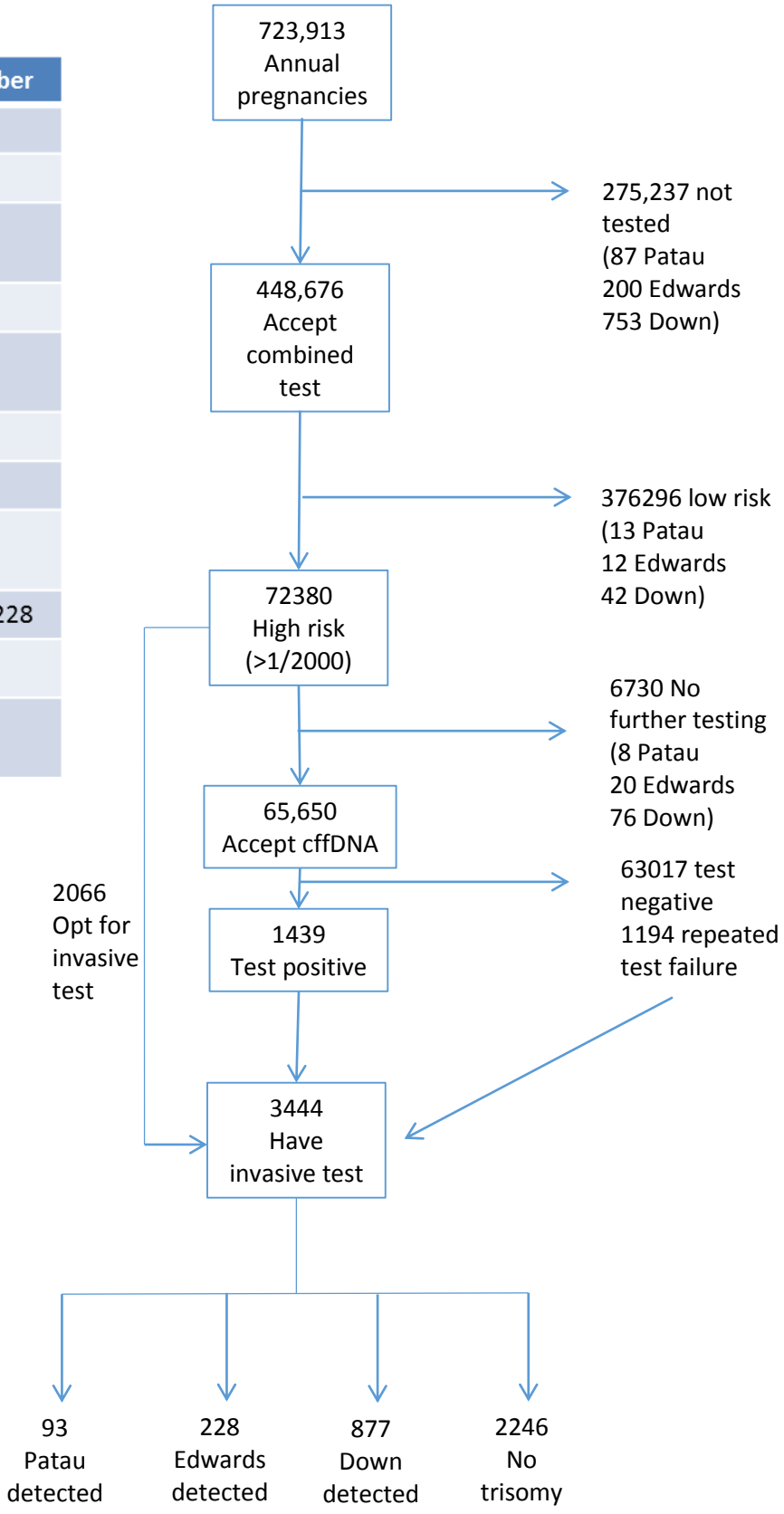


Figure 24 Simplified flow chart illustrating the economic model of combined test at threshold 1/2000 followed by cfDNA testing, using higher estimates of test accuracy without zero cell correction for TP and FN

Outcome	Number
Test-related miscarriage	6
Downs cases identified	962
Downs cases undetected (missed)	1019
Edwards cases identified	250
Edwards cases undetected (missed)	276
Patau cases identified	109
Patau cases undetected (missed)	120
Invasive tests/trisomy detected	1.7
Cost per case detected	£81,517
Total cost (£millions)	108
cffDNA initial test failures	13,410

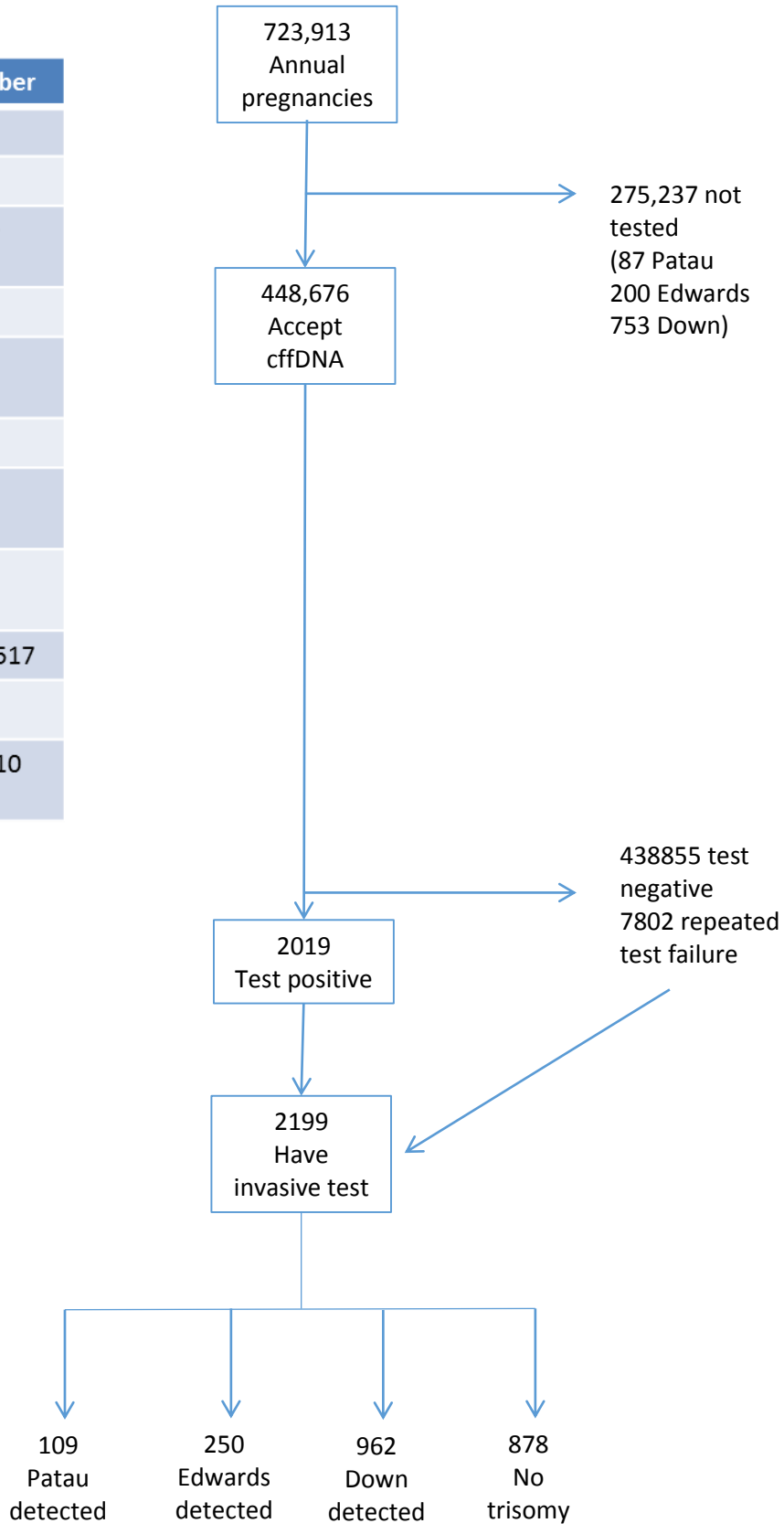


Figure 25 Simplified flow chart illustrating the economic model of cfDNA as the primary screen, using higher estimates of test accuracy without zero cell correction for TP and FN

5.2.2 Costs

We assume a cost per cfDNA testing of £232 in these calculations, but it is difficult to predict the actual cost to the NHS of cfDNA testing if adopted. The impact of varying this cost can be evaluated by considering the number of tests required. For example, the predicted number of cfDNA tests given a threshold of 1/150 is 10,262. If the cost of cfDNA testing were to fall to £100 per test, the total annual saving from introducing cfDNA testing would therefore be £1.25m. If the true cost of cfDNA testing were £500, however, this would add £2.8m to the annual cost of screening, so that the total additional costs of introducing cfDNA testing would be around £3m per year. There are also multiple sources of evidence on the cost to the NHS of invasive testing. Our reference case uses bottom-up costing by Chung (2012)[50] as this is likely to be a reasonable reflection of the economic cost. An alternative source is the NHS reference cost dataset provided by the Department of Health. The 2013/2014 day case reference cost for ante-natal diagnostic testing (including amniocentesis or CVS) is £515, which is £132 and £196 higher than our costs for amniocentesis and CVS respectively. The introduction of cfDNA testing is projected to reduce the number of such tests by 3095 (amniocentesis) and 3380 (CVS), so that using reference cost estimates would increase the annual projected savings in invasive testing from introducing cfDNA testing by £1.1m.

5.2.3 Risk of miscarriage from invasive tests

Table 10 illustrates the impact of changing the data on which our risk of test-related miscarriages are based. Using Danish registry data[52] rather than Australian Medicare data increases the expected risk, compared with the reference case (1.4% vs 0.6% for amniocentesis and 1.9% vs 0.7% for CVS). As a result, this change increases the predicted number of early miscarriages avoided in healthy pregnancies from 40-45 per year to around 105 per year. Conversely, data from Akolekar et al. (2015)[53] suggests the risks of miscarriage are much lower – 0.1% for amniocentesis and 0.2% for CVS. Assuming these risks reduces the number of miscarriages avoided to 10 per year.

Table 10 Impact of varying the risk of miscarriage associated with invasive testing on the total number of miscarriages avoided.

	Combined test alone	**cfDNA testing if combined test result >1/150**	**cfDNA testing if combined test result >1/1000**	**cfDNA testing alone**
Reference case	46.05 (29.98,68.85)	2.779 (1.169,5.739)	9.658 (4.45,17.74	7.462 (3.795,12.99)
Danish Registry Data	114.1 (107.4,121)	6.873 (3.288,12.85)	23.96 (12.8,39.04)	18.48 (11.01,27.6)
Akolekar et al. (2015)[53]	10.37 (0.6133,61.06)	0.5896 (0.03145,3.589)	2.108 (0.1175,12.03)	1.608 (0.09557,9.681)

5.2.4 Choices following the combined test results

Table 11 illustrates the impact of varying the source of data used to predict choices following a positive combined test. Gill et al. (2015)[48] found that only 57% accepted cfDNA testing when offered after a combined test result above the risk threshold, with 40% opting for invasive testing. If we base predicted decisions following combined testing on this study, the predicted number of trisomies detected by cfDNA testing following a combined test result of >1/150 rises from 1019 to 1214. This is because the proportion opting for no further testing is 2.7% compared with 6.4%. The predicted number of miscarriages of healthy fetuses rises from 2.8 to 30.6, driven by the numbers opting for an invasive test after combined testing.

Table 11 Predictions for annual FASP performance in England and Wales assuming cfDNA testing acceptance rates from Gil et al. (2015)[48] 95% credible intervals for mean estimates are provided in brackets

	Combined test alone	**cfDNA testing if combined test result >1/150**	**cfDNA testing if combined test result >1/1000**	**cfDNA testing alone**
Number of combined tests performed	448676	448676	448676	0
Number of initial cfDNA tests performed	0	6227 (5546,6904)	25330 (22610,27980)	448676
Number of cfDNA test failures	0	242 (184.2,320.1)	814.3 (693.6,947.4)	13410 (12230,14600)
Number of cfDNA retests performed	0	138.3 (98.24,191.6)	465.2 (360.6,581.2)	7646 (6622,8724)
Number of amniocenteses performed	3781 (3683,3879)	2764 (2313,3464)	9384 (8095,10730)	4092 (3234,5337)
Number of CVS performed	4129 (4022,4236)	3019 (2526,3783)	10250 (8840,11720)	4469 (3531,5828)
Total cost (£m)	14.93 (14.86,15)	15.66 (15.39,16.17)	25.01 (24.5,25.61)	108.9 (108.2,109.8)
T13 cases detected by testing	73.7 (64.4,83.5)	81.1 (61.39,99.25)	97.86 (75.23,118.2)	96.62 (56.52,124.1)
T18 cases detected by testing	196.4 (180.0, 213.6)	227.5 (177,265.6)	257.9 (200.8,297.7)	248.4 (146.5,306.4)
T21 cases detected by testing	763.9 (730.4,797.7)	905.4 (706.4,1020)	992.7 (775.2,1116)	974.1 (577.1,1172)
Total trisomies detected	1031.65 (964.08,1102.51)	1214 (944.79,1384.85)	1348.46 (1051.23,1531.9)	1319.12 (780.12,1602.5)
T13 cases undetected after testing	155 (135.1,177.6)	147.8 (125,174.1)	131 (109.7,157.3)	132.3 (104.8,175.8)
T18 cases undetected after testing	330.5 (301.5,361.5)	298.4 (259.2,353.9)	268.2 (230.1,327.6)	277.1 (222.4,378.6)
T21 cases undetected after testing	1219 (1163,1277)	1077 (961.7,1279)	989.3 (870,1207)	1009 (816.6,1405)
Test-related miscarriage of healthy pregnancy	46.05 (29.98,68.85)	30.63 (18.64,48.63)	122.5 (77.37,186.7)	48.51 (29.09,78.94)
Cost per trisomy detected through testing (£/trisomy)	14472 (13605,15414)	12900 (11676,16289)	18547 (16718,23306)	82555 (68518,138697)

6. Discussion

6.1 Decision problem and objectives

The overall objective of this review was to undertake a clinical and cost-effectiveness analysis of integrating non-invasive prenatal testing for the three most common trisomies (T21, T18, T13) into the UK prenatal screening pathway assessing three possible scenarios:

- cfDNA testing after the combined test
- cfDNA testing as a primary screen
- cfDNA testing in combination with the combined test

We aimed to systematically review the evidence on the diagnostic performance of cfDNA testing for the three trisomies and develop a cost-consequence model.

6.2 Summary of Methods and Findings

Systematic review

We searched MEDLINE, EMBASE and the Cochrane library from 1997 to 9th February 2015 and included 52 eligible articles which were heterogeneous in terms of study population, testing strategy and study quality. Fifty-one articles reported cfDNA testing performance for T21, 46 articles for T18, and 38 articles for T13. Seven studies included only first trimester pregnancies directly addressing the research question.

None of the 52 included articles were of optimal quality. The QUADAS-2 results are suggestive of a high risk of bias introduced particularly by exclusions from analysis and by the role of the sponsor. There were also significant concerns regarding applicability of the included patient spectrum to cfDNA testing introduction in the first trimester, as most studies had significant parts of their populations tested in the second or third trimester. Another concern for the head to head comparison of cfDNA testing performance with the first-trimester combined test was that all studies addressing this review question excluded multiple pregnancies and two of five studies used different screening tests to the combined test in some or all pregnant women.

Meta-analysis

Out of the 52 included articles, 41 qualified for the inclusion in the quantitative analysis. 2x2 data of diagnostic accuracy were extracted from the studies and pairs of sensitivity and specificity calculated which were subsequently pooled for T21, T18 and T13 using bivariate random-effects regression models. We used sensitivity, subgroup and meta-regression analyses to explore potential sources of heterogeneity in test accuracy estimates across studies including the following variables: study design, risk of trisomy pregnancy, multi-parity, gestational age, test type, and publication year. A series of sensitivity analyses were conducted to check the robustness of the results.

Pooled sensitivity was 99.4% (CI 98.9% to 99.6%) for T21, 97.4% (CI 95.8% to 98.4%) for T18, and 97.4% (CI 86.2% to 99.6%) for T13. The pooled specificity was 99.9% (CI 99.9% to 100%) for all three trisomies. These are without the zero cell correction for TP/FN and may be overestimates due to publication bias, risk of bias in included studies, and inclusion of second and third trimester pregnancies where there is a higher fetal fraction enhancing cfDNA test performance. We applied estimates of sensitivity and specificity without zero cell corrections to a theoretical cohort of 10,000 pregnant women from a high risk population as described in Table 6. Population prevalence was determined as the median prevalence for the studies included in high risk groups. In a high risk population with prevalence 3% for T21, 1.5% for T18 and 0.5% for T13 the positive predictive value was 91% for T21, 84% for T18 and 87% for T13, indicating that even when using the most favourable meta-analysis results which may be overestimates, cfDNA testing must not be considered a diagnostic test.

Economic Model

We constructed a decision tree to estimate the impact of alternative antenatal screening strategies on the annual performance of the England and Wales FASP. Our model predicts that implementing cfDNA testing after the current first trimester combined screening test with a 1/150 risk threshold would reduce the number of test-related miscarriages from 46 to 3 per year, but result in detection of 13 fewer trisomies and cost an extra £120,000 per year. Relaxing the threshold of the combined test to 1:1000 increases the number of trisomies detected by 93 but results in an additional cost to the NHS of £7million. Implementing cfDNA testing as the primary screening test for trisomy 21, 18 and 13 would cost an extra £105 million, including the £12 million cost of continuing the combined test for other reasons.

6.3 Strengths and Limitations

Systematic Review

We undertook comprehensive systematic searches for relevant evidence and updated our searches until April 2015 to maximise inclusion of emerging evidence in this fast moving field of research. We selected the papers for inclusion based on study size and methods most closely matching our research question. We had to rely on the reporting of the studies to indicate whether studies included participants already assessed in a previous study and as some studies were hospital based, and others used commercial laboratory based cohorts it is possible that there may be some overlapping samples in the included studies, but it was impossible to identify which studies (if any) this may apply to.

The findings of our review are in line with the results from previous reviews stating that cfDNA testing has high performance in terms of sensitivity and specificity,[102, 103] that specificity is slightly higher than sensitivity,[102] and that the test performance is greater for T21 than for T18 and T13.[104] However, in comparison to the recent meta-analysis by Gil et al. (2015)[104] our review included a greater number of studies, and meta-analysis resulted in slightly lower pooled sensitivities for T21, T18 and T13. Our review also found that cfDNA testing is less successful in twin pregnancies than in singleton pregnancies. This effect might be even more pronounced when twin pregnancies are not investigated separately in a retrospective study design.

The major limitation of this review is the lack of data to address some of our objectives. First and foremost there was no evidence on test accuracy when cfDNA testing and the first trimester screening tests are combined, so we were unable to model the integrated testing scenario. There was also a lack of evidence on cfDNA testing performance at the 1:150 risk cut-off from prior screening tests and no data was available of cfDNA testing performance for a range of thresholds. As studies either included heterogeneous high risk women, a mix of high and low risk women or women from the general obstetric population no analysis could be undertaken to investigate the impact of risk threshold on cfDNA testing performance.

Another limitation of this review is the limited evidence on head to head comparisons of cfDNA testing and first trimester combined testing. We were unable to pool data from the available three studies due to extensive heterogeneity and the low number of studies. The assessment of the comparative diagnostic performance (decision question 2) therefore is limited to the narrative review of the available evidence. These included some of the largest and most influential studies.

The main limitation when appraising the quality of included studies using the QUADAS-2 tool was that it relied on published data. Some studies may have received an unclear or high risk of bias classification due to unclear reporting, while the study itself may have been of good quality. We did not try to overcome this by contacting investigators to obtain more detailed information because of

the magnitude of additional work this would have created considering the number of included articles. Another concern regarding the reliability of the studies is the large proportion of studies that were sponsored by manufacturers of cfDNA testing which will inevitably bias the results.

Meta-analysis

The strengths of the meta-analysis included using rigorous methods of data analysis, including bivariate random-effects regression models and HSROC curve analysis. A particular strength of this review is the extensive investigation of heterogeneity within the meta-analysis using sensitivity and subgroup analyses to test the robustness of our pooled estimates of sensitivity and specificity. 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 in accuracy estimates. However considerable statistical heterogeneity was observed in the pooled accuracy estimates, especially for sensitivity estimates. Despite our attempts to explain this, substantial heterogeneity remained unexplained. For some of the subgroup analyses, a relatively small number of studies were available limiting 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 and this tends to underestimate test accuracy. This was necessary to enable model convergence. Therefore we produced two versions of our economic model for test accuracy estimates with and without the zero cell corrections for TP/FN. Accuracy estimates with the zero cell correction may be underestimates, but estimates without may be overestimates due to high risk of bias, and inclusion of studies in the later trimesters of pregnancy.

The consideration of the potential influence of test failures and indeterminate results on the diagnostic performance of cfDNA testing in additional sensitivity analyses is a further strength of the review. The investigation of the potential impact was hampered by inconsistent reporting of test failures and the lack of reference standard outcomes for many cases. 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. Whilst this may be an inaccurate assumption, it is preferable to ignoring the indeterminate results and test failures altogether. A further challenge was to identify the categories of 'test failures' that are important in terms of influencing test accuracy if excluded from the analysis and those that are only important from the economic point of view, i.e. requirement to retest.

The finding that the pooled sensitivity was lower among recent studies published between 2014 and 2015 investigating T18 compared with studies published between 2007 and 2013 (89.6% versus 95.4%; p=0.056) is slightly surprising as it was expected that the accuracy of cfDNA testing might have increased in recent years. However, this might be explained by a greater proportion of studies reporting clinical experience in the latter years compared to retrospective and controlled study designs in earlier years, producing study outcomes which are closer to the truth.

The accuracy of cfDNA testing for T21 and T18 in the general population and in high risk pregnancies is similar. This is in line with findings from other studies which compared performance of cfDNA testing in the general obstetrical population with test performance in high-risk women.[24, 67, 76, 79, 90] It has been suggested that this is due to similar fetal-fraction distributions in the low-risk and high-risk subgroups[67] and others have concluded that the ability to detect trisomies with cfDNA testing is dependent on assay precision and fetal fraction in the sample rather than the prevalence of the disease in the study population.[76] However, our meta-analysis did find statistically significantly lower sensitivity for T13 in the general obstetric population in comparison to high risk groups.

As the objective of this review was to investigate the diagnostic performance of a 'generic' cfDNA test we have pooled all studies using different testing strategies. The testing strategies are however

substantially different from each other and our method of overall pooling may be considered inappropriate due to heterogeneity. However, sub-group analyses showed no systematic differences between performances of different test types.

Economic model

Limitations of the economic model include the assumption that the sensitivity and specificity of cfDNA testing are the same regardless of the population tested. In our meta-analysis we found differences in performance in high risk compared to general obstetric population, but these were not statistically significant for T21 and T18. As a result of this T13 outcomes from the model scenario with NIPT as the primary screen should be interpreted with caution.

Furthermore, we used failure rates from the largest study reporting failure rates by trisomy. This study reported a failure rate of 10% in trisomic pregnancies and 3% in pregnancies without trisomy. However, there were only 6 cases in total in this study that were test failures in trisomic pregnancies. Generally, there were large differences in reported failure rates in different studies, and we assume in the economic model that failure rate does not differ between the different types of trisomy or that the failure rate increases with increasing prevalence / risk. This would be of importance if failure is more likely in trisomy cases.

In our economic model we assumed that the delay caused by cfDNA testing does not change the proportion of invasive tests that are amniocentesis rather than CVS. Although adding cfDNA testing after the combined test would delay the timing of a potential subsequent invasive test, and amniocentesis is performed later in pregnancy than CVS we did not include an adjustment for this. The rationale is that the risk of fetal miscarriage associated with amniocentesis is slightly lower than that for CVS, and therefore the longer the delay in our model, the greater proportion of invasive tests would be amniocentesis, and so the lower the risk of procedure related miscarriage. This would create an advantage in our model for the addition of any test that incurs a delay, even a useless one, which is undesirable, particularly as it is preferable to produce results as soon as possible.

Costs that are not included in the modelling include start-up costs, midwife time and counselling. However, these are the key things that a follow up impact assessment would need to consider.

The choices women will make when offered cfDNA testing in routine practice are difficult to predict in advance. In our economic model we used data from a survey of women at antenatal clinics that 91% would take cfDNA testing if offered it after a combined test risk score of >1/150, just 3% would opt to go straight to an invasive test, and 6% would opt for no further testing. These were opinions in response to a survey rather than actual choices and may be heavily dependent on the information given about cfDNA testing. Therefore we completed a sensitivity analysis using choices made in a study offering cfDNA testing in the NHS 46 where 57% accepted cfDNA testing, 40% opted to go straight to the invasive test and just 3% opted for no further testing. We did not use this in the primary analysis because these choices were based on cfDNA testing presented as part of a research study and not part of routine practice.

Decisions about whether to have cfDNA testing or go straight to invasive may be dependent on the risk score from combined test. There were no reliable data on which to base this so it was not included in the model. The effect of our simplification of this could be an overestimate of test-related miscarriages in the models including the cfDNA test because women receiving higher risk scores from the combined test may be more likely to elect to go directly to an invasive test than women with lower risk scores, which was not reflected in the model. We also assumed decisions would be identical following cfDNA test failure.

Finally, we were unable to model the scenario in which cfDNA testing is integrated into the first trimester combined test due to lack of evidence demonstrating test accuracy after implementing of such approach. We remain therefore uncertain if combining the conventional screening tests with the cfDNA test would offer an increased accuracy.

6.4 Implications for policy and practice

There is agreement that cfDNA testing achieves high sensitivity and specificity in detecting the three trisomies. However, there are several considerations that need to be taken into account before implementation into clinical practice is feasible. These include technical considerations, clinical considerations, time, handling of FP and FN test results, test performance measure in clinical practice, the testing sequence and women's choices.

Technical considerations

Two technical considerations are test failures and multiplexing. Test failures are often associated with low fetal fractions because DNA levels are the limiting factor for successful analysis. Fetal fraction varies between different individuals and throughout pregnancy.[60] It has also been shown that the fetal fraction is lower in cases of discordant results between cfDNA testing and invasive testing than in cases with concordant results[90] and that the frequency of trisomy may be higher in samples with low fetal fraction.[46, 79] To overcome this limitation, samples should be either retested or invasive tests offered.[79] Test failures due to human error need to be minimised by effective quality assurance procedures to reduce costs and delays associated with retesting. The studies included in this review reported a wide range of test failure rates, and minimising test failures would be key to any implementation of cfDNA testing.

The level of multiplexing affects the sequencing depth and an optimal level of multiplexing needs to be determined before consideration of cfDNA testing for implementation. While increased levels of multiplexing (lower sequencing depth) can reduce costs[60] it has also been shown to decrease the accuracy of the MPSS technology.[57] Therefore, a balance should be found between optimal test performance and costs of testing.[60] Liao et al. (2014)[60] therefore pointed out that it is important to determine the fetal fraction in the sample to be analysed and the sequencing depth required for analysis.

Clinical considerations

There is some indication in the literature that test failures might be associated with high BMI and trisomy risk, however, the evidence is currently insufficiently to formally assess this issue.

Time

Our model showed that adding cfDNA testing to the current screening path could potentially delay the diagnosis of 927 trisomies.

Handling of FP and FN test results

High sensitivity is desirable in screening tests in order to reduce the number of missed cases (FN). However, sensitivity of cfDNA testing is not 100% and while the number of missed trisomies when cfDNA testing follows the combined testing might be more than off-set by the avoided test-related miscarriages in cfDNA testing true negatives from the public health point of view it presents a real problem for the management of individual patients. Further research is needed to investigate whether FNs from cfDNA testing are associated with a defined combined test risk cut-off or whether initial conclusions hold that FPs and FNs are unrelated to trisomy risk which will help shape the management options for patients with positive combined screening results and negative cfDNA testing outcomes.

The number of FP is also of great concern in the case of trisomy detection as patients will be exposed to increased anxiety and pregnancies put at risk by unnecessary invasive procedures. Our model showed that cfDNA testing reduces the number of invasive procedures by approximately 85%. This is in agreement with the notion that the major benefit of cfDNA testing is the reduction of unnecessary invasive procedures by reducing the number of false positives being referred for diagnostic testing. However, studies seem to disagree on the scale of this benefit. While Dan et al. (2012)[68] estimated that over 98% of invasive tests could be avoided because only 1.17% (190/11,105) of the study population who were classified as high risk for trisomy 21 or 18 needed invasive testing, Comas et al. (2014)[85] predicted a reduction by 33%. Because trisomies are rare events, a slight change in the specificity can have a huge impact on the number of FPs.[89] In addition to lab errors, FPs can be due to confined placental mosaisism because the source of cff-DNA is the placenta.[89] Furthermore, early fetal demise of an affected twin can lead to a FP result of the unaffected fetus.[59, 89] The low fetal fraction on the other hand is not believed to be a major contributor to FP or FN results.[24] Before consideration of implementation it is therefore of importance to consider the uncertainty around the pooled estimate of the specificity.

Test performance measures in clinical practice

As previously discussed, prevalence may have some limited effect on test performance in terms of sensitivity and specificity. However, the utility of these diagnostic measures in clinical practice needs to be considered as these are not the only important measures of test performance and they are less frequently used in clinical practice. Clinically more intuitive are positive and negative predictive values. These however are dependent on the prevalence of trisomy in the population tested. Predictive values will therefore be considerably lower when the tested population is the general obstetric population than if cfDNA testing is performed in high risk women following the combined screening test. Using only studies in high-risk populations positive predictive values were 91%, 84%, 87% for T21, T18 and T13 respectively. Others reported PPVs of 45.5% for T21 and 40.0% for T18 for a general obstetric population with low trisomy prevalence.[67] A positive predictive value for other risk groups can only be estimated for theoretical values of prevalence as the true prevalence (as well as the true risk score for individual patients) is unknown in the testing population. The estimated PPVs for different prevalences will therefore only illustrate the trend the PPV would take with increasing prevalence but are not informative for the management of individual patients.

Sequence of testing

It seems unlikely that combined testing will be replaced by cfDNA testing not only due to increased costs but also because of other benefits of combined screening including: dating pregnancies, detection of other fetal defects (39% of the abnormal karyotypes identified by invasive testing a were abnormalities other than T21 and T18 [78]) and pregnancy complications.[54] Rather, the role of cfDNA testing is seen as part of a comprehensive risk assessment.[74] Therefore, our model of cfDNA testing as the primary screen included costs for the combined test as in practice it is unfeasible to remove the NT scan and possibly the blood tests. However, this does not affect the conclusions of this review as this strategy is not considered cost-effective and the combined test only contributed £12 million to costs, with £93million attributed to the cost of cfDNA test.

When considering whether cfDNA testing should be an add-on test or integrated into the combined test, performance, accessibility and costs should be taken into account. The add-on strategy has got the advantage that cfDNA testing could be restricted to an intermediate risk group, whereby low risk women receive no further testing, the very high risk group are offered invasive testing and the intermediate risk group will get cfDNA testing with the aim to classify women as very high or very low risk.[69] However, we do not have evidence in support of this strategy that would suggest the cut-offs for the three risk groups and adding a test to the current screening pathway might delay obtaining diagnostic results further into the second trimester and subsequently reduce time for

decision making on whether to keep or terminate an affected pregnancy. However, integration of results will require a tested algorithm to calculate an overall combined risk score and guidance how this risk score should be interpreted by clinicians.

The economic model does not investigate the delays to the diagnostic process caused by implementing cfDNA testing, but we have modelled the number delayed. This may be an important consideration in a screening programme.

Women's choices

It was outside the scope of this review to investigate women's choices and ethical issues following the introduction of cfDNA testing into routine clinical practice. Our model does not include the option of women to undergo elective abortion following a positive cfDNA test result and declining confirmation by invasive testing. If this becomes a real possibility with implemented cfDNA testing, the benefit of avoided test-related miscarriages is likely to be completely reversed. Our meta-analysis indicated that in populations with prevalence 3%, 1.5%, and 0.5% for T21, T18 and T13 respectively around 9%, 16% and 48% of pregnancies that test positive using cfDNA would not actually have a trisomy. Therefore, effective communication to clinicians and pregnant women that cfDNA testing is not a diagnostic test is of fundamental importance for implementation.

7. Conclusions

Evidence suggests that cfDNA testing can offer substantial benefits in the trisomy screening pathway, but that it is not appropriate for use as a diagnostic test. cfDNA testing is highly accurate in a number of relevant populations and is likely to be valuable if offered to those that test positive on the combined test. Using cfDNA testing as the primary screen is likely to be prohibitively expensive and not offer the same benefit to the total numbers of false positives when compared to a contingent strategy.

8. Tables

Table 12 Study characteristics

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 cfDNA testing or comparison of cfDNA testing with CT
Alberti 2015[66] 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 fetal T21. N=225 in case-control for sequencing Mean age (SD): 35.2 (6.7) years. Mean gestational age (SD) 14 (2) weeks 1st and 2nd trimester	N=0 from cohort. N=751 (76.9%): Not included in case-control study	T21	All high risk for fetal 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 fetal karyotype	none	cfDNA testing performance for T21 detection	Accuracy of cfDNA testing
Ashoor 2012[54] UK Study start date: NR	Nested case-control of stored maternal samples: Controls matched with T21/T18 cases for sample storage time in 3:1 ratio. Number of centres: 1	N=400 (50 T21, 50 T18, 300 euploid) Singleton pregnancies High-risk women; 1st trimester 100%; All 11-13 weeks' gestation Ethnicity: White 89%, 'Afro Caribbean' 5%, South/ East Asian 6%, Mixed 0.5%	Pregnant by IVF or multiple pregnancy N=NR	T21, T18	All high risk: Combined 1st trimester screen risk >1:300	DANSR FORTE Aria Diagnostics (USA)	Karyotyping after CVS	None	FORTE risk score for aneuploidies, sensitivity and specificity for detection of T21 and T18	Accuracy of cfDNA testing
Beamon 2014[83] USA	Prospective cohort	N=208 High-risk pregnancies who chose cfDNA	Multiple Pregnancy N=NR	T21, T18, T13	All high-risk: AMA: 148 (71.2%),	MPS (whole genome)	Karyotyping after amniocent	None	Test performance for T13, T18	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
Study start date: January 2012	Number of centres: 1	testing as triage test, singleton or dichorionic twin gestations, ≥10 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: 1st: 111 (53.4%), 2nd: 95 (45.7%), 3rd: 2 (1%)			AMA alone: 121 (58.2%), 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: 1 (0.5%)	Sequenom Center for Molecular Medicine (USA) (n=163, 78.4%) or Verinata Health (USA) (n=45, 21.6%)	esis, cordocentesis or CVS, phenotype of newborn		and T21 detection	
Bevilacqua 2015[84] 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 1st trimester: 68.5%	Criteria for exclusion from study NR	T21, T18, T13	Mixed risk: High risk for fetal trisomy by 1st-trimester combined test or 2nd-trimester triple/quadruple test or ultrasound or cfDNA testing 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) cfDNA testing performance for T13, T18 and T21 detection in	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
									twins	
Bianchi 2012[55] USA Study start date: NR	Nested case-control Controls un matched in 4:1 ratio (Part of MELISSA prospective cohort). Number of centres: 53 (of 60)	N=2,882 in cohort. N=534 in nested case-control study Singleton pregnancies, high risk. Mean age (SD), range: 35.2 (6.40), 18 – 46 years Mean gestational age (SD), range: 15.1 (3.16), 10 – 23 weeks Trimester: 1st 165 (30.9%), 2nd 369 (69.1%). Ethnicity: White 72.7%, African American 10.9%, Asian 9.9%, Native American or Alaska Native 0.9%, Multiracial 5.6%	257/2,882 (8.9%) from MELISSA cohort: 85 multiple pregnancies, 45 no karyotype information, 127 ineligible blood sample	T21, T18, T13	All high risk: Advanced maternal age only (age >38 years) 152 (28.5%); Positive screen risk 91 (17.0%); Ultrasound abnormality 122 (22.8%); Prior aneuploidy pregnancy 15 (2.8%); More than 1 risk 154 (28.9%)	MPS (whole genome) Verinata-Illumina (USA)	Karyotyping after CVS	None	1) MPS performance (sensitivity and specificity) for T21, T18 and T13 detection 2) Sex chromosome classification and Monosomy X detection	Accuracy of cfDNA testing
Bianchi 2014[67] 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: 1st: 759 (39.7%), 2nd: 610 (31.9%),	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 No risk	MPS (whole genome) Verifi Verinata-Illumina (USA)	Newborn phenotype (97.0%), Karyotyping (3.0%)	Standard prenatal aneuploidy screening produced by accredited clinical laboratories. Cutoff values as	1) Comparison of false positive rates of cfDNA testing with conventional screening for T21 and T18	Comparison of cfDNA testing 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 cfDNA testing or comparison of cfDNA testing with CT
		3rd: 545 (28.5%) Mean gestational age (SD), range: 20.3 (8.6), 8.0 – 39.4 weeks Mean age (SD), range: 29.6 (5.54), 18.0 – 48.6 years Assisted conception 66 (3.4%)			thresholds given			used by individual laboratories 1st-trimester: Combined test (PAPP-A, β-hCG, NT) N=739 (38.6%) 2nd-trimester: Quadruple (MS-AFP, β-hCG, estriol and inhibin A) N=439 (22.9%); Quadruple + combined test N= 53 (2.8%); Quadruple + 1st-trimester serum markers only N=164 (8.6%); Sequential: 1st-trimester screen results reported before final report in 2nd trimester N=519 (27.1%)	2) Comparison of false positive rates for T13. Comparison of fetal fractions in low-risk with high-risk patients	
Chen 2011[56] Hong Kong, UK, Netherlands, China Study start date: NR	Case-control of stored samples and prospectively recruited women	N=392 (N=140 archived plasma samples with and without aneuploidy matched for gestational age; N=252 prospectively recruited.)	NR	T18, T13	All high risk based on clinical indicators as per the existing obstetric practice of each recruitment unit	MPS (whole genome) Sequenom (USA)	Karyotyping after CVS or amniocentesis	None	Diagnostic performance of MPS for T13 and T18 detection.	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
	Number of centres: 10	344/392 samples analysed in a previous study [57], 48 cases newly recruited Singleton pregnancy undergoing CVS/amniocentesis								
Chiu 2011[57] Hong Kong, UK, Netherlands, China Study start date: October 2008	Case-control of stored samples and prospectively recruited women Number of centres: 10	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 samples), N=764 included. Singleton pregnancies Median age: 35.4 years Median gestational age: 13+1 weeks 1st trimester: 74%	N=60 (7.3%): 14 failed recruitment criteria (2 twin pregnancies, 12 without full karyotyping); 46 compromised blood sample (3 samples collected after invasive obstetric procedure, 2 delayed blood processing, 3 with ambiguous information, 12 haemolysed, 26 inadequate volume)	T21	High risk by conventional screening (>1:300): 582 (77%), Median risk for T21: 1 in 43 Intermediate risk by conventional screening (1:300-1:1000) 39 (5%), Median risk for T21: 1 in 502 Other indications (previous T21 pregnancy, ultrasound abnormalities, risk for monogenic diseases).	MPS (whole genome) Sequenom (USA)	Full karyotyping after amniocentesis (18%) or CVS (82%)	None	Diagnostic sensitivity, specificity, PPV & NPV for T21 detection	Accuracy of cfDNA testing
Comas 2014[85] Spain Study start date: January 2013	Prospective cohort Number of centres: 1	N=333 Singleton pregnancies who chose to have cfDNA testing Mean maternal age	Multiple pregnancies, ultrasound anomalies or high risk of congenital malformation	T21, T18, T13	Routine general population in a real clinical setting 83.5% Low-risk	DANSR FORTE (Harmony Prenatal Test), Ariosa	Invasive testing and karyotyping, newborn phenotype	None	1) cfDNA testing performance for T13, T18, and T21	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
		(range): 37 (21-46) years Mean gestational age (range): 14.6 (9.5-23.5) weeks 1st and 2nd trimester	N=NR		by conventional screenings but unable to alleviate their anxiety 16.5% High-risk from CT or referred for AMA with no prior screening	Diagnostics (USA) (n=120, 36.0%) or SNP- and NATUS (Panorama) Natera Inc. (USA) (n=213, 64.0%)			2) Comparison of Harmony and Panorama tests, factors influencing fetal fraction	
Dan 2012[68] China Study start date: 1st 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 Median gestational age (range): 20 (9-28) weeks. 2nd trimester: >74% 42/49 centres offered test to high-risk pregnant women identified by a conventional T21 screening test, 7/49 centres enrolled participants regardless of prior risk assessment	N=79 (0.7%): 55 unqualified gestational age, 14 multiple pregnancies, 10 fetal 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 more other risk factors (≥ 35 y), family history of aneuploidies, ultrasound abnormalities): 2,770 (24.9%) No – without any risk factors: 1,387 (12.5%)	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 cfDNA testing
Dar 2014[86]	Retrospectiv	N=31,030 samples	N=325 (1.0%):	T21, T18,	Mixed high- and	SNP and	Karyotype	None	1) Foetal	Accuracy 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 cfDNA testing or comparison of cfDNA testing with CT
USA Study start date: March 2013	e analysis of prospective lab cohort Number of centres: NR (1 lab)	received for commercial cfDNA testing, N=30,705 accepted. Singleton pregnancies ≥ 9 weeks' gestation Mean age (SD): 33.3 (6.0) years; Median age (range): 35.0 (14.0-60.0) years Mean gestational age (SD): 14.0 (4.4) weeks; Median gestational age (range): 12.6 (3.1-40.9) weeks Trimester: 1st: 20,001 (64.5%), 2nd: 10,479 (33.8%), 3rd: 550 (1.8%)	127 Insufficient serum/plasma, 70 <9 weeks gestational age, 45 Test cancelled, 28 Sample collection date too old, 11 Missing information, 4 Sample damaged, 4 Wrong tube, 8 Multiple gestation, 1 Egg donor, 1 Surrogate, 26 Other	T13	low-risk: AMA (≥ 35 years) 51.4%. ICD-9 codes in 5,468/28,739 (19.0%) of women: Low-risk 16.6%, High-risk based only on AMA 44.1%, High-risk codes 39.3%.	NATUS Panorama Natera Inc. (USA)	information or at-birth clinical evaluation		fraction and its associations. 2) cfDNA testing results, PPV.	cfDNA testing
Del Mar Gil 2014[69] UK Study start date: NR	Retrospective cohort of stored samples Number of centres: 1	N=207 Twin pregnancies undergoing first-trimester screening for trisomies by combined test Age range: 26 – 41 years Gestational age, range: 11 - 13 weeks 1st trimester: 100% Ethnicity:	Singleton pregnancies N=NR	T21, T18, T13	NR	DANSR FORTE Harmony Ariosa Diagnostics (USA)	Known birth outcome	None	Performance of Harmony Test in twin pregnancies only	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
		Caucasian 70.0%, Afro-Caribbean 23.7%, South/East Asian 1.0%, Mixed 5.3%								
Dhallan 2007[97] USA Study start date: January 2004	Prospective observational study (cohort?) Number of centres: 10	N=60 Women ≥ 18 years, singleton pregnancy Mean age (range): *32.8 (18-43** years, Mean gestational age (range): *19+6 (8+1 - 38+6)* weeks, 1st trimester: 8 (13%)	N=NR	T21	Mostly high risk. Definition unspecified.	SNP allelic ratio Ravgen Inc. (USA)	Amniocentesis or newborn reports	None	Performance of SNP method in detecting T21	Accuracy of cfDNA testing
Ehrich 2011[58] 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 3rd-party database Pregnancies at increased risk for fetal aneuploidies with scheduled invasive diagnostic procedure Median age (range): 37 (18 -47) years Median gestational age (range): 16 (8-36) weeks	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 cfDNA testing
Fang 2015[95] China Study start date: October 2012	Prospective cohort Number of centres: 1	N=25,149 for prenatal screening, N=1,512 with cfDNA testing Pregnant women < 35 years and high-risk 2nd-trimester dual	N=23,637 (94.0%): No cfDNA testing.	T21, T18	For cfDNA testing: All high-risk for fetal aneuploidies: Maternal age ≥ 35 years (high and	MPS (whole genome) Company NR	Amniocentesis and karyotyping or follow-up after delivery	none	Concordance between cfDNA testing-positive cases and invasive testing	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
		serum screening (≥1:270 for T21 or ≥ 1:350 for T18) or as first screening test for pregnant women ≥ 35 years who are refusing direct invasive prenatal testing All 2nd trimester			critical risk) 526 (34.8%); Maternal age < 35 years and high risk on 2nd-trimester dual serological screening (≥ 1:270 for T21 or ≥ 1:350 for T18) 986 (65.2%)					
Futch 2013[70] USA Study start date: February 2012	Cohort of lab submissions Number of centres: NR	N=6,123 samples submitted for commercial cfDNA testing N=6,017 accepted. Singleton pregnancies with a gestational age of ≥ 10 weeks 0 days accepted Mean age (SD), range: 35.0 (5.7), 14.6-51.7 years. Mean gestational age (SD), range: 15.6 (4.6), 5-37 weeks 1st trimester: 2,883 (47.2%)	N=106 (1.7%): 4 multiple pregnancies, 9 gestational age <10 weeks, 3 duplicates, 26 improper / no labelling, 5 transport issue, 15 cancelled by referring agency, 43 inadequate blood volume, 1 wrong sample type	T21, T18, T13	Population met high risk criteria (389 with indication of cystic hygroma) Threshold and criteria NR	MPS (whole genome) Verifi® prenatal test Verinata Health Inc (now Illumina) (USA)	Karyotype or birth outcome	None	Test performance in clinical laboratory setting	Accuracy of cfDNA testing
Hall 2014[59] USA Study start date: March 2012	Nested case-control (selected from a cohort of >1000 women, all	N=68 (17 T13, 51 euploid) High-risk pregnancy couples, women ≥ 18 years, singleton pregnancy	N=1/>1,000 (<0.1%) from cohort: 1 known fetal mosaicism	T13	High-risk for fetal aneuploidy (positive serum screen, ultrasound abnormality or maternal age of greater than 35	SNP- and NATUS Natera Inc. (USA)	CVS, amniocentesis or genetic testing of cord blood,	None	1) Test performance for T13 detection 2) Specificity of T18, T21	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
	T13 cases matched 1:3 on gestational age) Number of centres: NR	Median gestational age (range): 16.0 (12.1-22.7) weeks, 1st trimester: 23 (35.9%)			years)		buccal, saliva, or products of conceptio n		and Monosomy X detection	
Huang 2014[71] China (Denmark, Hong Kong) Study start date: NR	Prospective, multicentre cohort Number of centres: 7	N=189 Twin pregnancies requiring invasive procedure (CVS/ amniocentesis) Median age (range): 31 (22-44) years Median gestational age (range): 19 (11-36) weeks 1st trimester: ≥ 2.1%, 2nd trimester: ≥ 74%	N=NR Intrauterine death, without fetal karyotype	T21, T18	All high risk Threshold and risk establishment NR	MPS (whole genome) NIFTY test BGI-Shenzen (China)	Full karyotypin g from CVS (2.1%), amniocent esis (94.2%), or cordocent esis (3.7%)	None	Test performance for T18 and T21 detection in twin pregnancies	Accuracy of cfDNA testing
Jensen 2013[98] USA Study start date: April 2009	Unclear study design (Part of an international clinical validation study with 4,385 women, NCT008772 92) Number of centres: 27	N=4,664 enrolled in cohort, N=1,269 included in this study Women ≥ 18 years, singleton pregnancy, 1st or 2nd trimester (≤ 21 weeks 6 days), at increased risk for fetal aneuploidy scheduled for diagnostic invasive testing. Samples overlap with two studies by Palomaki et al.[61, 62]	N=279/4,664 (6.0%) from cohort: 116 sample not adequate, 112 multiple gestation / fetal death, 51 no karyotype /outcome available; N=3,116/4,385 (71.1%): Not included in this study	T21, T18, T13	All high risk (screen positive for Down syndrome or other trisomy by serum and/or ultrasound testing, maternal age of ≥38 years at delivery or a family history of aneuploidy)	MPS (whole genome) Sequenom, Inc. (USA)	Karyotypi ng from CVS or amniocent esis	None	Test performance for T21, T18 and T13 detection from high-throughput assay analysed using new bioinformatics algorithm	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
Jeon 2014[87] 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), range: 30.73 (4.99), 19-43 years Trimester: 1st: <18.1%, 2nd: >55.5%	NR	T21, T18	High risk of fetal defects by standard aneuploidy screening with individual risk scores and interpretations produced by accredited clinical laboratories	MPS (whole genome) Semiconductor sequencing	Amniocentesis and fetal karyotyping	None	T18 and T21 detection by semiconductor sequencer Ion Proton (PPV, NPV)	Accuracy of cfDNA testing
Jiang 2012[72] 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 fetal fraction and gestational age	Accuracy of cfDNA testing
Korostolev 2014[88] Russia Study start date: 2012	Prospective cohort Number of centres: NR (Moscow private clinics)	N=1,968 included, N=1,728 for cfDNA testing Women with singleton pregnancies, high risk for aneuploidies, >9 weeks' gestation	N=240 (12.2%): Ultrasound abnormality (increased NT, heart defects, malformations, fetal growth retardation) or presence of balanced	T21, T18, T13	Mixed risk: High risk result of combined FTS 87%, AMA ≥ 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 assessmen	None	cfDNA testing and/or invasive test based on CMA for chromosomal abnormalities diagnostics	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
		Mean age (range): 34.4 (26-45) years Mean gestational age (range): 14 (9-33) weeks. 1st trimester: "about 50%"	chromosomal rearrangements in the parents				t (n=624), TOP and molecular study (n=1)			
Lau 2012[73] Hong Kong, China, Japan Study start date: NR	Prospective cohort Number of centres: 1	N=108 Pregnant women undergoing CVS or amniocentesis (possibly singleton pregnancies but NR) Mean age (SD): 37 (4.3) years, Median gestational age (range): 12+5 (11+4 – 28+0) weeks 1st trimester: 97 (89.8%)	NR	T21, T18, T13	Mostly high risk: Positive 1st trimester screening 47.2%, positive 1st trimester sonographic markers 22.2%, other structural anomalies 1.5%, previous T21 0.9%, maternal anxiety 11.1%	MPS (whole genome) BGI-Shenzhen (China)	Conventional karyotyping from CVS (94.4%) or amniocentesis (5.6%)	None	Diagnostic accuracy of novel z-score method with internal reference chromosome.	Accuracy of cfDNA testing
Lau 2014[74] 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 cfDNA testing, regardless of whether they had undergone any previous T21 screening Mean age (SD),	NR	T21, T18 T13	Prenatal diagnosis centre accepted referral of any pregnant woman for cfDNA testing: Previous trisomy / Family history 53 (2.7%) No prior screening test: 669 (34.2%) Prior screening	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 cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
		range: 36 (4.35), 20-46 years. Median gestational age: 14.5 weeks. 1st trimester: 56.25% Ethnicity: Chinese 90.91%, Caucasian 5.21%, Other 3.88%			test 1,290 (65.8%): High risk 593/1,290 (46.0%), Low risk 368/1,290 (28.5%), Result not available yet 329/1,290 (25.5%)					
Liang 2013[75] 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. 1st 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 fetal karyotyping	None	Test accuracy for detection of fetal aneuploidies for all 24 chromosomes in one single sequencing event	Accuracy of cfDNA testing
Liao 2014[60] China, USA Study start date: NR	Retrospective case-control (unclear how matched) Number of centres: 2	N=2,275 included in study N=515 with karyotyping in retrospective case-control (55 T21, 16 T18, 3 T13, 15 SCA, 426 euploid) Pregnancies with karyotyping results	Exclusion criteria: IVF, blood transfusion or immunotherapy within 12 months N=NR N=1,760/2,275 (77.4%):	T21, T18 T13	All high risk: Advanced maternal age (>30 years); History of previous miscarriage; Positive serum marker screening; Abnormal fetal	MPS (whole genome) Semiconductor sequencing	Full karyotyping	None	Test performance for detection of T21, T18, T13 and SCA	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
		No other demographic information	No karyotyping results		ultrasound results					
McCullough 2014[89] USA Study start date: August 2012	Retrospective study of prospective lab cohort Number of centres: NR (1 lab)	N=100,000 samples submitted for commercial cfDNA testing Singleton and multiple pregnancies, ≥ 10 weeks' gestation, high-risk for fetal aneuploidies Mean age: 35.1 years Mean gestational age: 15+3 weeks Trimester: 1st 54.1%, 2nd: 43.4%, 3rd: 2.5% Singleton pregnancies: 96,470 (96.5%), Multiple pregnancies: 3,530 (3.5%)	N=870 (0.9%): Tests cancelled "No indication for testing"	T21, T18, T13	All high-risk: Maternal age ≥ 35 years for singleton pregnancies, ≥ 32 years for twins, ≥ 27 years for triplets and more: 59.7%; Ultrasound findings 13.9%; Positive serum screening 11.3%; Personal or family history 4.0%; Multiple indications 10.1%	MPS (whole genome) MaterniT21 ® PLUS Sequenom Inc. (USA)	Karyotype or phenotypic assessment at birth	None	Operational laboratory performance (turnaround time, non-reportable rate), cfDNA testing autosomal aneuploidy positivity rate by indication, test accuracy estimation and modelled test performance at various fetal fractions	Accuracy of cfDNA testing
Nicolaides 2012[76] UK Study start date: October 2010	Retrospective cohort of stored samples Number of centres: 1	N=2,230 original cohort, N=2,049 eligible cases Women with singleton pregnancies attending for first-trimester combined screening for aneuploidies and ultrasound (general	N=181 (8.1%): 74 no fetal karyotype, 7 abnormal karyotype other than T21 or T18, 29 inadequate sample volume, 1 wrongly labelled 70 lab mixed	T21, T18	General obstetric population undergoing first-trimester screening for aneuploidies as part of their routine antenatal care All had 1st-	DANSR FORTE Harmony Prenatal Test Ariosa Diagnostics (USA)	86 (4.2%) CVS or amniocentesis and fetal karyotyping. 1963 (95.8%) phenotypic newborn	First-trimester CT (free β-hCG, PAPP-A, NT) with or without additional ultrasound markers (nasal bone, tricuspid regurgitation, reversed a-wave in ductus	1) Performance of screening by cfDNA testing for trisomies 21 and 18 2) Comparison of cfDNA	Comparison of cfDNA testing 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 cfDNA testing or comparison of cfDNA testing with CT
		obstetric population) Median age (IQR): 31.8 (27.7 – 35.4) years, Gestational age, range: 11+0 – 13+6 weeks, 1st trimester: 100% Ethnicity: Caucasian 69.8%, African 20.6%, South Asian 4.0%, East Asian 2.8%, Mixed 2.8%	samples together		trimester combined test: Median estimated T21 risk (range) 1:8,469 (1:2–1:23,527), Median estimated T18 risk (range) 1:14,894 (1:2-1:47,472)		examination	venosus) Risk threshold ≥1:150 (0.67%) for T21 and T18	testing with detection rate and false positive rate of 1st-trimester CT with or without additional ultrasound markers	
Nicolaides 2013[77] 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 1st trimester: 100%	NR	T21, T18, T13	High risk for aneuploidies or sickle cell disease: 1st-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 or T13 by CT (range): 1:75 (1:2–1:12,433)	SNP- and NATUS Natera Inc. (USA)	CVS and karyotyping	None	Performance of cfDNA testing to detect T21, T18, T13, SCA and triploidy	Accuracy of cfDNA testing
Norton 2012[78]	Prospective,	N=4,002 enrolled,	Exclusion	T21, T18	Undergoing	DANSR	Karyotypi	None	1) Harmony	Accuracy 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 cfDNA testing or comparison of cfDNA testing with CT
USA, Netherlands, Sweden Study start date: August 2010	multicentre cohort study (NICE study) Number of centres: 48	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 Ethnicity: Caucasian 49.6%, African American 6.4%, Asian 13.4% , Hispanic 22.7%, Other 7.9%	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 criteria, 84 insufficient sample volume, 20 incorrect sample labelling		invasive testing for any indication (primarily high risk women)	FORTE Harmony Prenatal Test Ariosa Diagnostics (USA)	ng, FISH or QF-PCR from amniocentesis (74.7%) or CVS (25.3%)		Test performance for T21 and T18 at 1% risk cut-off 2) Foetal fraction. Test performance at different risk cut-off values	cfDNA testing
Norton 2015[46] USA, Sweden Study start date: March 2012	Prospective multicentre cohort (NEXT study) Number of centres: 35	N=18,955 enrolled. N=18,510 met I/E criteria Women with singleton pregnancies, ≥ 18 years of age, presenting for aneuploidy screening at 10-14 weeks of gestation (cfDNA testing and 1st-trimester CT) Mean age (range): 31	N=450 (2.4%): 229 did not meet inclusion criteria or met exclusion criteria, 31 had twins discovered on NT testing, 121 had unknown ovum-donor status, 64 withdrew or were withdrawn by investigator	T21, T18, T13	General obstetric population (unselected)	DANSR FORTE Harmony Prenatal Test Ariosa Diagnostics (USA)	Invasive prenatal testing (135 CVS, 422 amniocentesis), 52 postnatal genetic testing, 16 testing on products of conceptio	First-trimester CT (cut-off ≥1:270 for T21, ≥1:150 for T18 and T13)	1) Area under ROC curve for T21 screening with cfDNA testing versus standard screening 2) Evaluation of cfDNA testing and standard screening to assess the risk	Comparison of cfDNA testing 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 cfDNA testing or comparison of cfDNA testing with CT
		(18-48) years Mean gestational age (range): 12.5 (10.0-14.3) weeks. 100% 1st trimester					n, all other examination of the newborn		for T18 and T13 Performance of cfDNA testing in low-risk patients	
Palomaki 2011[61] USA Study start date: Trial submission 6th April 2009 cfDNA testing Jan to March 2011	Nested case-control in a cohort (Part of an international clinical validation study, NCT00877292). 7 euploid samples matched to each T21 case, based on gestational age, enrolment site, race, and time in the freezer (within one month). Number of centres: 27	N=4,664 in cohort, N=1,696 in nested case-control study (212 T21 and 1,484 euploid) Singleton pregnancies at high risk for T21 Mean age (SD): 36.7 (SD 5.1) years Mean gestational age (range): 15.0 (8.1-21.5) weeks. 1st trimester: 50%, 2nd trimester: 50% Ethnicity: 89% Caucasian, 2% Black, 7% Asian, 2% unknown	N=279/4,664 (6.0%) from cohort: 116 sample not adequate, 112 multiple gestation / fetal death, 51 no karyotype /outcome available N=2,689/4,385 (61.3%): Not selected for case-control study	T21	High risk for T21: 1st-trimester screening positive: 22%, 2nd-trimester screening positive: 7.6%, Integrated test positive: 13.8%, Ultrasound anomaly: 10.7%, AMA ≥ 38 years: 33.4%, 2 or more indications 8.9%, Family history of aneuploidy: 2.6%, Other /unknown 1.1%	MPS (whole genome) Sequenom, Inc. (USA)	Amniocentesis (53.1%), CVS (46.8%) or examination of products of conception (0.1%) and karyotyping (n=1,694) or QF-PCR (n=1) or FISH (n=1)	None	Testing performance for T21 detection	Accuracy of cfDNA testing
Palomaki 2012[62] USA	Nested case-control in a cohort (Part of an	N=4,664 in cohort, N=293 case-control study (62 T18, 12 T13, 219 euploid)	N=279/4,664 (6.0%) from cohort: 116 sample not	T21, T18, T13	High risk for T21: 1st-trimester screening positive: 7.2%,	MPS (whole genome) Sequenom	Amniocentesis (48.5%) or CVS	None	Correct identification of T21, T18 & T13	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
Study start date: Trial submission 6th April 2009	international clinical validation study, NCT00877292) Each pregnancy with T18 and T13 matched with 3 controls based on the gestational age, enrolment site, race, and time in freezer (within 1 month) Number of centres: 27	plus 212 T21 and 1,483 matched controls reported earlier[61] N=1,988 for cfDNA testing Singleton pregnancies at high risk for T21 Mean age (SD): 37.2 (5.0)* years. Median gestational age (range): 14.6 (9-22) weeks*. 1st trimester: 52%, 2nd trimester: 48% Ethnicity: Caucasian 84.7%, Black 4%, Asian 5.4%, Unknown 5.4%	adequate, 112 multiple gestation / fetal death, 51 no karyotype /outcome available N=2,397/4,385 (54.7%): Not selected for case-control study		2nd-trimester screening positive: 4.4%, Integrated test positive: 10.2%, Ultrasound anomaly: 19.5%, AMA ≥ 38 years: 41.6%, 2 or more indications: 12.6%, Family history of aneuploidy: 3.4%, Other /unknown: 1.0%	Inc. (USA)	(51.5%) and karyotyping			
Pergament 2014[79] USA Study start date: NR	Prospective international multicentre cohort Number of centres: 36	N=1,064 enrolled, N=1,051 for testing (926 euploid, 67 T21, 32 T18, 14 T13, 12 Monosomy X) Singleton pregnancies of at least 7 weeks of gestation Mean age (SD), range: 30.3 (7.4),	N=13 (1.2%): 6 triploidy, 3 fetal mosaic, 2 47,XXY, 1 47,XXX, 1 47,XYY	T21, T18, T13	543 (51.0%) High risk: abnormal serum screen, ultrasound abnormality, maternal age ≥ 35 years 521 (49.0%) Low risk: maternal age < 35 years and lacking	SNP- and NATUS Natera Inc. (USA)	Amniocentesis/CVS (44.1%) and karyotyping/FISH; genetic testing of cord blood, buccal sample or	None	Performance of single-nucleotide polymorphism-based test on both high- and low-risk pregnant women	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
		18-47 years Mean gestational age (SD), range: 17.0 (4.1), 7.6-40.6 weeks			any reported high-risk indications		saliva (13.2%) or products of conception (42.8%)			
Porreco 2014[80] USA Study start date: September 2009	Prospective multicentre cohort (NCT00847990) Number of centres: 31	N=4,170 enrolled, N=3,430 for testing Singleton pregnancies, high risk for fetal aneuploidy 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%	N=740 (17.7%): 320 insufficient sample volume, 120 outside 6h lab processing window, 270 used as lab quality control set, 24 incomplete case report forms, 6 no amniocentesis / CVS	T21, T18, T13	High risk for fetal aneuploidy: Abnormal NT 104 (3%), Abnormal Triple/ quad screen 289 (8.4%), Abnormal ultrasound 492 (14.3%), AMA ≥ 35 years 1,417 (41.3%), Multiple indications 929 (27.1%), Previous or family history of aneuploidies 98 (2.9%)	MPS (whole genome) MaterniT21 ® PLUS Sequenom, Inc. (USA)	Amniocentesis (75.5%) or CVS (24.5%) and karyotype	None	Clinical performance of MPS to test for T21, T18, T13, fetal sex and SCA	Accuracy of cfDNA testing
Quezada 2015[90] UK Study start date: October 2012	Prospective cohort Number of centres: 1	N=2,905 Women with singleton pregnancies undergoing routine first-trimester screening for the major trisomies by cfDNA testing and by the combined test	N=NR	T21, T18, T13	No prior screening, general obstetric population, AMA ≥ 35 years 1,958 (67.4%)	DANSR & FORTE Harmony Ariosa Diagnostics (USA)	CVS or amniocentesis and fetal karyotyping, post-mortem examinati	First-trimester CT for T21 (PAPP-A, free β-hCG, nuchal translucency) Risk threshold ≥ 1/100 for T21	1) Numbers and concordance of results of cfDNA testing and 1st-trimester combined screen	Comparison of cfDNA testing 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 cfDNA testing or comparison of cfDNA testing with CT
		Mean age (range): 36.9 (20.4–51.9) years Median gestational age (range): 10+4 (10+0 -11+6) weeks. 1st trimester: 100% Ethnicity: Caucasian 2,570 (88.5%), South Asian 173 (6.0%), East Asian 96 (3.3%), Afro-Caribbean 21 (0.7%), Mixed 45 (1.5%)					on and karyotyping, newborn phenotype		2) Discordant results between cfDNA testing and fetal karyotype	
Sago 2014[91] 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 cfDNA testing Mean age (range): 38.3 (21-48) years Mean gestational age (range): 13.3 (10.0-19.9) weeks 1st and 2nd trimester	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 fetal karyotyping, fetal death and karyotyping or birth phenotype	NR	PPV for T21, T18 and T13.	Accuracy of cfDNA testing
Sehnert 2011[63] USA	Training set: Prospective case-control	N=1,014 in cohort, 946 singleton pregnancies with fetal	N=68/1,014 (6.7%) from cohort:	T21, T18, T13	906/946 (96%) showed at least 1 clinically	MPS (whole genome)	CVS or amniocentesis and	None	Test performance for T21, T18,	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
Study start date: April 2009	(all fetuses with abnormal karyotype as well as a random selection of non-affected individuals) Validation set: Prospective case-control or case series Number of centres: 13	karyotype Mean age (SD), range: 35.6 (5.66), 17-47 years. Mean gestational age (range): 15+4 (6+1 - 38+1) weeks. Trimester NR Ethnicity: 62.7% Caucasian 16.5% Hispanic 6.2% Asian, 5.2% multi-ethnic Selected for training set: 71/435, Selected for validation set: 48/575	Unspecified (no karyotype?) From training set N=6 (8.5%): 4 twin gestations, 1 contaminated during preparation, 1 69, XXX From validation set N=1 (2.1%): 1 twin gestation		recognized risk factor for aneuploidy: AMA ≥35 years 52.1%, Screen positive 18.6%, Increased NT 4.5%, Other congenital abnormality 9.0%, Other maternal risk 7.4%	Verinata Health (USA)	fetal karyotype		T13, gender and Monosomy X classification	
Shaw 2014[81] Taiwan, China Study start date: June 2012	Prospective cohort Number of centres: 11	N=201 Pregnant women > 12 weeks' gestation. High risk (n=100): Mean age (SD): 35.1 (3.2) years. Mean gestational age (SD) 17.3 (2.1) weeks. 98 singleton, 2 twin pregnancies Low risk (n=100): Mean age (SD): 34.6 (2.6) years. Mean gestational age (SD) 16.1 (3.0)	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: 1:3,179	MPS (whole genome) Berry Genomics (China)	Amniocentesis and karyotyping or birth outcome	None	Test performance for detection of all fetal autosomal and sex chromosome aneuploidies	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
		weeks. 98 singleton, 2 twin pregnancies								
Song 2013[23] China Study start date: April 2011	Prospective cohort Number of centres: 2	N=1,916 Singleton pregnancies, women <35 years undergoing routine antenatal screening Mean age (SD), range: 29.03 (2.7), 20 - 34 years Mean gestational age (SD), range: 16.57 (1.56), 11 - 21+6 weeks. 1st trimester: 3.4%, 2nd trimester: 96.6% Assisted conception 14 (0.8%)	N=NR	T21, T18 T13	General obstetric population < 35 years High risk 275/1,741 (15.8%): Positive serum screening >1:270: 249 (14.3%), Increased NT: 10 (0.6%), Other indications 16 (0.9%) Low risk 1,466/1,741 (84.2%)	MPS (whole genome) Berry Genomics (China)	CVS, amniocentesis or cordocentesis and karyotyping or birth phenotype	2nd trimester triple serum screening (α-fetoprotein, free β-hcg, unconjugated estriol) Cutoff ≥ 1:270 for T21 and T18	cfDNA testing performance for detection of T21, T18, T13 and SCA Comparison of cfDNA testing and serum screening performance	Comparison of cfDNA testing with CT
Song 2015[96] China Study start date: May 2012	Prospective cohort Number of centres: 1	N=213 Women with singleton pregnancies, ≥ 35 years, 8+0 – 12+6 weeks' gestation, high-risk of fetal aneuploidies, presenting for cfDNA testing Mean age (range): 37.25 (35-45) years Mean gestational age (range): 9+6	N=1 (0.5%): 1 with quality control failure (haemolysis)	T21, T18, T13	All high-risk for fetal aneuploidies due to advanced maternal age ≥ 35 years	MPS (whole genome) Berry Genomics (China)	CVS or amniocentesis and karyotyping (n=178) or newborn phenotypic examination (n=34)	none	1) Clinical performance of cfDNA testing in the first trimester 2) Relationship between fetal DNA fraction and early gestational age	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
		(8+0 – 12+6) weeks. 100% 1st trimester								
Sparks 2012[64] USA Study start date: NR	Prospective case-control Number of centres: NR	N=298 252 average-risk controls, 46 confirmed cases (39 T21, 7 T18). Singleton pregnancies Median age (range): 31 (18-44) years Median gestational age (range): 13.4 (7-35.4) weeks	NR	T21, T18	Mixed risk Controls: average risk (no invasive testing) T18/T21 cases: possibly high-risk as invasive testing was performed	DANSR & standard z-test Aria Diagnostics (USA)	Invasive testing with FISH and/or karyotype analysis Average-risk women without reference standard	None	Detection of T21 and T18	Accuracy of cfDNA testing
Sparks 2012[25] USA Study start date: NR	Prospective case-control Number of centres: NR	Number enrolled unclear Singleton pregnancies, women ≥ 18 years, ≥10 weeks' gestation, high risk for fetal trisomies undergoing invasive testing. Subset of N=338 (250 euploid, 72 T21, 16 T18) randomised into <u>Validation set</u> (n=167) (36 T21, 8 T18, 123 euploid): Mean age (SD), range: 33.5 (7.1), 18-51 years. Mean gestational age (SD), range: 18.6	NR	T21, T18	High risk for fetal trisomy	DANSR and z statistic or FORTE Aria Diagnostics (USA)	Invasive testing with FISH and/or karyotype analysis	None	Detecting fetal aneuploidy using DANSR and z statistic or FORTE	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
		(4.0), 11.0-36.1 weeks <u>Training set</u> (n=171) (36 T21, 8 T18, 127 euploid): Mean age (SD), range: 34.5 (6.3), 18-44 years. Mean gestational age (SD), range: 17.6 (4.4), 10.3-33.0 weeks								
Stumm 2014[22] 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 fetal 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 fetal karyotyping	None	1) Diagnostic accuracy for fetal T21 detection (using DAP.21). 2) Diagnostic accuracy for fetal T13 and T18 detection (using DAP.plus) and comparison of algorithms for T21	Accuracy of cfDNA testing
Verweij 2013[82] Netherlands, Norway, Sweden, USA Study start date: May	Multicentre international prospective cohort (EU-NITE study) Number of centres: 6	N=595 enrolled, N=520 eligible Women undergoing invasive testing, singleton pregnancy, ≥10 weeks' gestation	N=75 (12.6%): 21 failed I/E criteria (non-invasive procedure performed, twin pregnancy, no blood sample);	T21	Mostly increased risk for T21 based on 1st trimester screening (serum screening, NT and/or maternal age), detection of fetal anomalies	DANSR FORTE Harmony Ariosa Diagnostics (USA)	CVS (54%) or amniocentesis (46%) and karyotyping or quantitativ	None	Test performance for T21 detection by shipping whole blood samples from Europe to a	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
2011	(4 Dutch, 2 Swedish)	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%, Black 1.3%, Other 4.6%	19 insufficient plasma volume; 11 logistical problems - shipping difficulties; 24 chromosome abnormalities other than T21		on ultrasound, previous affected pregnancy or family history) 8.8% other indications (psychosocial or anxiety reasons)		e fluorescen t PCR		laboratory in the USA	
Wang 2014[92] USA Study start date: NR	Lab cohort Number of centres: NR (1 diagnostics lab)	N=109 samples with prior cfDNA testing referred for confirmation by cytogenetic studies Age, gestational age, demographics NR	NR	T21, T18, T13	NR	Only reported for 42/109 (38.5%) samples Panorama, Natera Inc. (USA); Harmony, Ariosa Diagnostics (USA); MaterniT21, Sequenom Inc. (USA); Verifi, Illumina (USA)	Amniocen tesis/CVS, postnatal material & standard karyotypin g, FISH, and/or oligo-single-nucleotide polymorp hism microarra ys	None	Concordance of cfDNA testing and cytogenetic results	Accuracy of cfDNA testing
Wax 2015[93] USA Study start date: June 2012	Retrospectiv e review of prospective cohort Number of	N=1,046 eligible for cfDNA testing, N=166 high-risk pregnant women with singleton pregnancies opted for cfDNA	Multiple pregnancy N=NR; N=880 (84.1%) chose not to have cfDNA testing	T21, T18, T13	All high-risk: AMA ≥ 35 years 742 (70.9%), Ultrasound abnormality 280 (26.8%),	MPS (whole genome) Manufacture r: NR	Amniocen tesis (n=56) or CVS (n=50) and	None	Difference in genetic counselling utilisation, invasive procedures	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
	centres: 1	testing Mean age (SD): 34.6 (5.5) years Gestational age: range 10+0 – 21+6 weeks 1st and 2nd trimester			Positive screen 115 (11.0%), Prior trisomy 15 (1.4%), Parental translocation 1 (0.1%)		karyotyping, postnatal karyotyping of neonatal blood, birth phenotype from records		and T21 detection before and after cfDNA testing implementation	
Willems 2014[47] Netherlands, Belgium Study start date: March 2013	Lab cohort Number of centres: NR (GENDIA Genetic Diagnostic Network, Antwerp, Belgium)	N=3,000 from a larger cohort of more than 4,000. Consecutive samples submitted for commercial cfDNA testing. Pregnant women ≥ 10 weeks' gestation. Mean age (SD), range: 36 (3), 18-49 years. Mean gestational age (SD), range: 13 (2), 10-30 weeks	N=NR	T21, T18 T13	Mixed (women who had chosen to have cfDNA testing): Positive 1st trimester screening (>1/200 Netherlands, >1/300 Belgium): 22%, AMA >37 years only: 40.06%, Previous or family history of chromosomal anomalies: 3.27%. No indication (fear for T21and/or fear for invasive procedures and/or fear for false-negative standard screen): 34.73%.	DANSR FORTE Harmony Ariosa Diagnostics (USA)	CVS or anmiocentesis and genetic analysis or birth outcome	None	Detection of T21, T18, T13	Accuracy of cfDNA testing
Yu 2014[65]	Retrospectiv	First sample set:	NR	T21, T18,	Clinical	MPS (whole	CVS or	None	T21, T13 and	Accuracy 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 cfDNA testing or comparison of cfDNA testing with CT
Hong Kong, Netherlands, UK Study start date: NR	e case-control of stored maternal samples Number of centres: 3	N=144 (60 euploid, 36 T21, 27 T18, 21 T13) previously analysed by[56] and[57] Singleton pregnancies with clinical indications for CVS or amniocentesis and full karyotyping results Median gestational age (IQR): 13.0 (12.6-14.0) weeks		T13	indications for CVS or amniocentesis and full karyotyping results (from [57])	genome) DNA fragment size (paired-end MPS)	amniocentesis and karyotyping		T18 detection	cfDNA testing
Zhang 2015[24] China, Hong Kong (Denmark) Study start date: January 2012	Prospective multicentre cohort Number of centres: 508	N=147,314 samples received for cfDNA testing 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: 1st (9-13 wks): 4.21%, 2nd (14-27 wks): 94.13%, 3rd (≥ 28 wks): 1.47%,	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 cfDNA testing in detecting T21, T18, and T13 2) cfDNA testing performance in twin pregnancies. cfDNA testing performance for T21 detection in high-risk and low-risk subjects. Factors contributing to cfDNA	Accuracy of cfDNA testing

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 cfDNA testing or comparison of cfDNA testing with CT
		Unknown: 0.18% 99.45% singleton pregnancies, 0.55% twins							testing false-positive and false-negative results	
Zhou 2014[94] 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 1st and 2nd 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-Shenzen, China	Amniocentesis and karyotyping (n=54), postnatal karyotype (n=2) or birth outcome (n=3,894)	none	1) cfDNA testing performance for detection of trisomies 13, 18, and 21 2) Confirming care flow path	Accuracy of cfDNA testing
Zimmermann 2012[26] USA Study start date: NR	Prospective case-control Unblinded proof-of-principle study Number of centres: NR	N=166 (11 T21, 3 T18, 2 T13, 2 45X, 2 47XXY, 146 putatively euploid) Singleton pregnancies, women ≥ 18 years, ≥ 9 weeks' gestation Median gestational age: 17.0 and 17.5 weeks for euploid and aneuploid samples, respectively	NR	T21, T18 T13	Mixed: Aneuploidy samples from pregnant women with invasive prenatal testing Putative euploid samples from average-risk women without known risk indicators	SNP-based, Parental Support (PS) algorithm Natera Inc. (USA)	Invasive testing and FISH and/or karyotype in aneuploid samples, 62/146 putative euploid samples comfirmed by karyotyping of post-birth child tissue	None	Detection of fetal aneuploidies at chromosomes 13, 18, 21, X, and Y	Accuracy of cfDNA testing

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; cfDNA, cell-free deoxyribonucleic acid; 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.

Table 13 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[66] 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[55] 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 ≤ NCV ≤ 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[67] 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	NCV ≥ 4.0 affected, NCV ≤ 3.0 unaffected, 3.0 < 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[56] Hong Kong, Netherlands, UK, China	5-10 ml / Before invasive test	All fragments mapping to Chr13 or Chr18 (no markers)	Genome Analyzer IIx (Illumina)	2-plex	z-score > 3 for T13 and T18; 103 independent male euploid samples as controls	Total GC-corrected read counts from a sample	GC correction (LOESS regression) / non-repeat masked	Hg18 NCBI.36	Short Oligonucleotide Alignment Programme 2 (SOAP2); no mismatch
Chiu 2011[57] 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[68] 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	Total number of unique reads. Then normalisation by average *k*-	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**
					positive, t >2.5 or L >1: test positive, t <2.5 and L <1: test negative	mer coverage of the 22 autosomes			
Ehrich 2011[58] USA	10 ml / Before invasive test	All fragments aligned to Chr21 (no markers)	Genome Analyzer IIx (Illumina)	4-plex (1-plex for resequencing when fetal fraction ≤ 3.9%)	z-score > 2.5 for T21; used 24 independent euploid reference samples; iterative censoring to adjust for biased control group	All sequence reads excluding chr X and Y	no / non-repeat masked	UCSC hg19 human reference genome	CASAVA version 1.6 / up to 1 mismatch
Fang 2015[95] China	10 ml / Before invasive testing	NR	NR	NR (high-throughput)	NR	NR	NR	NR	NR
Futch 2013[70] USA	NR / NR	All fragments mapping to Chr13, Chr18 or Chr21 (no markers)	Illumina HiSeq 2000 (from [55])	6-plex (from [55])	NCV > 4.0 aneuploid, NCV < 2.5 euploid, 2.5 ≤ NCV ≤ 4.0 unclassified (from [55])	Normalising chromosome denominators not specified	Normalising chr denominators / NR	Human genome assembly hg18 (from [55])	Bowtie short read aligner (version 0.12.5) / ≤ 2 mismatches (from [55])
Huang 2014[71] 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 [68])	4-plex or 12-plex (from [68])	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
Jensen 2013[98] USA	20-50 ml / Before invasive test	All fragments mapping to chr13, chr18 and chr21 (no markers)	Illumina HiSeq2000	12-plex	FC-robust z-score > 3 for T21; Robust z-score > 3.95 for T13 and T18. 0%, 4% and 13% T21 DNA as control samples.	Counts for all 22 autosomes (from [61])	GC correction / repeat-masked	February 2009 build of the human genome (hg19)	Bowtie 2 / 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**
Jeon 2014[87] South Korea, China	10 ml / Before invasive testing	All fragments mapping to Chr18 or Chr21 (no markers)	Ion Proton™ 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[72] China	5 ml / NR	All fragments mapping to chr13, chr18 and chr21 (no markers)	Illumina GAIIx and Illumina HiSeq 2000	multiplex	Binary hypothesis t-test and logarithmic LR between the two t-tests (NIFTY): $\lvert t_{i;j;first} \rvert > 3$ and $\lvert t_{i;j;second} \rvert < 3$ as warning criteria. Autosomal aneuploidy if $L_{i;j} > 1$.	For k-mer coverage: total number of unique reads. Then normalisation by average k-mer coverage for the 22 autosomes	GC correction (Losses regression) / NR	hg18, NCBI build 36	NR / 0 mismatch
Lau 2012[73] 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[74] 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 [72])	Multiplex (from [72])	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. Threshold t-value NR	For k-mer coverage: total number of unique reads. Then normalisation by average k-mer coverage for the 22 autosomes (from [72])	GC correction (Losses regression) / NR	Hg18, NCBI build 36	NR / 0 mismatch
Liang 2013[75] China	5 ml / Before invasive test	All fragments mapping to chr13, chr18	Illumina HiSeq 2000	8-plex or 12-plex	z-score > 3 for T21, z-score > 5.91 for T18,	Total count of sequences uniquely	GC correction (slope of simple linear regression) /	Unmasked human reference	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**
		and chr21 (no markers)			z-score > 5.72 for T13; reference set of 50 independent female euploid samples	mapped to all autosomal chromosomes	non-repeat masked	genome (hg19)	
Liao 2014[60] China, USA	5-10 ml / Before invasive test	All fragments mapping to Chr13, Chr18 or Chr21 (no markers)	Semiconductor sequencing platform (Ion Proton sequencer, Life Technologies)	NR (13-15 samples per run)	z-score > 3, cross-validation using 90% of samples as reference and 10% of samples for validation with 1,000 repeats	Intrarun normalisation: total corrected reads number of all autosomes	GC correction (LOESS regression, intrarun normalisation, linear model regression) / NR	Hg19	Burrows–Wheeler Aligner (BWA) / NR
McCullough 2014[89] USA	NR / Before invasive testing	All fragments mapping to chromosomes 13, 18, 21, and Y (no markers)	Illumina HiSeq2000	Multiplex	Robust z-scores z > 3 for chromosome 21 and z > 3.95 for chromosomes 18 and 13 (from [98])	Counts for all 22 autosomes (from [98])	GC correction / repeat-masked (from [98])	February 2009 build of the human genome (hg19) (from [98])	Bowtie2 / Perfect matches within the seed sequence (from [98])
Palomaki 2011[61] USA	20-50 ml / Before invasive test	All fragments mapping to chr21 (no markers)	Illumina HiSeq 2000	4-plex	FC-robust z-score > 3 for T21; Historical reference ranges[58] adjusted after review of first results	Counts for all 22 autosomes	no / non-repeat masked; (GC correction / repeat masked in post hoc analysis)	UCSC hg19 human reference genome	CASAVA version 1.6 / 0 mismatches
Palomaki 2012[62] 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 ≥ 3 for T21, T18 and T13. Euploid pregnancies considered to be controls for each chromosome	Counts for all 22 autosomes (from [61])	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 [61])	CASAVA version 1.6 / 0 mismatches (from [61])
Porreco 2014[80] 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 ≥ 3 for T21, Flow cell-robust z-score ≥ 3.95 for T18 and T13.	Counts for all 22 autosomes (from [98])	GC correction / repeat-masked (from [98])	UCSC hg19	Bowtie version 2 / 0 mismatch (from [98])
Sago 2014 [91]	20 ml /	NR	NR	NR	NR	NR	GC correction /	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**
Japan	Before invasive testing	(**MaterniT21 Plus, Sequenom:** All fragments mapping to chr13, chr18 and chr21 (no markers))	(Illumina HiSeq 2000[98])	(12-plex[98])	(Robust z-scores z > 3 for chromosome 21 and z > 3.95 for chromosomes 18 and 13[98])	(Counts for all 22 autosomes (from [98]))	repeat-masked (from [98])	(UCSC hg19 (from [98]))	(Bowtie2 / Perfect matches within the seed sequence (from [98]))
Sehnert 2011[63] 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 ≤ NCV ≤ 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[81] 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 [75])	GC correction (slope of simple linear regression[75]) / non-repeat masked	hg19	SOAP2 / NR
Song 2013[23] 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 [75])	GC correction (slope of simple linear regression[75]) / non-repeat masked	hg19	BWA / NR
Song 2015[96] China	NR / Before invasive testing	All fragments mapping to Chr13, Chr18 or Chr21 (no markers)	Illumina HiSeq 2000	12-plex (from [23])	z-score ≥ 3 for trisomy; Used 50 independent female euploid samples as reference set (from [23])	Total count of sequences uniquely mapped to all the autosomal chromosomes (from [75])	GC correction (slope of simple linear regression[75]) / non-repeat masked(from [23])	hg19	BWA / NR (from [23])
Stumm 2014[22] Germany, Switzerland	7-10 ml / Before invasive procedure	All fragments mapping to chr13, chr18 and chr21 (no	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%,	Total counts of all autosomes, X and Y	DAP.21 for T21: no / repeat-masked (after unblinding DAP.plus for T13,	DAP.21: hg18?, DAP.plus: hg19	ELAND / 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**
		markers)			20% or 40% T21 DNA control samples in each FC		T18 and T21 with GC correction (LOWESS))		
Wax 2015[93] USA	NR / Before invasive testing	NR (Single commercial laboratory using MPSS)	NR	NR	NR	NR	NR / NR	NR	NR / NR
Yu 2014[65] Hong Kong, UK, Netherlands	5-10 ml (from[56 57]) / Before invasive test	All fragments mapping to chr13, chr18 and chr21 (no markers)	Genome Analyzer IIx (Illumina) (paired-end MPS)	2-plex	Size-based Z score > 3 for trisomies; size cutoff of 150 bp; randomly assigned 20 cases with euploid fetus in the first sample set as reference controls	Difference in proportion of short (≤ 150 bp) DNA fragments in target chr and the reference chr (all autosomes except chr21, 18 and 13)	no / non-repeat-masked	NCBI Build 36.1/hg 18	SOAP2 / 0 mismatch
Zhang 2015[24] 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 [68, 72]). Threshold NR	Total number of unique reads. Then normalisation by average k-mer coverage of the 22 autosomes (from [68, 72])	GC correction (Losses regression) / NR (from [68, 72])	hg18, NCBI build 36	NR / 0 mismatch (from [68, 72])
Zhou 2014[94] China	NR / Before invasive testing	All fragments mapping to chr13, chr18 and chr21 (no markers) (from [68])	Illumina GAIIx or Illumina HiSeq 2000 (from [68])	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 [68])	Total number of unique reads. Then normalisation by average k-mer coverage of the 22 autosomes (from [68])	GC correction (Losses regression) / NR (from [68])	hg18, NCBI build 36 (from [68])	NR / 0 mismatch (from [68])

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.

Table 14 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[54] 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[84] Belgium, UK, Spain	20 ml / Before invasive testing	576 nonpolymorphic loci on each chr13, chr18 and chr21 (from [25, 69])	Illumina HiSeq 2000 (from [25, 69])	96-plex (from [25, 69])	FORTE risk score (threshold NR, **Harmony™ 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 [25, 69])	Median polish on log-transformed counts / NA (from [25, 69])	Expected locus sequences (from [25, 69])	NR / <3 mismatches (from [25, 69])
Del Mar Gil 2014[69] 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 / NA	Expected locus sequences	NR / <3 mismatches
Nicolaides 2012[76] 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[78] 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[46] USA, Sweden	NR / Before invasive testing	**Harmony™ Prenatal test:** 576 nonpolymorphic loci on each chr13, chr18 and chr21 for	Illumina HiSeq 2000 (from [25])	96-plex (from [25])	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 [25])	Median polish on log-transformed counts / NA (from [25])	Genome Reference Consortium human build 37	NR / <3 mismatches (from [25])

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**
		chromosome proportion.							
Quezada 2015[90] UK	20 ml / Before invasive testing	**Harmony™ Prenatal test**: 576 nonpolymorphic loci on each chr13, chr18 and chr21 for chromosome proportion.	Illumina HiSeq 2000 (from [25, 54])	96-plex (from [25, 54])	FORTE risk score (threshold NR, usually 1% cutoff).	Sum of mean cfDNA counts of the loci for chr13, chr18 and chr21 (from [25, 54])	Median polish on log-transformed counts / NA (from [25, 54])	Expected locus sequences (from [25, 54])	NR / <3 mismatches (from [25, 54])
Sparks 2012[64] USA	8 ml on average / NR	384 nonpolymorphic loci on each chr18 and chr21	Illumina HiSeq 2000	96-plex	Standard Z-test of proportions; iterative censoring, z-score > 3	20% trimmed mean count across all loci (both chr21 and 18)	Median polish on log-transformed counts / NA	Expected locus sequences	NR / <3 mismatches
Sparks 2012[25] 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[82] 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
Willems 2014[47] Belgium, Netherlands	20 ml / Before invasive test	576 nonpolymorphic loci on each chr13, chr18 and chr21	Illumina HiSeq 2000	96-plex	FORTE risk score, threshold NR (possibly 1%)	Sum of mean cfDNA counts of the loci for chr13, 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.

Table 15 Test characteristics – Single-nucleotide polymorphism-based cfDNA testing (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**
Dar 2014[86] USA	>13 ml / Before invasive testing	19,488 polymorphic loci covering chromosomes 21, 13, 18, X, and Y	Illumina GAIIx or HiSeq sequencer (from [26])	19,488-plex targeted PCR	Maximum likelihood estimate generated by the NATUS algorithm combined with maternal and gestational age prior risks. Risk score ≥1/100: high risk for aneuploidy, Risk score < 1/100: low risk.	Around 30% of patients given the opportunity	NA / NR	NR	Proprietary algorithm adapted from the Novoalign (Novocraft, Selangor, Malaysia) commercial software package / NR (from [26])
Hall 2014[59] 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 fetal genotype and fetal 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 [26])
Korostelev 2014[88] Russia	NR / Before invasive testing	>19,000 polymorphic loci covering chromosomes 21, 13, 18, X, and Y	NR (Illumina GAIIx or HiSeq sequencer (from [26]))	NR (19,488-plex targeted PCR (from [79])	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 [26]))
Nicolaides, 2013[77] UK	20 ml / Before invasive test	19,488 SNPs on chromosomes 21, 13, 18, X, and Y	Illumina GAIIx or HiSeq sequencer (from [26])	19,488-plex targeted PCR	NATUS: calls fetal genotype and fetal 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 [26])
Pergament 2014[79] USA	NR / 93% before invasive test,	19,488 SNPs on chromosomes 21, 13, 18, X,	Illumina GAIIx or HiSeq	19,488-plex targeted PCR	NATUS: calls fetal genotype and fetal fraction with	yes for 48.1% of samples	NA / NR	NR	Proprietary algorithm adapted from Novoalign

	7% at least 4 days after	and Y	sequencer (from [26])		maximum likelihood, calculates copy number call accuracy, threshold NR				(Novocraft, Selangor, Malaysia) / NR (from [26])
Zimmermann 2012[26] 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 fetal genotype and fetal 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; cfDNA, cell-free deoxyribonucleic acid; NR, not reported; PCR, polymerase chain reaction; PS, Parental Support™ algorithm; SNP, single-nucleotide polymorphism

Table 16 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[97] 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 fetal 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

Table 17 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[83] 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 fetuses. NCV between 2.5 and 4.0 as "unclassified" **Sequenom:** NR (Robust z-scores, cutoff NR)	NR	NR / NR	NR	NR / NR
Comas 2014[85] Spain	≤ 20 ml / Before invasive testing **Panorama:** Paternal genetic sample in 51% of samples.	NR **(Harmony test:** 576 nonpolymorphic loci on each chr13, chr18 and chr21 **Panorama test:** 19,488 polymorphic loci covering chromosomes 21, 13, 18, X, and Y)	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 Threshold NR)	NR (**Harmony:** Sum of mean cfDNA counts of the loci for chr13, chr18 and chr21 (from [25, 54]) **Panorama:** NA)	NR / NR	NR	NR / NR
Wang 2014[92] USA	NR / Before invasive testing	NR (only reported for 42/109 samples: **Panorama** (Natera, USA), **Harmony** (Ariosa, USA), **MaterniT21 (**Sequenom, USA),	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**
		Verifi (Illumina – formerly Verinata, USA)							

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

Table 18 Study quality according to QUADAS-2[27]

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[66]	High	High	Low	High	Low	Unclear	Low	Low
Ashoor 2012[54]	High	Unclear	Low	High	Unclear	Low	Low	Low
Beamon 2014[83]	High	Low	Low	High	Low	High	Low	Low
Bevilacqua 2015[84]	Unclear	Low	Low	High	High	High	Low	Low
Bianchi 2012[55]	High	Low	Low	High	High	High	Low	Low
*Bianchi 2014[67]	Unclear	Unclear	Low	High	High	High	High	Low
Chen 2011[56]	High	Unclear	Low	Low	High	Unclear	Low	Low
Chiu 2011[57]	High	Low	Low	High	High	High	Low	Low
Comas 2014[85]	High	Low	Low	High	High	High	Low	Low
Dan 2012[68]	Unclear	Unclear	Low	High	High	High	Low	Low
Dar 2014[86]	High	Low	Low	High	High	High	Low	Low
Del Mar Gil 2013[69]	Unclear	Unclear	Unclear	High	Unclear	Low	Low	Low
Dhallan 2007[97]	High	Unclear	Low	Low	High	High	High	Low
Ehrich 2011[58]	High	High	Low	High	High	High	Low	Low
Fang 2015[95]	Unclear	Unclear	Low	High	Low	High	Unclear	Low
Futch 2013[70]	High	Low	Low	High	High	High	Low	Low
Hall 2014[59]	High	Unclear	Low	High	High	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
Huang 2014[71]	Unclear	Unclear	Low	Low	High	High	Low	Low
Jensen 2013[98]	High	High	Low	Low	High	High	Low	Low
Jeon 2014[87]	Unclear	High	Low	Low	Low	High	Low	Low
Jiang 2012[72]	Unclear	Unclear	Low	Low	High	High	Low	Low
Korostolev 2014[88]	Unclear	Low	Low	High	Low	High	Low	Low
Lau 2012[73]	Unclear	Low	Low	Low	Unclear	Low	Low	Low
Lau 2014[74]	Low	Unclear	Low	High	Unclear	High	Low	Low
Liang 2013[75]	Unclear	Low	Low	High	Low	High	Low	Low
Liao 2014[60]	High	High	Low	Unclear	Low	High	Low	Low
McCullough 2014[89]	High	Low	Low	High	High	High	Low	Low
*Nicolaides 2012[76]	Unclear	Low	Low	High	Unclear	High	Low	Low
Nicolaides 2013[77]	Unclear	Unclear	Low	High	Unclear	Low	Low	Low
Norton 2012[78]	Unclear	Low	Low	High	High	High	Low	Low
*Norton 2015[46]	Unclear	Low	Low	High	High	High	Low	Low
Palomaki 2011[61]	High	Low	Low	High	High	High	Low	Low
Palomaki 2012[62]	High	High	Low	High	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
Pergament 2014[79]	Unclear	High	Low	High	High	High	Low	Low
Porreco 2014[80]	High	Low	Low	High	High	High	Low	Low
*Quezada 2015[90]	Unclear	Low / High$	Low	High	Unclear	High	Low	Low
Sago 2014[91]	High	Low	Low	High	Unclear	Unclear	Low	Low
Sehnert 2011[63]	High	Low	Low	High	High	High	Low	Low
Shaw 2014[81]	Unclear	Low	Low	Low	Unclear	High	Low	Low
*Song 2013[23]	Unclear	Low	Low	High	Low	High	High	Low
Song 2015[96]	Unclear	Low	Low	High	Unclear	Low	Low	Low
Sparks 2012[64]	High	High	Low	High	High	High	Low	Low
Sparks 2012[25]	High	High	Low	High	High	High	Low	Low
Stumm 2014[22]	Low	Low for DAP.21 High for DAP.plus**	Low	High	High	High	Low	Low
Verweij 2013[82]	Low	Low	Low	High	High	High	Low	Low
Wang 2014[92]	High	Unclear	Low	High	High	High	Unclear	Low
Wax 2015[93]	Low	Unclear	Low	High	Low	Unclear	Low	Low
Willems 2014[47]	High	Low	Low	High	High	High	Low	Low
Yu 2014[65]	High	High	Low	High	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
Zhang 2015[24]	Unclear	Unclear	Low	High	High	High	Low	Low
Zhou 2014[94]	Unclear	Unclear	Low	High	Unclear	High	Low	Low
Zimmermann 2012[26]	High	High	Low	High	High	High	High	Low

* Studies comparing cfDNA testing 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

Table 19 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[66] France	20.11 (mean among 43 male euploid fetuses) 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[54] 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 failed amplification / 50 T18 cases excluded from T21 performance analysis and vice versa → 53 (13.3%) excluded
		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	
Beamon 2014[83] 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. 38 without birth outcome. 1 fetal demise without karyotype and normal cfDNA testing, 2 fetal demises without karyotype and unclassified for T13 and/or T21. 2 unclassified for T21, 1 unclassified for T13 → 46 (22.1%), 44 (21.2%), and 45 (21.6%) excluded from T21, T18, and T13 analysis, respectively
		T18	*2*	*160*	*1*	*1*	*66.7 (12.5-98.2)*	*99.4 (96.1-99.97)*	*66.7 (12.5-98.2)*	*99.4 (96.1-99.97)*	NR	
		T13	*0*	*162*	*1*	*0*	NA	*99.4 (96.1-99.97)*	NA	*100 (97.1-100)*	NR	
		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	
Bevilacqua 2015[84] 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	175/515 (34%) without cfDNA testing result (n=16) and/or reference standard result (n=164: 7 miscarriage or stillbirth without unknown karyotype, 19 pregnancies still
		T18	5	335	0	0	*100 (46.3-100)*	*100 (98.6-100)*	*100 (46.3-100)*	*100 (98.6-100)*	NR	
		T13	0	340	0	0	NA	*100 (98.6-100)*	NA	*100 (98.6-100)*	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						
												continuing, 138 lost to follow-up) excluded
Bianchi 2012[55] USA	NR	T21	89	404	0	0	100 (95.9-100)	100 (99.1-100)	*100* *(94.8-100)*	*100* *(98.8-100)*	NR	Test failure: 2 sample tracking issues, 16 no fetal DNA Inconclusive: 7 for T21, 5 for T18, 2 for T13 Censored complex karyotype: 19 for T21, 18 for T18, 18 for T13 (Overlap of 3 censored and no fetal DNA.) → 41 (7.7%) for T21, 38 (7.1%) for T18 and 35 (6.6%) for T13 excluded
		T18	35	460	0	1	97.2 (85.5-99.9)	100 (99.2-100)	*100* *(87.7-100)*	*99.8* *(98.6-99.99)*	NR	
		T13	11	485	0	3	78.6 (49.2-95.3)	100 (99.2-100)	*100* *(67.9-100)*	*99.4* *(98..1-99.8)*	NR	
Bianchi 2014[67] USA. cfDNA testing	NR	T21	5	1941	6	0	100 (47.8-100)	99.7 (99.3-99.9)	45.5 (16.7-76.6)	100 (99.8-100)	FP rate, %: 0.3	72 no clinical outcome; 17 no result on cfDNA testing; 38 no result on standard screening; 1 no result on either cfDNA testing or standard screening → 90 (4.4%) for T21 and T18, 128 (6.3%) for T13 cfDNA testing performance excluded For standard screening performance and T21 FP rate in either test: Another 2 uninterpretable results on standard screening excluded For standard screening performance and T18 FP rate in either test: Another 2 uninterpretable and 6 without results on
		T18	2	1947	3	0	100 (15.8-100)	99.8 (99.6-100)	40.0 (5.3-85.3)	100 (99.8-100)	0.2	
		T13	*1*	*1910*	*3*	*0*	*100* *(5.5-100)*	*99.8* *(99.5-99.96)*	*25.0* *(13.2-78.1)*	*100* *(99.7-100)*	0.1	
Standard screening	NR	T21	3	1840	69	0	100 (29.2-100)	96.4 (95.4-97.2)	4.2 (0.9-11.7)	100 (99.8-100)	FP rate, %: 3.6	
		T18	1	1894	11	0	100 (2.5-100)	99.4 (99.0-99.7)	8.3 (0.2-38.5)	100 (99.8-100)	0.6	
		T13	NR	NR	6	0	NR	*99.3*	NR	NR	0.7	

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						
												standard screening excluded. For T13 FP rate: Another 1015 without standard screening results excluded from either test
Chen 2011[56] Hong Kong, UK, Netherlands, China	NR	T18	34	247	5	3	91.9 *(77.0-97.9)*	98.0 *(95.2-99.3)*	87.2 *(71.8-95.2)*	98.8 *(96.2-99.7)*	NR	0 test failures / 0 inconclusive results / no exclusions.
		T13	25	261	3	0	100.0 *(83.4-100)*	98.9 *(96.4-99.7)*	89.3 *(70.6-97.2)*	100.0 *(98.2-100)*	NR	
Chiu 2011[57] Hong Kong, UK, Netherlands, China	NR	T21[(8)]	68	565	6	18	79.1 *(68.7-86.8)*	98.9 *(97.6-99.6)*	91.9 *(82.6-96.7)*	96.9 *(95.1-98.1)*	NR	11 failed quality control for sequencing; 96 (8-plex)/82 (2-plex) euploid male fetuses used as reference controls 439 not analysed in 2-plex. → 107 (14.0%) for 8-plex and 532 (69.6%) for 2-plex excluded
		T21[(2)]	86	143	3	0	100 *(94.7-100)*	97.9 *(93.6-99.5)*	96.6 *(89.8-99.1)*	100 *(96.7-100)*	NR	
Comas 2014[85] Spain	Mean 12.7 (Range 4.2-27.9)	T21	*4*	*308*	*0*	*0*	*100 (39.6-100)*	*100 (98.5-100)*	*100 (39.6-100)*	*100 (98.5-100)*	NR	4 without cfDNA testing result (test failure). 18 pregnancies in progress (1 overlap) → 21 (6.3%) excluded
		T18	*0*	*312*	*0*	*0*	NA	*100 (98.5-100)*	NA	*100 (98.5-100)*	NR	
		T13	*0*	*312*	*0*	*0*	NA	*100 (98.5-100)*	NA	*100 (98.5-100)*	NR	
Dan 2012[68] China, Hong Kong	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	79 failed quality control for sequencing; 3581 no reference standard. → 3660 (32.7%) excluded
		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	
Dar 2014[86] USA	Mean 10.8 SD (4.4), Median 10.1 (Range 3.7-50.0) (n=17,885)	T21	*140*	*68*	*14*	*2*	*98.6 (94.5-99.8)*	*82.9 (72.7-90.0)*	90.9 *(84.9-94.8)*	*97.1 (89.1-99.5)*	NR	1966 samples failed quality control metrics 10,854 cases from partner laboratories, clinical follow-up not available 17,661 cases without clinical follow-up → 30,481/30,705 (99.3%) excluded.
		T18	*27*	*195*	*2*	*0*	*100 (84.5-100)*	*99.0 (96.0-99.8)*	93.1 *(75.8-98.8)*	*100 (97.6-100)*	NR	
		T13	*8*	*203*	*13*	*0*	*100 (59.8-100)*	*94.0 (89.7-96.6)*	38.1 *(19.0-61.3)*	*100 (97.7-100)*	NR	
		All + MX	184	0	38	2	*98.9 (95.8-99.8)*	NA	82.9 *(77.1-87.5)*	NA	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						
Del Mar Gil 2014[69] UK	9.8 (7.4-12.1) in 193 euploid pregnancies	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	15 test failures. → 15 (7.2%) excluded
		T18	0	192	0	0	*NA*	*100 (97.6-100)*	*NA*	*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[97] USA	32.5 (range 17.0-93.8)	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	0 test failures / 0 inconclusive results / no exclusions
Ehrich 2011[58] 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 → 31 (6.5%) excluded
Fang 2015[95] China	NR	T21	11	NR	0	NR	NR	NR	*100 (67.9-100)*	NR	NR	Test failures and inconclusive results NR / Only 16 cfDNA testing-positive cases with invasive diagnostic testing Unclear if remaining 1496 cases (98.9%) had cfDNA testing-negative result or cfDNA testing failure and if all had follow-up after birth
		T18	2	NR	0	NR	NR	NR	*100 (19.8-100)*	NR	NR	
Futch 2013[70] USA	NR	T21	*52*	*40*	*2*	*2*	*96.3 (86.2-99.4)*	*95.2 (82.6-99.2)*	*96.3 (86.2-99.4)*	*95.2 (82.6-99.2)*	NR	43 test failures; 173 unclassifiable (chr21: 60, chr18: 50, chr13: 60, chr13 and chr21: 3); 5705 without reference standard. → 5921 (98.4%) excluded
		T18	*13*	*76*	*6*	*1*	*92.9 (64.2-99.6)*	*92.7 (84.2-97.0)*	*68.4 (43.5-86.4)*	*98.7 (92.0-99.9)*	NR	
		T13	*7*	*84*	*5*	*0*	*100 (56.1-100)*	*94.4 (86.8-97.9)*	*58.3 (28.6-83.5)*	*100 (94.6-100)*	NR	
Hall 2014[59] 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. → 4 (5.9%) excluded
		T18	0	64	0	0	NA	100 (94.4-100)	NA	*100 (92.9-100)*	NR	
		T13	15	49	0	0	100 (78.2-100)	100 (98.2-100)	*100 (74.7-100)*	*100 (90.9-100)*	NR	
Huang 2014[71] China, Denmark, Hong	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 / no exclusions
		T18	1	187	0	1	50	100	*100*	*99.5*	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						
Kong							*(2.7-97.3)*	*(97.5-100)*	*(5.5-100)*	*(96.6-99.97)*		
Jensen 2013[98] USA	14 (range 4 - 46)	T21	134	1134	1	0	100 *(96.5-100)*	99.91 *(99.4-100)*	*99.3* *(95.3-99.96)*	*100* *(99.6-100)*	FP rate, %: 0.09	0 test failures / 0 inconclusive results / no exclusions
		T18	36	1233	0	0	100 *(88.0-100)*	100 *(99.6-100)*	*100* *(88.0-100)*	*100* *(99.6-100)*	<0.01	
		T13	6	1262	1	0	100 *(51.7-100)*	99.92 *(99.5-100)*	*85.7* *(42.0-99.2)*	*100* *(99.6-100)*	0.08	
Jeon 2014[87] South Korea, China	NR	T21	11	144	0	0	*100.0* *(67.9-100.0)*	*100.0* *(96.8-100.0)*	100.0 (71.5-100.0)	100.0 (97.5-100.0)	NR	0 test failures / 0 inconclusive results / no exclusions from analysis
		T18	5	150	0	0	*100.0* *(46.3-100.0)*	*100.0* *(96.9-100.0)*	100.0 (47.8-100.0)	100.0 (97.6-100.0)	NR	
		T21+ T18	16	139	0	0	*100.0* *(75.9-100.0)*	*100.0* *(96.6-100.0)*	100.0 (79.4-100.0)	100.0 (97.4-100.0)	NR	
Jiang 2012[72] China	NR	T21	16	887	0	0	100 *(75.9-100)*	100 *(99.5-100)*	*100* *(75.9-100)*	*100* *(99.5-100)*	NR	0 test failures / 0 inconclusive results / no exclusions
		T18	12	890	1	0	100 *(69.9-100)*	99.9 *(99.3-100)*	*92.3* *(62.1-99.6)*	*100* *(99.5-100)*	NR	
		T13	2	901	0	0	100 *(19.8-100)*	100 *(99.5-100)*	*100* *(19.8-100)*	*100* *(99.5-100)*	NR	
Korostelev 2014[88] Russia	NR	T21	47	*635*	0	0	*100* *(90.6-100)*	*100* *(99.3-100)*	*100* *(90.6-100)*	*100* *(99.3-100)*	NR	0 test failures / 1 inconclusive result for gender & SCA / 1046 without reference standard → 1046/1728 (60.5%) excluded
		T18	2	*680*	0	0	*100* *(19.8-100)*	*100* *(99.3-100)*	*100* *(19.8-100)*	*100* *(99.3-100)*	NR	
		T13	3	*678*	0	1	*75.0* *(21.9-98.7)*	*100* *(99.3-100)*	*100* *(31.0-100)*	*99.85* *(99.0-99.99)*	NR	
Lau 2012[73] Hong Kong, China, Japan	NR	T21	11	97	0	0	*100* *(67.9-100)*	*100* *(95.3-100)*	*100* *(67.9-100)*	*100* *(95.3-100)*	NR	0 test failures / 0 inconclusive results / no exclusions
		T18	10	98	0	0	*100* *(65.5-100)*	*100* *(95.3-100)*	*100* *(65.5-100)*	*100* *(95.3-100)*	NR	
		T13	2	106	0	0	*100* *(19.8-100)*	*100* *(95.6-100)*	*100* *(19.8-100)*	*100* *(95.6-100)*	NR	
Lau 2014[74] Hong Kong, USA, China	NR	T21	*23*	*1659*	*0*	*0*	*100* *(82.2-100)*	*100* *(99.7-100)*	*100* *(82.2-100)*	*100* *(99.7-100)*	NR	1 unclassifiable result / 299 without reference standard → 300 (15.1%) excluded
		T18	*4*	*1678*	*0*	*0*	*100* *(39.6-100)*	*100* *(99.7-100)*	*100* *(39.6-100)*	*100* *(99.7-100)*	NR	
		T13	*2*	*1680*	*0*	*0*	*100*	*100*	*100*	*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						
							(19.8-100)	*(99.7-100)*	*(19.8-100)*	*(99.7-100)*		
Liang 2013[75] China	NR	T21	40	372	0	0	100 *(89.1-100)*	100 *(98.7-100)*	100 *(89.1-100)*	100 *(98.7-100)*	NR	12 test failures, 11 failed karyotyping → 33 (7.6%) excluded
		T18	14	398	0	0	100 *(73.2-100)*	100 *(98.8-100)*	100 *(73.2-100)*	100 *(98.8-100)*	NR	
		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	
Liao 2014[60] China, USA	NR	T21	NR	NR	NR	NR	99.94	99.46	NR	NR	NR	NR
		T18	NR	NR	NR	NR	100	99.24	NR	NR	NR	
		T13	NR	NR	NR	NR	100	100	NR	NR	NR	
McCullough 2014[89] USA. All	NR	T21	NR	NR	4	6	99.6 (97.3-99.9)	99.9 (99.8-100.0)	NR	NR	NR	870 (0.9%) cancelled tests. 1,330 (1.3%) without reportable test result after redraw Reference standard for only 67/100,000 samples (0.07%) of which only 37 (0.04%) are specified and usable for 2x2 Sensitivity and specificity were calculated under the assumption that if the lab was not contacted by the clinician, then the results were not discordant
		T18	NR	NR	5	4	99 (93.3-100.0)	99.9 (99.3-99.9)	NR	NR	NR	
		T13	NR	NR	13	2	98.9 (64.6-98.5)	99.9 (99.5-99.9)	NR	NR	NR	
Multiple gestation	NR	T21	NR	NR	1	0	>99.9	99.9	NR	NR	NR	
		T18	NR	NR	0	1	95.20	>99.9	NR	NR	NR	
		T13	NR	NR	1	0	>99.9	99.9	NR	NR	NR	
Nicolaides 2012[76] UK. cfDNA testing	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) (46 fetal fraction <4%, 54 assay failure) → 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	

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						
Nicolaides 2013[77] UK	≥3.95	T21	25	204	0	0	100 (86.3-100)	100 (98.2- 100)	*100 (83.4-100)*	*100 (97.7-100)*	NR	13 test failures. → 13 (5.4%) excluded
		T18	3	226	0	0	*100 (31.0-100)*	*100 (97.9-100)*	*100 (31.0-100)*	*100 (97.9-100)*	NR	
		T13	1	228	0	0	*100 (5.5-100)*	*100 (97.9-100)*	*100 (5.5-100)*	*100 (97.9-100)*	NR	
Norton 2012[78] 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; 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
		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	
Norton 2015[46] USA, Sweden. cfDNA testing	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, %: 0.06 (0.03-0.11) AUC: 0.999	384 had sample handling errors, 308 did not have standard-screening result, 488 did not have cfDNA result, 1,489 were lost to follow-up. → 2,669 (14.4%) excluded from either test For T13, patients enrolled before September 2012 were excluded from either test (n=4,656)
		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)	FP rate, %: 0.01 (0-0.04)	
		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)	
Combined FTS (≥1:270 for T21, ≥1:150 for T13 and T18)	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 2011[61] USA	Geometric mean 13.4	T21	209	1468	3	3	*98.6 (95.6-99.6)*	*99.8 (99.4-99.9)*	*98.6 (95.6-99.6)*	*99.8 (99.4-99.9)*	FP rate, %: 0.20 (<0.1-0.6)	13 repeated test failures. → 13 (0.8%) excluded
Palomaki 2012[62] USA	4-50% accepted	T21	210	1758	1	2	*99.1 (96.3-99.8)*	*99.9 (99.6-100)*	*99.5 (97.0-99.98)*	*99.9 (99.5-99.98)*	FP rate, %: 0.1 (<0.1-0.3)	17 repeated test failures → 17 (0.9%) 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						
		T18	59	1907	5	0	*100 (92.4-100)*	*99.7 (99.4-99.9)*	*92.2 (82.0-97.1)*	*100 (99.7-100)*	0.3 (0.1-0.7)	
		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[79] 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 failed quality control, 8 failed quality control for 1/5 chromosomes → 88 (8.4%) for T21, 87 (8.3%) for T13, 86 (8.2%) for T13 excluded
		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	
		T13	12	953	0	0	100 (73.5 -100)	100 (99.6-100)	*100 (69.9-100)*	*100 (99.5-100)*	NR	
Porreco 2014[80] 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	52 failed quality control; 54 complex karyotypes; 2 with insufficient quality criteria and complex karyotype → 108 (3.1%) excluded
		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	
		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	
Quezada 2015[90] UK. cfDNA testing	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 without cfDNA testing result / 69 without reference standard (48 miscarriages or stillbirths with unknown karyotype; 21 Lost to follow up.), overlap of 3 patients without cfDNA testing and reference standard result → 120 (4.1%) excluded
		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	
		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	
		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	
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*[#]	69 without reference standard (48 miscarriages or stillbirths with unknown karyotype; 21 Lost to follow up.), 12 of which had also no combined FTS result → 69 (2.4%) excluded [#] T13, T18 and non-trisomic cases with positive FTS result classified as FP
		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	

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						[$] Only non-trisomic cases with positive FTS result classified as FP
Sago 2014[91] 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 3 withdrawals (TOP) after positive cfDNA testing; 9 fetal deaths after positive cfDNA testing without karyotype; 5956 women with negative cfDNA testing but no birth outcome → 5972 (77%) excluded
		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	
		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	
		All	NR	NR	NR	NR	NR	NR	NR	NR	FN rate, %: <0.1	
Sehnert 2011[63] 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	1 twin sample removed; 1 inconclusive result for T13 (T13 case) → 1 (2.1%) for T21 and T18 excluded, 2 (4.2%) for T13 excluded
		T18	8	39	0	0	*100 (59.8-100)*	*100 (88.8-100)*	*100 (59.8-100)*	*100 (88.8-100)*	NR	
		T13	0	46	0	0	*NA*	*100 (90.4-100)*	*NA*	*100 (90.4-100)*	NR	
Shaw 2014[81] 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% FN rate 0%	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)*	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[23] China. cfDNA testing	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 FN rate, %: 0.00	102 no birth outcome, 64 failed quality control, 9 no birth follow-up and failed quality control criteria. → 175 (9.1%) excluded for either test
		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%	
Serum screening (≥1:270 for T18	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%	

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						
and T21)												
Song 2015[96] China	8.54 (range 2.69-18.75) (n=100 male fetuses)	T21	2	202	0	0	*100 (19.8-100)*	*100 (97.7-100)*	*100 (19.8-100)*	*100 (97.7-100)*	NR	1 quality control failure (haemolysis), 8 without reference standard (2 intrauterine fetal death, 1 TOP, 5 spontaneous miscarriages without karyotyping) → 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[64] USA	NR	T21	*39*	*7*	*0*	*0*	*100 (88.8-100)*	*100 (56.1-100)*	*100 (88.8-100)*	*100 (56.1-100)*	NR	252 (84.6%) putative non-trisomy cases without reference standard excluded
		T18	*7*	*39*	*0*	*0*	*100 (56.1-100)*	*100 (88.8-100)*	*100 (56.1-100)*	*100 (88.8-100)*	NR	
Sparks 2012[25] USA. Training 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 For both sets: T18 cases excluded from T21 test performance and vice versa → 15 (8.8%) for T21 and 43 (25.1%) for T18 excluded from training set 8 (4.8%) for T21 and 36 (21.6%) for T18 excluded from validation set
		T18	7	121	0	0	100 *(56.1-100)*	100 *(96.2-100)*	*100 (56.1-100)*	*100 (96.2-100)*	NR	
Validation set	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	*8*	*122*	*1*	*0*	100 *(59.8-100)*	99.2 *(94.9-99.96)*	*88.9 (50.7-99.4)*	*100 (96.2-100)*	NR	
Stumm 2014[22] 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 (14 failed QC, 18 failed libraries) → 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[82] 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	7 low fetal fraction (<4%); 9 assay failure; 24 other chromosomal abnormalities besides T21; 11 logistical problems (shipping time > 5 days or

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						
												incorrect labelling); 19 insufficient plasma volume → 70 (12.2%) excluded
Wang 2014[92] USA	NR	T21	38	NR	3	1	NA	NA	93	NA	FP rate (1-PPV), %: 7	No information on cfDNA testing failures or inconclusive results / no information on number and outcome of cfDNA testing-negative cases without invasive testing / no information on number and outcome of cfDNA testing-positive cases who refused invasive testing
		T18	16	NR	9	0	NA	NA	64	NA	FP rate (1-PPV), %: 36	
		T13	7	NR	9	0	NA	NA	44	NA	FP rate (1-PPV), %: 56	
Wax 2015[93] USA	NR	T21	3	161	0	0	*100 (31.0-100)*	*100 (97.1-100)*	*100 (31.0-100)*	*100 (97.1-100)*	NR	No test failures reported / no inconclusive results reported / 1 miscarriage without karyotype, 1 IUFD without karyotype → 2 (1.2%) excluded
		T18	1	163	0	0	*100 (5.5-100)*	*100 (97.1-100)*	*100 (5.5-100)*	*100 (97.1-100)*	NR	
		T13	0	164	0	0	NA	*100 (97.1-100)*	NA	*100 (97.1-100)*	NR	
Willems 2014[47] Netherlands, Belgium	> 4%	T21	NR	>2000	0	1	NR	NR	NR	NR	NR	32 test failures; 10/57 positive cfDNA testing results unconfirmed; Unclear how many cfDNA testing negative results with birth outcome; used 2000 for performance analysis
		T18	NR	>2000	0	1	NR	NR	NR	NR	NR	
		T13	NR	>2000	0	0	NR	NR	NR	NR	NR	
		All	*47*	*>2000*	*0*	*2*	*95.9 (84.9-99.3)*	*100 (99.8-100)*	*100 (90.6-100)*	*99.9 (99.6-99.98)*	NR	
Yu 2014[65] Hong Kong, Netherlands, UK	NR	T21	36	88	0	0	100 *(88.0-100)*	100 *(94.8-100)*	*100 (88.0-100)*	*100 (94.8-100)*	NR	20 cases with euploid fetus used as reference controls; No test failures → 20 (13.9%) excluded
		T18	27	97	0	0	100 *(84.5-100)*	100 *(95.3-100)*	*100 (84.5-100)*	*100 (95.3-100)*	NR	
		T13	20	102	1	1	95.2 *(74.1-99.8)*	99.0 *(93.9-99.9)*	*95.2 (74.1-99.8)*	*99.0 (93.9-99.9)*	NR	
Zhang 2015[24] China, Hong Kong	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)	FP rate, %: 0.05	211 inappropriate samples (inadequate volume, contamination, sample
		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)	FP rate, %: 0.05	

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						
(Denmark). Overall performance (n=112,669)		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)	FP rate, %: 0.04	obtained < 9 weeks' gestation or improper labelling), 145 test failures, 34,289 without karyotyping or clinical follow-up → 34,645 (23.5%) excluded.
		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)	FP rate, %: 0.14	
Twins only (n=404)	NR	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)	NR	
Zhou 2014[94] China. cfDNA testing implementation study	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)*	FP rate, %: 0.05 (0.02-0.10)	4 test failures / 0 inconclusive results / 5 TOP without karyotype, 5 IUFD without karyotype, 3741 cfDNA testing-negative cases without follow-up → 3755 (48.7%) excluded
		T18	10	3938	2	0	100 *(65.5-100)*	99.9 *(99.8-99.99)*	83.3 *(50.9-97.1)*	*100 (99.9-100)*	FP rate, %: 0.05 (0.02-0.10)	
		T13	2	3946	2	0	100 *(19.8-100)*	99.9 *(99.8-99.99)*	50.0 *(9.2-90.8)*	*100 (99.9-100)*	FP rate, %: 0.05 (0.02-0.10)	
Zimmermann 2012[26] 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 68/126 putative euploid samples without reference standard → 89/166 (53.6%) excluded
		T18	*3*	*74*	*0*	*0*	*100 (31.0-100)*	*100 (93.9-100)*	*100 (31.0-100)*	*100 (93.9-100)*	NR	
		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 fetal 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.

Table 20 Test failure rates, reasons for failure and repeat tests

Reference	cfDNA testing failure		Failure rate				Repeat tests	Repeat tests successful	Failure rate of combined test
	Reason	Frequency	8 weeks	10 weeks	12 weeks	14 weeks			
Alberti 2015[66] France	Insufficient fetal fraction; Assay failure	6/225 (2.7%) 5/225 (2.2%)	NR				NR	NR	NA
Ashoor 2012[54] UK	Pre-analytic failures: < 2 ml plasma, Labelling errors, Sample mixing or cross contamination Failed amplification and sequencing	25/400 (6.25%) 8 (2.0%) 5 (1.25%) 12 (3.0%) 3/400 (0.75%)	NR – All samples 11-13 weeks.				0/400	0/0	NA
Beamon 2014[83] USA	Foetal fraction below the threshold. Unclassifiable result: unclassified for T21, unclassified for T13, unclassified for T21 and T13	3/208 (1.4%) 5/45 (11.1%) samples tested by Verinata 3 1 1	NR				1/3 (33.3%)	0/1 (0%)	NA
Bevilacqua 2015[84] Belgium, UK, Spain	Low fetal fraction (lower fetal fraction of the 2 fetuses)	After 1st cfDNA testing: 29/515 (5.6%). After 2nd cfDNA testing: 16/515 (3.1%)	Before 14 weeks: first sample failure in 16/353 (4.5%).				26/29 (89.7%)	13/26 (50%)	NA
Bianchi 2012[55] USA	Sample tracking issues; No fetal DNA; Inconclusive results	2/534 (0.4%) 16/532 (3.0%) 14/532 (2.6%)	NR				NR	NR	NA
Bianchi 2014[67] USA	Pre-analytic failures: Insufficient blood volume, Late receipt Testing failure occurred during cfDNA extraction or sequencing No clear biological reasons	8/2052 (0.4%) 7 (0.3%) 1 (0.05%) 18/2042 (0.9%)	NR				12 inconclusive results re-sequenced in monoplex	12/12 (100%)	2 uninterpretable results for T21, 2 uninterpretable results for T18, 6 no results for T18. 39/2042 incomplete.

Reference	cfDNA testing failure		Failure rate				Repeat tests	Repeat tests successful	Failure rate of combined test
	Reason	Frequency	8 weeks	10 weeks	12 weeks	14 weeks			
Chen 2011[56] Hong Kong, UK, Netherlands, China	NA	0/289	NA				NR	NR	NA
Chiu 2011[57] Hong Kong, UK, Netherlands, China	Compromised blood sample: Samples collected after invasive obstetric procedure, Delayed blood processing, Ambiguous information, Haemolysed, Inadequate volume Failed QC: Failed DNA extraction, library preparation, or sequencing	46/810 (5.7%) 3 (0.4%) 2 (0.25%) 3 (0.4%) 12 (1.5%) 26 (3.2%) 11/764 (1.4%)	NR				NR	NR	NA
Comas 2014[85] Spain	Excluding chromosomal anomalies, success rate was dependent on cffDNA fraction (mean of 12.8% successful vs 7.3% unsuccessful) and maternal age (37.1 yrs vs 39.5 yrs, respectively)	1st cffDNA testing: 9/333 (2.7%) After 2nd cffDNA testing: 4/333 (1.2%)	NR				6/9 (67%)	5/6 (83.3%)	NA
Dan 2012[68] China, Hong Kong	Failed QC: Failed DNA extraction, library preparation or sequencing	79/11,184 (0.7%)	NR				97/11,105 (0.9%)	NR	NA
Dar 2014[86] USA	Pre-analytic failures: Insufficient serum/plasma, Sample collection date too old, Missing information, Sample damaged, Wrong tube	174/31030 (0.6%) 127 (0.4%) 28 (0.09%) 11 (0.035%) 4 (0.01%) 4 (0.01%)	NR				Redraw analysis in 21,292 cases with single case identification: 888/1572 (56.5%)	571/888 (64.3%)	NA

Reference	cfDNA testing failure		Failure rate				Repeat tests	Repeat tests successful	Failure rate of combined test
	Reason	Frequency	8 weeks	10 weeks	12 weeks	14 weeks			
	Failed quality metrics: Low fetal fraction, Labchip QC failed, Contamination, Laboratory error, Unexplained bad model fit, Insufficient DNA, Uninformative SNP pattern of unknown origin, Multiple sequencing failures, Suspected egg donor/surrogate, Maternal loss of heterozygosity, Foetal loss of heterozygosity, Suspected maternal mosaicism, Suspected fetal mosaicism	1966/30,705 (6.4%): 1667 (84.8%) 48 (2.4%) 42 (2.1%) 34 (1.7%) 24 (1.2%) 17 (0.9%) 13 (0.7%) 9 (0.5%) 60 (3.1%) 38 (1.9%) 12 (0.6%) 1 (0.05%) 1 (0.05%)							
Del Mar Gil 2014[69] UK	Low fetal fraction, laboratory processing issues	11/207 (5.3%) 4/207 (1.9%)	NR - All samples 11-13 weeks.				0/207	0/0	NA

Reference	cfDNA testing failure		Failure rate				Repeat tests	Repeat tests successful	Failure rate of combined test
	Reason	Frequency	8 weeks	10 weeks	12 weeks	14 weeks			
Dhallan 2007[97] USA	NA	0/60	NA				NR	NR	NA
Ehrich 2011[58] USA	Pre-analytic failures: Plasma volume < 3.5 ml, Sample dropped during DNA extraction, Samples mixing, Tube broke during centrifugation Failed QC: Foetal percentage ≤ 3.9% (only for 4-plex), total DNA ≤556 copies, Library conc. ≤32.2 nmol, Unique counts ≤3M for 4-plex or ≤12M for monoplex)	13/480 (2.7%) 9 (1.9%) 1 (0.2%) 2 (0.4%) 1 (0.2%) 18/467 (4%)	NR				10 samples with >3.9% fetal DNA re-sequenced in tetraplex, 10 samples with lower fetal DNA re-sequenced in monoplex	2/20 (10%)	NA
Fang 2015[95] China	NR	NR	NR				NR	NR	NR
Futch 2013[70] USA	Pre-analytic failures: Inadequate blood volume, Improper labelling or unlabelled, Samples received beyond stability or compromised in transit, Wrong sample type Interfering substances (excess cffDNA), Insufficient DNA extracted Unclassifiable result	75/6123 (1.2%) 43 (0.7%) 26 (0.4%) 5 (0.08%) 1 (0.02%) 33/6017 (0.5%) 10/6017 (0.17%) 173/6017 (2.9%)	NR				NR	NR	NA
Hall 2014[59] USA	Failed QC (fetal fraction < 4%)	4/68 (5.9%)	NR				NR	NR	NA
Huang 2014[71] China, Denmark, Hong Kong	NA	0/189	NA				NR	NR	NA
Jensen 2013[98] USA	Pre-analytic failures: Sample not adequate	116/4664 (2.5%)	NA				NR	NR	NA

Reference	cfDNA testing failure		Failure rate				Repeat tests	Repeat tests successful	Failure rate of combined test
	Reason	Frequency	8 weeks	10 weeks	12 weeks	14 weeks			
	cfDNA testing failure	0/1269							
Jeon 2014[87] South Korea, China	NA	0/155 (0%)	NR				NR	NR	NA
Jiang 2012[72] China	NA	0/903	NA				NR	NR	NA
Korostelev 2014[88] Russia	Non-informative pattern obtained from sex chromosomes did not allow determining gender of the fetus and the probability of chromosomal abnormalities associated with it	1/1728 (0.06%)	NR				NR	NR	NA
Lau 2012[73] Hong Kong, China, Japan	NA	0/108	NA				NR	NR	NA
Lau 2014[74] Hong Kong, USA, China	Initial test failure Repeat test failure: Low fetal fraction, vanished twin	23/1982 (1.16%) 1/1982 (0.05%)	NR				23/1982 (1.16%)	22/23 (95.7%)	NA
Liang 2013[75] China	Failed QC: < 2M uniquely mapped tags; > 6% tags redundant; < 60% sequence tags uniquely mapped to the unmasked genome; GC content of mapped reads <40%	12/435 (2.8%) 7/435 (1.6%) 3/435 (0.7%) 1/435 (0.2%) 1/435 (0.2%)	NR				NR	NR	NA
Liao 2014[60] China, USA	NR	NR	NR				NR	NR	NA
McCullough 2014[89] USA	Insufficient fetal DNA (<4.0% ccffDNA or <100 copies of ccffDNA); Other not reportable aetiologies (samples failing all other laboratory quality metric including library and sequencing passing criteria; generally technical or	842 (0.9%) (0.54% after redraw); 1086 (1.0%) (0.1% after redraw).	NR				For insufficient fetal DNA: 571 (67.2%); For other not reportable aetiologies: 493 (47.9%); Total redraws: 1064	416/571 (72.9%); 443/493 (89.9%) 859/1064 (80.7%).	NA

Reference	cfDNA testing failure		Failure rate				Repeat tests	Repeat tests successful	Failure rate of combined test
	Reason	Frequency	8 weeks	10 weeks	12 weeks	14 weeks			
	mechanical failures); Not reportable after redraw	 1,330 (1.3%)							
Nicolaides 2012[76] UK	Pre-analytic failures: Inadequate sample volume, Wrongly labelled, Lab mixed samples together Foetal fraction < 4%; Assay failure.	100/2149 (4.7%) 29 (1.3%) 1 (0.05%) 70 (3.3%) 46/2049 (2.2%) 54/2049 (2.6%)	NR – All samples 11-13 weeks.				NR	NR	0/2049
Nicolaides 2013[77] UK	Failed QC: Low fetal fraction (<3.5%), insufficient total DNA after preparation (<1500 input genomic DNA units), high noise levels	13/242 (5.4%)	NR – All samples 11-13 weeks.				0/242	0/0	NA
Norton 2012[78] USA, Sweden, Netherlands	Pre-analytic failures: Insufficient sample volume, Incorrect sample labeling. Low (<4%) fetal fraction, Assay failure (inability to measure fraction of fetal cfDNA, unusually high variation in cfDNA counts, failed sequencing)	104/4002 (2.6%) 84 (2.1%) 20 (0.5%) 57/3228 (1.8%) 91/3228 (2.8%)	NR				NR	NR	NA
Norton 2015[46] USA, Sweden	Pre-analytic failures: Sample-collection or labelling error Foetal fraction < 4%, Foetal fraction that could not be measured, High assay variance or assay failure	 384/18,510 (2.1%) 192/16,329 (1.2%) 83/16,329 (0.5%) 213/16,329 (1.3%)	NR				NR	NR	308/18,955 (1.6%)
Palomaki 2011[61] USA	Pre-analytic failures: Sample not adequate Initial test failure Failed QC after repeat test:	 116/4664 (2.5%) 90/1696 (5.3%) 13/1696 (0.8%)	NR				90/1696 (5.3%)	77/90 (85.6%)	NA

Reference	cfDNA testing failure		Failure rate				Repeat tests	Repeat tests successful	Failure rate of combined test
	Reason	Frequency	8 weeks	10 weeks	12 weeks	14 weeks			
	Foetal fraction < 4%, fetal fraction > 50%, failed library, multiple failures	6/1696 (0.4%) 1/1696 (0.06%) NR							
Palomaki 2012[62] USA	Pre-analytic failures: Sample not adequate Initial test failure Repeat test failure: (most common reason fetal fraction <4%)	 116/4664 (2.5%) 110/1988 (5.5%) 17/1988 (0.9%)	NR				110/1988 (5.5%)	93/110 (84.5%)	NA
Pergament 2014[79] USA	Failed QC: Low fetal fraction; low input DNA; contamination; loss of heterozygosity; poor model fit	85/1051 (8.1%) 64/1051 (6.1%) 12/1051 (1.1%) 6/1051 (0.6%) 2/1051 (0.2%) 1/1051 (0.1%)	< 9 weeks: 26/95 (27.4%) 9.0-9.9 weeks: 6/50 (12.0%) ≥ 10 weeks: 53/900 (5.9%)				0/1051	0/0	NA
Porreco 2014[80] USA	Pre-analytic failures: Insufficient sample volume, Outside 6 hour lab processing window Failed QC: Foetal fraction <4.0%, fetal fraction > 50%, fetal DNA > 26 copies, library concentration < 7.5 nmol, number of autosomal aligned reads < 9 million	440/4170 (10.6%) 320 (7.7%) 120 (2.9%) 54/3430 (1.6%)	NR				NR	NR	NA
Quezada 2015[90] UK	Initial test failures Total test failures after redraw: Sample not received by lab, Low fetal fraction (<4 %), Assay failure	123/2905 (4.2%) 54/2905 (1.9%) 1 38 15	All 10-11 weeks' gestation.				110/123 (89.4%)	69/110 (62.7%)	12/2905 (0.4%)
Sago 2014[91] Japan	Not reportable at first cfDNA testing Not reportable after repeat cfDNA testing	18/7740 (0.2%) 4/7740 (0.05%)	NR				16/18 (88.9%)	14/16 (87.5%)	NA

Reference	cfDNA testing failure		Failure rate				Repeat tests	Repeat tests successful	Failure rate of combined test
	Reason	Frequency	8 weeks	10 weeks	12 weeks	14 weeks			
Sehnert 2011[63] USA	1 inconclusive result for T13 ("no call")	Test set: 1/47 (2.1%)	NR				NR	NR	NA
Shaw 2014[81] Taiwan, China	1 case excluded due to early gestational age (10 weeks)	1/201 (0.5%)	NR				NR	NR	NA
Song 2013[23] China	Failed QC: DNA QC, Sequencing QC	73/1916 (3.8%) 48/1916 (2.5%) 25/1916 (1.3%)	NR				No re-draw permitted.	0/0	NR
Song 2015[96] China	Pre-analytic quality control failure: Haemolysis (Foetal fraction < 4%)	1/213 (0.5%) (5/100 (5%) male samples)	NR				0/213	0/0	NA
Sparks 2012[64] USA	NA	0/298	NR				NR	NR	NA
Sparks 2012[25] USA	Failed QC: Low count, fetal fraction <3%, and/or evidence from SNPs of a nonsingleton pregnancy)	Training set: 8/171 (4.7%), Validation set: 0/167	NR				NR	NR	NA
Stumm 2014[22] Germany, Switzerland	Failed QC: Failed sequencing QC (<10 million reads), failed libraries	32/504 (6.3%) 14/504 (2.8%) 18/504 (3.6%)	NR				0/504	0/0	NA
Verweij 2013[82] Netherlands, Sweden, USA	Pre-analytic failures: Insufficient plasma volume; Logistical problems - shipping time > 5 days or incorrect labelling Initial test failures Repeat test failures: Low fetal fraction (<4%), assay failure	30/595 (5.0%) 19 (3.2%) 11 (1.8%) 51/520 (9.8%) 16/520 (3.1%) 7/520 (1.3%) 9/520 (1.7%)	NR				51/520 (9.8%)	35/51 (68.6%)	NA
Wang 2014[92] USA	NR	NR	NR				NR	NR	NR
Wax 2015[93] USA	NA	0/166 (0%)	NR				NR	NR	NA
Willems 2014[47] Netherlands,	Initial test failures (fetal fraction < 4%)	55/3000 (1.8%)	NR				50/3000 (1.7%) 50/55 (90.9%)	23/50 (46.0%)	NA

Reference	cfDNA testing failure		Failure rate				Repeat tests	Repeat tests successful	Failure rate of combined test
	Reason	Frequency	8 weeks	10 weeks	12 weeks	14 weeks			
Belgium	Repeat test failures: Foetal fraction < 4% in 1st and repeated blood sample(s); Foetal fraction < 4% in 1st sample and no 2nd attempt.	32/3000 (1.1%) 27/3000 (0.9%) 5/3000 (0.17%)							
Yu 2014[65] Hong Kong, Netherlands, UK	NA	0/124	NR				NR	NR	NA
Zhang 2015[24] China, Hong Kong, (Denmark)	Pre-analytic failures: Inadequate volume, contamination, sample obtained < 9 weeks' gestation or improper labelling. Initial test failures Repeat test failures: Quality control failure, assay failure, or low fetal fraction (<3.5%).	 211/147,314 (0.14%) 3213/147,103 (2.2%) 145/147,103 (0.098%)	NR				3,213/147,103 (2.18%)	3,068/ 3,213 (95.5%)	NA
Zhou 2014[94] China	1st failure: Low fetal DNA fraction and failure in quality control criteria. Failure after repeat cfDNA testing: persistently low fetal fraction.	141/7705 (1.8%) 4/7705 (0.05%)	NR				141/141 (100%)	137/141 (97.2%)	NA
Zimmermann 2012[26] USA	Failed DNA QC: <4.0% fetal fraction or DNA quality metric (quality of the plasma sequence data, noise levels, how well the data corresponded to the statistical model, and the calculated accuracies) below threshold.	21/166 (12.7%)	NR				NR	NR	NA

Ccff, circulating cell-free fetal; cff, cell-free fetal; cf, cell-free; DNA, deoxyribonucleic acid; NA, not applicable; NR, not reported; QC, quality control; SNP, single-nucleotide polymorphism

Table 21 Test failure rates and inconclusive results by karyotype

Reference	cfDNA testing failure	Karyotype / Birth outcome	Inconclusive cfDNA testing result	Karyotype / Birth outcome
Alberti 2015[66]	11 test failures	5 euploid, 1 T21, 5 NR	No unclassified results	NA
Ashoor 2012[54]	3 test failures	3/300 euploid, 0/50 T21, 0/50 T18	No unclassified results	NA
Beamon 2014[83]	3 failed initial cfDNA testing, 1/1 failed repeat cfDNA testing	3 obese patients, karyotype NR NR	5 unclassifiable results	1 T18, 2 euploid, 2 IUFD without karyotype
Bevilacqua 2015[84]	16 test failures	NR "Univariable regression analysis demonstrated that significant predictors of failure of the test were twin pregnancy, higher maternal weight and conception by IVF, but not gestational age at test, cigarette smoking or origin of oocyte."	No unclassified results	NA
Bianchi 2012[55]	16 No fetal DNA detected	3/9 69,XXX "After unblinding, there were no distinguishing clinical features for these samples"	7 unclassified for T21 5 unclassified for T18 2 unclassified for T13	1 affected, 6 unaffected (1 T13) 2 affected, 3 unaffected (1 T21) 2 affected
Bianchi 2014[67]	18 test failures	"No clear biologic reasons"	No unclassified results	NA
Chen 2011[56]	No test failures	NA	No unclassified results	NA
Chiu 2011[57]	11 test failures	NR	No unclassified results	NA
Comas 2014[85]	4 test failures	3 euploid (1 had early vanishing twin), 1 preeclampsia and IUGR (karyotype NR)	No unclassified results	NA
Dan 2012[68]	97 needed resampling, 79 test failures.	NR NR	No unclassified results	NA
Dar 2014[86]	1966 test failures	NR	No unclassified results	NA
Del Mar Gil 2014[69]	15 test failures	0/10 T21, 1/1 T18, 2/3 T13, 12/193 euploid	No unclassified results	NA
Dhallan 2007[97]	No test failures	NA	No unclassified results	NA
Ehrich 2011[58]	31 test failures: 13 preanalytic failures 18 failed sequencing	3/42 T21, 28/438 euploid: 1/42 T21, 12/438 euploid 2/42 T21, 16/438 euploid	No unclassified results	NA
Fang 2015[95]	NR	NR	NR	NR
Futch 2013[70]	106 administrative cancellations 43 technical cancellations	NR NR	173 unclassifiable results	50/173 with outcome: 3 T18, 1 45,XX (presumed balanced Robertsonian translocation), 46 euploid.
Hall 2014[59]	4 test failures	2/17 T13, 2/51 euploid	No unclassified results	NA

Reference	cfDNA testing failure	Karyotype / Birth outcome	Inconclusive cfDNA testing result	Karyotype / Birth outcome
Huang 2014[71]	No test failures	NA	No unclassified results	NA
Jensen 2013[98]	No test failures	NA	No unclassified results	NA
Jeon 2014[87]	No test failures	NA	No unclassified results	NA
Jiang 2012[72]	No test failures	NA	No unclassified results	NA
Korostelev 2014[88]	No test failures	NA	1 inconclusive result for SCA/gender	Microdeletion of chrX and microduplication of chrY
Lau 2012[73]	No test failures	NA	No unclassified results	NA
Lau 2014[74]	23 repeat blood samples, 1 test failure after repeat testing	NR 1 euploid, vanished twin	No unclassified results	NA
Liang 2013[75]	12 test failures	NR	No unclassified results	NA
Liao 2014[60]	No test failures	NA	No unclassified results	NA
McCullough 2014[89]	1330 not reportable after redraw	NR	No unclassified results	NA
Nicolaides 2012[76]	100 test failures	1/3 T18, 0/8 T21, 99/2038 euploid	No unclassified results	NA
Nicolaides 2013[77]	13 test failures	2/27 T21, 0/3 T18, 0/1 T13, 11/204 euploid	No unclassified results	NA
Norton 2012[78]	148 test failures: 57 low fetal fraction, 91 assay failure	NR "Assay failure rate was comparable in normal vs. trisomy cases. The reasons for assay failure, including low fetal fraction, are not known. We did not find assay failure associated with gestational age, maternal age, race/ethnicity or fetal karyotype."	No unclassified results	NA
Norton 2015[46]	488 test failures	13 aneuploidies (3/41 T21, 1/11 T18, 2/8 T13, 4 triploidy, 1 T16 mosaic, 1 with deletion 11p, and 1 with a structurally abnormal chromosome), 475/15773 euploid "Prevalence of aneuploidy in this group (1 in 38 [2.7%]) is higher than the prevalence of 1 in 236 (0.4%) in the overall cohort (P<0.001)."	No unclassified results	NA
Palomaki 2011[61]	90 initial test failures 13 repeat test failures	NR 0/212 T21, 13/1484 euploid	No unclassified results	NA
Palomaki 2012[62]	110 initial test failures, 17 repeat test failures after repeat cfDNA testing	NR 0/12 T13, 3/62 T18, 0/212 T21, 14/1702 euploid	No unclassified results	NA

Reference	cfDNA testing failure	Karyotype / Birth outcome	Inconclusive cfDNA testing result	Karyotype / Birth outcome
		"Similar aneuploidy risk in the 17 women with repeat test failures (17.6%) to the 15% occurring in the population with successful testing."		
Pergament 2014[79]	85 test failures	8/67 T21, 7/32 T18, 2/14 T13, 2/12 45X, 66/926 euploid "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 call (105/966 [10.9%], P=0.004)."	1 unclassified for T21	T21
Porreco 2014[80]	54 test failures	NR	No unclassified results	NA
Quezada 2015[90]	123 initial test failures 54 repeat test failures	49 non-trisomic, 2 T21, 72 unknown 49 non-trisomic, 2 T21, 3 miscarriages without karyotype	No unclassified results	NA
Sago 2014[91]	18 failed initial cfDNA testing, 2 failed repeat cfDNA testing	NR NR	No unclassified results	NA
Sehnert 2011[63]	No test failures	NA	1 "no call" (inconclusive: NCV=3) result	1/1 T13
Shaw 2014[81]	No test failures	NA	No unclassified results	NA
Song 2013[23]	73 test failures	64 non-trisomy, 9 no follow-up.	No unclassified results	NA
Song 2015[96]	No test failures	NA	No unclassified results	NA
Sparks 2012[64]	No test failures	NA	No unclassified results	NA
Sparks 2012[25]	Training set: 8 test failures Validation set: No test failures	1/8 T18, 1/36 T21, 6/127 euploid NA	No unclassified results No unclassified results	NA NA
Stumm 2014[22]	32 test failures	1/43 T21, 0/8 T18, 0/5 T13, 31/448 non-trisomic	No unclassified results	NA
Verweij 2013[82]	51 failed initial cfDNA testing, 16 failed repeat cfDNA testing (7 low fetal fraction, 9 assay failure)	NR NR	No unclassified results	NA
Wang 2014[92]	No test failures	NA	No unclassified results	NA
Wax 2015[93]	No test failures	NA	No unclassified results	NA
Willems 2014[47]	55 failed initial cfDNA testing, 27 failed repeat cfDNA testing	NR 1 triploidy	No unclassified results	NA
Yu 2014[65]	No test failures	NA	No unclassified results	NA

Reference	cfDNA testing failure	Karyotype / Birth outcome	Inconclusive cfDNA testing result	Karyotype / Birth outcome
Zhang 2015[24]	3213 initial test failures, 145 test failures after redraw	NR NR	No unclassified results	NA
Zhou 2014[94]	141 initial test failures, 4 test failures after redraw	NR NR	No unclassified results	NA
Zimmermann 2012[26]	21 test failures	0/11 T21, 0/3 T18, 0/2 T13, 1/2 45X, 20/146 putative euploid	No unclassified results	NA

IUFD, intrauterine fetal death; IUGR, intrauterine growth restriction; NA, not applicable; NR, not reported

9. References

1. Nicolaides KH, Syngelaki A, Poon LC, Gil MM, Wright D. 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):185-92.
2. Coppus AM, Evenhuis HM, Verberne GJ, Visser FE, Oostra BA, Eikelenboom P*, et al.* Survival in elderly persons with Down syndrome. *J Am Geriatr Soc*. 2008;56(12):2311-6.
3. Edwards' syndrome (trisomy 18) NHS Choices; 25/09/2014 [cited 05/05/2015]; Available from: http://www.nhs.uk/conditions/edwards-syndrome/Pages/Introduction.aspx.
4. Morris JKS, A. The National Down Syndrome Cytogenetic Register for England and Wales: 2012 Annual Report. 2014.
5. Willacy H. Edwards' Syndrome (Trisomy 18). Patient.co.uk; 18/03/2011 [cited 05/05/2015]; Available from: http://www.patient.co.uk/doctor/edwards-syndrome-trisomy-18.
6. Tidy C. Patau's Syndrome (Trisomy 13). 28/03/2013 [cited 05/05/2015]; Available from: http://www.patient.co.uk/doctor/pataus-syndrome-trisomy-13.
7. Patau syndrome. NHS National Genetics and Genomics Education Centre 01/08/2013 [cited 05/05/2015]; Available from: http://www.geneticseducation.nhs.uk/genetic-conditions-54/691-patau-syndrome-new.
8. Statistical bulletin: Births in England and Wales, 2013. Office for national statistics; 2014 [cited 12/05/2015]; Available from: http://www.ons.gov.uk/ons/rel/vsob1/birth-summary-tables--england-and-wales/2013/stb-births-in-england-and-wales-2013.html.
9. Carothers AD, Hecht CA, Hook EB. International variation in reported livebirth prevalence rates of Down syndrome, adjusted for maternal age. *J Med Genet*. 1999;36(5):386-93.
10. Morris JK, Mutton DE, Alberman E. Revised estimates of the maternal age specific live birth prevalence of Down's syndrome. *J Med Screen*. 2002;9(1):2-6.
11. Crane E, Morris JK. Changes in maternal age in England and Wales--implications for Down syndrome. *Downs Syndr Res Pract*. 2006;10(1):41-3.
12. Wald N, Rodeck C, Hackshaw A, McGuire A. Antenatal screening for Down's syndrome. *Health Technology Assessment*. 1998;2(1).
13. Ward PA, Gray MJ. National Down's Syndrome screening programme for England - A handbook for staff. London: UK National Screening Committee. 2004 04 June 2012 [cited; Available from: fetalanomaly.screening.nhs.uk/getdata.php?id=10659.
14. NICE. Antenatal care: NICE clinical guideline 62. 2008 04 June 2012 [cited; Available from: http://www.nice.org.uk/guidance/cg62.
15. NHS Fetal Anomaly Screening Programme: Possible screening results. NHS Screening Programmes; [cited 05/05/2015]; Available from: http://www.fetalanomaly.screening.nhs.uk/cms.php?folder=2951.
16. Wright C. Cell-free fetal nucleic acids for non-invasive prenatal diagnosis: Report of the UK expert working group. 2009.
17. Sparkes R JJ-A, Langlois S. New molecular techniques for the prenatal detection of chromosomal aneuploidy. *Journal of Obstetrics and Gynecology Canada*. 2008;30(7):617-21.
18. 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.
19. Simpson JL, Elias S. Isolating fetal cells from maternal blood: Advances in prenatal diagnosis through molecular technology. *JAMA (Journal of the American Medical Association)*. 1993;270(19):2357-61.
20. Lo YMD, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CWG*, et al.* Presence of fetal DNA in maternal plasma and serum. *The Lancet*. 1997;350(9076):485-7.
21. Norwitz ER, Levy B. Noninvasive prenatal testing: the future is now. *Revue Obstetricale et Gynecologique*. 2013;6(2):48-62.

22. 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*. Prenatal Diagnosis*. 2014;34(2):185-91.
23. 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*. Prenatal Diagnosis*. 2013;33(7):700-6.
24. 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.
25. 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*. American Journal of Obstetrics & Gynecology*. 2012;206(4):319.e1-9.
26. 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*. Prenatal Diagnosis*. 2012;32(13):1233-41.
27. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB*, et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies*. Ann Intern Med*. 2011;155(8):529-36.
28. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance*. J Clin Epidemiol*. 2003;56(11):1129-35.
29. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy*. Ann Intern Med*. 2008;149(12):889-97.
30. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews*. J Clin Epidemiol*. 2005;58(10):982-90.
31. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations*. Stat Med*. 2001;20(19):2865-84.
32. Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests*. BMJ*. 2001;323(7305):157-62.
33. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method*. Med Decis Making*. 1993;13(4):313-21.
34. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy*. J Clin Epidemiol*. 1995;48(1):119-30; discussion 31-2.
35. Dwamena BA. MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies. Boston, Massachusetts: Boston College Department of Economics; 2008.
36. 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.
37. Shinkins B, Thompson M, Mallett S, Perera R. Diagnostic accuracy studies: how to report and analyse inconclusive test results*. Bmj*. 2013;346:f2778.
38. Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests*. Stat Med*. 2002;21(11):1525-37.
39. 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(9):882-93.
40. Harbord RM, Higgins JPT. Meta-regression in Stata*. Stata Journal*. 2008;8(4):493-519.
41. Harbord RM, Whiting P. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression*. Stata Journal*. 2009;9(2):211-29.
42. White IR. Multivariate random-effects meta-regression: Updates to mvmeta*. Stata Journal*. 2011;11(2):255-70.

43. Rowe R, Puddicombe D, Hockley C, Redshaw M. Offer and uptake of prenatal screening for Down syndrome in women from different social and ethnic backgrounds. *Prenat Diagn*. 2008;28(13):1245-50.
44. O'Leary P, Maxwell S, Murch A, Hendrie D. Prenatal screening for Down syndrome in Australia: costs and benefits of current and novel screening strategies. *Aust N Z J Obstet Gynaecol*. 2013;53(5):425-33.
45. Lewis C, Hill M, Silcock C, Daley R, Chitty LS. Non-invasive prenatal testing for trisomy 21: a cross-sectional survey of service users' views and likely uptake. *BJOG: An International Journal of Obstetrics & Gynaecology*. 2014;121(5):582-94.
46. 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.
47. Willems PJ, Dierickx H, Vandenakker E, Bekedam D, Segers N, Deboulle K*, et al.* The first 3,000 Non-Invasive Prenatal Tests (NIPT) with the Harmony test in Belgium and the Netherlands. *Facts Views & Vision in Obgyn*. 2014;6(1):7-12.
48. Gil MM, Giunta G, Macalli EA, Poon LC, Nicolaides KH. UK NHS pilot study on cell-free DNA testing in screening for fetal trisomies: factors affecting uptake. *Ultrasound Obstet Gynecol*. 2015;45(1):67-73.
49. Hill M, Taffinder S, Chitty LS, Morris S. Incremental cost of non-invasive prenatal diagnosis versus invasive prenatal diagnosis of fetal sex in England. *Prenatal Diagnosis*. 2011;31(3):267-73.
50. Chung NL. Decision Planning Tool: Methodology Report. 2012.
51. Unit Costs of Health and Social Care 2014. PSSRU; [cited 03/06/2015]; Available from: http://www.pssru.ac.uk/project-pages/unit-costs/2014/.
52. Tabor A, Vestergaard CH, Lidegaard O. Fetal loss rate after chorionic villus sampling and amniocentesis: an 11-year national registry study. *Ultrasound Obstet Gynecol*. 2009;34(1):19-24.
53. Akolekar R, Beta J, Picciarelli G, Ogilvie C, D'Antonio F. Procedure-related risk of miscarriage following amniocentesis and chorionic villus sampling: a systematic review and meta-analysis. *Ultrasound Obstet Gynecol*. 2015;45(1):16-26.
54. Ashoor G, Syngelaki A, Wagner M, Birdir C, Nicolaides KH. Chromosome-selective sequencing of maternal plasma cell-free DNA for first-trimester detection of trisomy 21 and trisomy 18. *American Journal of Obstetrics & Gynecology*. 2012;206(4):322.e1-5.
55. 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.
56. 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 [Electronic Resource]*. 2011;6(7):e21791.
57. 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.
58. 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. *American Journal of Obstetrics & Gynecology*. 2011;204(3):205.e1-11.
59. 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 [Electronic Resource]*. 2014;9(5):e96677.
60. Liao C, Yin AH, Peng CF, Fu F, Yang JX, Li R*, et al.* Noninvasive prenatal diagnosis of common aneuploidies by semiconductor sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(20):7415-20.
61. 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*. 2011;13(11):913-20.

62. Palomaki GE, Deciu C, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM*, et al.* DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international collaborative study. *Genetics in Medicine*. 2012;14(3):296-305.
63. 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. *Clinical Chemistry*. 2011;57(7):1042-9.
64. 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. *Prenatal Diagnosis*. 2012;32(1):3-9.
65. 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. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(23):8583-8.
66. 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.
67. Bianchi DW, Parker RL, Wentworth J, Madankumar R, Saffer C, Das AF*, et al.* DNA sequencing versus standard prenatal aneuploidy screening. *New England Journal of Medicine*. 2014;370(9):799-808.
68. 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. *Prenatal Diagnosis*. 2012;32(13):1225-32.
69. 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 Diagnosis & Therapy*. 2014;35(3):204-11.
70. 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. *Prenatal Diagnosis*. 2013;33(6):569-74.
71. 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. *Prenatal Diagnosis*. 2014;34(4):335-40.
72. 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).
73. 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. *Journal of Maternal-Fetal & Neonatal Medicine*. 2012;25(8):1370-4.
74. 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 in Obstetrics & Gynecology*. 2014;43(3):254-64.
75. 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. *Prenatal Diagnosis*. 2013;33(5):409-15.
76. Nicolaides KH, Syngelaki A, Ashoor G, Birdir C, Touzet G. Noninvasive prenatal testing for fetal trisomies in a routinely screened first-trimester population. *American Journal of Obstetrics & Gynecology*. 2012;207(5):374.e1-6.
77. 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. *Prenatal Diagnosis*. 2013;33(6):575-9.
78. 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.
79. 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. *Obstetrics & Gynecology*. 2014;124(2 Pt 1):210-8.
80. 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.
81. 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 Diagnosis & Therapy*. 2014;35(1):13-7.
82. 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. *Prenatal Diagnosis*. 2013;33(10):996-1001.
83. Beamon CJ, Hardisty EE, Harris SC, Vora NL. A single center's experience with noninvasive prenatal testing. *Genetics in Medicine*. 2014;16(9):681-7.
84. 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.
85. 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.
86. 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. *American Journal of Obstetrics & Gynecology*. 2014;211(5):527.e1-.e17.
87. 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 [Electronic Resource]*. 2014;9(10):e110240.
88. 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.
89. 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 [Electronic Resource]*. 2014;9(10):e109173.
90. 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.
91. 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.
92. 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.
93. 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.
94. 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. *Prenatal Diagnosis*. 2014;34(11):1061-5.
95. 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 Biochem Biophys*. 2015.

96. 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 in Obstetrics & Gynecology*. 2015;45(1):55-60.
97. 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.
98. 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 [Electronic Resource]*. 2013;8(3):e57381.
99. Gil MM, Akolekar R, Quezada MS, Bregant B, Nicolaides KH. Analysis of cell-free DNA in maternal blood in screening for aneuploidies: meta-analysis. *Fetal Diagnosis & Therapy*. 2014;35(3):156-73.
100. Gratacos E, Nicolaides K. Clinical perspective of cell-free DNA testing for fetal aneuploidies. *Fetal Diagn Ther*. 2014;35(3):151-5.
101. Wright D, Wright A, Nicolaides KH. A unified approach to risk assessment for fetal aneuploidies. *Ultrasound Obstet Gynecol*. 2015;45(1):48-54.
102. Mersy E, Smits LJ, van Winden LA, de Die-Smulders CE, South-East Netherlands NC, Paulussen AD*, 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. *Human Reproduction Update*. 2013;19(4):318-29.
103. Verweij EJ, van den Oever JM, de Boer MA, Boon EM, Oepkes D. Diagnostic accuracy of noninvasive detection of fetal trisomy 21 in maternal blood: a systematic review. *Fetal Diagn Ther*. 2012;31(2):81-6.
104. Gil MM, Quezada MS, Revello R, Akolekar R, Nicolaides KH. Analysis of cell-free DNA in maternal blood in screening for fetal aneuploidies: updated meta-analysis. *Ultrasound Obstet Gynecol*. 2015;45(3):249-66.

10. Appendix 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)

11. Appendix 2 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. *Am J Med Genet A* 2014;164A(6):viii-ix.	Commentary
2. Anonymous. Trisomy 21 DNA test (MaterniT21) for detecting Down syndrome in the first trimester. *Manag Care* 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. Ultrasound Obstet Gynecol 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. Obstetrical and Gynecological Survey. 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. Prenatal Diagnosis, 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. Proceedings of the National Academy of Sciences of the United States of America, 2008. 105(51): p. 20458-63.	Case-control study: < 15 cases
7. Deng, Y.H., et al., Non-invasive prenatal diagnosis of trisomy 21 by reverse transcriptase multiplex ligation-dependent probe amplification. Clinical Chemistry & Laboratory Medicine, 2011. 49(4): p. 641-6.	Not cff DNA (cell-free fetal RNA)
8. 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 referring centre. Journal of Prenatal Medicine. 2014;8(1-2):31-5.	Case series: < 15 cases
9. 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
10. 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
11. Feenstra, H., et al., Complexity of noninvasive prenatal screening and diagnostic testing for an unbalanced translocation involving chromosomes 5 and 18.	Case report

Prenatal Diagnosis, 2014. 34: p. 195-198.	
12. 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
13. 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
14. 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)
15. 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
16. Guex, N., et al., A robust second-generation genome-wide test for fetal aneuploidy based on shotgun sequencing cell-free DNA in maternal blood. Prenatal Diagnosis, 2013. 33: p. 707-710.	Letter
17. 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
18. Hayes Inc., Harmony? Prenatal Test (Structured abstract). Health Technology Assessment Database, 2012.	Abstract of review
19. Hayes Inc., Noninvasive Prenatal Testing (NIPT) for fetal aneuploidy (Structured abstract). Health Technology Assessment Database, 2013.	Abstract of review
20. 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
21. Hyett J. Non-invasive prenatal testing for down syndrome. Australian Prescriber. 2014;37(2):51-5.	Review
22. 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
23. 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
24. 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
25. 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	No additional diagnostic accuracy data to Palomaki 2011[61]

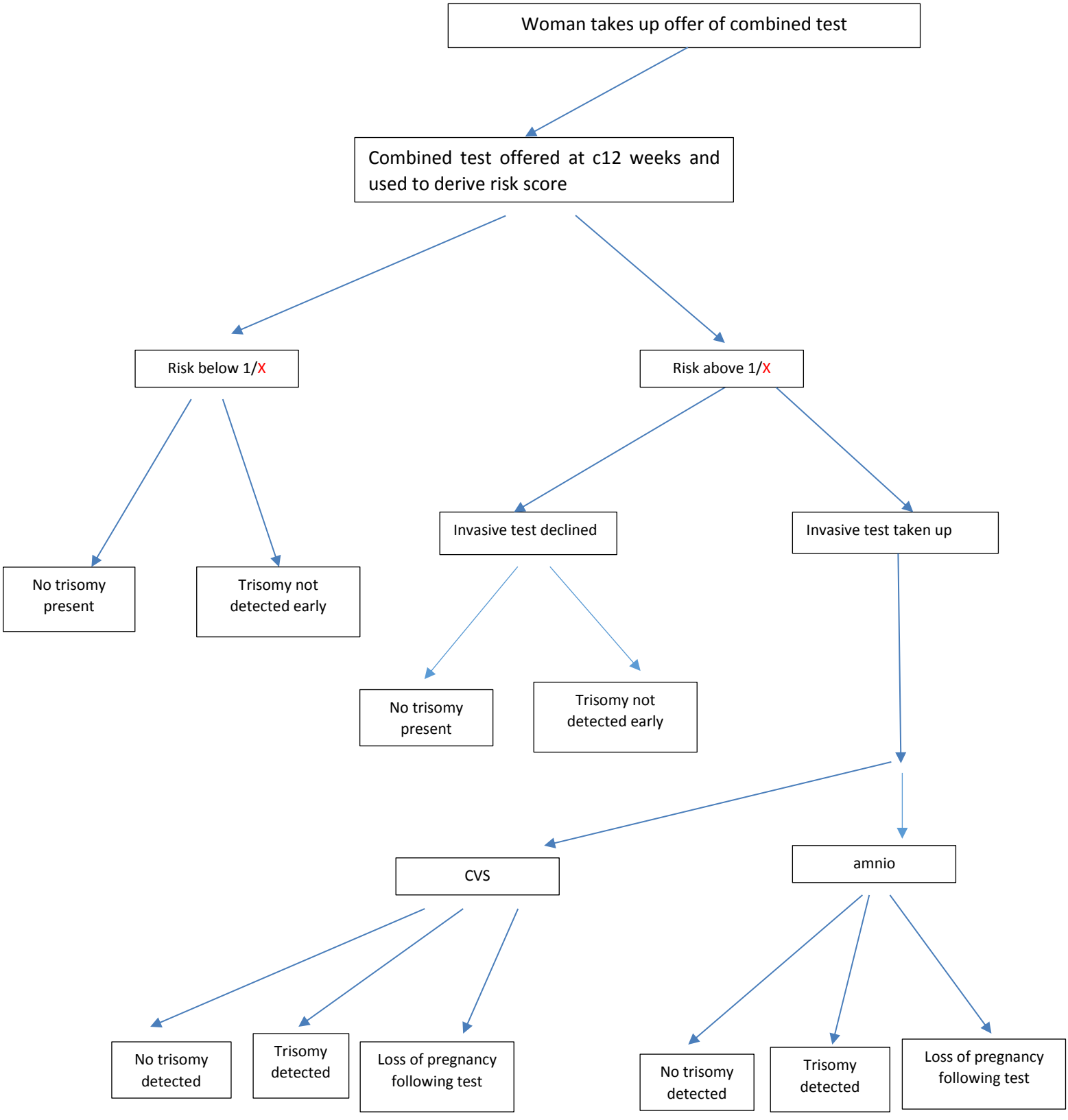
assisted reproductive technologies. Genetics in Medicine. 2014;16(5):419-22.	
26. Larion S, Warsof SL, Romary L, Mlynarczyk M, Peleg D, Abuhamad AZ. Uptake of noninvasive prenatal testing at a large academic referral center. American Journal of Obstetrics & Gynecology. 2014;211(6):651.e1-7.	No diagnostic accuracy data
27. Lee da, E., et al., Non-invasive prenatal testing of trisomy 18 by an epigenetic marker in first trimester maternal plasma. PLoS ONE [Electronic Resource], 2013. 8(11): p. e78136.	Nested case-control study: < 15 cases
28. Levy B, Norwitz E. Non-invasive prenatal aneuploidy testing: technologies and clinical implication. *MLO Med Lab Obs* 2013;45(6):8, 10, 12 passim; quiz 16.	Review
29. Lim, J.H., et al., Disease specific characteristics of fetal epigenetic markers for non-invasive prenatal testing of trisomy 21. BMC Medical Genomics [Electronic Resource], 2014. 7: p. 1.	Method development study
30. Lim, J.H., et al., Non-invasive detection of fetal trisomy 21 using fetal epigenetic biomarkers with a high CpG density. Clinical Chemistry & Laboratory Medicine, 2014. 52(5): p. 641-7.	Nested case-control study: < 15 cases
31. Lim, J.H., et al., Non-invasive epigenetic detection of fetal trisomy 21 in first trimester maternal plasma. PLoS ONE [Electronic Resource], 2011. 6(11): p. e27709.	Epigenetic approach
32. Lo KK, Boustred C, Chitty LS, Plagnol V. RAPIDR: an analysis package for non-invasive prenatal testing of aneuploidy. Bioinformatics. 2014;30(20):2965-7.	No information on population and reference standard
33. Louis-Jacques, A., et al., Effect of commercial cell-free fetal DNA tests for aneuploidy screening on rates of invasive testing. Obstetrics & Gynecology, 2014. 123 Suppl 1: p. 67S.	Abstract
34. Louis-Jacques, A., et al., Use of commercial tests for aneuploidy screening using cell-free fetal DNA in clinical practice. Obstetrics & Gynecology, 2014. 123 Suppl 1: p. 154S.	Conference abstract
35. Manegold-Brauer, G., et al., A new era in prenatal care: non-invasive prenatal testing in Switzerland. Swiss Medical Weekly, 2014. 144: p. w13915.	Cohort study: < 50 women
36. 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
37. 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
38. 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
39. 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

40. 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
41. 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
42. 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
43. 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
44. Shaw, S.W., C.P. Chen, and P.J. Cheng, From Down syndrome screening to noninvasive prenatal testing: 20 years' experience in Taiwan. Taiwanese Journal of Obstetrics & Gynecology, 2013. 52(4): p. 470-4.	Review
45. 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. *Clin Chem* 2013;59(8):1151-9.	Interview
46. Shi X, Zhang Z, Cram DS, Liu C. Feasibility of noninvasive prenatal testing for common fetal aneuploidies in an early gestational window. Clinica Chimica Acta. 2015;439:24-8.	Cohort study: < 50 women with index and reference test result
47. 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. Journal of Obstetrics & Gynaecology, 2003. 23(3): p. 228-32.	Maternal plasma samples after first trimester spontaneous miscarriage vs. genetic analysis of evacuated products of the uterus
48. Struble CA, Syngelaki A, Oliphant A, Song K, Nicolaides KH. Fetal fraction estimate in twin pregnancies using directed cell-free DNA analysis. Fetal Diagnosis & Therapy. 2014;35(3):199-203.	No diagnostic accuracy data
49. Stumm, M., et al., Noninvasive prenatal detection of chromosomal aneuploidies using different next generation sequencing strategies and algorithms. Prenatal Diagnosis, 2012. 32(6): p. 569-77.	Method development
50. Tong, Y.K., et al., Noninvasive prenatal detection of fetal trisomy 18 by epigenetic allelic ratio analysis in maternal plasma: Theoretical and empirical considerations. Clinical Chemistry, 2006. 52(12): p. 2194-202.	Case series: < 50 women
51. Tong, Y.K., et al., Noninvasive prenatal detection of trisomy 21 by an epigenetic-genetic chromosome-dosage approach. Clinical Chemistry, 2010. 56(1): p. 90-8.	Case-control study: < 15 cases
52. Tsaliki, E., et al., MeDIP real-time qPCR of maternal peripheral blood reliably identifies trisomy 21. Prenatal Diagnosis, 2012. 32(10): p. 996-1001.	Epigenetic approach

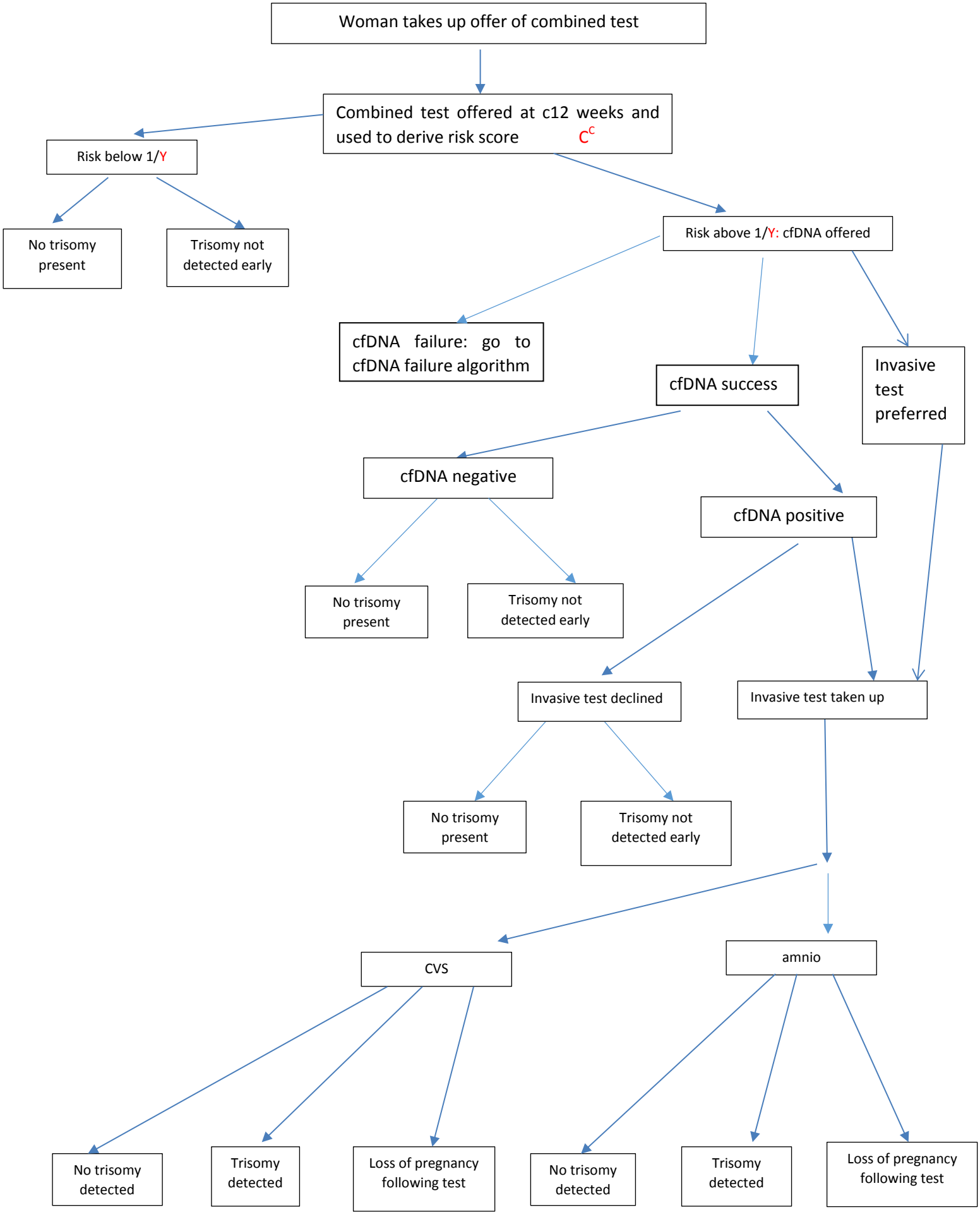
53. van den Oever, J.M., et al., Single molecule sequencing of free DNA from maternal plasma for noninvasive trisomy 21 detection. Clinical Chemistry, 2012. 58(4): p. 699-706.	Case control: < 15 cases
54. van den Oever, J.M., et al., Successful noninvasive trisomy 18 detection using single molecule sequencing. Clinical Chemistry, 2013. 59(4): p. 705-9.	Case control: < 15 cases
55. Wu, D., et al., Prenatal diagnosis of Down syndrome using cell-free fetal DNA in amniotic fluid by quantitative fluorescent polymersase chain reaction. Chinese Medical Journal, 2014. 127(10): p. 1897-901. (Is Wu the first name or surname??? Surname might be “Dan”.)	Not cff DNA (amniotic fluid)
56. 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

12. Appendix 3 Economic Model Decision Tree

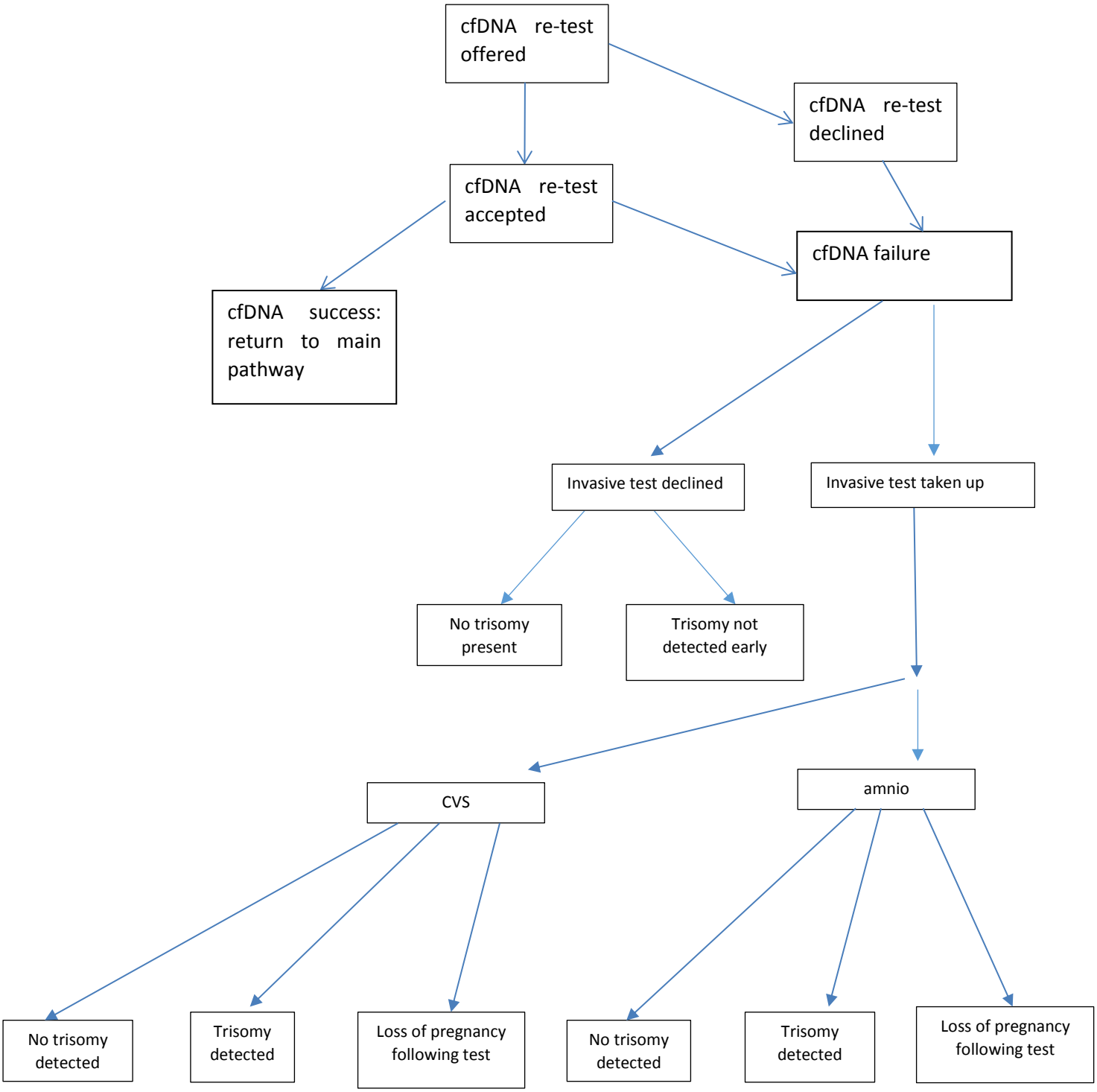
Trisomy screening pathway without cfDNA



Trisomy screening pathway with cfDNA after combined test



Sub-pathway for cfDNA failure



Trisomy screening pathway with cfDNA as primary test