UK National Screening Committee

Update of a systematic review on prenatal cell-free DNA testing for fetal trisomies 21, 18 and 13 (twin/multiple pregnancies and DNA microarray technology)

External review against programme appraisal criteria for the UK National Screening Committee

Version: FINAL

Author: Julia Geppert, Chris Stinton, Samantha Johnson, Aileen Clarke, Dimitris Grammatopoulos, Sian Taylor-Phillips

Date: February 2019

The UK National Screening Committee secretariat is hosted by Public Health England.

About the UK National Screening Committee (UK NSC)

The UK NSC advises ministers and the NHS in the 4 UK countries about all aspects of population screening and supports implementation of screening programmes. Conditions are reviewed against evidence review criteria according to the UK NSC's evidence review process.

Read a complete list of UK NSC recommendations.

UK NSC, Floor 5, Wellington House, 133-155 Waterloo Road, London, SE1 8UG
www.gov.uk/uknsc
Twitter: @PHE_Screening Blog: phescreening.blog.gov.uk

For queries relating to this document, please contact: phe.screeninghelpdesk@nhs.net

Published April 2019

Contents

Plain English summary

Screening and previous/current recommendations:

In England, screening for Down's syndrome, Edwards' syndrome and Patau's syndrome is available to all women. Women are offered an *extra* test if there is a higher chance (equal or greater than 1 in 150) that the baby would be born with Down's syndrome, Edwards' syndrome or Patau's syndrome. The extra test is invasive and there is a small chance of miscarriage.

In 2018/2019 a new Non-Invasive Prenatal Test (NIPT) is being rolled-out in the NHS, with the aim to reduce the number of women who have invasive tests. All women will continue to be offered the current screening tests but women with a chance equal or greater than 1 in 150 of having a baby with Down's syndrome, Edwards' syndrome or Patau's syndrome will be offered the additional option of NIPT. NIPT involves taking a blood sample from the mother and there is no risk of miscarriage.

Updating the review

This review looks at

- whether NIPT works in twin or multiple pregnancies;
- a new method of doing NIPT called DNA microarray. There are several versions of the NIPT test. Microarray is an example of a change to one step in the testing process.

The review focused on evidence published between February 2015 and July 2018.

This review found:

a) NIPT in twin or multiple pregnancies:

1. NIPT might be less accurate in detecting Down's syndrome, Edwards' syndrome and Patau's syndrome for babies in multiple pregnancies. In addition, NIPT might have a higher failure rate.

2. NIPT found 97.5% (72/74) of the babies who had Down's syndrome in multiple pregnancies. It correctly identified 99.9% of the babies who did not have Down's syndrome in multiple pregnancies.

3. NIPT detected 89.5% (17/19) of babies with Edwards' syndrome from multiple pregnancies. It correctly identified all the babies who did not have Edwards' syndrome.

4. NIPT found 87.5% (14/16) of the babies who had Patau's syndrome from multiple pregnancies. It correctly identified 99.9% of the babies who did not have Patau's syndrome in multiple pregnancies.

b) DNA Microarray tests:

Five studies looked at the performance of NIPT using DNA Microarray-based tests. Over 3,000 blood samples were tested. The test detected 99.5% (186/187) of the babies who had Down's syndrome, 97.7% (42/43) of babies who had Edwards' syndrome and 100% (19/19) of babies who had Patau's syndrome. The review found no evidence that DNA microarray-based testing for detecting Down's syndrome, Edwards' syndrome and Patau's syndrome was different to the previous test version.

Recommendations:

a) NIPT in twin / multiple pregnancies

Whilst there is limited evidence of the NIPT performance to detect the Down's syndrome, Edwards' syndrome, and Patau's syndromes in twin pregnancies, the estimates obtained in this review are lower than in singleton pregnancies, in particular for Edwards' syndrome and Patau's syndrome. The number of babies with these conditions in the studies was small and the studies are at high risk of bias. This review did not compare the performance of NIPT to the currently used screening tests in twin pregnancies. It is "uncertain" if NIPT should be offered to women pregnant with twins.

b) DNA microarray-based NIPT:

There is evidence that NIPT using DNA Microarray has very similar test accuracy to other NIPT methods in detecting Down's syndrome, Edwards' syndrome and Patau's syndrome. But there is only a small amount of evidence for NIPT using DNA Microarray and it is also at high risk of bias.

DNA Microarray-based NIPT could be offered as one of the possible processes for NIPT testing, whilst also monitoring the accuracy of the NIPT test in practice in the UK as the different versions of the test are rolled out.

Executive summary

Purpose of the review

This review updates the scientific evidence on prenatal cell-free DNA (cfDNA) screening for Down's syndrome (T21), Edwards' syndrome (T18) and Patau's (T13) syndrome in the fetus that has been published since February 2015. It covers test accuracy and test failures in twin or higher order multiple pregnancies as well as in cfDNA testing approaches using DNA microarray technology for DNA quantification (cfDNA-MA).

Background

The Fetal Anomaly Screening Programme (FASP) offers screening tests to pregnant women to assess their chance of having a baby with Down's syndrome, Edwards' syndrome and Patau's syndrome.

These conditions are usually caused by an extra (third) chromosome in an otherwise typical (euploid) karyotype of 23 pairs of chromosomes. Individuals with one of these conditions carry an aneuploidy karyotype characterised by:

- 3 copies of chromosome 21 in Down's syndrome (Trisomy 21)
- 3 copies of chromosome 18 in Edwards' syndrome (Trisomy 18)
- 3 copies of chromosome 13 in Patau's syndrome (Trisomy 13)

The extra genetic material causes the phenotypic characteristics and symptoms of trisomy. In partial trisomies, only a part of the additional chromosome is present in the cells, rather than a whole additional chromosome. In mosaicism, the third chromosome is only present in a proportion of cells in the body. Mosaic and partial trisomies are associated with milder symptoms than full trisomy. Translocation Down's syndrome can be inherited from an unaffected parent who carries a rearrangement of genetic material between chromosome 21 and another chromosome. This rearrangement without gain or loss of chromosomal material is called 'balanced translocation' and usually causes no symptoms. However, if the translocation is passed on to the baby, it can become unbalanced and the extra genetic material from chromosome 21 can cause Down's syndrome.

The current primary screening test for Down's syndrome, Edwards' syndrome and Patau's syndrome in the fetus is performed between 10 and 14 weeks of pregnancy. This involves a maternal serum test (for the identification of specific biomarkers), an ultrasound scan to measure fetal nuchal translucency thickness and maternal and clinical characteristics (the so called "combined test"). The outcomes of each assessment are fed into an algorithm to give a total chance of chromosomal fetal anomaly. If it is not possible to measure nuchal translucency thickness or booking

between 14–20 weeks of gestation, the “quadruple test” (involving a maternal blood test alone) is used to screen for Down’s syndrome. A mid-pregnancy scan is offered to women presenting after 14 weeks of gestation to look for structural anomalies and 11 rare conditions, including Edwards’ syndrome and Patau’s syndrome. Up until recently, if the screening test showed a chance equal or greater than 1 in 150 for one of the 3 trisomies, women were offered an invasive diagnostic test (amniocentesis or chorionic villus sampling). However, these screening tests have a high false positive rate; about 3–5% of pregnancies without one of these conditions will receive a false-positive screening test result. As invasive testing carries a procedure-related risk of miscarriage, it is desirable to reduce the number of invasive procedures. Cell-free DNA testing, sometimes known as Non-Invasive Prenatal Testing (or NIPT), for fetal trisomies has introduced the possibility of reducing the number of invasive tests arising from the FASP screening pathway. NIPT analyses the DNA fragments present in maternal plasma during pregnancy, the so-called cell-free DNA. Most of this cell-free DNA comes from the mother, but around 10–20% of it comes from the unborn baby (more precisely from the placenta). Several testing strategies have been developed and are commercially available.

Focus of the review

The aim of this review is to examine the test performance of cfDNA testing for fetal T21, T18 and T13 in (a) twin/multiple pregnancies and (b) in tests using cfDNA-MA approaches for DNA quantitation.

This relates to UK NSC criterion 4:
‘*There should be a simple, safe, precise and validated screening test.*’

Specific research questions for the review are shown below:

1. cfDNA testing as follow-on test

1.1 What is the accuracy of cfDNA testing in predicting T21, T18 and T13 in pregnant women with a pre-defined higher-chance result (≥1:150) following a combined test
a) in twin/multiple pregnancies, and
b) using cfDNA testing with cfDNA-MA approach?

1.2 How does changing the threshold for defining a higher-chance result following a combined test affect the accuracy of cfDNA testing
a) in twin/multiple pregnancies, and
b) using cfDNA testing with cfDNA-MA approach?

2. cfDNA testing as replacement test

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
a) in twin/multiple pregnancies, and

b) using cfDNA testing with cfDNA-MA approach?

3. cfDNA testing as an add-on test
What test accuracy is achievable by integrating cfDNA testing into the combined test
a) in twin/multiple pregnancies, and
b) using cfDNA testing with cfDNA-MA approach?

4. cfDNA test failure rate
What is the rate of cfDNA testing failure (number of inconclusive and excluded samples / total number of samples)
a) in twin/multiple pregnancies, and
b) using cfDNA testing with cfDNA-MA approach?

The reviewers used a systematic review approach and updated the searches from the previous review [(1)]. Database searches were conducted on 9 July 2018 and limited to articles published since February 2015 (i.e. the final search date of the previous review) and English language articles. Bibliographic databases included: MEDLINE; MEDLINE In-Process & Other Non-Indexed Citations; EMBASE; Web of Science, and the Cochrane Library: Cochrane Database of Systematic Reviews. In addition, the reviewers contacted experts in the field for suggestion of additional studies, and screened reference lists of included studies and relevant systematic reviews.

Recommendation under review

In 2015, the UK National Screening Committee (UK NSC) reviewed the evidence on NIPT using cfDNA for T21, T18 and T13 in the fetus to inform their decision on the introduction of this test into the current fetal anomaly screening programme in the UK. In January 2016, following a public consultation on the review, the UK NSC recommended an evaluative roll-out of cfDNA testing to assess what impact it would have on the existing NHS Fetal Anomaly Screening Programme. NIPT based on cfDNA is about to be introduced on the NHS in 2018/2019 as an evaluative roll-out. All women will continue to have the first-trimester combined or second-trimester quadruple test as primary screening test. Women who received a higher-chance result following the primary screen (screening test result shows chance for T21, T18 or T13 equal or higher than 1 in 150) will then be offered cfDNA testing as follow-on screening test. The roll-out of cfDNA testing will happen gradually over the next 3-year period.

The previous review found limited evidence of the cfDNA test accuracy in twin pregnancies (4 studies for the detection of T21 and T18 and 3 studies for T13) and a subgroup analysis indicated lower pooled sensitivity for T21 and T18 in twin compared to singleton pregnancies. The studies included in the previous review also all relied upon next generation sequencing (NGS) to analyse cfDNA fragments using 3 different

approaches: (1) random whole-genome assay, (2) targeted sequencing of selected nonpolymorphic regions (Digital Analysis of Selected Regions, DANSR), or (3) single-nucleotide polymorphisms. Since then, another cfDNA quantitation method based on DNA microarray has been developed and was first published in 2014. DNA microarray imaging is a rapid process that might reduce the turnaround time for sample quantitation, allowing greater sample throughput and lower costs. This update review therefore focusses on these 2 areas.

Findings and gaps in the evidence of this review

A total of 1,891 unique records were identified, of which 18 met the inclusion criteria. Two additional articles were identified from previous searches and expert suggestions, resulting in 15 relevant articles on twin/multiple pregnancies and 5 relevant articles on cfDNA-MA testing.

Overall test accuracy

a) cfDNA testing in twin/multiple pregnancies

This update review identified 16 studies (published in 15 articles) on the performance of cfDNA screening tests in twin or higher order multiple pregnancies. The vast majority of cfDNA tests were performed in twin pregnancies. Risk of bias was high in at least one domain in 15 out of 16 studies. Eight studies were considered at high risk of bias in 2 or more domains, and 7 studies in one domain. One study scored low or unclear risk of bias in all domains. The unclear ratings were due to limited reporting of information. Study flow (exclusions from analysis) and patient selection presented the areas with the greatest risk of bias, with only 2 studies each being classified as low risk of bias in these domains.

Down's syndrome (T21)

In total, 6 studies identified in the previous review and 11 studies identified in this update review had complete 2x2 tables (numbers of true positive, false positive, true negative and false negative test results reported) for the detection of trisomy 21. Fourteen studies included twin pregnancies only, while one study included one triplet pregnancy, and in the remaining study, it was unclear if 2 of the enrolled cases were twin or higher order multiple pregnancies. Four studies were excluded from the bivariate meta-analysis as there were no cases of trisomy 21. Bivariate meta-analysis gave summary estimates across 13 studies using cfDNA testing for the detection of Down's syndrome in twin pregnancies of 97.5% (95%CI 88.2% to 99.5%) for sensitivity and 99.93% (95%CI 99.3% to 99.99%) for specificity. The equivalent estimates for majority singleton pregnancies without zero cell correction from the previous review were sensitivity of 99.4% (95%CI 98.9% to 99.6%) and specificity of 99.9% (95%CI 99.9% to 100%). The sensitivity in 50 dichorionic twin pregnancies with Down's syndrome derived from 10 studies was 96.0% (48/50; 95%CI 85.1% to 99.3%), and the sensitivity in 3 monochorionic twin pregnancies was 100% (3/3; 95%CI 31.0% to 100%).

Edwards' syndrome (T18)

In total, 6 studies identified in the previous review and 10 studies identified in this update review had complete 2x2 tables for the detection of trisomy 18. Summing across all studies gave a pooled sensitivity of 89.5% (95%CI 68.6% to 97.1%), and a pooled specificity of 100% (95%CI 99.9% to 100%). For comparison, the meta-analysis estimates for majority singleton pregnancies without zero cell correction from the previous review were sensitivity of 97.4% (95%CI 95.8% to 98.4%) and specificity of 99.9% (95%CI 99.9% to 100%). The sensitivity in 12 dichorionic twin pregnancies with Edwards' syndrome derived from 10 studies was 91.7% (11/12), and the sensitivity in one monochorionic twin pregnancy was 0% (0/1).

Patau's syndrome (T13)

In total, 5 studies identified in the previous review and 10 studies identified in this update review had complete 2x2 tables for the detection of trisomy 13. Summing across all studies gave a pooled sensitivity of 87.5% (95%CI 64.0% to 96.5%), and a pooled specificity of 99.92% (95%CI 99.7% to 99.98%). For comparison, the meta-analysis estimates for majority singleton pregnancies without zero cell correction from the previous review were sensitivity of 97.4% (95%CI 86.2% to 99.6%) and specificity of 99.9% (95%CI 99.9% to 100%). The sensitivity in 3 dichorionic twin pregnancies with Patau's syndrome derived from 9 studies was 33.3% (1/3), and there was no case of Patau's syndrome in a monochorionic twin pregnancy.

b) cfDNA-MA testing

This update review identified 5 studies on the performance of cfDNA-MA screening tests. One study was identified in the previous review, and 4 were identified in the update. Risk of bias was high in all studies, with 3 out of 5 studies considered high risk in 2 or more domains and 2 out of 5 studies in one domain of the tailored QUADAS-2. Study flow (exclusions from analysis) was the domain which introduced the greatest risk of bias into the studies as all 5 studies excluded women from analysis due to test failures and/or non-available results of newborn examination or genetic testing (pre- or postnatal). Another issue was the role of sponsor, with only one of the 5 studies stating that the role of sponsor played no part in design, conduct or publication. These potential biases could result in overestimation of test accuracy and underestimation of test failures. There were also some applicability concerns.

A single study directly comparing a cfDNA test based on Digital ANalysis of Selected Regions (DANSR) and sequencing with a newer version of the test based on DANSR and DNA microarrays in the same samples found sensitivity and specificity estimates for both tests of 100% for all 3 trisomies. However, sample selection in this study, undertaken in-house by the manufacturer, was unclear: the study sample appears enriched with extra trisomies, test failures were removed and sample selection appears to have been made with knowledge of the reference standard, so there was a high risk

of bias. It is also unclear whether the blood samples were taken before or at least 7 days after invasive testing. For trisomy 21, the difference in sensitivity between the 2 tests was 0% (95%CI -1.4% to +1.4%), and the difference in specificity between the 2 tests was 0% (95%CI -0.3% to 0.3%). For trisomy 18, the difference in sensitivity between the 2 tests was 0% (95%CI -7.7% to +7.7%), and the difference in specificity between the 2 tests was 0% (95%CI -0.3% to 0.3%). For trisomy 13, the difference in sensitivity between the 2 tests was 0% (95%CI -14.3% to +14.3%), and the difference in specificity between the 2 tests was 0% (95%CI -0.3% to 0.3%). For all trisomies, the McNemar's test statistic of 1 indicated no statistically significant difference between the technologies in either sensitivity or specificity. This should be interpreted as no evidence of a difference rather than evidence of no difference.

For trisomy 21, summary estimates across the 4 studies included in the meta-analysis using the cfDNA-MA test (one study with no trisomy 21 cases was excluded) were 99.5% (186/187; 95%CI 96.3% to 99.9%) for sensitivity and 100% for specificity. The confidence interval for specificity was not estimable in the meta-analysis, so using the Wilson score for the summed studies it was estimated to be 99.87% to 100%. Meta-analysis was not possible for trisomies 18 or 13 because 3 out of the 5 studies contained no cases of trisomy 18 or trisomy 13. Summing across all 5 studies for trisomy 18 gave a sensitivity of 97.7% (42/43; 95%CI 87.9% to 99.6%) and specificity of 99.97% (95%CI 99.81% to 99.99%). For trisomy 13, summing across all 5 studies gave a sensitivity of 100% (19/19; 95%CI 83.2% to 100%), and a specificity of 99.97% (95%CI 99.81% to 99.99%).

For comparison, bivariate meta-analysis of the 7 studies using a cfDNA test based on DANSR and sequencing performed in (mainly) singleton pregnancies and with complete 2x2 tables from the previous review gave estimates across studies for trisomy 21 detection of 99.7% (95%CI 82.1% to 99.9%) for sensitivity and 99.95% (95%CI 99.91% to 99.97%) for specificity. For trisomy 18 detection, summing across 6 studies evaluating this target condition gave a sensitivity of 96.6% (95%CI 91.6% to 98.7%) and a specificity of 99.95% (95%CI 99.92% to 99.97%). For trisomy 13 detection, summing across 2 studies evaluating this target condition gave summary estimates for sensitivity of 57.1% (95%CI 25.0% to 84.2%) and for specificity of 99.97% (95%CI 99.93% to 99.99%).

Indirect comparison of the 7 studies identified in the previous review (DANSR with sequencing) with the 5 studies identified in this update review (DANSR with DNA microarray) found no evidence of a difference in sensitivity (p=0.81) or specificity (p=0.15) for trisomy 21 detection between the 2 technologies. This represents no evidence of a difference rather than evidence of no difference.

Findings for the specific research questions

Research question 1.1: What is the accuracy of cfDNA testing in predicting T21, T18 and T13 in pregnant women with a pre-defined higher-chance result (≥1:150) following a combined test?

a) cfDNA testing in twin/multiple pregnancies

There were no studies reporting the performance of cfDNA testing after the first-trimester combined test at threshold 1:150 in twin or multiple pregnancies.

b) cfDNA-MA testing

There were no studies reporting the performance of cfDNA-MA testing after the first-trimester combined test at threshold 1:150.

Research question 1.2: How does changing the threshold for defining a higher-chance result following a combined test affect the accuracy of cfDNA testing?

a) cfDNA testing in twin/multiple pregnancies

Two studies were identified in the previous review and 5 studies were identified in this update review that carried out cfDNA testing in pregnant women with twin/multiple pregnancies at higher chance of fetal trisomies defined in a variety of ways. It was not possible to present cfDNA testing performance in twin/multiple pregnancies at different chance cut-offs following a first trimester combined test ranging from very high to low chance or to present an optimal chance cut-off to maximise cfDNA testing performance in clinical practice.

b) cfDNA-MA testing

One study reported the cut-off for a higher-chance result from the first trimester combined screening test (FTCS) prior to cfDNA-MA testing. In 54 women with singleton pregnancy, a FTCS chance for T21 greater than 1 in 250 and no fetal anomalies detected on ultrasound, one out of a total of one cases of T21 was detected with no false positive and no false negative results. The remaining 4 studies did not contribute to the research question. It is therefore not possible to present cfDNA-MA testing performance at different chance cut-offs ranging from very high to low chance or present an optimal chance cut-off to maximise cfDNA-MA testing performance in clinical practice.

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?

a) cfDNA testing in twin/multiple pregnancies

It was not possible to include a comparison of test performance of cfDNA testing for T21 and the combined test in the meta-analysis, due to a lack of studies making the comparison in twin or multiple pregnancies.

b) cfDNA-MA testing

It was not possible to include a comparison of test performance of cfDNA-MA testing for T21 and the combined test in the meta-analysis, due to a lack of studies making the comparison. Two individual studies have provided evidence that the specificity of the combined test is lower than that of cfDNA-MA testing (97.5% [95% CI 96.0% to 98.5%] and 93.3% [95% CI 89.5% to 95.8%], respectively, for the combined test versus 100% [95% CI 99.6% to100%]) and 100% [95% CI 99.3% to 100%], respectively, for cfDNA-MA testing). Sensitivity was 100% for both tests in one study, and could not be calculated in the other study as neither of the tests had any true positive or false negative results.

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

a) cfDNA test accuracy in twin/multiple pregnancies

No study was identified that reported test accuracy of cfDNA testing as add-on test to the first trimester combined test in twin or higher order multiple pregnancies.

b) cfDNA-MA testing

No study was identified that reported test accuracy of cfDNA-MA testing as add-on test to the first trimester combined test.

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

a) cfDNA test accuracy in twin/multiple pregnancies

The rate of initial analytic failure (failure of the initial cfDNA testing) ranged from 0% to 9.4% in 10 studies performing cfDNA testing in twin/multiple pregnancies. In 5 of these studies, it was unclear if the reported failure rate was after initial or repeat testing. Summing across the 5 studies with clear reporting gives 72 initial test failures among 1,939 analysed samples from twin pregnancies (3.7%). One study found a 3-times higher failure rate after first sampling in twin pregnancies (9.4%, 41/438) than in singleton pregnancies (2.9%, 316/10,698) ($p<0.0001$). Fetal fraction was lower in dichorionic than in monochorionic twin pregnancies (Median 7.7% vs 10.1%; $p<0.0001$).

b) cfDNA-MA testing

The rate of initial analytic failure (failure of the initial cfDNA testing) ranged from 0.9% to 1.9% in 4 studies using cfDNA-MA testing. Repeat tests using a second blood sample were successful in one out of one and 5 out of 11 (45.5%) women. The main reason for cfDNA-MA test failure was insufficient circulating fetal DNA in 8 out of 8 and 10 out of 11 samples, respectively. The remaining study only included samples that met the laboratory's quality control thresholds and did not report the number of test failures.

Recommendations on screening

UK NSC criterion 4: *"There should be a simple, safe, precise and validated screening test"*

a) cfDNA testing for fetal trisomies 21, 18 and 13 in twin/multiple pregnancies

Whilst there is limited evidence of test accuracy of cfDNA testing for fetal trisomies in twin or higher order multiple pregnancies, based on few cases, sensitivity point estimates are lower than in singleton pregnancies, particularly for Edwards' syndrome and Patau's syndrome. The available evidence was mostly at high risk of bias and there were applicability concerns, in particular regarding the timing of cfDNA testing. Findings from one study suggested that the initial failure rate after sampling in the first trimester is higher in twin pregnancies than in singleton pregnancies. There is insufficient evidence to assess if the test accuracy (especially sensitivity) is lower and test failure rate higher in dichorionic twin pregnancies than in monochorionic twin pregnancies or singleton pregnancies. This review did not identify any head-to-head studies comparing the performance of cfDNA testing and the first-trimester combined test in twin/multiple pregnancies. It was not commissioned to indirectly compare the performance of cfDNA testing with the currently used screening practice in twin pregnancies, nor to assess the costs and consequences for the current NHS screening programme when cfDNA testing is used in sequence with the combined test, as a replacement for the combined test or in combination with (i.e. alongside) the combined test in twin or higher order multiple pregnancies. Therefore, it is "**UNCERTAIN**" if cfDNA testing should be offered to twin or higher order multiple pregnancies.

b) Microarray-based cfDNA testing for fetal trisomies 21, 18 and 13

This review found no evidence of difference in test accuracy between cfDNA-MA testing and sequencing-based cfDNA testing including:

- a head-to-head study which reported equivalent test accuracy of the 2 approaches for T21, T18 and T13;
- a meta-analysis of cfDNA-MA testing which produced comparable test accuracy estimates to cfDNA testing in the previous review;

- an indirect comparison of the accuracy of sequencing and microarray approaches using DANSR for detection of T21. This was also seen in summary values calculated for T18 and T13.

However, this demonstrates no evidence of a difference rather than evidence of no difference.

The initial test failure rate might be lower with cfDNA-MA testing than with sequencing-based cfDNA testing but this indirect comparison has to be treated with caution as cfDNA tests were performed in different populations and during different time periods.

Compared to the previous review of all cfDNA testing technologies the evidence base for microarray-based cfDNA testing is more limited in terms of volume of studies and participants. There is considerable statistical uncertainty for detection of trisomies 18 and 13. However, the respective evidence bases for different sequencing methods, when considered in isolation, also each had a limited volume when the previous review was undertaken.

Similarly, the assessments risk of bias and applicability of the studies using cfDNA-MA testing suggests that the evidence base is limited in terms of quality and generalisability. A further limitation of this review is that it did not update the previous review's assessment of sequencing-based cfDNA testing.

These issues do represent real limitations on the robustness of the estimates of test performance generated by this review's statistical analyses. However, a number of factors should be borne in mind:

- Compared to the already accepted sequencing-based cfDNA test technologies, it is important to note that, while the DNA quantitation method using DNA microarray is technically distinct from sequencing-based cfDNA testing, the other stages of the workflow (e.g. plasma separation, cfDNA isolation, assays to select clinical relevant regions of the human genome) and bioinformatics analysis remain comparable to sequencing strategies.
- The test will be offered to a group of women considered to be at higher chance of fetal trisomies.
- The test accuracy estimates for cfDNA-MA testing are similar to those used in the economic model in the previous review, which found cfDNA testing to be cost effective. Further, the lower limit of the confidence interval for cfDNA-MA test accuracy is similar to the more conservative test accuracy estimates used in the economic model.
- The test would be applied in an evaluative roll-out of cfDNA testing which aims to assess the performance of cfDNA testing and further information could be

generated in that process. In this context, adding microarray-based cfDNA testing could contribute positively to understanding of different versions of NIPT.

For these reasons microarray-based cfDNA testing meets the criterion for test accuracy.

This review has highlighted a number of important issues including:

- There is heterogeneity in types of cfDNA tests, with a more limited evidence base for each individual subtype of test, thus increasing statistical uncertainty and concerns regarding risk of bias and applicability to each of the questions.
- Secondly, there is potential for ongoing change in the tests (in algorithms and in materials selected and methods used) which may easily occur after publication of the included studies or in the future after implementation in the NHS. The ongoing changes may affect test accuracy or failure rate.
- Thirdly, there is greater statistical uncertainty in accuracy to detect trisomies 18 and 13 than in trisomy 21.

The reviewers would therefore recommend the following to decrease the risks associated with these issues:

1. Initial and ongoing measurement of test accuracy in practice as part of quality assurance. This would require establishing trisomy status using the reference standards outlined in this review or equivalents.
2. All changes to any part of the test post roll-out should be notified to the UK NSC to consider whether accuracy should be re-evaluated.
3. Implementation of robust quality assurance processes to ensure that changes to the tests post roll-out can be monitored.

Limitations

a) cfDNA testing in twin/multiple pregnancies

The evidence on the overall test accuracy of cfDNA testing for common fetal trisomies in twin/multiple pregnancies comes from 17 studies (6 from the previous review and 11 from the update review) that in total successfully analysed around 2,800 maternal blood samples for T21 and T18 and around 2,600 samples for T13. The identified evidence is at high risk of bias and there were applicability concerns.

For one key study comparing directly 438 twin pregnancies and nearly 10,700 singleton pregnancies from the same UK population, the reviewers had to use unpublished data received by personal communication on the number of false positives and true negatives in twin pregnancies. Unfortunately, the reviewers were unable to get a complete 2x2 table for the singleton pregnancies, and this most applicable UK study with relative high

number of samples tested could not be included in the direct comparison of the test accuracy of twin/multiple and singleton pregnancies.

The indirect comparison of the cfDNA test accuracy in twin pregnancies (identified from the previous and the update review) versus (mainly) singleton pregnancies (identified from the previous review only) might be biased as this review did not update the evidence on cfDNA testing in singleton pregnancies published since February 2015, and recently published studies might have been missed. It should be regarded as exploratory and should not be used to draw conclusions.

Differences in test performance between monochorionic and dichorionic twin pregnancies could not be assessed due to the low number of affected cases in monochorionic twin pregnancies (n=4). The risk of publication bias was not assessed.

This review did not identify any head-to-head studies comparing the performance of cfDNA testing and the first-trimester combined test in twin/multiple pregnancies. It was not commissioned to indirectly compare the performance of cfDNA testing with currently used screening practice in twin pregnancies, nor to assess the costs and consequences for the current NHS screening programme when cfDNA testing is used in sequence with the combined test, as a replacement for the combined test or in combination with (i.e. alongside) the combined test in twin or higher order multiple pregnancies.

b) cfDNA-MA testing

As the cfDNA test methodology was not sufficiently reported in most of the potentially relevant publications, the reviewers had to rely on information provided by the manufacturer and provider of the commercially available test. Studies with unclear proportions of samples analysed using the DNA microarray technology and/or without possibility to disaggregate test accuracy results were excluded.

For the key study that directly compared the 2 DNA quantitation technologies (sequencing versus DNA microarrays) in the same population, the reviewers had to use unpublished data for the subgroup of women with appropriate reference standard. The indirect comparison of the 2 technologies in different populations might be biased as this review did not update the evidence on sequencing-based cfDNA testing published since February 2015, and recently published studies might have been missed. It should be regarded as exploratory and should not be used to draw conclusions.

Taken together, the evidence on the test accuracy of cfDNA-MA testing for common fetal trisomies comes from 5 studies that in total successfully analysed just over 3,000 maternal blood samples. The identified evidence is at high risk of bias and concerns regarding applicability of the tested women (e.g. time point of cfDNA testing, prior chance of fetal trisomy, trisomy prevalence) are mostly high.

Evidence uncertainties

a) cfDNA testing in multiple pregnancies

The number of blood samples from twin or higher order multiple pregnancies that were screened for the common fetal trisomies using cfDNA testing is still low (around 2,600 to 2,900 samples in total). The studies included in the meta-analysis contained a low number of trisomy cases, especially for T18 (19 cases) and T13 (16 cases), resulting in wide confidence intervals for the sensitivity estimates. The identified evidence is at high risk of bias and concerns regarding applicability of the tested women (e.g. time point of cfDNA testing, prior testing and/or chance of fetal trisomy, trisomy prevalence) are mostly high. Due to missing information on chorionicity in the majority of studies and small number of trisomy cases in monochorionic twin pregnancies, possible differences in the cfDNA test accuracy (especially sensitivity) between mono- and dichorionic twin pregnancies could not be assessed. Further large scale prospective studies are recommended to verify the performance of cfDNA testing as screening test for the common trisomies in twin pregnancies.

b) cfDNA-MA testing

The volume of evidence concerning prenatal cfDNA-MA testing for fetal trisomies is currently small, with only one single head-to-head study comparing its performance against sequencing-based approaches. Future research should include (1) additional studies making direct comparisons between DNA microarray-based and sequencing-based cfDNA testing within the same sample of women, and (2) an evaluation of the test accuracy of cfDNA-MA testing after it has been implemented in a routine public antenatal screening programme.

Introduction and approach

Background

The Fetal Anomaly Screening Programme (FASP) offers screening tests to pregnant women to assess their chance of having a baby with Down's syndrome, Edwards' syndrome and Patau's syndrome [(2)].

These conditions are usually caused by an extra (third) chromosome in an otherwise typical (euploid) karyotype of 23 pairs of chromosomes. Individuals with one of these conditions carry an aneuploidy karyotype characterised by:

- 3 copies of chromosome 21 in Down's syndrome [Trisomy 21 (T21)]
- 3 copies of chromosome 18 in Edwards' syndrome [Trisomy 18 (T18)]
- 3 copies of chromosome 13 in Patau's syndrome [Trisomy 13 (T13)]

The extra genetic material causes the phenotypic characteristics and symptoms of trisomy. In partial trisomies, only a part of the additional chromosome is present in the cells, rather than a whole additional chromosome. In mosaicism, the third chromosome is only present in a proportion of cells in the body. Mosaic and partial trisomies are associated with milder symptoms than full trisomy. Translocation Down's syndrome can be inherited from a parent without the condition who carries a rearrangement of genetic material between chromosome 21 and another chromosome. This rearrangement without gain or loss of chromosomal material is called 'balanced translocation' and usually causes no symptoms. However, if the translocation is passed on to the baby, it can become unbalanced and the extra genetic material from chromosome 21 can cause Down's syndrome.

A person with Down's syndrome will have some level of learning disability. This means they will find it harder than most people to understand and to learn new things. They may have communication challenges and difficulty managing some everyday tasks. People with Down's syndrome have distinctive facial features but they do not all look the same. Most children with Down's syndrome attend mainstream schools but will require additional support. Some health problems are more common in people with Down's syndrome. These include heart conditions and problems with hearing and vision. Many health problems can be treated but unfortunately around 5% of babies will not live past their first birthday. For babies without serious health problems survival is similar to that of other children and most people with Down's syndrome will live into their 60s or longer. People with Down's syndrome can have a good quality of life and most say they enjoy their lives. With support, many more people with Down's syndrome are able to get jobs, have relationships and live semi-independently in adulthood.

Babies with Edwards' syndrome have an extra copy of chromosome 18 in all or some cells. Babies with Patau's syndrome have an extra copy of chromosome 13 in all or some cells. Sadly the survival rates are low and of those babies born alive only around 10% live past their first birthday. Some babies may survive to adulthood but this is rare. All babies born with Edwards' syndrome and Patau's syndrome will have a learning disability and a wide range of physical challenges, which can be extremely serious. They may have problems with their heart, limbs, kidneys and digestive system. Around half of babies with Patau's syndrome will have a cleft lip and palate. Babies with Edwards' syndrome and Patau's syndrome will have a low birthweight. Despite their difficulties, children can slowly make progress in their development. Older children with either condition would need to attend a specialist school.

Down's syndrome is the most common chromosomal disorder and was prenatally diagnosed in about 2.7 in 1,000 pregnancies in England and Wales in 2013. Due to the rate of spontaneous miscarriage/stillbirth or termination following prenatal diagnosis, the live birth prevalence is lower at 1.1 per 1,000 live births. [5] Trisomy 21 occurs in all ethnicities.[8] The chance of Down's syndrome is related to maternal age and rises from about 1:1,300 in 25 year olds to 1:380 in 35 year olds, and further to 1:28 in women aged 45 years.[9] The birth prevalence of Edwards' syndrome and Patau's syndrome in England and Wales was 0.7 per 1,000 births and 0.3 per 1,000 births, respectively, in 2013. [5]

Due to the rate of spontaneous miscarriage/stillbirth or termination following prenatal diagnosis, the prevalence at the time of screening is higher than the birth prevalence for all 3 trisomies, in particular for T18 and T13.

Current policy context and previous reviews

The current primary screening test for T21, T18, and T13 is performed between 10 and 14 weeks of pregnancy. This involves a maternal serum test (for the identification of specific biomarkers), an ultrasound scan to measure fetal nuchal translucency (NT) thickness and maternal and clinical characteristics (the so called "combined test"). The outcomes of each assessment are fed into an algorithm to give a total chance of chromosomal fetal anomaly. If it is not possible to measure nuchal translucency thickness or booking between 14–20 weeks of gestation, the "quadruple test" (involving a maternal blood test alone) is used to screen for Down's syndrome. A mid-pregnancy scan is offered to women presenting after 14 weeks of gestation to look for structural anomalies and 11 rare conditions, including Edwards' syndrome and Patau's syndrome. If the screening test shows a chance equal or greater than 1 in 150 for one of the 3 trisomies, women are offered an invasive diagnostic test [amniocentesis or chorionic villus sampling (CVS)] [2]. However, these screening tests have a false positive rate (defined as 1-specificity) of about 3–5% [10]. As invasive testing carries a small

procedure-related risk of miscarriage [11], it is desirable to reduce the number of invasive procedures.

In 2015, the UK NSC reviewed the evidence on prenatal cfDNA testing for Down's syndrome (T21), Edwards' syndrome (T18) and Patau's (T13) syndrome in the fetus to inform their decision on introduction of this test into current fetal anomaly screening programme in the UK [1]. In January 2016, following a public consultation on the review, the UK NSC recommended an evaluative roll-out of cfDNA testing to assess what impact it would have on the existing NHS FASP [2]. The proposed change is for cfDNA testing to be offered to women who are deemed to have a higher chance (≥1:150) of having a baby with Down's syndrome, Edwards' syndrome or Patau's syndrome following the current primary screen.

At the time of the 2015 review, 3 main cfDNA testing strategies had been developed for trisomy screening. These relied upon next generation sequencing to quantify cfDNA: random whole-genome sequencing (MPSS), targeted sequencing of selected nonpolymorphic regions (digital analysis of selected regions, DANSR) or targeted sequencing of single-nucleotide polymorphisms (SNPs).

The result of cfDNA testing for an individual woman in most genome-wide methods is calculated as a z-score or normalized chromosome value (NCV). The z-score reflects the number of standard deviations that the percentage of aligned reads originated from the chromosome of interest (here chr 21, chr 18 or chr 13) among the total reads from all other chromosomes sequenced from a test sample is above the mean acquired from unaffected controls. The z-score approach can be further optimised in a number of ways to improve the accuracy for the specific chromosomes tested (e.g. a correction factor for the guanine-cytosine content) or by using information on chromosomal variations within the sample set. The NCV is similar to the z-score but compares the reads from a chromosome of interest to the reads from reference chromosomes with similar biochemical behaviours (so the denominator is different). Analogous to z-scores, the ratio is then normalised by the mean and standard deviation acquired from unaffected controls [12].

In the DANSR method, the z-score approach is replaced by the Fetal-fraction Optimized Risk of Trisomy Evaluation (FORTE) algorithm which considers information on the fetal fraction and maternal age to report an individualised chance for trisomy [13]. It is the first approach to incorporate different risk factors with the outcome of cfDNA testing and does not require the information from, and testing against, external reference samples.

The sequencing data obtained from the SNP-based approach is analysed using a proprietary statistical algorithm called Next-generation Aneuploidy Test Using SNPs (NATUS). It takes into account the actual DNA from the mother obtained from the buffy coat (the fraction of an anticoagulated blood sample that contains most of the white

blood cells and platelets following density gradient centrifugation of the whole blood), as well as the mixed maternal and fetal cfDNA (obtained from the plasma) and it uses Bayesian statistics and Maximum Likelihood Estimation to identify the most likely combination of fetal genotype and fetal fraction. NATUS does not require the information from and testing against external reference samples. It calculates a sample-specific accuracy for each interrogated chromosome, which represents the likelihood that the copy number call is correct, and is expressed as a proportion of the maximum value of 1 (100%) [14].

The studies included in the previous review all relied upon these next generation sequencing-based approaches to the quantitation and interpretation of cfDNA.

Another DNA quantitation method, using DNA microarray for trisomy screening, was first published in 2014 by Juneau et al. [15] This study was not included in the previous review. This was because the microarray-based index test met the review's inclusion criteria but the reference standard did not. This was because not all samples had a complete outcome evaluation by invasive testing or newborn examination and it was not possible to disaggregate the results.

DNA microarray is a technology in which thousands of nucleic acids are bound to a surface and are used to measure the relative concentration of target nucleic acid sequences in a mixture via hybridisation and subsequent detection of the hybridisation events [16]. DNA microarray imaging is a rapid process and might allow greater sample throughput and lower costs [15].

The workflow of the different approaches to cfDNA testing is presented in Figure 1. The red oval marks the single step that has changed in the DNA microarray-based cfDNA test method compared to the sequencing-based methods which were included in the previous review.

It is important to note that, while the DNA quantitation method is technically distinct from sequencing-based cfDNA testing, the other stages of the workflow and analysis remain comparable to sequencing strategies. For example, the microarray-based version of the Harmony Prenatal Test (Ariosa Diagnostics, Inc., San Jose, CA) uses the same methods for plasma separation, cfDNA isolation, and DANSR assays (to select regions from the human genome with clinical relevance). It also uses the FORTE algorithm to analyse the quantified DANSR products and to generate a trisomy risk score in the same way as the sequencing-based version of the test.

During the preparations for the evaluative roll-out of cfDNA testing in England, the UK NSC was asked to commission a review of the accuracy of microarray-based cfDNA testing. The main purpose was to explore the reported performance of this more recent approach compared with that reported for sequencing-based cfDNA testing in the

previous review. To enable comparison with the previous review, this was undertaken using the same, systematic review, methods to address the same research questions. Studies of microarray-based cfDNA testing have been grouped with sequencing-based approaches in recent systematic reviews and meta-analyses of clinically relevant outcomes of cfDNA testing [17; 18]. However, to enable a rapid turnaround, this review focuses solely on studies of microarray-based cfDNA testing rather than updating the whole body of evidence relating to cfDNA testing which would include many recently published studies of sequencing-based cfDNA testing. This approach imposed limitations on the analysis which are described in the course of this review.

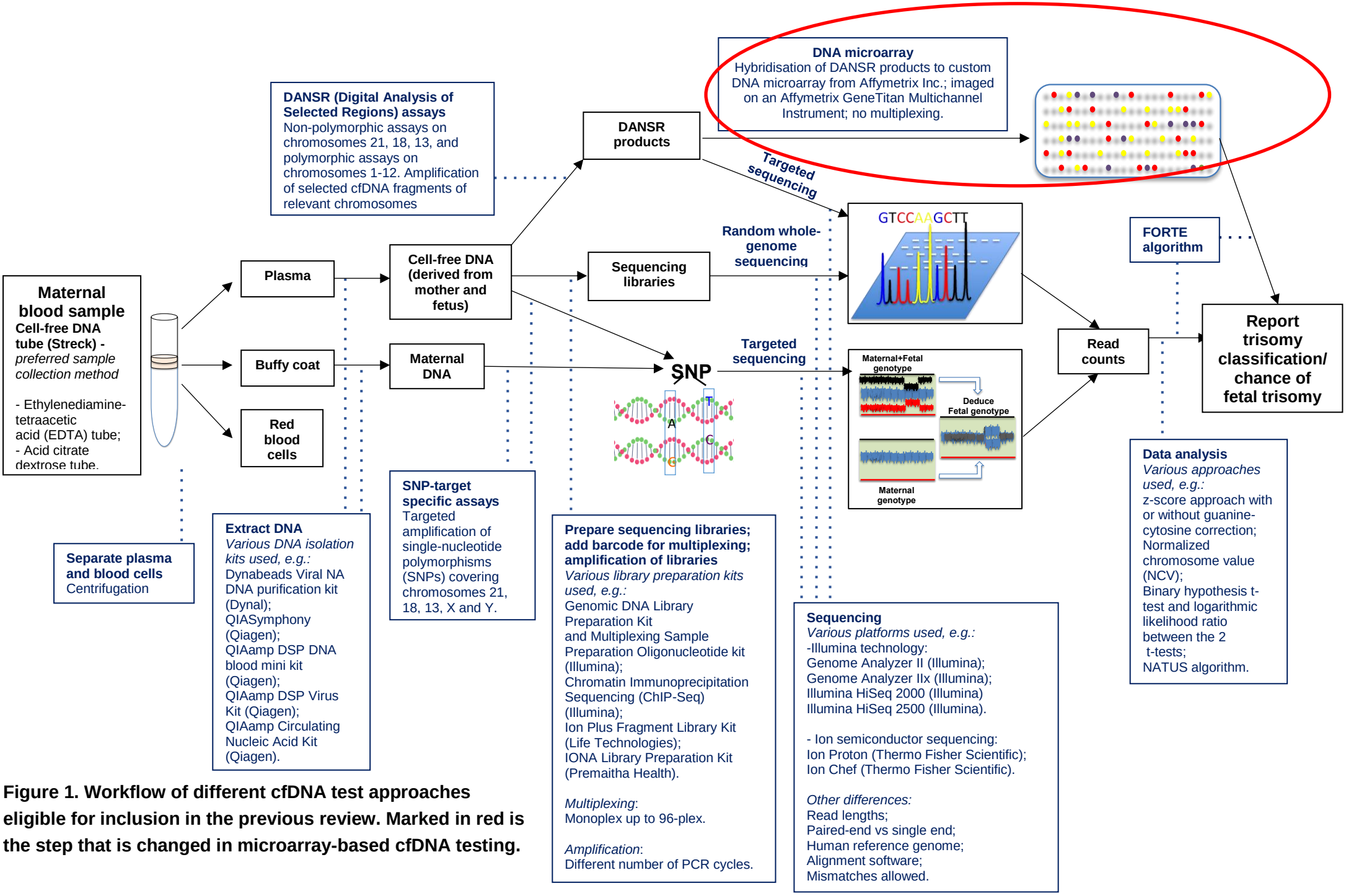


Figure 1. Workflow of different cfDNA test approaches eligible for inclusion in the previous review. Marked in red is the step that is changed in microarray-based cfDNA testing.

Objectives

This review examined the scientific evidence on prenatal cfDNA screening for fetal T21, T18 and T13 that has been published since February 2015 in relation to test accuracy and test failures in twin/multiple pregnancies and in cfDNA testing approaches using DNA microarray technology for DNA quantification (abbreviated cfDNA-MA). The specific questions for this update review are the same as in the original review [1] but limited to

a) twin or higher order multiple pregnancies, and

b) cfDNA tests using DNA microarray technology; namely:

Table 1. Key questions for the evidence summary, and relationship to UK NSC screening criteria

	Criterion	Key questions	# Studies included
	THE TEST		
4	There should be a simple, safe, precise and validated screening test.	**What is the overall test accuracy of cfDNA testing in predicting fetal T21, T18 and T13**	
		a) in twin/multiple pregnancies, and	15
		b) using cfDNA testing with DNA microarray-approach?	5
		Specific research questions	
		1. cfDNA testing as follow-on test	
		1.1 What is the accuracy of cfDNA testing in predicting T21, T18 and T13 in pregnant women with a pre-defined higher-chance result (≥1:150) following a combined test	
		a) in twin/multiple pregnancies, and	0
		b) using cfDNA testing with DNA microarray-approach?	0
		1.2 How does changing the threshold for defining a higher-chance result following a combined test affect the accuracy of cfDNA testing	
		a) in twin/multiple pregnancies, and	5
		b) using cfDNA testing with DNA microarray-approach?	1
		2. cfDNA testing as replacement test	
		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	
		a) in twin/multiple pregnancies, and	0
		b) using cfDNA testing with DNA microarray-approach?	2
		3. cfDNA testing as an add-on test	
		What test accuracy is achievable by integrating cfDNA testing into the combined test	
		a) in twin/multiple pregnancies,	0
		b) using cfDNA testing with DNA microarray-approach?	0
		4. cfDNA test failure rate	
		What is the rate of cfDNA testing failure (number of inconclusive and excluded samples / total number of samples)	
		a) in twin/multiple pregnancies, and	10
		b) using cfDNA testing with DNA microarray-approach?	4

Methods

The current review was commissioned by the UK National Screening Committee. It was conducted by the University of Warwick applying the same full systematic review methodology as was used in the previous 2015 review [(1)].

Databases/sources searched

The reviewers updated the searches from the previous review [(1)]. A copy of the search strategies that were used in the major databases is provided in Appendix 1. Database searches were conducted on 9 July 2018 to identify studies relevant to the questions detailed in

Table 1. Searches were limited to articles published since February 2015 (i.e. the final search date of the previous review) and English language articles.

Bibliographic databases included: MEDLINE (Ovid); MEDLINE In-Process & Other Non-Indexed Citations (Ovid); EMBASE (Ovid); Web of Science (Ovid), and Cochrane Library: Cochrane Database of Systematic Reviews.

The search strategy comprised the following main elements:
1. Searching of electronic bibliographic databases,
2. Contacting experts in the field, and
3. Scrutiny of references of included studies and relevant systematic reviews.

Eligibility for inclusion in the review

Two reviewers independently screened the titles/abstracts of records identified by the searches. Any disagreements were resolved by consensus or retrieval of the full publication. Full copies of all studies deemed potentially relevant were obtained and 2 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.

General eligibility criteria for part a) and part b) of the research questions are presented in Table 2 below.

Table 2. General inclusion and exclusion criteria for part a) and b) of the review questions.

Key question	Inclusion criteria							Exclusion criteria
	Population	**Target condition**	**Intervention**	**Reference standard**	**Compa-rator**	**Outcome**	**Study type**	
a) Test accuracy and test failures in twin/multiple gestations	Women with twin/multiple pregnancies (higher chance or general population). If only part of the population are twin/multiple pregnancies, the study will be included if outcome data are reported separately.	Trisomies 21, 18 and 13 in the fetus, also including translo-cation or mosaicism.	cfDNA testing using cfDNA derived from maternal blood (serum, plasma, whole blood) using next generation sequencing or DNA microarray as DNA quantitation method.	Genetic verification through amniocentesis, chorionic villus sampling, cordocentesis, fetal pathologic examination after abortion, or postnatal phenotypic assessment.	No compara-tor, any "conven-tional" screening test, or different cfDNA test methods.	Any type of test performance estimates including: accuracy, detection rate, sensitivity, specificity, predictive values, likelihood ratios, diagnostic odds ratios, receiver operating characteristic curves, numbers of true positive, false positive, true negative and false negative results, numbers of inconclusive, indeterminate and excluded samples.	All study designs will be included, including randomised controlled trials, cross-sectional test accuracy studies, cohort studies and case-control studies.	Studies reporting the quantification of fetal cells, measuring total DNA levels in maternal blood, or using epigenetic markers as a screening/diagnostic tool. Case-control studies with < 15 cases and cohort studies with < 50 pregnant women. Letters, reviews, editorials and communications containing insufficient information on methods and no numerical outcomes data. Grey literature and conference abstracts. Articles not available in the English language.
b) Test accuracy and test failures in microarray-based cfDNA testing	All pregnant women (singleton and/or multiple pregnancies, higher chance or general population).	As in a)	cfDNA testing using cfDNA derived from maternal blood (serum, plasma, whole blood) using DNA microarray as DNA quantitation method.	As in a)	As in a)	As in a)	As in a)	As in a)

Studies were then organised by the specific research question addressed. Table 3 shows the criteria for inclusion into the analyses for the specific questions 1–4.

Table 3. Inclusion criteria for research questions 1–4 analyses

Population	
Questions 1a) and 3a)	Women with twin/multiple pregnancies with higher chance of one (or more) of T21, T18 or T13 according to a combined test outcome. If only part of the population are twin/multiple pregnancies, the study will be included if outcome data are reported separately.
Questions 1b) and 3b)	All pregnant women with higher chance of one (or more) of T21, T18 or T13 according to a combined test outcome.
Question 2a)	Women with twin/multiple pregnancies in the general obstetric population. If only part of the population are twin/multiple pregnancies, the study will be included if outcome data are reported separately.
Question 2b)	All pregnant women in the general obstetric population.
Question 4a)	Women with twin/multiple pregnancies. If only part of the population are twin/multiple pregnancies, the study will be included if outcome data are reported separately.
Question 4b)	All pregnant women.
Intervention	
Questions 1–4, part a) (Multiple pregnancies)	cfDNA testing using cfDNA derived from maternal blood (serum, plasma, whole blood) using next generation sequencing or DNA microarray as DNA quantitation method.
Questions 1–4, part b) (DNA microarray)	cfDNA testing using cfDNA derived from maternal blood (serum, plasma, whole blood) using DNA microarray as DNA quantitation method.
Comparator	
Question 1, part a)	No comparator.
Question 1, part b)	No comparator or head-to-head comparison with cfDNA testing using cfDNA derived from maternal blood (serum, plasma, whole blood) using next generation sequencing as DNA quantitation method.
Question 2	Any “conventional” screening test.
Questions 3 and 4	No comparator.
Outcome	
Questions 1–3	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.
Question 4	Data on inconclusive, indeterminate and excluded samples to determine test failure rates.

Data extraction

Data were extracted by one reviewer, using a piloted, data extraction form. A second reviewer checked the extracted data and any disagreements were resolved by consensus or discussion with a third reviewer.

The reviewers contacted the corresponding authors of potentially relevant articles if it was unclear from the publication if DNA microarrays were used for DNA quantitation for all analysed samples. As some of the corresponding authors did not know the cfDNA test methodology, the reviewers were directed by the authors to contact the laboratory that performed the cfDNA test (Ariosa Diagnostics Inc., San Jose, CA). The review authors have also received unpublished subgroup data for one study [15] from Ariosa Diagnostics Inc. (San Jose, CA). All correspondence was via email only and can be reviewed if requested.

Appraisal for quality/risk of bias tool

Quality appraisal of diagnostic accuracy studies was conducted using the same tailored QUADAS-2 tool [19] that was used in the previous 2015 review [1]. Quality assessment was undertaken by one reviewer and checked by a second reviewer. Any disagreements were resolved by iteration, discussion and consensus. If required, we consulted a third reviewer.

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.

Statistical Analysis

The reviewers extracted data from the primary studies to obtain the 4 cell values of a diagnostic 2x2 table in order to calculate test accuracy measures: sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Confidence intervals were calculated using the Wilson score interval with continuity correction.

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.

The previous review used a zero cell correction to enable model convergence. Test accuracy may have been overestimated in the previous review due to the high risk of bias in included studies, or may have been underestimated due to the zero cell correction. In this review no zero cell correction was used as there were fewer cases so such a correction would have been more influential. Therefore, in this review test accuracy estimates may be overestimates due to the high risk of bias in included studies.

a) cfDNA test accuracy in twin/multiple pregnancies

For twin/multiple pregnancies, cases were categorised as (1) 'true positive' (TP) if cfDNA test result was positive and matched the karyotype or birth outcome of at least one fetus/baby; (2) 'false positive' (FP) if the cfDNA test result was positive and did not match the karyotype or birth outcome of either fetus/baby; (3) 'true negative' (TN) if the cfDNA test result was negative and all fetuses/babies were determined to be unaffected by karyotyping or birth outcome; and (4) 'false negative' (FN) if the cfDNA test result was negative but at least one fetus/baby were determined to be affected by karyotyping or birth outcome. Studies with incomplete 2x2 tables (missing numbers for true positive, true negative, false positive or false negative cfDNA test results) were excluded from the meta-analysis.

Eligible studies on multiple gestations with complete 2x2 tables (number of true positive, true negative, false positive and false negative cfDNA test results reported) identified in this update review were combined with eligible studies on multiple gestations that were identified in the previous review [1]. Analyses were stratified according to condition (T21, T18 and T13). The reviewers estimated the overall accuracy of cfDNA testing for detecting Down's syndrome in twin/multiple pregnancies using bivariate meta-analysis. Studies with no cases of Down's syndrome were excluded from the meta-analysis.

The reviewers could not estimate the accuracy for detecting Edwards' syndrome or Patau's syndrome using bivariate meta-analysis as the models would not converge due to the small number of cases and studies. For these 2 conditions the methods of a recent Cochrane review by Badeau et al. [20] were followed to simply sum the numbers of true positive, false positive, true negative and false negative across all of the studies and calculate overall sensitivity and specificity with confidence intervals using the Wilson method.

The reviewers planned to repeat the meta-analysis including only cohort studies with either consecutive women or a random sample of women enrolled, but there were only 3 such studies with only 5 cases of Down's syndrome and no cases of Edwards' syndrome or Patau's syndrome, so this was not possible.

One study included a case of triplets, which did not have any of the 3 trisomies [21]. Another study used the term "multiple gestations" and it was unclear from the publication if this related only to twin pregnancies or also to higher order multiple gestations [22]. The 2 cases of "multiple gestations" in this study did not have any of the 3 trisomies. All 3 cases were included in the analysis.

To directly compare the accuracy in multiple pregnancies to singleton pregnancies the reviewers planned to meta-analyse studies reporting results for multiple pregnancies and singleton pregnancies in the same population, with twin/singleton as a covariate in a bivariate model. However, due to the low number of studies and events these models did not converge. Therefore, the reviewers followed the methods of Badeau et al. [20]. Only including studies that reported a complete 2x2 table for both multiple and singleton pregnancies in the same population, the reviewers summed the total number of true positive, false positive, true negative and false negative results for multiple and singleton gestations separately, and calculated sensitivity, specificity and then confidence intervals using the Wilson method.

b) cfDNA testing based on DNA microarrays (cfDNA-MA testing)

To estimate test accuracy using the cfDNA-MA test, the reviewers meta-analysed studies of test accuracy using a bivariate model. Only studies where the cfDNA test was

based on DNA microarray for all cases were included; studies where some samples were tested with sequencing and others with the DNA microarray-based version of the test were excluded. The reviewers excluded studies with no cases of trisomy 21, 18 or 13 from the meta-analysis. Where there was no heterogeneity in either sensitivity or specificity, for example no false positives or false negatives, meta-analysis cannot calculate confidence intervals. In this case, the reviewers followed the methods of the recently published Cochrane review by Badeau et al. [20] to simply sum all of the studies and calculate confidence intervals using the Wilson method.

To test whether the new DNA microarray-based version of a cfDNA test using the DANSR approach can be considered equivalent to the previous sequencing-based version, head-to-head test accuracy studies comparing both versions in the same cohort with a reference standard is the most informative study design. The reviewers planned to meta-analyse these studies with test type as a covariate in a bivariate model. However, as there was only one study of this type the reviewers compared sensitivity and specificity of the 2 test versions using McNemar's test. The reviewers also made this same comparison indirectly, by adding microarray or sequencing technologies as a covariate to a bivariate meta-analysis of studies using the targeted DANSR approach. There is bias in this indirect comparison because population may be a confounder in the analysis, and because the reviewers did not update the search to look for more recent papers using the sequencing-based version of the DANSR test, so this final analysis is exploratory and should not be used to draw conclusions.

Question level synthesis

Criterion 4

There should be a simple, safe, precise and validated screening test.

Questions 1–4, part a) — What is the test performance (accuracy and test failures) of cfDNA testing in twin/multiple pregnancies?

The previous review identified limited evidence of cfDNA test accuracy in twin pregnancies (4 studies for the detection of T21 and T18 and 3 studies for T13) and a subgroup analyses indicated lower pooled sensitivity for T21 and T18 in twin compared to singleton pregnancies [(1)]. This review therefore updated the published evidence on test performance of cfDNA testing in twin/multiple pregnancies.

Eligibility for inclusion in the review

For part a) of review questions 1–4, the reviewers included studies of pregnant women with twin or multiple pregnancies, who had been given non-invasive prenatal testing using cfDNA derived from maternal blood (serum, plasma, whole blood) as screening test for trisomies 21, 18 or 13 (including mosaicism and translocation), and a reference standard of either genetic verification through amniocentesis, CVS, cordocentesis and fetal pathologic examination after abortion or postnatal phenotypic assessment. Mixed (singleton and multiple pregnancies) populations were included if relevant outcome data were reported separately for the twin/multiple pregnancies. Studies with and without a comparator of any “conventional” screening test were included. Studies with any test accuracy or test failure information were included, but those studies from which a full 2x2 table could not be adequately constructed were excluded from the meta-analysis.

The reviewers excluded studies reporting the quantification of fetal cells, measurement of total DNA levels in maternal blood, or using 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.

Description of the evidence

Database searches yielded 1,891 unique results, of which 14 were judged to be relevant to part a) of the review questions. One additional article [23] was identified by expert suggestions, resulting in a total of 15 included articles reporting on 16 studies.

Appendix 2 contains a full PRISMA flow diagram (Figure 12), along with a table of the included publications and details of which questions these publications were identified as being relevant to (Table 10). A list of excluded studies at full text level with reasons is provided in Table 11.

Discussion of findings

Characteristics of included studies

A study-level summary of data extracted from each included publication is presented in 'Appendix 3 - Summary and appraisal of individual studies' (Table 12).

Of the 16 included studies, 14 were cohort studies [21; 22; 23; 24; 25; 26; 27; 28; 29; 30; 31; 32; 33; 34] and 2 were case-control studies [32; 35].

Four studies were multicentre studies conducted in hospitals or fetal medicine/prenatal diagnostic centres in England [31; 35], France [24] and the Netherlands [25]. Three studies are experience reports from laboratories providing cfDNA testing in Belgium [26] and the USA [28; 32]. One of these laboratories based in the USA also reported the results of a validation study in their article [32]. One study was performed in 4 cytogenetic laboratories in the USA and consisted of a review of all cases received for cytogenetic testing after NIPT [30]. Among the 6 remaining studies, 4 were single centre studies from Korea [21; 22], Taiwan [21], and the USA [29; 34], while the setting was unclear in 2 studies from China [27; 33] the affiliations from the authors suggest that they were conducted in single hospitals.

Eleven studies included mixed (singleton and multiple pregnancies) populations [21; 22; 24; 25; 26; 29; 30; 31; 32; 34; 35] with 2 [22] to 438 [31] twin/multiple pregnancies enrolled in the study. The remaining 5 studies included twin/multiple pregnancies only [23; 27; 28; 32; 33]; the study size ranged from 92 [27] to 565 [33] women with twin or multiple pregnancies. One study included a case of triplets [21]; another study used the term "multiple gestations" and it was unclear from the publication if this related only to twin pregnancies or also to higher order multiple gestations (2 cases) [22].

Chorionicity of the included twin pregnancies was not reported for all women in 11 studies (reported in 10 articles) [21; 22; 24; 26; 28; 29; 30; 32; 34; 35]. In the remaining 5 studies [23;

[25; 27; 31; 33], twin pregnancies were dichorionic in the majority of cases ranging from 53/92 (58%) [27] to 544/565 (96%) [33].

Six studies used samples from pregnant women primarily at a higher chance of fetal trisomies with a range of different indications for invasive testing [22; 24; 25; 28; 30; 35]. One study [33] performed cfDNA testing in the general obstetric population, and the remaining 9 studies (published in 8 articles) included women with mixed chances [21] or unclear [23; 26; 27; 29; 31; 32; 34] prior chances of fetal trisomies.

The vast majority of studies used cfDNA testing based on random whole-genome sequencing (11 studies published in 10 articles) [22; 23; 24; 25; 26; 27; 28; 32; 33; 35]). One study used the DANSR targeted approach combined with sequencing or DNA microarrays [31]. Three studies offered more than one cfDNA testing approach to their patients [29; 30; 34], and the cfDNA test methodology was not described in the remaining study [21].

Nine studies reported cfDNA test accuracy outcomes and test failure rates [21; 23; 24; 25; 27; 28; 31; 33; 35], 6 studies (published in 5 articles) reported test accuracy outcomes only [22; 26; 30; 32; 34], while the remaining study only reported the test failure rate of cfDNA testing [29].

Methodological quality of included studies

The methodological quality of the 16 included studies, assessed by the tailored QUADAS-2 is summarised in Figure 2, Figure 3 and Table 14. These illustrate the risk of bias regarding the 5 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.

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 was judged as low in 2 of the 16 studies [25; 26]. Five studies were classified as unclear risk of bias because it was not explicitly stated that patients were recruited randomly or consecutively [21; 23; 24], and exclusions from the study were not further described [21; 22; 23; 24; 31]. Nine studies were classified as being at high risk of bias because recruitment of women rather than samples was not random or consecutive [28; 30; 32], samples without reference standard were excluded from the study [27], >10% of eligible women did not have cfDNA testing [29; 33], women without follow-up cytogenetic testing were excluded [30], women with negative cfDNA test result were excluded [34], or a case-control design was used [32; 35].

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 7 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 10 studies [21; 23; 24; 25; 27; 30; 31; 32; 33; 34]. Risk of bias was judged as unclear in 3 studies as it was unclear if blood samples were taken before or at least 7 days after invasive testing [22; 28; 29]. Three studies were classified as high risk of bias in the index test domain as the analysis was not performed blinded to the results of the reference standard [32] and/or the threshold was not pre-specified and adjustment of the threshold was not avoided [26; 32; 35].

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. Prenatal or postnatal karyotyping or phenotypic newborn assessment were accepted as appropriate reference standard. Six studies were classified as unclear risk of bias, as standards other than the pre-specified reference standards were used [28; 32; 33; 36] or the reference standard was not clearly described [21; 23; 32]. One study reporting only on the cfDNA test failure rate did not use any reference standards [29]. The 9 remaining studies were at low risk of bias in this domain [22; 24; 25; 26; 27; 30; 31; 34; 35].

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. The risk of bias was judged as high for this domain in 11 studies [23; 24; 25; 27; 28; 30; 31; 32; 33; 34; 35] because cfDNA testing was not performed in all eligible women and/or failed to provide a result and/or not all women received a reference standard, and therefore not all women were included in the analysis. The risk of bias was judged as unclear in 3 studies as it was unclear if all patients received a reference standard [27], if all patients were included in the analysis [32], or due to the possibility of verification bias as the reference standards were not described [21; 32]. Two studies [22; 29] were judged as low risk of bias in this domain.

The risk of bias regarding the *role of sponsor* was considered as high if studies were funded by profit-making companies and if involvement of the sponsor in the design, conduct or publication of the study was stated and/or if the majority of authors or main authors were employees or shareholders of companies offering cfDNA testing and/or other conflicts of interest (i.e. patents, stock or stock options) were declared. The risk of bias regarding the role of sponsor was judged as low in 4 included studies [23; 26; 27; 29]. Risk of bias was classified as unclear in 6 studies as the funding source was not reported [21; 24; 30; 34], the article did not state the role of the sponsor and conflicts of interest [25], or the cost of collection and analysis of some of the study samples was covered by the company providing the cfDNA tests [31]. The remaining 6 studies (published in 5 articles) were judged as at high risk of bias in this domain as the study

was designed, performed and published by employees of companies providing cfDNA testing [22; 28; 32; 33; 35].

In summary, risk of bias was high in 15 out of 16 studies with 8 studies (published in 7 articles) considered at high risk of bias in 2 or more domains [26; 28; 30; 32; 33; 34; 35], and 7 studies in one domain [22; 23; 24; 25; 27; 29; 31]. One study scored low or unclear risk of bias in all domains [21] which was due to unclear reporting of the study that precluded a judgement. Figure 72 shows that the study flow domain (exclusions from analysis) and the patient selection domain presented the areas with the greatest risk of bias with only 2 studies each being classified as low risk of bias.

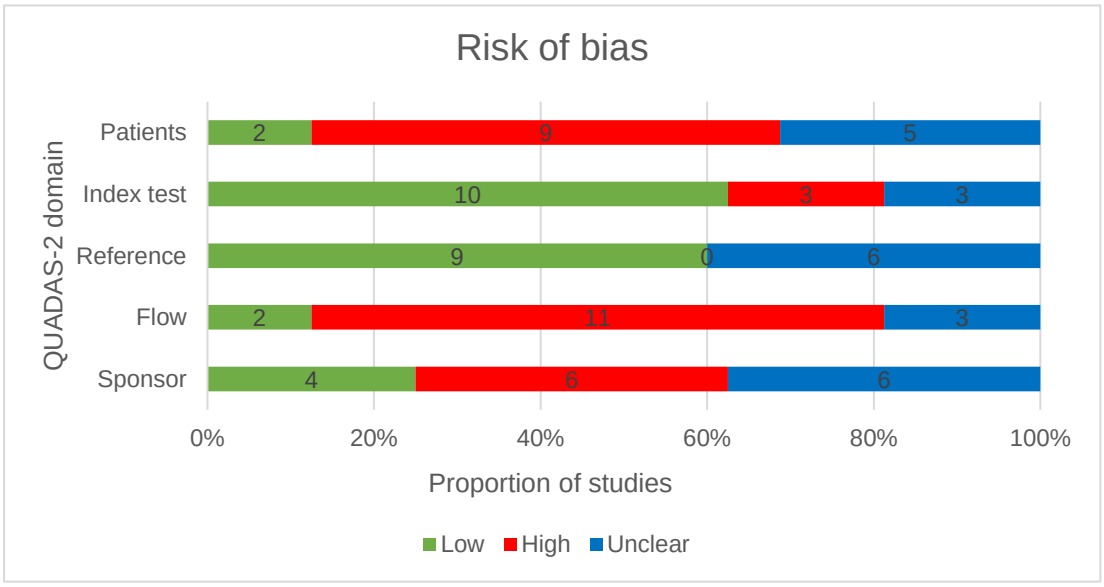


Figure 2. Proportion of studies with low, high and unclear risk of bias.
Risk of bias in the reference standard domain was rated in 15 studies only.

Applicability concerns

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

Only one study included 100% pregnant women in their first trimester and was classified as low levels of concern regarding applicability of the patient spectrum [31]. Four studies were classified as unclear concerns as the proportion of first trimester pregnancies was not reported [21; 30; 32; 34]. In the remaining 11 studies, the concerns regarding the applicability of the patients were high as less than 80% of women were tested in the first trimester [22; 23; 24; 25; 26; 27; 28; 29; 35], the trisomy prevalence in the study population was higher than expected even for a higher-chance population [32], or the study population

consisted solely of pregnancies after treatment with assisted reproductive technologies [33].

In terms of the *index test*, applicability concerns were classified as low in 14 out of 16 studies (published in 13 articles) [22; 23; 24; 25; 26; 27; 28; 29; 31; 32; 33; 34; 35]. The remaining 2 studies were classified as unclear concerns as there was no information on the cfDNA test methodology and its conduct for all [21] or a proportion of samples [30].

Applicability concerns regarding the *reference standard* were classed as low in 9 out of 16 studies as the pre-defined reference standards were used [22; 24; 25; 26; 27; 30; 31; 34; 35]. Applicability was unclear in 5 studies as standards other than the pre-specified reference standards were used in a proportion of women [28; 33] or the reference standard was not described in detail [21; 23; 32]. Concerns regarding the applicability were high in the clinical implementation study by Strom et al. [32] as ultrasound findings were accepted as confirmation of T13 and T18. Applicability concerns were not rated in one study [29] that only assessed failure rate and not test accuracy of cfDNA testing.

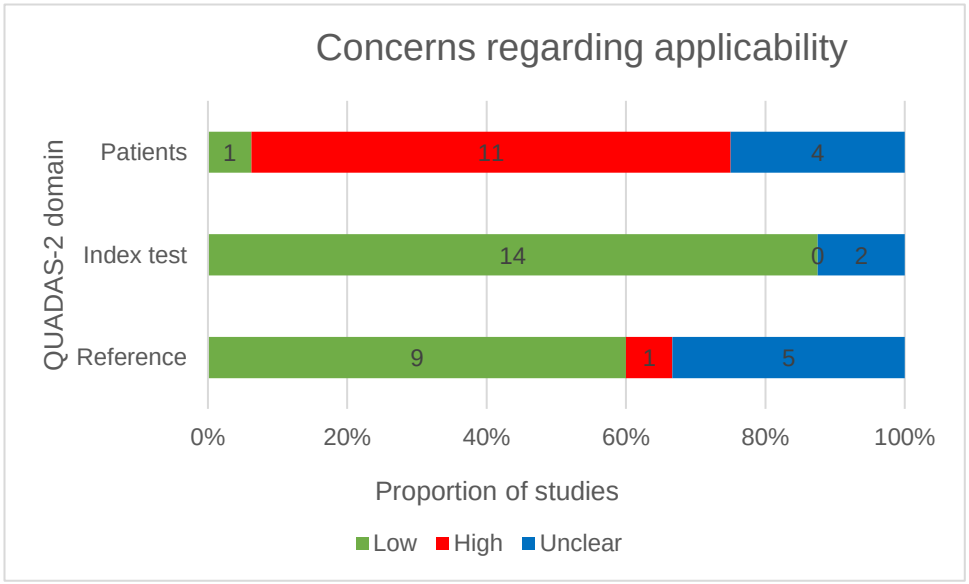


Figure 3. Applicability concerns in included studies.
The applicability concerns regarding the reference standard were rated in 15 studies only.

Analysis of the evidence

The accuracy of cfDNA testing in every included study is shown in Table 15. 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.

Meta-analysis for trisomy 21 (Down's syndrome)

Overall test accuracy of cfDNA testing in twin/multiple pregnancies

In total, 6 studies identified in the previous review [37; 38; 39; 40; 41; 42] and 11 studies identified in this update review [21; 22; 23; 24; 25; 27; 28; 31; 32; 33; 35] had complete 2x2 tables for the detection of trisomy 21. Figure 4 shows a forest plot of the sensitivity and specificity with 95% confidence intervals for each study included in the meta-analysis. There were 72 true positives, 2,782 true negatives, 3 false positives and 2 false negatives among the 17 studies. Included in the previous review, there were 34 true positives, 1,095 true negatives, 2 false positives, and 2 false negatives among 6 studies. Four studies [21; 22; 39; 40] were excluded from the bivariate meta-analysis as there were no cases of trisomy 21. Bivariate meta-analysis gave summary estimates across 13 studies using cfDNA testing for the detection of Down's syndrome syndrome in twin pregnancies of 97.5% (95%CI 88.2% to 99.5%) for sensitivity and 99.93% (95%CI 99.3% to 99.99%) for specificity, the summary receiver operating characteristic (ROC) plot is shown in Figure 5.

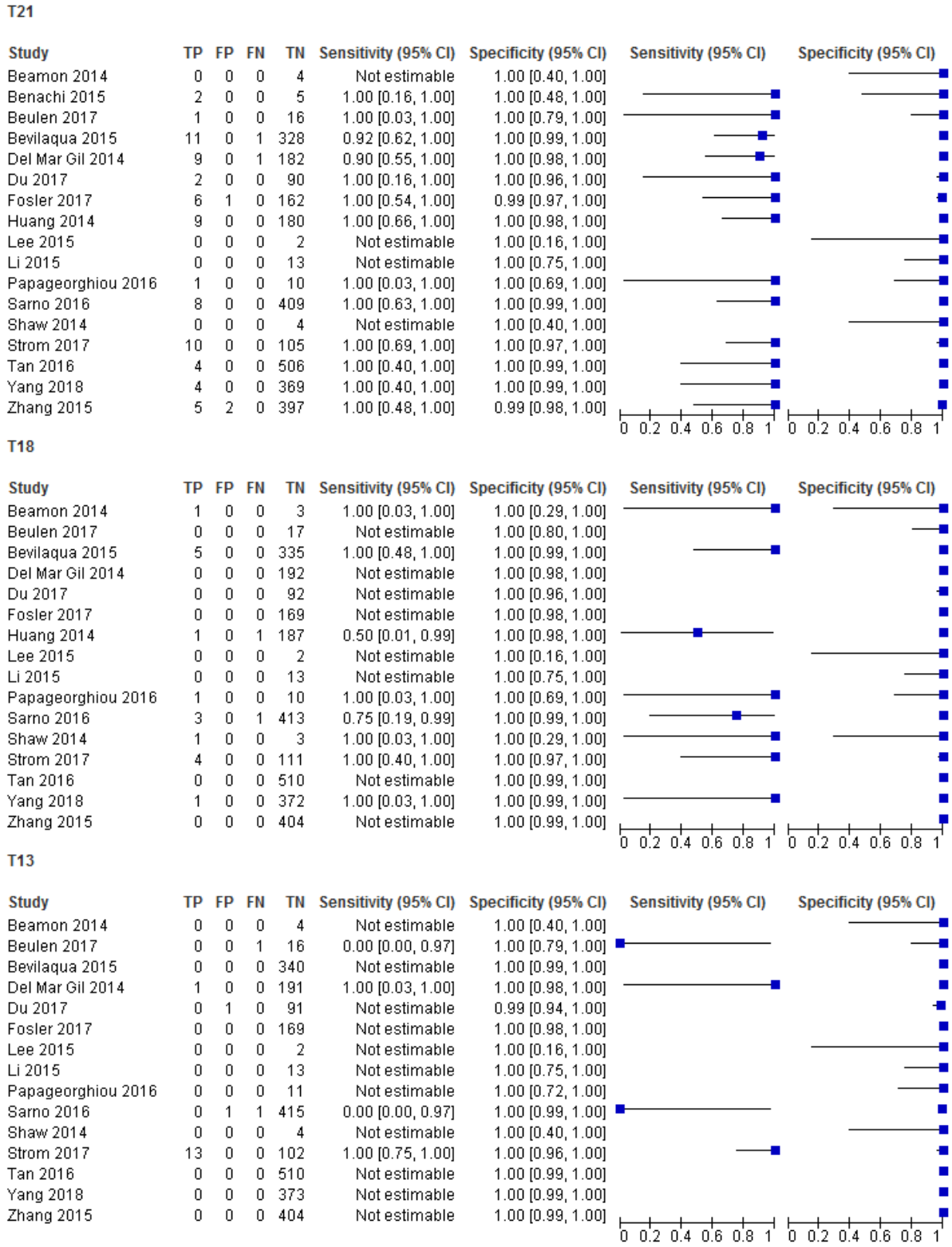


Figure 4. Sensitivity and specificity for the detection of trisomies 21, 18 and 13 in twin/multiple pregnancies for each study included in the meta-analysis.

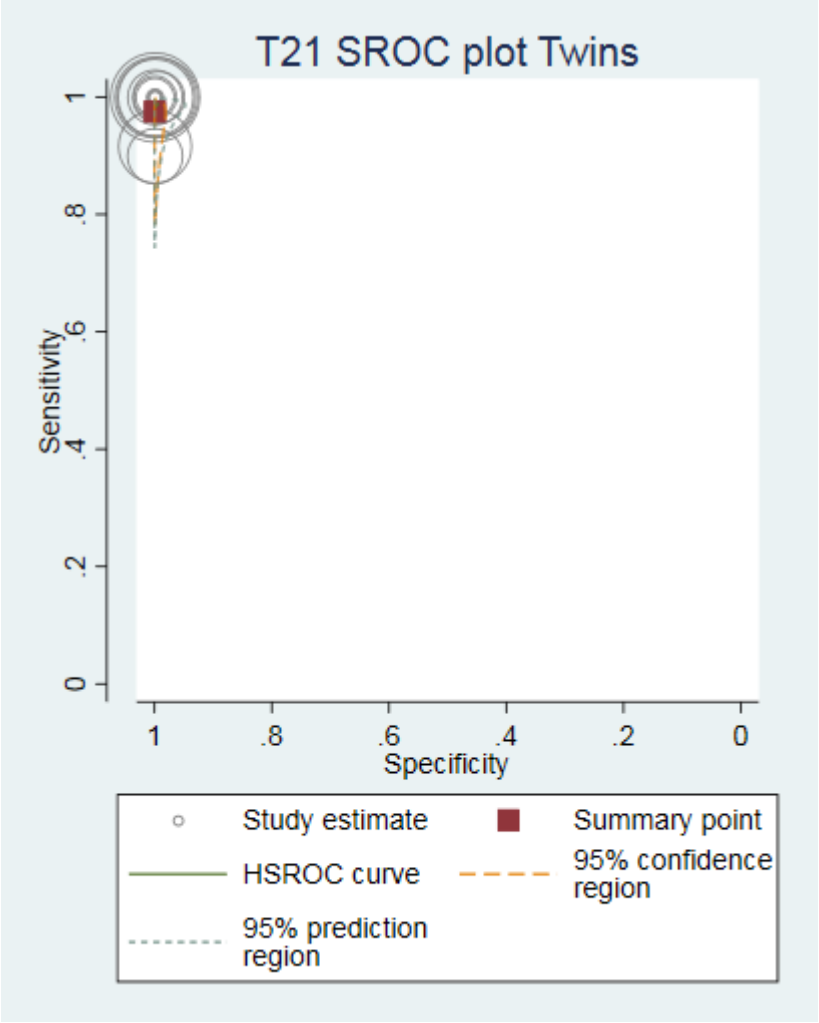


Figure 5. Summary ROC plot for detection of Down's syndrome in twin pregnancies (13 studies).

Comparison of cfDNA test accuracy in twin/multiple versus singleton pregnancies

The direct comparison in the same population is preferred as there are fewer confounders. There were 8 studies that reported cfDNA test accuracy in twin and singleton pregnancies in the same population [21; 22; 24; 25; 35; 39; 40; 42]. Of these only 4 studies [24; 25; 35; 42] contained at least one case of Down's syndrome in a twin pregnancy. Due to the low number of studies and event rates it was not possible to meta-analyse the direct comparisons, and compare test accuracy in twin and singleton pregnancies. However, using the methods of Badeau et al. [20], the summaries across these 8 studies were as follows: pooled sensitivity was 100% (9/9, 95%CI 70.1% to 100%) in twin pregnancies and 99.3% (863/869, 95%CI 98.5% to 99.7%) in singleton pregnancies. Pooled specificity was 99.6% (451/453, 95%CI 98.4% to 99.9%) in twin pregnancies and 99.9% (113,456/113,516, 95%CI 99.93% to 99.96%) in singleton pregnancies.

Indirect comparison via meta-analysis was not undertaken due to the differing time periods between reviews. In the previous review the test accuracy estimates for cfDNA

testing derived from all 40 studies with at least one true positive or one false negative Down's syndrome case [27 of which included singleton pregnancies only, while the others included twin pregnancies only (3 studies), mixed populations with a great majority of singleton pregnancies (4 studies) or unclear populations assumed to be singleton pregnancies only (6 studies)] using similar methods without zero cell correction were sensitivity of 99.4% (95%CI 98.9% to 99.6%) and specificity of 99.9% (95%CI 99.9% to 100%). This point estimate for sensitivity in mainly singleton pregnancies is higher than that for twin pregnancies in this review (97.5%) but the confidence intervals are overlapping (95%CI 88.2% to 99.5% for testing in twin pregnancies). The point estimates for specificity are similar in twin and (mainly) singleton pregnancies. This comparison should be treated with caution due to the differing time periods between reviews, population confounders, and they may both be an overestimate due to the risk of bias in included studies. The economic model in the previous review used a more conservative estimate of sensitivity of 97.1%.

Meta-analysis for trisomy 18 (Edwards' syndrome)

Overall test accuracy of cfDNA testing in twin/multiple pregnancies

In total, 6 studies identified in the previous review [37; 38; 39; 40; 41; 42] and 10 studies identified in this update review [21; 22; 23; 25; 27; 28; 31; 32; 33; 35] had complete 2x2 tables for the detection of trisomy 18. Figure 4 shows a forest plot of the sensitivity and specificity with 95% confidence intervals for each study included in the meta-analysis. There were 17 true positives, 2,833 true negatives, no false positives and 2 false negative among the 16 studies. Included in the previous review, there were 8 true positives, 1,124 true negatives, no false positives, and one false negatives among 6 studies. Summing across all studies gives a pooled sensitivity 89.5% (95%CI 68.6% to 97.1%), and pooled specificity 100% (95%CI 99.9% to 100%).

Comparison of cfDNA test accuracy in twin/multiple pregnancies versus singleton pregnancies

The direct comparison in the same population is preferred as there are fewer confounders. There were 7 studies which reported accuracy in twin and singleton pregnancies in the same population [21; 22; 25; 35; 39; 40; 42]. Of these only 3 studies [35; 39; 40] contained at least one case of Edwards' syndrome in a twin pregnancy. Due to the low number of studies and event rates it was not possible to meta-analyse the direct comparisons, and compare test accuracy in twin and singleton pregnancies. However, using the methods of Badeau et al. [20] the summaries across these 7 studies were as follows: pooled sensitivity was 100% (3/3, 95%CI 46.9% to 100%) in twin pregnancies and 98.0% (192/196, 95%CI 94.9% to 99.2%) in singleton pregnancies. Pooled specificity was 100% (452/452, 95%CI 99.2% to 100%) in twin pregnancies and 99.95% (113,260/113,312, 95%CI 99.94% to 99.97%) in singleton pregnancies.

Indirect comparison via meta-analysis was not undertaken due to the differing time periods between reviews. In the previous review the test accuracy estimates for cfDNA testing derived from all 33 studies with at least one true positive or one false negative case with Edwards' syndrome using different methods both without zero cell correction were sensitivity of 97.4% (95%CI 95.8% to 98.4%) and specificity of 99.9% (95%CI 99.9% to 100%). This point estimate for sensitivity in mainly singleton pregnancies is higher than that for twin pregnancies in this review (89.5%) but the confidence intervals are overlapping (95%CI 68.6% to 97.1% for twin pregnancies). The point estimates for specificity are similar in twin and (mainly) singleton pregnancies. This comparison should be treated with caution due to the very small number of cases in twin pregnancies, differing methods, differing time periods between reviews, population confounders, and they may both be an overestimate due to the risk of bias in included studies. The economic model in the previous review used a more conservative estimate of sensitivity of 93.1%.

Meta-analysis for trisomy 13 (Patau's syndrome)

Overall test accuracy of cfDNA testing in twin/multiple pregnancies

In total, 5 studies identified in the previous review [37; 39; 40; 41; 42] and 10 studies identified in this update review [21; 22; 23; 25; 27; 28; 31; 32; 33; 35] had complete 2x2 tables for the detection of trisomy 13. Figure 4 shows a forest plot of the sensitivity and specificity with 95% confidence intervals for each study included in the meta-analysis. There were 14 true positives, 2,645 true negatives, 2 false positive and 2 false negatives among the 15 studies. Included in the previous review, there were one true positive, 943 true negatives, no false positives, and no false negatives among 5 studies. Summing across all studies gives a pooled sensitivity of 87.5% (95%CI 64.0% to 96.5%), and pooled specificity of 99.92% (95%CI 99.7% to 99.98%).

Comparison of cfDNA test accuracy in twin/multiple pregnancies versus singleton pregnancies

The direct comparison in the same population is preferred as there are fewer confounders. There were only 7 studies which reported accuracy in twin pregnancies and singleton pregnancies in the same population [21; 22; 25; 35; 39; 40; 42]. Of these, only one study [25] contained a single case of Patau's syndrome in a twin pregnancy, and this was a false negative result. Due to the low number of studies and event rates it was not possible to meta-analyse the direct comparisons, and compare test accuracy in twin and singleton pregnancies. However, using the methods of Badeau et al. [20], the summaries across these 7 studies were as follows: pooled sensitivity was 0% (0/1, 95%CI 0% to 79.3%) in twin pregnancies and 97.1% (34/35, 95%CI 85.5% to 99.5%) in singleton pregnancies. Pooled specificity was 100% (454/454, 95%CI 99.2% to 100%) in twin pregnancies and 99.96% (113,426/113,472, 95%CI 99.95% to 99.97%) in singleton pregnancies.

Indirect comparison via meta-analysis was not undertaken due to the differing time periods between reviews. In the previous review the test accuracy estimates derived from all 24 studies with at least one true positive or one false negative case of Patau's syndrome using different methods (both without zero cell correction) were sensitivity of 97.4% (95%CI 86.2% to 99.6%) and specificity of 99.9% (95%CI 99.9% to 100%). This point estimate for sensitivity in mostly singleton pregnancies is higher than that for twin pregnancies in this review (87.5%) but the confidence intervals are overlapping (95%CI 64.0% to 96.5% for twin pregnancies). The point estimates for specificity are similar in twin and (mainly) singleton pregnancies. This comparison should be treated with caution due to the very small number of cases in twin pregnancies, differing methods, differing time periods between reviews, population confounders, and they may both be an overestimate due to the risk of bias in included studies. The economic model in the previous review used a more conservative estimate of sensitivity of 82.7%.

Monochorionic versus dichorionic twin pregnancies

Numbers of true positive and false negative cfDNA test results for the detection of the 3 trisomies were reported separately for monochorionic and dichorionic twin pregnancies in 10 studies [23; 25; 27; 28; 31; 33; 37; 38; 40; 41]. The sensitivity in 50 dichorionic twin pregnancies with Down's syndrome derived from 10 studies was 96.0% (48/50; 95%CI 85.1% to 99.3%), and the sensitivity in 3 monochorionic twin pregnancies was 100% (3/3; 95%CI 31.0% to 100%). The sensitivity in 12 dichorionic twin pregnancies with Edwards' syndrome derived from 10 studies was 91.7% (11/12), and the sensitivity in one monochorionic twin pregnancy was 0% (0/1). The sensitivity in 3 dichorionic twin pregnancies with Patau's syndrome derived from 9 studies was 33.3% (1/3), and there was no case of Patau's syndrome in a monochorionic twin pregnancy.

Interpreting meta-analysis results in a twin population at higher chance of fetal trisomies

The reviewers applied the estimates of sensitivity, specificity and test failure rate of cfDNA testing to a theoretical cohort of 10,000 pregnant women with twin pregnancies from a higher-chance population (Figure 6). Population prevalence was determined as the median prevalence for the studies included in higher chance groups as determined in the previous review (3.33% for Down's syndrome, 1.5% for Edwards' syndrome and 0.5% for Patau's syndrome) [1]. When all 10,000 pregnant women with a chance of 1 in 150 or higher after the first trimester combined test are undergoing cfDNA testing as follow-on test, there will be an estimated 371 women with an initial test failure. In the remaining 9,629 twin pregnancies with cfDNA test result, 313 (95%CI 283 to 319) cases of Down's syndrome will be detected and 8 (95%CI 2 to 38) will be missed by cfDNA testing. The positive predictive value for Down's syndrome detection in a higher-chance twin population would be 97.8% (313/320). For Edwards' syndrome, 129 (95%CI 99 to 140) cases will be detected and 15 (95%CI 4 to 45) cases be missed by cfDNA testing.

The positive predictive value would be 100% (129/129). For Patau's syndrome, 42 (95%CI 31 to 46) cases will be detected and 6 (95%CI 2 to 17) will be missed by cfDNA testing. The positive predictive value would be 84.0% (42/50). Of the 9,116 twin pregnancies with none of the babies affected by one of the 3 trisomies, 7 (95%CI 1 to 65) women will receive a false positive result for T21, no (95%CI 0 to 9) woman will receive a false positive result for T18, and 8 (95%CI 2 to 29) women will receive a false positive result for T13.

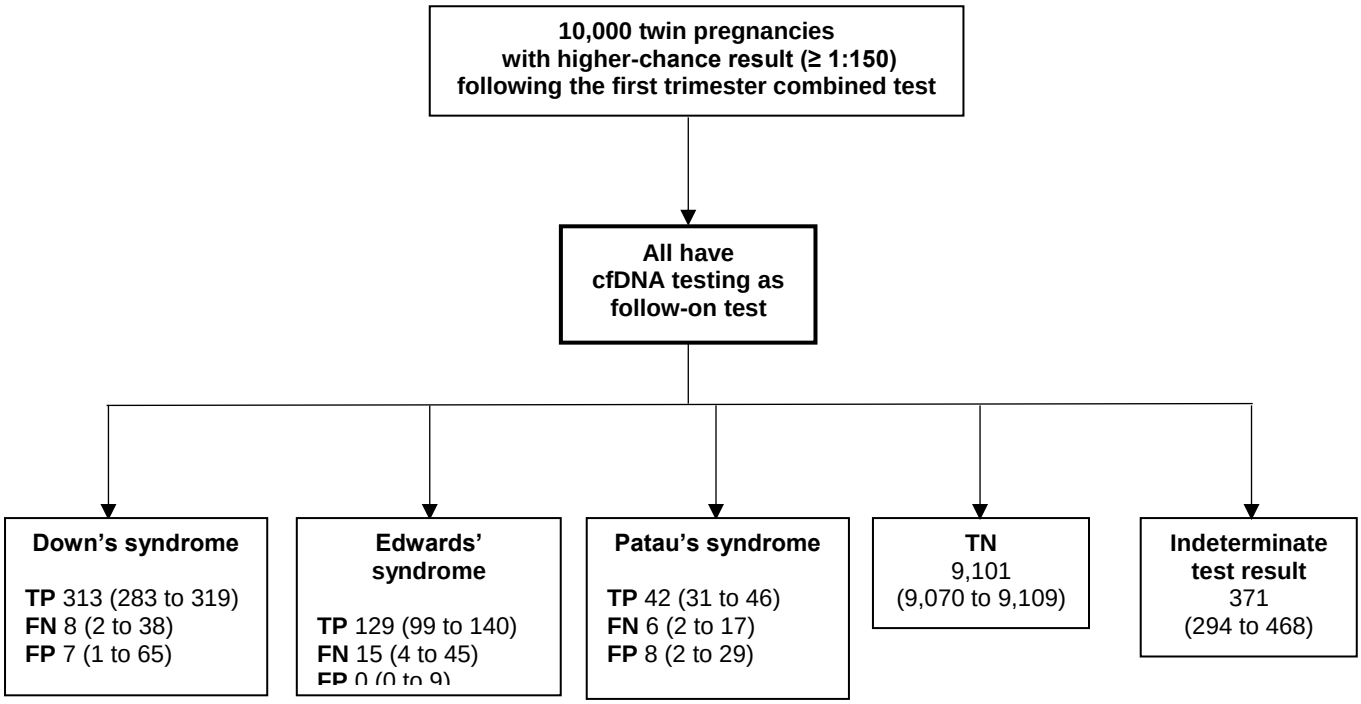


Figure 6. Accuracy in twin pregnancy applied to a hypothetical cohort of 10,000 twin pregnancies at higher chance of fetal trisomy after the first trimester combined test (95% confidence interval in brackets).

Analyses by specific research question addressed

Research question 1.1: What is the accuracy of cfDNA testing in predicting T21, T18 and T13 in women with twin/multiple pregnancies and a pre-defined higher-chance result (≥1:150) following a combined test?

No studies were identified that carried out cfDNA testing in pregnant women with a higher-chance test result threshold as typically used in the UK screening programme (1:150) when estimated by the first trimester combined screening test.

Research question 1.2: How does changing the threshold for defining a higher-chance result following a combined test affect the accuracy of cfDNA testing in women with twin/multiple pregnancies?

Two studies [38; 40] were identified in the previous review and 5 studies were identified in this update review [22; 24; 25; 28; 35] that carried out cfDNA testing in pregnant women with twin/multiple pregnancies at high chance of fetal trisomies. The studies used samples from (primarily) higher-chance pregnant women with a range of different indications which included higher-chance result following serum screening with unreported cut-offs or other indications like advanced maternal age or fetal ultrasound anomalies. It was not possible to present cfDNA testing performance in twin/multiple pregnancies at different chance cut-offs for the first trimester combined test ranging from very high to low chance or to present an optimal chance cut-off to maximise cfDNA testing performance in clinical practice.

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 with twin/multiple pregnancies?

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 in twin/multiple pregnancies.

Research Question 3: What diagnostic accuracy is achievable by integrating cfDNA testing into the combined test in twin/multiple pregnancies?

No study that reported test accuracy of cfDNA testing as add-on test to the combined test was identified.

Research Question 4: What is the rate of cfDNA testing failure in twin/multiple pregnancies (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% (0/7) [24] to 9.4% [31] in 10 studies performing cfDNA testing in twin/multiple pregnancies [21; 23; 24; 25; 27; 28; 29; 31; 33; 35] (see Table 16 for details). In 5 of these studies, it was unclear if the reported failure rate was after initial testing only or if it included initial and repeat testing [21; 23; 25; 27; 29]. Summing across the 5 studies with clear reporting gives 72 initial test failures among 1,939 test samples from twin/multiple pregnancies (3.7%) [24; 28; 31; 33; 35]. One study directly compared the test failure rate in 438 samples from twin pregnancies and 10,698 samples from singleton pregnancies [31]. The failure rate after first sampling was higher in twin pregnancies (9.4%, 41/438) than in singleton pregnancies (2.9%, 316/10,698) ($p<0.0001$). In this study, repeat testing was successful in 20/39 (51%) twin pregnancies and 148/235 (63%) singleton pregnancies. The median fetal fraction was higher in monochorionic than dichorionic twin pregnancies (10.1% [IQR, 7.6–14.5%] vs 7.7% [IQR, 5.8–10.0%]; $p<0.0001$). Chorionicity did not significantly contribute to the prediction of failed cfDNA testing in twin pregnancies ($p=0.078$).

The interquartile range (IQR) is a measure of variability. It contains the middle 50% of values (25^{th} to 75^{th} percentiles).

Questions 1–4, part b) – What is the test performance (accuracy and test failures) of cfDNA tests based on DNA microarray technology?

This review compares the published evidence on the test performance of cfDNA-MA testing with that of the performance of cfDNA testing based on next generation sequencing reported in the previous, 2015, review.

Eligibility for inclusion in the review

For part b) of review questions 1–4, the reviewers included studies of pregnant women, who had been given non-invasive prenatal testing using cfDNA derived from maternal blood (serum, plasma, whole blood) with microarray-based DNA quantitation as screening test for trisomies 21, 18 or 13 (including mosaicism and translocation), and a reference standard of either genetic verification through amniocentesis, CVS, cordocentesis and fetal pathologic examination after abortion or postnatal phenotypic assessment. Studies with and without a comparator of any “conventional” screening test and with or without a comparator of sequencing-based cfDNA tests were included. Studies with any test accuracy or test failure information were also included, but those studies from which a full 2x2 table could not be adequately constructed were excluded from the meta-analysis.

The reviewers excluded studies reporting the quantification of fetal cells, measurement of total DNA levels in maternal blood, or using epigenetic markers as a screening/diagnostic tool, studies using a cfDNA test not based on microarray, studies with unclear cfDNA test technology (e.g. use of several commercially available tests) and studies using not solely a cfDNA test with DNA microarray approach for DNA quantitation. Also excluded were 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.

Description of the evidence

Database searches yielded 1,891 unique results, of which 4 were judged to be relevant to part b) of the review questions. One additional article [15] was identified from the previous searches. This study was excluded from the previous review [1] as not all cfDNA test results were compared with a suitable reference standard and it was not possible to disaggregate test accuracy outcomes those which had been compared with a reference standard. As the emphasis of this review was to assess the test accuracy of cfDNA testing with DNA microarray-approach, the reviewers now contacted the corresponding author, and were referred to colleagues at Ariosa Diagnostics Inc. (San Jose, CA). The reviewers received unpublished test accuracy data for the cfDNA test results compared with suitable reference standard from this source.

Appendix 2 contains a full PRISMA flow diagram (Figure 12), along with a table of the included publications and details of which questions these publications were identified as being relevant to (Table 10). A list of excluded studies at full text level with reasons is provided in Table 11.

Discussion of findings

Characteristics of included studies

A study-level summary of data extracted from the 5 included publications is presented in ‘Appendix 3 - Summary and appraisal of individual studies’ (Table 12). An additional 7 studies (reported in 8 articles) used DNA microarrays or sequencing for DANSR product quantification in unknown proportions and no separate data for the subgroup analysed with the microarray-based test were available [31; 43; 44; 45; 46; 47; 48; 49]. These studies were excluded from the review. For additional information, a summary of the study characteristics and main findings of these 8 excluded articles can be found in Appendix 3 (Table 13).

Of the 5 included studies, one study used a randomised controlled trial (RCT) design [50], 3 studies were cohort studies [15; 51; 52] and the remaining study used an uncontrolled before-after design [36].

Two studies were single centre studies from Germany [50] and Spain [36]. One study included 3 centres in Canada [51], while the remaining 2 studies retrospectively analysed frozen samples provided by an unclear number of centres [15; 52].

The number of women included in the study ranged from 799 [52] to 6,011 [36] but cfDNA testing was not necessarily offered to all included women. The number of women who had cfDNA testing performed ranged from 54 [36] to 1,198 [51]. In total, 3,074 samples

were successfully tested using the cfDNA-MA test with a suitable reference standard in the 5 studies.

Four studies included women with singleton pregnancies only [15; 36; 50; 51], while the remaining study included a mixed population of women with singleton and twin pregnancies (759 and 40 women, respectively) [52].

Two studies compared the screening performance of cfDNA testing and standard screening approaches in first trimester pregnant women with low (no fetal ultrasound anomalies detected) [50] or general (no prior testing) [51] chance of fetal trisomies. One study reported the test performance of cfDNA testing in first trimester pregnant women at high chance of fetal trisomies (first trimester combined screening showed chance > 1:250 without fetal ultrasound anomalies) [36], while the 2 remaining studies retrospectively assessed the test accuracy of cfDNA-MA testing in populations with high trisomy prevalence and unclear proportion of first trimester pregnancies [15; 52].

All 5 studies used the Harmony Prenatal Test (Ariosa Diagnostics, Inc., San Jose, CA) which is based on the Digital Analysis of Selected Regions (DANSR) targeted approach and makes DANSR products from non-polymorphic assays on chromosomes 21, 18 and 13. DANSR assay products were quantified using custom DNA microarrays for all study samples. Analyses were performed by Ariosa Diagnostics Inc. (San Jose, CA) in 4 studies [15; 36; 51; 52] and by Cenata GmbH (Tuebingen, Germany) in one study [50]. The Fetal Fraction Optimized Risk of Trisomy Evaluation (FORTE) was used to estimate the chance of having a baby with Down's, Edwards' or Patau's syndrome. A higher-chance result was defined as a chance equal or greater 1in 100 in 4 studies [15; 36; 50; 52], while the cut-off was not reported in one study [51].

One study performed a head-to-head comparison of DNA microarray versus next-generation sequencing as DANSR assay product quantitation methods in the same sample [15].

Methodological quality of included studies

The methodological quality of the 5 included studies, assessed by the tailored QUADAS-2 is summarised in Figure 7, Figure 8 and Table 14. These illustrate the risk of bias regarding the 5 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 the 3 different cfDNA test implementation approaches (cfDNA testing as follow-on test, as replacement test or as add-on test).

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 was judged as low in only one study [50]. Two studies were classified as unclear risk of bias because it was not explicitly stated that patients were recruited randomly or consecutively [51; 52], and exclusions from the study were not further described [51; 52]. Two studies were classified as being at high risk of bias; in one study, more than 10% of women at high risk of fetal trisomies did not have a cfDNA test [36], while in the other study recruitment was not random or consecutive and samples that failed the quality control threshold were excluded from the study [15].

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 7 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 the 3 prospective cohort studies. [36; 50; 51] Risk of bias was judged as unclear in 2 studies as it was unclear if blood samples were taken before or at least 7 days after invasive testing [15; 52]. No study was classified as high risk of bias in the index test domain.

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. Prenatal or postnatal karyotyping or phenotypic newborn assessment were accepted as appropriate reference standard. One study was classified as unclear risk of bias, as standards other than the pre-specified reference standards were used in an unreported proportion of women [36]. The 4 remaining studies were at low risk of bias in this domain [15; 50; 51; 52].

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. The risk of bias was judged as

high for this domain in all 5 studies because cfDNA testing was not performed in all eligible women and/or failed to provide a result in all tested women and/or not all women received a reference standard and not all women were included in the analysis.

The risk of bias regarding the *role of sponsor* was considered as high if studies were funded by profit-making companies and if involvement of the sponsor in the design, conduct or publication of the study 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. The risk of bias regarding the role of sponsor was judged as low in only one of the 5 included studies [51]. Risk of bias was classified as unclear in 2 studies as the funding source was not reported [36] or the company Ariosa Diagnostic Inc. (San Jose, CA) provided the kits for the cfDNA test. In addition, one of the authors was a paid employee of Ariosa Diagnostic Inc. (San Jose, CA), and the role of the sponsor in the design, conduct and publication of the study was not stated [50]. The remaining 2 studies were judged as at high risk of bias in this domain as the study was designed, performed and published by employees of Ariosa Diagnostic Inc. (San Jose, CA) [15; 52].

In summary, risk of bias was high in all studies, with 3 out of 5 studies considered at high risk of bias in 2 or more domains [15; 36; 52] and in 2 out of 5 studies in one domain [50; 51]. No study scored low or unclear risk of bias in all domains. Figure 7 shows that the study flow (exclusions from analysis) presented the area with the greatest risk of bias as all 5 studies excluded women from analysis due to test failures and/or non-available results of newborn examination or genetic testing (pre- or postnatal). Another issue was the role of sponsor with only one out of 5 studies stating that the role of sponsor played no part in design, conduction and publication [51].

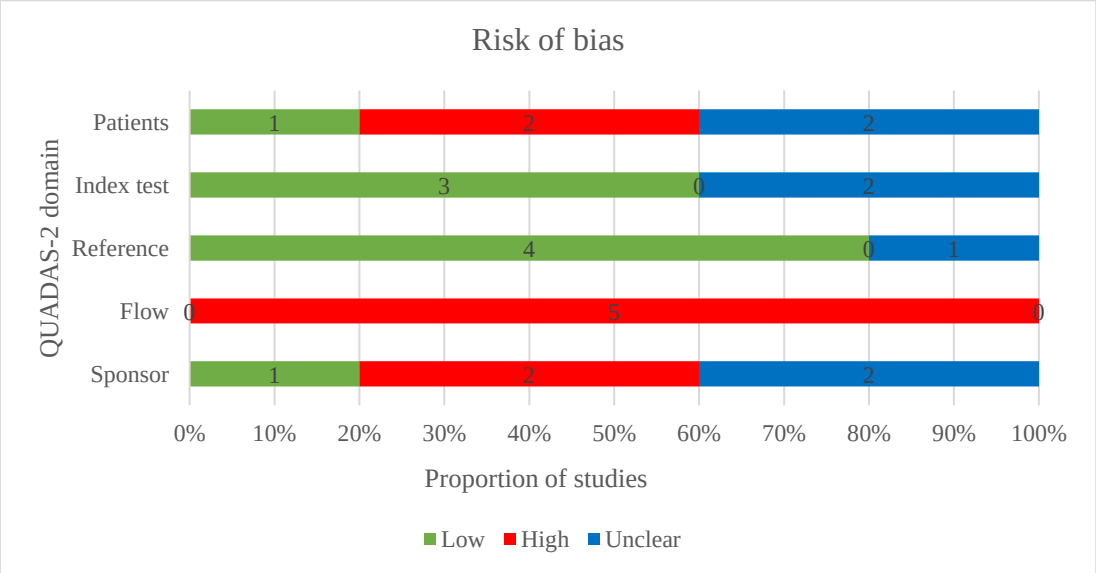


Figure 7. Proportion of studies with low, high or unclear risk of bias.

Applicability concerns

As the specific research questions aim to address cfDNA testing performance in the first trimester and in comparison with the first trimester combined test, applicability concerns of included patients should be regarded as high if less than 80% of women were recruited in the first trimester. In the context of cfDNA testing as follow-on test (research question 1) or add-on test (research question 3), cfDNA testing should have been carried out in pregnant women with prior aneuploidy screening using the first trimester combined test. For the use of cfDNA testing as replacement test (research question 2), cfDNA testing should have been carried out in pregnant women without prior aneuploidy screening (general obstetric population) and also include multiple gestations.

Only one study providing data on cfDNA test accuracy as follow-on test included 100% pregnant women in their first trimester at high chance of fetal trisomy following the first trimester combined test and was classified as low levels of concern regarding applicability of patient spectrum [36]. The other 2 studies investigating overall cfDNA test accuracy were rated as having high applicability concerns as they included less than 80% of pregnant women in their first trimester [15; 52] and had a very high prevalence of fetal trisomies (1:4.3 and 1:5.4, respectively) [15; 52]. Two studies included first trimester pregnant women for head-to-head comparison of cfDNA testing and standard screening test performance, but were judged as high concerns regarding patient applicability as both excluded multiple pregnancies [50; 51] or included lower-chance women without ultrasound anomalies only [50].

In terms of the *index test*, applicability concerns of the 2 studies comparing cfDNA testing to a standard screening test were classed as low in one study as the first trimester combined test was used in all women [50], and high in the other study which used the first trimester combined test in only 25% of women [51]. Concerns regarding applicability of the 3 studies evaluating cfDNA testing accuracy only were classified as low in all studies [15; 36; 52].

Applicability concerns regarding the *reference standard* were classed as low in 4 out of 5 studies as the pre-defined reference standards were used [15; 50; 51; 52]. Applicability was unclear in the remaining study as standards other than the pre-specified reference standards were used in an unreported proportion of women [36].

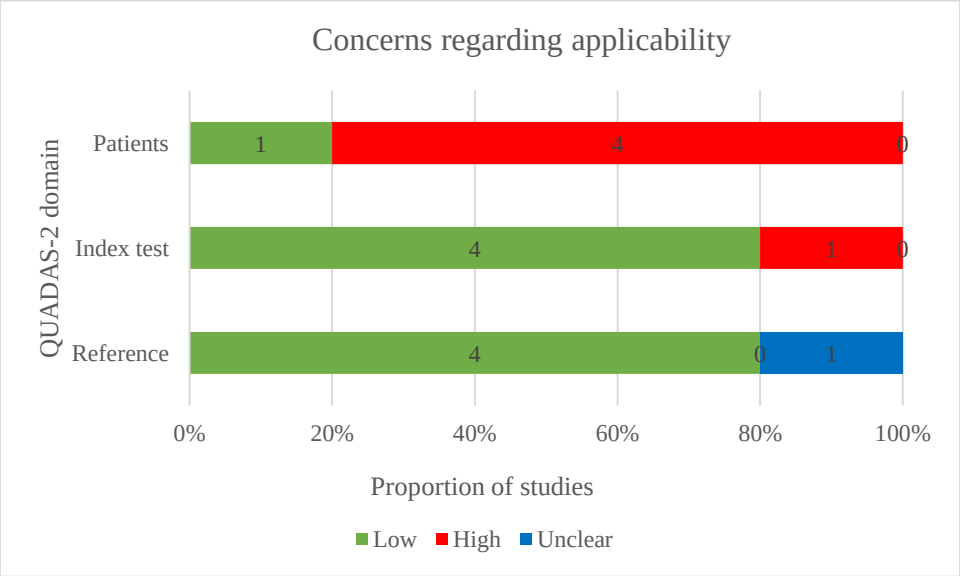


Figure 8. Applicability concerns in included studies

Analysis of the evidence

The accuracy of cfDNA testing in every included study is shown in Table 15. 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.

Meta-analysis results

Overall test accuracy of DNA microarray-based cfDNA testing (cfDNA-MA)
The reviewers included studies that provided a complete 2x2 table in the meta-analysis, even if they did not sample random or consecutive women, therefore the estimates from the meta-analysis may be subject to spectrum bias with atypical cases and controls selected. Studies with no cases of a certain trisomy were excluded from the meta-analysis for that particular condition.

There were 5 eligible studies that used cfDNA-MA testing only, including the Juneau paper [15] described above. One single centre study from Germany used a randomised controlled trial (RCT) design comparing the screening performance of cfDNA testing and first trimester combined screening in women with singleton pregnancy and no fetal ultrasound anomalies detected [50]. Langlois et al. report on a cohort study set in 3 Canadian centres that compared the screening performance of cfDNA testing and

standard screening approaches in first trimester pregnant women with singleton pregnancy with a general chance of fetal trisomy [51]. Gil et al. performed an uncontrolled before-after study at a single centre in Spain [36]. After the introduction of cfDNA testing, NIPT was offered to women with singleton pregnancies, a FTCS chance of >1:250 and no fetal ultrasound anomalies detected. The remaining 2 studies were designed and conducted by Ariosa Diagnostics Inc. (San Jose, CA) and included retrospective analyses of stored/frozen samples with high prevalence of trisomies provided from an unclear number of centres [15; 52]. Forest plots of the sensitivity and specificity for all 3 trisomies from the individual studies using cfDNA-MA testing are given in Figure 9.

Trisomy 21 (Down's syndrome)

For trisomy 21, there were a total of 186 true positives, 2,887 true negatives, no false positives and 1 false negative in the 5 included studies. One study [50] was excluded from the meta-analysis as there were no cases of T21. Summary estimates across the 4 studies were 99.5% (95%CI 96.3% to 99.9%) for sensitivity and 100% for specificity. The confidence interval for specificity was not estimable in the meta-analysis as there were no false positive results. Instead, the confidence interval for specificity was calculated by summing across studies and using the Wilson method (95% CI 99.87% to 100%). Figure 10 shows the summary receiver operating characteristic (ROC) plot for cfDNA-MA testing for detecting T21 (left part of the figure).

Trisomies 18 and 13 (Edwards' syndrome and Patau's syndromes)

Bivariate meta-analysis was not possible for trisomies 18 or 13 because 3 out of the 5 studies would have been excluded because they contained no cases of trisomy 18 or 13. For trisomy 18, there were in total 42 true positives, 3,030 true negatives, 1 false positive and 1 false negative reported in the 5 included studies. Summing across all studies gives a sensitivity of 97.7% (95%CI 87.9% to 99.6%), and specificity of 99.97% (95%CI 99.81% to 99.99%). For trisomy 13, there were in total 19 true positives, 3,054 true negatives, 1 false positive and no false negatives reported in the 5 studies. Summing across all studies gives a sensitivity of 100% (95%CI 83.2% to 100%), and a specificity of 99.97% (95%CI 99.81% to 99.99%).

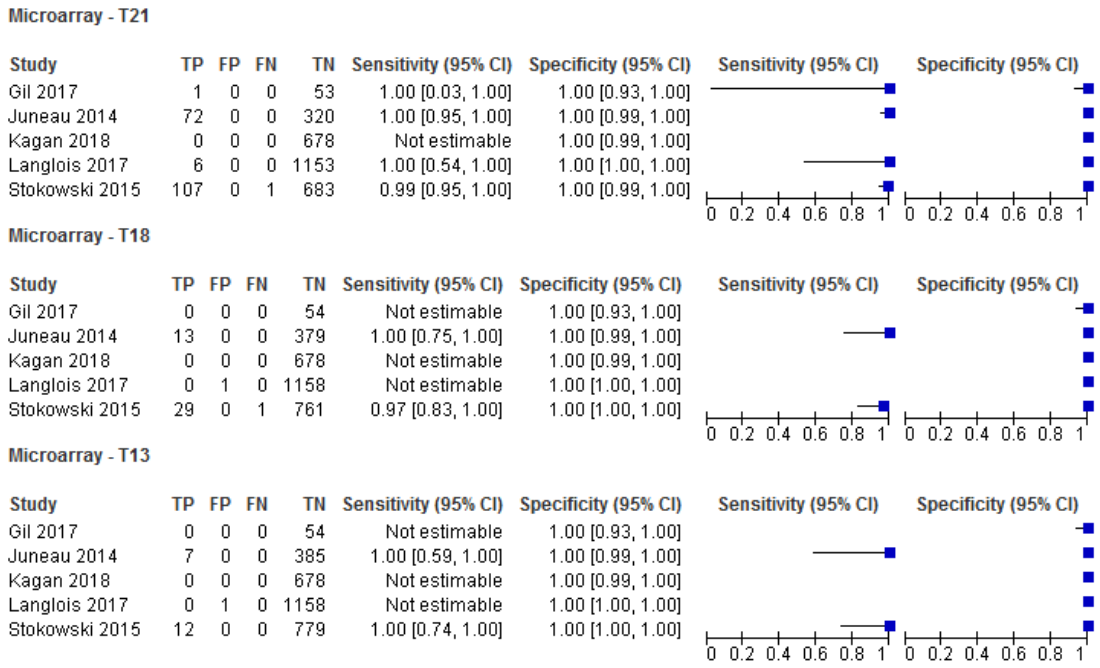


Figure 9. Sensitivity and specificity of prenatal cfDNA testing with DNA microarrays for the detection of fetal Down's syndrome (T21), Edwards' syndrome (T18) and Patau's syndrome (T13).

Direct comparison of DNA microarray-based versus sequencing-based cfDNA testing in head-to-head studies

Head-to-head studies are the most informative studies to inform whether the microarray-based version of a cfDNA test using the Digital ANalysis of Selected Regions (DANSR) approach can be considered equivalent to the sequencing-based version of the test. This is because biases arising from testing in differing populations between studies are not present. The reviewers only found one head-to-head test accuracy study with a suitable reference standard. This precluded conducting the planned meta-analysis of this type of study. Juneau et al. [15] included a subset of 392 out of 878 women for whom there was an acceptable reference standard (invasive genetic testing or postnatal newborn examination followed by detailed genetic analysis, when trisomy was suspected). Risk of bias in this study was rated as high in 3 of 5 domains of the tailored QUADAS-2 tool (i.e. patient selection, flow and timing and role of sponsor). There was an unclear risk of bias in the index test domain as it was unclear if the blood samples were taken before or at least 7 days after invasive testing.

For both DNA microarray and sequencing technologies, this study had 72, 13 and 7 true positives for T21, T18 and T13, respectively, with the remainder true negatives in the subgroup of 392 samples with a suitable reference standard (data received by personal communication with Roche Sequencing Solutions Inc. / Ariosa Diagnostics Inc., San

Jose, CA). Therefore, sensitivity and specificity estimates for both tests were 100% for all 3 trisomies. For T21, the difference in sensitivity between the 2 tests was 0% (95%CI -1.4% to +1.4%). The difference in specificity between the 2 tests was 0% (95%CI -0.3% to 0.3%). For T18, the difference in sensitivity between the 2 tests was 0% (95%CI -7.7% to +7.7%). The difference in specificity between the 2 tests was 0% (95%CI -0.3% to 0.3%). For T13, the difference in sensitivity between the 2 tests was 0% (95%CI -14.3% to +14.3%). The difference in specificity between the 2 tests was 0% (95%CI -0.3% to 0.3%). For all trisomies, the McNemar's test statistic of 1 indicated no statistically significant difference between the tests in either sensitivity or specificity. This should be interpreted as no evidence of a difference rather than evidence of no difference.

Indirect comparison of DNA microarray-based versus sequencing-based cfDNA testing

For the indirect comparison of DNA microarray-based cfDNA testing versus sequencing-based cfDNA testing (both using the DANSR approach) the reviewers first meta-analysed test accuracy studies of the sequencing-based cfDNA test alone, using the same methods as for the microarray-based cfDNA test alone used above. Then the authors combined these studies into a single bivariate meta-analysis with test type (microarray or sequencing) as a covariate.

Overall test accuracy of sequencing-based cfDNA testing with DANSR approach

The previous review [1] identified 11 studies using sequencing for DANSR assay product quantification [13; 37; 41; 53; 54; 55; 56; 57; 58; 59; 60]. The Digital ANalysis of Selected Regions (DANSR) approach was performed in all 11 studies by a single laboratory and differed in the number of non-polymorphic loci on chromosomes 13, 18, and 21 which were amplified and sequenced (384 loci [53] versus 576 loci [13; 37; 41; 54; 55; 56; 57; 58; 59; 60] on each chromosome) as well as in the threshold used to distinguish between euploid and aneuploid samples. A z-score of chromosome proportions with external reference set was used in the initial exploration of the DANSR assay [13; 53] whereas later studies applied the Fetal-fraction Optimized Risk of Trisomy Evaluation (FORTE) algorithm, which incorporates fetal fraction and the prior chance of aneuploidy associated with the subject's maternal and gestational age, and needs no external reference data set [13; 54; 55; 56; 57; 58; 59; 60]. For assessment of chance for trisomies in twin pregnancies, the lower fetal fraction contribution of the 2 fetuses was used by the FORTE algorithm [37; 41].

The reviewers excluded studies from the meta-analysis with incomplete 2x2 tables (n=2) [53; 55] and studies performed exclusively in women with twin pregnancies (n=2) [37; 41]. Bivariate meta-analysis of the remaining 7 studies using sequencing-based quantitation of DANSR products [13; 54; 56; 57; 58; 59; 60] gave estimates across studies for T21 detection of 99.7% (95%CI 82.1% to 99.9%) for sensitivity and 99.95% (95%CI 99.91% to 99.97%) for specificity. This is shown in the summary ROC plot in Figure 10 (right hand

side). For trisomy 18, there were 114 true positives, 23,850 true negatives, 11 false positives and 4 false negatives reported in 6 studies [13; 54; 56; 57; 58; 60]. Summing across all studies, this gives a sensitivity of 96.6% (95%CI 91.6% to 98.7%), and a specificity of 99.95% (95%CI 99.92% to 99.97%) for T18 detection. For trisomy 13, there were 4 true positives, 13,959 true negatives, 4 false positive and 3 false negatives reported in 2 studies [54; 60]. Summing across both studies, this gives summary estimates for sensitivity of 57.1% (95%CI 25.0% to 84.2%) and for specificity of 99.97% (95%CI 99.93% to 99.99%) for T13 detection.

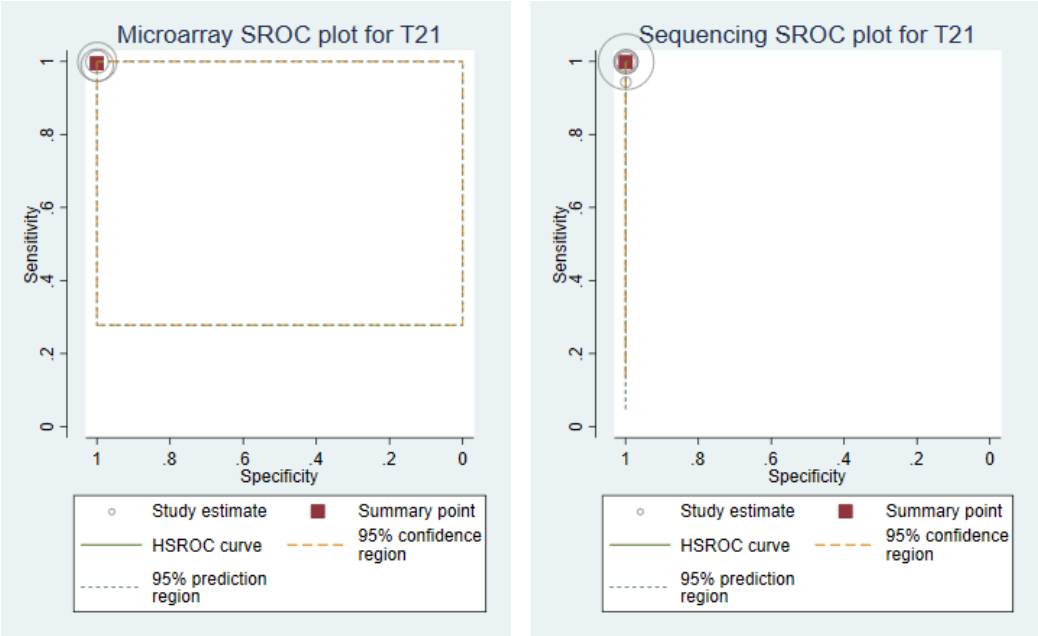


Figure 10. Summary ROC plot for the 4 DNA microarray-based cfDNA test accuracy studies (left) and the 7 sequencing-based cfDNA test accuracy studies from the previous review (right). The large 95% confidence region for DNA microarray is an artefact of the lack of variance due to 0 false positive cases in any studies.

Indirect comparison of DNA microarray-based and sequencing-based cfDNA testing with DANSR approach

The variance of random effects differed between the meta-analysis of sequencing-based cfDNA testing and the meta-analysis of DNA microarray-based cfDNA testing described above. This was primarily due to the lack of false positives in the meta-analysis of DNA microarray-based cfDNA testing. Therefore, we intended to set the variance associated with the specificity of DNA microarray to zero in the meta-analysis of studies using the DANSR approach, with microarray vs sequencing as a covariate. However, these models would not converge so the following analysis assumes equal variances.

Sequencing-based cfDNA testing had a summary sensitivity of 99.6% (95%CI 97.4% to 99.9%) and a summary specificity 99.95% (95%CI 99.91% to 99.97%) for T21 detection.

DNA microarray-based cfDNA testing had a summary sensitivity of 99.5% (95%CI 96.3% to 99.92%) and a summary specificity of 100% with inestimable confidence intervals due to lack of false positive results. There was no evidence of a difference in sensitivity (p=0.81) or specificity (p=0.15) between DNA microarray-based and sequencing-based cfDNA testing for T21 detection. This represents no evidence of a difference rather than evidence of no difference.

Interpreting meta-analysis results in a higher-chance population

The reviewers applied the estimates of sensitivity, specificity and test failure rate of microarray-based cfDNA testing to a theoretical cohort of 10,000 pregnant women deemed at higher-chance of fetal trisomies (
Figure **11**). Population prevalence was determined as the median prevalence for the studies enrolling higher-chance groups as determined in the previous review (3.33% for Down's syndrome [T21], 1.5% for Edwards' syndrome [T18] and 0.5% for Patau's syndrome [T13]) [(1)]. When all 10,000 pregnant women with a chance of 1 in 150 or higher after the first trimester combined test are undergoing cfDNA testing as follow-on test, there will be an estimated 111 with an initial test failure (95%CI 76 to 160). In the remaining 9,889 pregnancies with successful cfDNA test, 328 (95%CI 317 to 329) cases of Down's syndrome will be detected and 2 (95%CI 0 to 12) will be missed by cfDNA testing. The positive predictive value for Down's syndrome detection in a higher-chance obstetric population would be 100% (328/328). For Edwards' syndrome, 145 (95%CI 130 to 148) cases will be detected and 3 (95%CI 1 to 18) cases be missed by cfDNA testing. The positive predictive value would be 98.0% (145/148). For Patau's syndrome, 49 (95%CI 41 to 49) cases will be detected and none (95%CI 0 to 8) will be missed by cfDNA testing. The positive predictive value would be 94.2% (49/52). Of the 9,362 pregnancies not affected by one of the 3 trisomies, none (95%CI 0 to 12) will receive a false positive result for T21, 3 (95%CI 1 to 19) women will receive a false positive result for T18, and 3 (95%CI 1 to 19) women will receive a false positive result for T13.

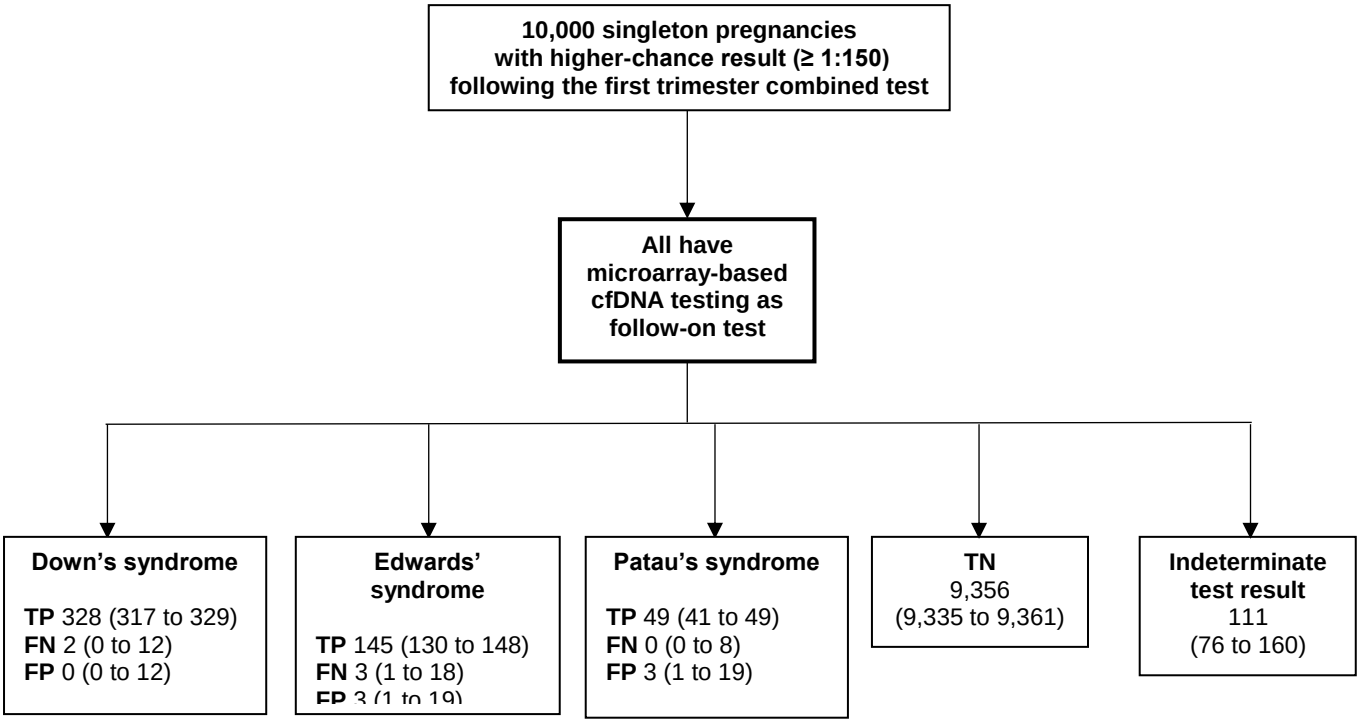


Figure 11. Findings for microarray-based cfDNA testing applied to a hypothetical cohort of 10,000 pregnancies at higher chance of fetal trisomy (95% confidence intervals in brackets).

Analyses by specific research question addressed

Research question 1.1: What is the accuracy of cfDNA-MA testing in predicting T21, T18 and T13 in pregnant women with a pre-defined higher-chance result (≥1:150) following a combined test?

No studies were identified that carried out cfDNA-MA testing in pregnant women with a higher-chance result threshold as typically used in the UK screening programme (1:150) when estimated by the first trimester combined screening test.

Research question 1.2: How does changing the threshold for defining a higher-chance result following a combined test affect the accuracy of cfDNA-MA testing?

No study was identified that reported a comparison of cfDNA-MA test accuracy following different thresholds for the definition of a higher-chance result from the combined test. There were also insufficient studies in populations at different thresholds used to define a higher-chance result to make indirect comparisons of performance between different studies. One study reported the threshold for a higher-chance result from the first trimester combined screening tests (FTCS) prior to cfDNA testing. In 54 women with singleton pregnancy and a FTCS chance for T21 higher than 1 in 250 and no fetal ultrasound anomalies, one out of a total of one cases of T21 was detected with no false positive and no false negative results.

Research Question 2: What is the most accurate primary prenatal screening tool for T21, T18 and T13 in the first trimester when cfDNA-MA 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. We only included studies that reported performance of the combined test in comparison to or in parallel with cfDNA-MA testing, and did not investigate the combined test performance alone. Therefore, a narrative summary of these papers is included. Two individual studies have provided evidence that the specificity of the combined test for T21 is lower than that of cfDNA testing [50; 51]. Kagan et al. found a specificity for T21 detection of 100% (95% CI 99.3% to 100%) for cfDNA testing and 97.5% (95% CI 96.0% to 98.5%) for FTCS (cut-off 1:100) in women with singleton pregnancy and without fetal ultrasound anomalies [50]. Langlois et al. reported a specificity for T21 detection of 100% (95% CI 99.6% to 100%) for cfDNA testing and 93.3% (95% CI 89.5% to 95.8%) for the FTCS test in women with singleton pregnancies and no prior screening test for fetal trisomies [51]. Sensitivity was 100% for both tests in one study [51], and could not be calculated in the other study as neither of the tests had any true positive or false negative results [50].

Research Question 3: What diagnostic accuracy is achievable by integrating cfDNA-MA 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 result 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. No study that reported test accuracy after implementing this approach was identified.

Research Question 4: What is the rate of cfDNA-MA 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 0.9% to 1.9% in 4 studies using cfDNA-MA testing (see Table 16 for details) [36; 50; 51; 52]. Repeat tests after a second blood sample were successful in one out of one and 5 out of 11 (45.5%) women [51]. The main reason for cfDNA test failure was insufficient circulating fetal DNA in 8 out of 8 [52] and 10 out of 11 [51] samples. The only paper directly comparing DNA microarray-based versus sequencing-based cfDNA testing using the DANSR approach included only samples in the study that met quality control thresholds for both quantitation methods [15]. Number of samples that failed quality control were not reported. Summing across all 4 studies which reported failures gives 30 initial test failures per 2,706 samples analysed using the DANSR with DNA microarrays approach (1.1%; 95% CI 0.8% to 1.6%). For comparison, the initial test failure rate in 9 studies using DANSR with sequencing-based cfDNA testing in women with mainly singleton pregnancies (as identified in the previous review [1]) ranged from 0% [13; 53] to 4.9% [57].

Summing across all 9 studies [13; 53; 54; 55; 56; 57; 58; 59; 60] gives a total of 933 initial test failures per 31,077 samples analysed using the DANSR with sequencing approach (3.0%; 95% CI 2.7% to 3.3%). However, this comparison must be treated with caution as these were all published in an earlier time period.

Summary of Findings Relevant to Criterion 4:

a) cfDNA testing in twin/multiple pregnancies

This review identified 16 studies (published in 15 articles) published since February 2015 on the performance of cfDNA testing in twin or higher order multiple pregnancies. It review found that:

1. Whilst there is limited evidence of test accuracy of cfDNA testing for fetal trisomies in twin or higher order multiple pregnancies, based on few cases, sensitivity point estimates are lower than in singleton pregnancies, particularly for Edwards' syndrome and Patau's syndrome, with wide confidence intervals. Due to missing information on chorionicity in the majority of studies and small number of trisomy cases in monochorionic twin pregnancies, possible differences in the cfDNA test accuracy (especially sensitivity) between mono- and dichorionic twin pregnancies could not be assessed.
2. Failure of the initial test in twin pregnancies ranged from 0% to 9.4%. Summing across 5 studies gives 72 initial test failures among 1,939 test samples from twin pregnancies (3.7%). One study found a significantly higher initial failure rate after first sampling in twin pregnancies (9.4%) than in singleton pregnancies (2.9%) tested in the first trimester of pregnancy.

Recommendations on screening

UK NSC criterion 4: *"There should be a simple, safe, precise and validated screening test"*

Sensitivity point estimates are lower in twin pregnancies than in singleton pregnancies, particularly for Edwards' syndrome and Patau's syndrome, with wide confidence intervals. Data from one study suggest that the initial test failure rate after sampling in the first trimester is higher in twin pregnancies than in singleton pregnancies. There is insufficient evidence to assess if the test accuracy (especially sensitivity) is lower and test failure rate higher in dichorionic twin pregnancies than in monochorionic twin pregnancies or singleton pregnancies. This review did not identify any head-to-head studies comparing the performance of cfDNA testing and the first-trimester combined test in twin/multiple pregnancies. It was not commissioned to indirectly compare the performance of cfDNA testing with the currently used screening practice in twin pregnancies, nor to assess the costs and consequences for the current NHS screening programme when cfDNA testing is used in sequence with the combined test, as a replacement for the combined test or in combination with (i.e. alongside) the combined test in twin or higher order multiple pregnancies. Therefore, it is "**UNCERTAIN**" if cfDNA testing should be offered to twin or higher order multiple pregnancies.

b) cfDNA-MA testing

Five studies reported on the test accuracy of cfDNA-MA testing for fetal T21, T18, and T13 with a total of 3,073 samples analysed. Risk of bias was high in all studies with 3 out of 5 studies considered high risk in 2 or more domains and 2 out of 5 studies in one domain of the tailored QUADAS-2.

A single study directly compared cfDNA-MA testing to sequencing-based cfDNA testing in 392 samples. Sensitivity and specificity were 100% for both approaches and no evidence of a difference between the 2 technologies was found. Overall test accuracy of cfDNA-MA testing for fetal trisomy 21 detection derived from summing across all 5 studies gave a pooled sensitivity of 99.5% (95%CI 96.3% to 99.9%) and a pooled specificity of 100% (95% CI 99.87% to100%). For fetal trisomy 18 detection, summing across all 5 studies gave a pooled sensitivity of 97.7% (95%CI 87.9 to 99.6%), and a pooled specificity of 99.97% (95%CI 99.81 to 99.99). For fetal trisomy 13 detection, summing across all studies gave a sensitivity of 100% (95%CI 83.2 to 100%), and a specificity of 99.97% (95%CI 99.81 to 99.99). An indirect comparison of cfDNA-MA testing and sequencing-based cfDNA testing with indicated no significant difference in sensitivity or specificity for fetal trisomy 21 detection.

Failure of the initial test ranged from 0.9% to 1.9% in 4 studies that used cfDNA-MA testing. Summing across all 4 studies gives 30 initial test failures per 2,706 samples analysed using cfDNA-MA testing (1.1%).

Recommendations on screening

UK NSC criterion 4: *“There should be a simple, safe, precise and validated screening test”*

This review found no evidence of difference in test accuracy between cfDNA-MA testing and sequencing-based cfDNA testing including:

- a head-to-head study which reported equivalent test accuracy of the 2 approaches for T21, T18 and T13;
- a meta-analysis of cfDNA-MA testing which produced comparable test accuracy estimates to cfDNA testing in the previous review;
- an indirect comparison of the accuracy of sequencing and microarray approaches using DANSR for detection of T21. This was also seen in summary values calculated for T18 and T13.

However, this demonstrates no evidence of a difference rather than evidence of no

difference.

The initial test failure rate might be lower with cfDNA-MA testing than with sequencing-based cfDNA testing but this indirect comparison has to be treated with caution as cfDNA tests were performed in different populations and during different time periods.

Compared to the previous review of all cfDNA testing technologies the evidence base for microarray-based cfDNA testing is more limited in terms of volume of studies and participants. There is considerable statistical uncertainty for detection of trisomies 18 and 13. However, the respective evidence bases for different sequencing methods, when considered in isolation, also each had a limited volume when the previous review was undertaken.

Similarly, the assessments risk of bias and applicability of the studies using cfDNA-MA testing suggests that the evidence base is limited in terms of quality and generalisability. A further limitation of this review is that it did not update the previous review's assessment of sequencing-based cfDNA testing.

These issues do represent real limitations on the robustness of the estimates of test performance generated by this review's statistical analyses. However, a number of factors should be borne in mind:

- Compared to the already accepted sequencing-based cfDNA test technologies, it is important to note that, while the DNA quantitation method using DNA microarray is technically distinct from sequencing-based cfDNA testing, the other stages of the workflow (e.g. plasma separation, cfDNA isolation, assays to select clinical relevant regions of the human genome) and bioinformatics analysis remain comparable to sequencing strategies.
- The test will be offered to a group of women considered to be at higher chance of fetal trisomies.
- The test accuracy estimates for cfDNA-MA testing are similar to those used in the economic model in the previous review, which found cfDNA testing to be cost effective. Further, the lower limit of the confidence interval for cfDNA-MA test accuracy is similar to the more conservative test accuracy estimates used in the economic model.
- The test would be applied in an evaluative roll-out of cfDNA testing which aims to assess the performance of cfDNA testing and further information could be generated in that process. In this context, adding microarray-based cfDNA testing could contribute positively to understanding of different versions of NIPT.

For these reasons microarray-based cfDNA testing meets the criterion for test accuracy.

Review summary

Conclusions and implications for policy

This report assesses prenatal screening using cfDNA testing for Down's syndrome, Edwards' syndrome, and Patau's syndrome in the fetus against selected UK NSC criteria for appraising the viability, effectiveness and appropriateness of a screening programme.

This review assessed 4 key questions relating to test accuracy and test failures of prenatal cfDNA testing for Down's syndrome, Edwards' syndrome, and Patau's syndrome in (a) twin/multiple gestations, and (b) using DNA microarray-based cfDNA testing.

Sixteen studies were identified in this update review that examined the test performance of cfDNA tests in twin or higher order multiple pregnancies; 11 of which had complete 2x2 tables and were included in the meta-analysis. Combined with 6 studies enrolling twin/multiple pregnancies identified in the previous review, 17 studies provided complete test accuracy outcomes for this population. Although there is a relatively large number of studies (n = 17), the total number of samples tested in twin pregnancies is small (2,600–2,900 depending on trisomy), and high risks of bias are common. Sensitivity point estimates are lower than in singleton pregnancies, particularly for Edwards' syndrome and Patau's syndrome, with wide confidence intervals. Due to missing information on chorionicity in the majority of studies and small number of trisomy cases in monochorionic twin pregnancies, possible differences in the cfDNA test accuracy (especially sensitivity) between mono- and dichorionic twin pregnancies could not be assessed.

Five studies were identified that examined the test accuracy of cfDNA tests that used DNA microarrays to quantify the targeted DANSR products. There is some evidence that NIPT based on cfDNA-MA testing may be as accurate as sequencing-based cfDNA testing, both using the DANSR approach. However, the volume of evidence is currently small and at high risk of bias. There is one head-to-head test accuracy study indicating that the microarray-based version of the test had the same sensitivity (± 1.4%) and specificity (± 0.3%) as the sequencing-based version for Down's syndrome detection, but it was at high risk of bias and had wider confidence intervals for Patau's syndrome and Edwards' syndrome.

Recommendations on screening

UK NSC criterion 4: *"There should be a simple, safe, precise and validated screening test"*

a) cfDNA testing for fetal trisomy in twin/multiple pregnancies

Sensitivity point estimates are lower in twin pregnancies than in singleton pregnancies, particularly for Edwards' syndrome and Patau's syndrome, with wide confidence intervals. Data from one study suggest that the initial test failure rate after sampling in the first trimester is higher in twin pregnancies than in singleton pregnancies. There is insufficient evidence to assess if the test accuracy (especially sensitivity) is lower and test failure rate higher in dichorionic twin pregnancies than in monochorionic twin pregnancies or singleton pregnancies. This review did not identify any head-to-head studies comparing the performance of cfDNA testing and the first-trimester combined test in twin/multiple pregnancies. It was not commissioned to indirectly compare the performance of cfDNA testing with the currently used screening practice in twin pregnancies, nor to assess the costs and consequences for the current NHS screening programme when cfDNA testing is used in sequence with the combined test, as a replacement for the combined test or in combination with (i.e. alongside) the combined test in twin or higher order multiple pregnancies. Therefore, it is "**UNCERTAIN**" if cfDNA testing should be offered to twin or higher order multiple pregnancies.

Further large scale prospective studies are recommended to verify the performance of cfDNA testing as screening test for the common trisomies in twin pregnancies.

b) DNA microarray-based cfDNA testing (cfDNA-MA)

This review found no evidence of difference in test accuracy between cfDNA-MA testing and sequencing-based cfDNA testing including:

- a head-to-head study which reported equivalent test accuracy of the 2 approaches for T21, T18 and T13;
- a meta-analysis of cfDNA-MA testing which produced comparable test accuracy estimates to cfDNA testing in the previous review;
- an indirect comparison of the accuracy of sequencing and microarray approaches using DANSR for detection of T21. This was also seen in summary values calculated for T18 and T13.

However, this demonstrates no evidence of a difference rather than evidence of no difference.

The initial test failure rate might be lower with cfDNA-MA testing than with sequencing-based cfDNA testing but this indirect comparison has to be treated with caution as cfDNA tests were performed in different populations and during different time periods.

Compared to the previous review of all cfDNA testing technologies the evidence base for microarray-based cfDNA testing is more limited in terms of volume of studies and participants. There is considerable statistical uncertainty for detection of trisomies 18 and 13. However, the respective evidence bases for different sequencing methods, when considered in isolation, also each had a limited volume when the previous review was undertaken.

Similarly, the assessments risk of bias and applicability of the studies using cfDNA-MA testing suggests that the evidence base is limited in terms of quality and generalisability. A further limitation of this review is that it did not update the previous review's assessment of sequencing-based cfDNA testing.

These issues do represent real limitations on the robustness of the estimates of test performance generated by this review's statistical analyses. However, a number of factors should be borne in mind:

- Compared to the already accepted sequencing-based cfDNA test technologies, it is important to note that, while the DNA quantitation method using DNA microarray is technically distinct from sequencing-based cfDNA testing, the other stages of the workflow (e.g. plasma separation, cfDNA isolation, assays to select clinical relevant regions of the human genome) and bioinformatics analysis remain comparable to sequencing strategies.
- The test will be offered to a group of women considered to be at higher chance of fetal trisomies.
- The test accuracy estimates for cfDNA-MA testing are similar to those used in the economic model in the previous review, which found cfDNA testing to be cost effective. Further, the lower limit of the confidence interval for cfDNA-MA test accuracy is similar to the more conservative test accuracy estimates used in the economic model.
- The test would be applied in an evaluative roll-out of cfDNA testing which aims to assess the performance of cfDNA testing and further information could be generated in that process. In this context, adding microarray-based cfDNA testing could contribute positively to understanding of different versions of NIPT.

For these reasons microarray-based cfDNA testing meets the criterion for test accuracy.

This review has highlighted a number of important issues including:

- There is heterogeneity in types of cfDNA tests, with a more limited evidence base for each individual subtype of test, thus increasing statistical uncertainty and concerns regarding risk of bias and applicability to each of the questions.
- Secondly, there is potential for ongoing change in the tests (in algorithms and in materials selected and methods used) which may easily occur after publication of the included studies or in the future after implementation in the NHS. The ongoing changes may affect test accuracy or failure rate.
- Thirdly, there is greater statistical uncertainty in accuracy to detect trisomies 18 and 13 than in trisomy 21.

The reviewers would therefore recommend the following to decrease the risks associated with these issues:

1. Initial and ongoing measurement of test accuracy in practice as part of quality assurance. This would require establishing trisomy status using the reference standards outlined in this review or equivalents.
2. All changes to any part of the test post roll-out should be notified to the UK NSC to consider whether accuracy should be re-evaluated.
3. Implementation of robust quality assurance processes to ensure that changes to the tests post roll-out can be monitored.

Limitations

a) cfDNA testing in twin/multiple pregnancies

The evidence on the overall test accuracy of cfDNA testing for common fetal trisomies in twin/multiple pregnancies comes from 17 studies (6 from the previous review and 11 from the update review) that in total successfully analysed around 2,800 maternal blood samples for T21 and T18 and around 2,600 samples for T13. The identified evidence is at high risk of bias and concerns regarding applicability of the tested women (e.g. time point of cfDNA testing, prior testing and/or chance of fetal trisomy, trisomy prevalence) are mostly high.

The number of blood samples from twin or multiple pregnancies that were analysed using cfDNA testing as screening test for the common fetal trisomies is still low (around 2,600 to 2,900 samples in total). The studies included in the meta-analysis contained a low number of trisomy cases, especially for T18 (19 cases) and T13 (16 cases), resulting in wide confidence intervals for the sensitivity estimates.

For one key study comparing directly 438 twin pregnancies and nearly 10,700 singleton pregnancies from the same UK population, the reviewers had to use unpublished data received by personal communication on the number of false positives and true negatives in twin pregnancies. Unfortunately, the reviewers were unable to get a complete 2x2 table for the singleton pregnancies, and this most applicable UK study with relative high number of samples tested could not be included in the meta-analysis that directly compared the test accuracy of twin/multiple and singleton pregnancies.

A limitation of this review is the lack of data to address some of the key questions. There was no evidence on test accuracy in twin/multiple pregnancies when cfDNA testing and the first trimester screening tests are combined in an integrated testing scenario. There was also a lack of evidence on cfDNA testing performance in twin/multiple pregnancies at the 1:150 cut-off for defining a higher-chance result from prior screening tests and no data was available on cfDNA testing performance for a range of thresholds. There was no evidence comparing directly the test performance of cfDNA testing and the first trimester combined test as primary screening test in twin/multiple pregnancies. This review was not commissioned to indirectly compare the performance of cfDNA testing with the currently used screening practice in twin pregnancies, nor to assess the costs and consequences for the current NHS screening programme when cfDNA testing is used in sequence with the combined test, as a replacement for the combined test or in combination with (i.e. alongside) the combined test in twin or higher order multiple pregnancies.

The indirect comparison of the cfDNA test accuracy in twin or higher order multiple pregnancies (identified from the previous and the update review) versus (mainly) singleton pregnancies (identified from the previous review only) might be biased as this review did not update the evidence on cfDNA testing in singleton pregnancies published since February 2015, and recently published studies might have been missed. It should be regarded as exploratory and should not be used to draw conclusions.

Differences in test performance between monochorionic and dichorionic twin pregnancies could not be assessed due to the low number of affected cases in monochorionic twin pregnancies (n=4).

This review did not assess publication bias. It is possible that the published studies overestimated the test accuracy of cfDNA testing for the common fetal trisomies in multiple pregnancies.

b) cfDNA-MA testing

Whilst assessing the full texts of potentially relevant studies the reviewers noticed that many papers did not describe the cfDNA testing methodology used in sufficient detail to allow a decision on inclusion or exclusion. The reviewers therefore contacted the

corresponding authors for clarification and realised that the majority of them sent the study samples to providers of commercially available cfDNA testing without knowing details of the testing method. The reviewers therefore contacted the laboratory concerned (Ariosa Diagnostics Inc., San Jose, CA) for information on the technologies used to analyse these particular study samples and have to rely on the reliability of the manufacturer's information. The reviewers received contradictory information for the cfDNA test methodology used in the study published by Miltoft et al. (2018) [48] and excluded it as Ariosa Diagnostics Inc. (San Jose, CA) confirmed that sequencing was used for DNA quantitation in a proportion of study samples. In total, we excluded 7 studies (published in 8 articles) from this review that, according to information from Ariosa Diagnostics Inc. (San Jose, CA), used DNA microarray as DNA quantitation method in an unknown proportion of study samples. These "mixed technology" studies might provide additional useful information on the test performance. The reviewers therefore summarised them in an extra table for information purposes only (Table 13).

The key study by Juneau et al. (2014) that directly compared test accuracy of sequencing-based versus DNA microarray-based cfDNA testing was sponsored and performed by the manufacturer, which inevitably carries a high risk of bias. The reviewers also had to use unpublished data provided by the manufacturer for the subgroup of samples that had a suitable reference standard.

This review did not update the evidence published since February 2015 on the test performance of sequencing-based cfDNA tests with DANSR approach but used the test accuracy as reported in the previous UK NSC report from 2015 for the indirect comparison of the 2 technologies for DANSR product quantification. The manufacturer (Ariosa Diagnostics Inc., San Jose, CA) confirmed that from 10 November 2014, they used DNA microarrays as the main technology for quantifying DANSR products, so there might have been some recent publications on the test performance of the sequencing-based version of the test that might have been missed. This could have introduced bias in the indirect comparison of the sequencing-based versus DNA microarray-based version of the cfDNA test and the analysis should be regarded as exploratory and should not be used to draw conclusions.

The evidence identified on test performance of cfDNA-MA testing was limited. Only 5 studies were identified. These included in total just over 3,000 maternal blood samples. None of the 5 included articles was 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 in 4 out of 5 studies, as 2 of the 3 studies investigating cfDNA test accuracy only included less than 80% of pregnant women in their first trimester and had a very high prevalence of fetal trisomies in the tested populations. The other 2 studies included first trimester pregnant women for head-to-head comparison of cfDNA-MA testing and standard

screening test performance, but were judged as high concerns regarding patient applicability as both excluded multiple pregnancies and one of them only included lower-chance women without ultrasound anomalies only.

A limitation of this review is the lack of data to address some of the key questions. There was no evidence on test accuracy when cfDNA-MA testing and the first trimester screening tests are combined in an integrated testing scenario. There was also a lack of evidence on cfDNA-MA testing performance at the 1:150 cut-off for defining a higher-chance result from prior screening tests and no data was available on cfDNA-MA testing performance for a range of thresholds.

Appendix 1 — Search strategy

Electronic databases

The search strategy included searches of the databases shown in Table 4 (MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print, Embase, Cochrane Library, and Web of Science).

Table 4. Summary of electronic database searches and dates

Database	Platform	Searched on date	Date range of search
MEDLINE	Ovid SP	9th July 2018	February 2015 to July 2018
MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print	Ovid SP	9th July 2018	February 2015 to July 2018
Embase	Ovid SP	9th July 2018	February 2015 to July 2018
The Cochrane Library, including: - Cochrane Database of Systematic Reviews (CDSR) - Cochrane Central Register of Controlled Trials (CENTRAL) Database of Abstracts of Reviews of Effects (DARE)	Wiley Online	9th July 2018	February 2015 to July 2018
Web of Science	Ovid SP	9th July 2018	Years 2015 - 2018

Search Terms

Search terms included combinations of free text and subject headings (Medical Subject Headings [MeSH] for MEDLINE, and Emtree terms for Embase), grouped into the following categories:

- Disease area: Trisomy, aneuploidy, Down's syndrome, Edwards' syndrome, Patau's syndrome.
- Intervention: NIPT, cfDNA testing.

Search terms for MEDLINE are shown in Table 5, for Medline In-Process & Other Non-Indexed Citations in Table 6, for Embase in Table 7, for the Cochrane Library in Table 8, and for Web of Science (Science and Social Science databases, Science and Social Sciences Conferences) in Table 9.

Table 5. Search strategy for MEDLINE (Ovid)

Database: Ovid MEDLINE(R) <1946 to June Week 5 2018>

Search Strategy:

---

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. (29480)

2 (NIPD or NIPT).mp. (482)

3 (cf?DNA or cff?DNA or ccff?DNA or cell?free?DNA).mp. (782)

4 (DNA adj1 (cell or free or cell?free or f?etal)).mp. (8908)

5 (maternal adj1 (blood or plasma or DNA)).mp. (10377)

6 (MPS or DANSR or parental support or MateriT21 or Verify or Harmony or Panorama*).mp. (59679)

7 1 or 2 or 3 or 4 or 5 or 6 (106210)

8 Trisomy/ (11695)

9 trisom*.mp. (20398)

10 Aneuploidy/ (11637)

11 aneuploid*.mp. (21616)

12 Down Syndrome/ (23063)

13 (down* adj1 syndrom*).mp. (26889)

14 (edward* adj1 syndrom*).mp. (269)

15 (Patau adj1 syndrom*).mp. (134)

16 ("T21" or "T18" or "T13").mp. (1325)

17 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 (59578)

18 7 and 17 (1593)

19 limit 18 to ed=20150209-20180709 (607)

20 limit 19 to english language (551)

Table 6. Search strategy for Medline In-Process & Other Non-Indexed Citations

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. (4198)
2 (NIPD or NIPT).mp. (224)
3 (cf?DNA or cff?DNA or ccff?DNA or cell?free?DNA).mp. (409)
4 (DNA adj1 (cell or free or cell?free or f?etal)).mp. (1195)
5 (maternal adj1 (blood or plasma or DNA)).mp. (769)
6 (MPS or DANSR or parental support or MateriT21 or Verify or Harmony or Panorama*).mp. (14682)
7 1 or 2 or 3 or 4 or 5 or 6 (20494)
8 Trisomy/ (5)
9 trisom*.mp. (1213)
10 Aneuploidy/ (3)
11 aneuploid*.mp. (1575)
12 Down Syndrome/ (18)
13 (down* adj1 syndrom*).mp. (1587)
14 (edward* adj1 syndrom*).mp. (27)
15 (Patau adj1 syndrom*).mp. (25)
16 ("T21" or "T18" or "T13").mp. (191)
17 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 (3798)
18 7 and 17 (293)
19 limit 18 to ed=20150209-20180709 (41)
20 limit 19 to english language (40)

Table 7. Search strategy for Embase

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. (49472)
2 (NIPD or NIPT).mp. (1240)
3 (cf?DNA or cff?DNA or ccff?DNA or cell?free?DNA).mp. (2627)
4 (DNA adj1 (cell or free or cell?free or f?etal)).mp. (23379)
5 (maternal adj1 (blood or plasma or DNA)).mp. (17543)
6 (MPS or DANSR or parental support or MateriT21 or Verifi* or Harmony or Panorama*).mp. (173918)
7 1 or 2 or 3 or 4 or 5 or 6 (259158)
8 Trisomy/ (9503)
9 trisom*.mp. (29540)
10 Aneuploidy/ (21780)
11 aneuploid*.mp. (30257)
12 Down Syndrome/ (30581)
13 (down* adj1 syndrom*).mp. (34020)
14 (edward* adj1 syndrom*).mp. (665)
15 (Patau adj1 syndrom*).mp. (321)
16 ("T21" or "T18" or "T13").mp. (2184)
17 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 (82880)
18 7 and 17 (3909)
19 limit 18 to english language (3617)
20 limit 19 to dc=20150209-20180709 (1454)
21 limit 19 to em=201502-201828 (1203)
22 20 or 21 (1503)

Table 8. Search strategy for Cochrane Library

#1	((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*)):ti,ab,kw or (NIPD or NIPT):ti,ab,kw or (cfDNA or cffDNA or ccffDNA or "cell free DNA"):ti,ab,kw or (DNA near/3 (cell or free or cell?free or f?etal)):ti,ab,kw or (maternal near/3 (blood or plasma or DNA)):ti,ab,kw (Word variations have been searched)	S	3549
#2	(MPS or DANSR or parental support or MaterniT21 or Verify or Harmony or Panorama*):ti,ab,kw (Word variations have been searched)	S	9706
#3	#1 or #2		13211
#4	MeSH descriptor: [Trisomy] explode all trees	M	31
#5	trisom*:ti,ab,kw (Word variations have been searched)	S	237
#6	MeSH descriptor: [Aneuploidy] explode all trees	M	170
#7	Aneuploid*:ti,ab,kw (Word variations have been searched)	S	391
#8	MeSH descriptor: [Down Syndrome] explode all trees	M	420
#9	(down* near/1 syndrom*):ti,ab,kw (Word variations have been searched)	S	674
#10	(edward* near/1 syndrom*):ti,ab,kw (Word variations have been searched)	S	5
#11	(patau near/1 syndrom*):ti,ab,kw (Word variations have been searched)	S	1
#12	t21 or t18 or t13:ti,ab,kw (Word variations have been searched)	S	48
#13	#4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12		1251
#14	#3 and #13		113

Table 9. Search strategy for Web of Science (Science and Social Science databases, Science and Social Sciences Conferences)

# 5	987	#2 AND #1 **Refined by: PUBLICATION YEARS:** (2018 OR 2017 OR 2016 OR 2015) AND **LANGUAGES:** (ENGLISH) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years*
# 4	1,017	#2 AND #1 **Refined by: PUBLICATION YEARS:** (2018 OR 2017 OR 2016 OR 2015) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years*
# 3	3,434	#2 AND #1 *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years*
# 2	58,737	**TOPIC:** (trisom* or aneuploid* or "down syndrome") *OR* **TOPIC:** (down NEAR/1 syndrom*) *OR* **TOPIC:** (edward* NEAR/1 syndrom*) *OR* **TOPIC:** (patau* NEAR/1 syndrom*) *OR* **TOPIC:** (T21 or T18 or T13) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years*
# 1	621,942	**TOPIC:** (((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*))) *OR* **TOPIC:** ((NIPD or NIPT)) *OR* **TOPIC:** ((cfDNA or cffDNA or ccffDNA or "cell free DNA")) *OR* **TOPIC:** ((DNA near/3 (cell or free or cell?free or f? etal))) *OR* **TOPIC:** ((maternal near/3 (blood or plasma or DNA))) *OR* **TOPIC:** ((MPS or DANSR or "parental support" or MateriniT21 or Verifi* or Harmony or Panorama)) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years*

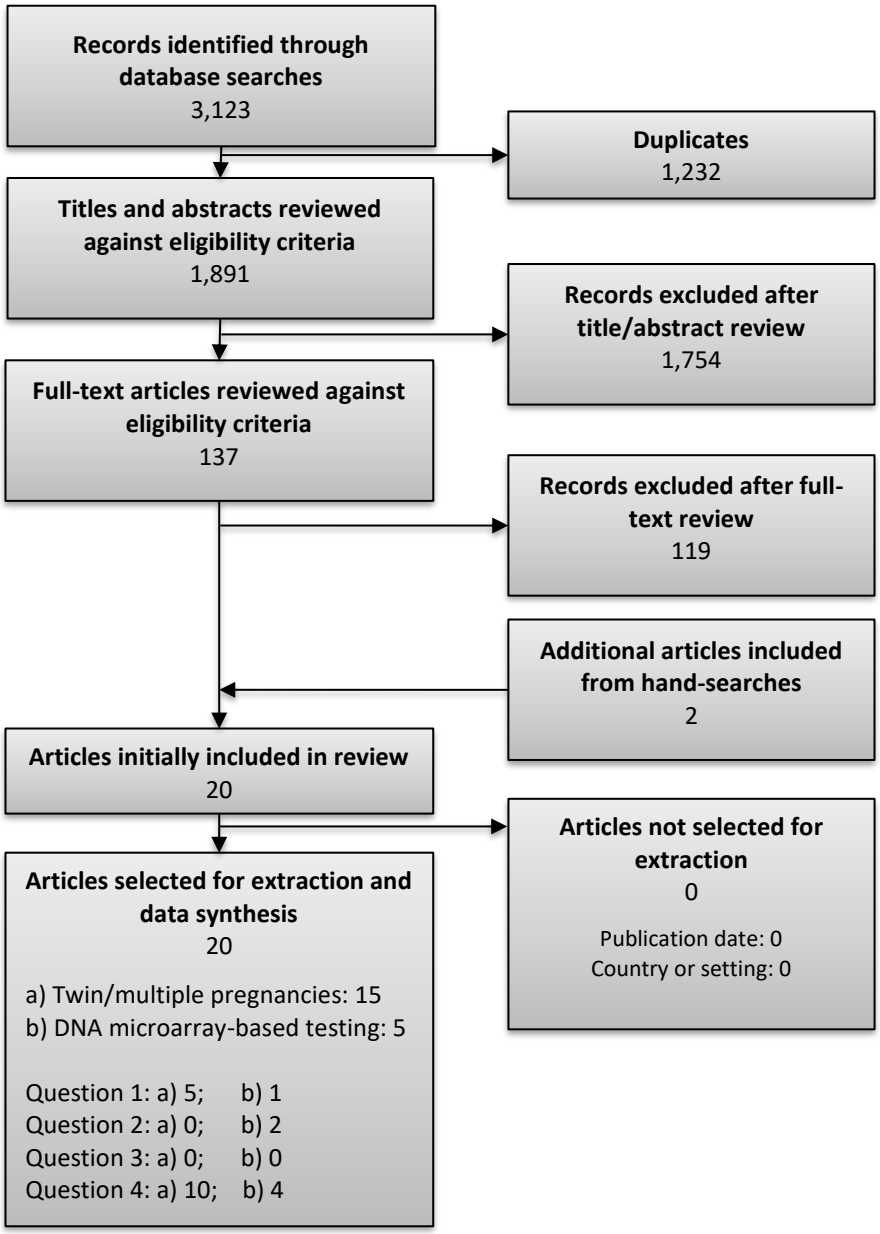
Results were imported into EndNote and de-duplicated.

Appendix 2 — Included and excluded studies

PRISMA flowchart

Figure 12 summarises the volume of publications included and excluded at each stage of the review. Twenty publications (15 on twin/multiple pregnancies and 5 on DNA microarray-based cfDNA testing) were ultimately judged to be relevant to one or more review questions and were considered for extraction. Publications that were included or excluded after the review of full-text articles are detailed below. Study characteristics and relevant findings of 7 excluded studies (published in 8 articles) using DNA microarray-based cfDNA testing in an unknown proportion of samples are described in Table 13.

Figure 12. Summary of publications included and excluded at each stage of the review



Publications included after review of full-text articles

The 20 publications included after review of full-texts are summarised in

Table 10 below.

Table 10. Summary of publications included after review of full-text articles, and the question(s) each publication was identified as being relevant to.

Study	The condition	The test	The intervention	The screening programme	Implementation criteria	Comments
Benachi 2015 [24]	-	a) Overall test accuracy 1a) 4a)	-	-	-	7 twin pregnancies
Beulen 2017 [25]	-	a) Overall test accuracy 1a) 4a)	-	-	-	21 twin pregnancies
Brison 2018 [26]	-	a) Overall test accuracy	-	-	-	1 TP T21, 1 FN T18, 1 TP T13 in twin pregnancies
Du 2017 [27]	-	a) Overall test accuracy 4a)	-	-	-	92 twin pregnancies
Fosler 2017 [28]	-	a) Overall test accuracy 1a), 4a)	-	-	-	Study B) 487 twin pregnancies
Gil 2017 [36]	-	b) Overall test accuracy 1b) 4b)	-	-	-	DNA microarray-based NIPT
Juneau 2014 [15]	-	b) Overall test accuracy	-	-	-	Direct comparison of DNA microarray- and sequencing-based NIPT
Kagan 2018 [50]	-	b) Overall test accuracy 2b) 4b)	-	-	-	DNA microarray-based NIPT

Study	The condition	The test	The intervention	The screening programme	Implementation criteria	Comments
Langlois 2017 [51]		b) Overall test accuracy 2b) 4b)				DNA microarray-based NIPT
Lee 2015 [22]	-	a) Overall test accuracy 1a)	-	-	-	2 twin pregnancies
Li 2015 [21]	-	a) Overall test accuracy 4a)	-	-	-	12 twin pregnancies, 1 set of triplets
Livergood 2017 [29]	-	4a)	-	-	-	72 multiple gestations
Meck 2015 [30]	-	a) Overall test accuracy	-	-	-	2 FP for T21 in twin pregnancies
Papageorghiou 2016 [35]	-	a) Overall test accuracy 1a) 4a)	-	-	-	11 twin pregnancies
Sarno 2016 [31]	-	a) Overall test accuracy 4a)	-	-	-	438 twin pregnancies
Stokowski 2015 [52]	-	b) Overall test accuracy 4b)	-	-	-	DNA microarray-based NIPT
Strom 2017 [32] (2 studies)	-	a) Overall test accuracy	-	-	-	Assay validation and clinical implementation
Tan 2016 [33]	-	a) Overall test accuracy 4a)	-	-	-	565 twin pregnancies
Valderramos 2016 [34]	-	a) Overall test accuracy	-	-	-	8 NIPT-positive twin pregnancies (TP and FN)

Study	The condition	The test	The intervention	The screening programme	Implementation criteria	Comments
Yang 2018 [23]	-	a) Overall test accuracy 4a)	-	-	-	432 twin pregnancies

FN, false negative; FP, false positive; NIPT, non-invasive prenatal testing (here: cfDNA testing); TN, true negative; TP, true positive.

Publications excluded after review of full-text articles

Of the 137 publications included after the review of titles and abstracts, 119 were ultimately judged not to be relevant to this review. These publications, along with reasons for exclusion, are listed in Table 11.

Table 11. Publications excluded after review of full-text articles with reason

a) Twin/multiple gestations; b) DNA microarray-based cfDNA testing.

Reference		Reason
1.	Barbu, M., et al. (2017). First Trimester Screening Options after the Introduction of NIPT - our Experience.	a & b) Exclude as full text not available.
2.	Bayindir, B., et al. (2015). "Noninvasive prenatal testing using a novel analysis pipeline to screen for all autosomal fetal aneuploidies improves pregnancy management." European Journal of Human Genetics 23(10): 1286-1293.	a) Exclude as unclear if multiple gestations were included; if yes, no separate outcome data reported. b) Exclude as sequencing not microarray for DNA quantification.
3.	Belloin, C., et al. (2016). "The noninvasive prenatal testing for Down's Syndrome. Retrospective study of 8821 patients." Journal de Gynecologie Obstetrique et Biologie de la Reproduction 45(9): 1127-1132.	a & b) Exclude as in French language.
4.	Benn, P. and F. R. Grati (2018). "Genome-wide non-invasive prenatal screening for all cytogenetically visible imbalances." Ultrasound in Obstetrics & Gynecology 51(4): 429-433	a & b) Exclude as editorial.
5.	Bestwick, J. P. and N. J. Wald (2016). "Antenatal reflex DNA screening for trisomy 18 and trisomy 13 in addition to Down's syndrome." Journal of Medical Screening 23(4): 171-174.	a & b) Exclude as simulation/modelling.
6.	Bevilacqua, E., et al. (2018). "Cell-Free DNA Analysis in Maternal Blood: Differences in Estimates between Laboratories with Different Methodologies Using a Propensity Score Approach." Fetal Diagnosis and Therapy: 1-10.	a) Exclude as only singleton pregnancies included. b) Exclude as MPSS or compared to Harmony Prenatal Test with both sequencing and microarray approach. See Table 13 for details.
7.	Bevilacqua, E., et al. (2017). "Screening for Sex Chromosome Aneuploidy by Cell-Free DNA Testing: Patient Choice and Performance." Fetal Diagnosis and Therapy. 23.	a) Exclude as unclear if multiple gestations were included, no separate test performance or test failures data reported. b) Exclude as Harmony Prenatal test was used, unclear if microarray technology, no test performance data for T21, T18 and T13.

8.	Bianchi, D. W., et al. (2015). "Noninvasive Prenatal Testing and Incidental Detection of Occult Maternal Malignancies.[Summary for patients in JAMA. 2015 Jul 14;314(2):198; PMID: 26172909]." JAMA 314(2): 162-169.	a) Exclude as unclear if multiple pregnancies were included; if yes, no separate outcome data reported. b) Exclude as MPSS was used, not microarray.
9.	Bianchi, D. W., et al. (2015). "Fetal sex chromosome testing by maternal plasma DNA sequencing: clinical laboratory experience and biology." Obstetrics & Gynecology 125(2): 375-382.	a) Exclude as unclear if multiple pregnancies were included; no separate test performance/failure data. b) Exclude as MPSS methodology was used, not microarray.
10.	Bjerregaard, L., et al. (2017). "The rate of invasive testing for trisomy 21 is reduced after implementation of NIPT." Danish Medical Journal 64(4).	a) Exclude as only singleton pregnancies included. b) Harmony Prenatal Test was used, but unclear if sequencing or microarray technology. Study period: 1 March 2013 to 1 February 2015. Agreed to exclude as study period overlaps November 2014. Ariosa/Roche confirmed that Harmony Prenatal Test with both sequencing and microarray approach was used for analysis. See Table 13 for details.
11.	Blackwell, S., et al. (2015). "#36: Prenatal aneuploidy screening using cell-free DNA." American Journal of Obstetrics and Gynecology 212(6): 711-716.	a & b) Exclude as no primary research article.
12.	Chitty, L. S., et al. (2016). "Uptake, outcomes, and costs of implementing non-invasive prenatal testing for Down's syndrome into NHS maternity care: prospective cohort study in eight diverse maternity units." BMJ 354: i3426.	a) Exclude as only singleton pregnancies included. b) Exclude as sequencing-based, not microarray-based NIPT.
13.	Chu, T., et al. (2017). "Comparative evaluation of the Minimally-Invasive Karyotyping (MINK) algorithm for non-invasive prenatal testing." PLoS ONE [Electronic Resource] 12(3): e0171882.	a) Exclude as unclear if multiple gestations were included; if yes, no separate outcome data reported. b) Exclude as MPSS not microarray.
14.	Cirigliano, V., et al. (2017). "Performance of the neoBona test: a new paired-end massively parallel shotgun sequencing approach for cell-free DNA-based aneuploidy screening." Ultrasound in Obstetrics & Gynecology 49(4): 460-464.	a) Exclude as only singleton pregnancies included. b) Exclude as MPSS-based NIPT, not microarray.
15.	Crea, F., et al. (2017). "The IONA Test: Development of an Automated Cell-Free DNA-Based Screening Test for Fetal Trisomies 13, 18, and 21 That Employs the Ion Proton Semiconductor Sequencing Platform." Fetal Diagnosis & Therapy 42(3): 218-224.	a) Exclude as unclear if multiple gestations were included (possibly not); if yes no separate data reported. b) Exclude as NGS on an Ion Proton sequencing platform, not microarray.
16.	Dahl, F., et al. (2018). "Imaging single DNA molecules for high precision NIPT." Scientific Reports 8.	a) Exclude as only singleton pregnancies included. b) Exclude as no DNA microarray was used: novel molecular probe technology.
17.	Dheedene, A., et al. (2016). "Implementation of non-invasive prenatal testing by semiconductor sequencing in a genetic laboratory." Prenatal Diagnosis 36(8): 699-707.	a) Exclude; 17 twin samples tested but only test accuracy data for 2 of them in supplemental table. b) Exclude as Ion Proton sequencing not microarray.

18.	Dobson, L. J., et al. (2016). "Patient choice and clinical outcomes following positive noninvasive prenatal screening for aneuploidy with cell-free DNA (cfDNA)." Prenatal Diagnosis 36(5): 456-462.	a) Exclude: 9 screen-positive twins reported but unable to calculate PPV. b) Exclude as "The commercial enterprise performing the cfDNA was at the discretion of the obstetrical provider and represented three different companies." Unclear if microarray, but no separate data available.
19.	Ehrich, M., et al. (2017). "Genome-wide cfDNA screening: Clinical laboratory experience with the first 10,000 cases." Genetics in Medicine 19(12): 1332-1337.	a) Exclude as unclear if multiple gestations were included; if yes so separate outcome data were reported. b) Exclude as whole genome sequencing, not microarray.
20.	Eiben, B., et al. (2015). "Single Nucleotide Polymorphism-Based Analysis of Cell-Free Fetal DNA in 3000 Cases from Germany and Austria." Ultrasound International Open 1(1): E8-E11.	a) Exclude as only singleton pregnancies were included. b) Exclude as SNP-based sequencing (Natera) not microarray.
21.	El Khattabi, L. A., et al. (2016). "Could Digital PCR Be an Alternative as a Non-Invasive Prenatal Test for Trisomy 21: A Proof of Concept Study." PLoS ONE [Electronic Resource] 11(5): e0155009.	a) Exclude as unclear if multiple pregnancies were included (possibly not); if so no separate outcome data were reported. b) Exclude as digital PCR not microarray.
22.	Ellison, C. K., et al. (2016). "Using Targeted Sequencing of Paralogous Sequences for Noninvasive Detection of Selected Fetal Aneuploidies." Clinical Chemistry 62(12): 1621-1629.	a) Exclude as unclear if multiple gestations were included; if so data are not reported separately. b) Exclude as sequencing not microarray.
23.	Fiorentino, F., et al. (2017). "The clinical utility of genome-wide non invasive prenatal screening." Prenatal Diagnosis 37(6): 593-601.	a) Exclude as only singleton pregnancies were included. b) Exclude as sequencing not microarray.
24.	Flock, A., et al. (2017). "Non-invasive prenatal testing (NIPT): Europe's first multicenter post-market clinical follow-up study validating the quality in clinical routine." Archives of Gynecology & Obstetrics 296(5): 923-928.	a) Exclude as only singleton pregnancies were included. b) Exclude as random massively parallel sequencing, not microarray.
25.	Gerundino, F., et al. (2017). "Validation of a method for noninvasive prenatal testing for fetal aneuploidies risk and considerations for its introduction in the Public Health System." Journal of Maternal-Fetal & Neonatal Medicine 30(6): 710-716.	a) Exclude as only singleton pregnancies included. b) Exclude as whole-genome MPS-based NIPT method not microarray.
26.	Gil, M. M., et al. (2016). "Clinical implementation of routine screening for fetal trisomies in the UK NHS: cell-free DNA test contingent on results from first-trimester combined test." Ultrasound in Obstetrics & Gynecology 47(1): 45-52.	a) Exclude as only singleton pregnancies were included. b) Harmony Prenatal Test was used; unclear if with sequencing or microarray technology. Study period: October 2013 and February 2015. Agreed to exclude as study period overlaps November 2014. See Table 13 for details.
27.	Gill, L. A. and T. L. Prosen (2017). "Indications for Invasive Prenatal Testing before and after Noninvasive Prenatal Screening." American Journal of Perinatology 34(11): 1084-1087.	a) Exclude as no information if multiple gestations were included and no separate test accuracy/failure data. b) Exclude as no test accuracy/failure data reported, no information on NIPT methodology used.

28.	Gomez-Manjon, I., et al. (2018). "Noninvasive Prenatal Testing: Comparison of Two Mappers and Influence in the Diagnostic Yield." BioMed Research International 2018 (no pagination)(9498140).	a) Exclude as only singleton pregnancies were included. b) Exclude as MPSS not microarray.
29.	Hartwig, T. S., et al. (2018). "Non-Invasive Prenatal Testing (NIPT) in pregnancies with trisomy 21, 18 and 13 performed in a public setting - factors of importance for correct interpretation of results." European Journal of Obstetrics Gynecology and Reproductive Biology 226: 35-39.	a) Exclude as unclear if multiple pregnancies included (possibly not); if so no separate outcome data reported. b) Exclude as MPS not microarray-based NIPT.
30.	Hu, H., et al. (2016). "Clinical Experience of Non-Invasive Prenatal Chromosomal Aneuploidy Testing in 190,277 Patient Samples." Current Molecular Medicine 16(8): 759-766.	a) Exclude as no separate test failure or test accuracy data for the 1,332 twins. b) Exclude as semiconductor sequencing was used, not microarray.
31.	Huang, S., et al. (2016). "Identifying Robertsonian Translocation Carriers by Microarray-Based DNA Analysis." Fetal Diagnosis & Therapy 40(1): 59-62.	a & b) Though it used microarray, exclude as case control study with only 7 pregnant cases; maternal Robertsonian translocation carriers, not T21/Robertsonian translocation in the fetus.
32.	Johansen, P., et al. (2016). "Open source non-invasive prenatal testing platform and its performance in a public health laboratory." Prenatal Diagnosis 36(6): 530-536.	a) Exclude as only singleton pregnancies were included. b) Exclude as whole-genome sequencing on the Ion Proton™ platform, not microarray.
33.	Jones, K. J., et al. (2018). "Targeted cell-free DNA analysis with microarray quantitation for assessment of fetal sex and sex chromosome aneuploidy risk." Ultrasound in Obstetrics & Gynecology 51(2): 275-+.	a &b) Exclude as letter.
34.	Kagan, K. O., et al. (2015). "Screening for chromosomal abnormalities by first trimester combined screening and noninvasive prenatal testing." Ultraschall in der Medizin 36(1): 40-46.	a & b) Exclude as modelling study, no NIPT performed, just assumed performance from the literature.
35.	Kane, S. C., et al. (2017). "Chorionic villus sampling in the cell-free DNA aneuploidy screening era: careful selection criteria can maximise the clinical utility of screening and invasive testing." Prenatal Diagnosis 37(4): 399-408.	a) No separate test performance/failure data for multiple gestations. b) Exclude as unclear if NIPT methodology involved microarray.
36.	Ke, W.-L., et al. (2015). "Detection of fetal cell-free DNA in maternal plasma for Down syndrome, Edward syndrome and Patau syndrome of high risk fetus." International journal of clinical and experimental medicine 8(6): 9525-9530.	a) Exclude as only singleton pregnancies were included. b) Exclude as sequencing not microarray.
37.	Kim, S., et al. (2016). "Comparison of two high-throughput semiconductor chip sequencing platforms in noninvasive prenatal testing for Down syndrome in early pregnancy." BMC Medical Genomics [Electronic Resource] 9(1): 22.	a) Exclude as it was unclear if multiple gestations were included; if so, no separate outcome data were reported. b) Exclude as semiconductor sequencing comparing Ion Torrent PGM and Proton platforms.

38.	Kornman, L., et al. (2017). "Non-Invasive Prenatal Testing for Sex Chromosome Aneuploidy in Routine Clinical Practice." Fetal Diagnosis and Therapy. 06.	a) Exclude as only singleton pregnancies were included. b) Harmony Prenatal Test was used with reference to Sparks (2012), so no microarray technology used. Study period: March 2013 and August 2014. Agreed to exclude as study period before November 2014.
39.	Kou, K. O., et al. (2016). "Effect of non-invasive prenatal testing as a contingent approach on the indications for invasive prenatal diagnosis and prenatal detection rate of Down's syndrome." Hong Kong Medical Journal 22(3): 223-230.	a) Exclude as only singleton pregnancies were included. b) Exclude as no mention of microarray, just "shotgun" and "targeted" DNA sequencing.
40.	Koumbaris, G., et al. (2016). "Cell-Free DNA Analysis of Targeted Genomic Regions in Maternal Plasma for Non-Invasive Prenatal Testing of Trisomy 21, Trisomy 18, Trisomy 13, and Fetal Sex." Clinical Chemistry 62(6): 848-855.	a) Exclude as only singleton pregnancies included. b) Exclude as targeted sequencing was used, not microarray.
41.	Krishna, I., et al. (2016). "Adverse perinatal outcomes are more frequent in pregnancies with a low fetal fraction result on noninvasive prenatal testing." Prenatal Diagnosis 36(3): 210-215.	a) Exclude as only singleton pregnancies included. b) Exclude as targeted sequencing or SNP sequencing was used; not microarray.
42.	Le Conte, G., et al. (2018). "Cell-free fetal DNA analysis in maternal plasma as a screening test for trisomy 21 in twin pregnancies." Gynecologie Obstetrique Fertilite et Senologie.	a & b) Exclude as article in French language.
43.	Lebo, R. V., et al. (2015). "Discordant circulating fetal DNA and subsequent cytogenetics reveal false negative, placental mosaic, and fetal mosaic cfDNA genotypes." Journal of Translational Medicine 13: 260.	a) Exclude as unclear if multiple pregnancies included; if so no separate data reported. b) Exclude as NIPT performed at (1) Sequenom testing with MaterniT21, or (2) Ariosa Diagnostics testing with Harmony at Integrated Genetics. Unclear if microarray technology.
44.	Lee, S. Y., et al. (2018). "A new approach of digital PCR system for non-invasive prenatal screening of trisomy 21." Clinica Chimica Acta 476: 75-80.	a) Exclude as only singleton pregnancies included. b) Exclude as NIPT using digital PCR not microarray for DNA quantification.
45.	Lee, S. Y., et al. (2015). "New application methods for chromosomal abnormalities screening test using digital PCR." Biochip Journal 9(4): 339-352.	a) Exclude as case control study with <15 T21 cases; unclear if any multiples included. b) Exclude as case control study with <15 T21 cases; digital PCR, not microarray, used for cfDNA quantification.

46.	Lee, T. J., et al. (2018). "Cell-free fetal DNA testing in singleton IVF conceptions." Human Reproduction 33(4): 572-578.	a) Exclude as only singleton pregnancies included. b) Harmony Prenatal Test was used with reference to Sparks (2012) method which is based on targeted sequencing not microarray. Study period: April 2013 and November 2016. Agreed to exclude as study period overlaps November 2014. Ariosa/Roche confirmed that Harmony Prenatal Test with both sequencing and microarray approach was used for analysis. See Table 13 for details.
47.	Lefkowitz, R. B., et al. (2016). "Clinical validation of a noninvasive prenatal test for genomewide detection of fetal copy number variants." American Journal of Obstetrics & Gynecology 215(2): 227.e221-227.e216.	a) Exclude as unclear if multiple gestations were included; if so, no separate data reported. b) Exclude as sequencing, not microarray-based NIPT.
48.	Li, B., et al. (2016). "Applicability of first-trimester combined screening for fetal trisomy 21 in a resource-limited setting in mainland China." BJOG: An International Journal of Obstetrics and Gynaecology 123(Supplement 3): 23-29.	a) Exclude as no separate test performance/failure data for multiple gestations. b) Exclude as cfDNA test methodology not reported.
49.	Li, R., et al. (2016). "Detection of fetal copy number variants by non-invasive prenatal testing for common aneuploidies." Ultrasound in Obstetrics & Gynecology 47(1): 53-57.	a) Exclude as only singleton pregnancies included. b) Exclude as semiconductor sequencing (MPSS), not microarray-based NIPT.
50.	Li, S. W., et al. (2015). "The assessment of combined first trimester screening in women of advanced maternal age in an Asian cohort." Singapore Medical Journal 56(1): 47-52.	a) Exclude as singleton pregnancies only and no NIPT performed. b) Exclude as no NIPT performed.
51.	Lo, K. K., et al. (2016). "Limited Clinical Utility of Non-invasive Prenatal Testing for Subchromosomal Abnormalities." American Journal of Human Genetics 98(1): 34-44.	a&b) Exclude as no testing for common trisomies T13, T18 or T21.
52.	Lu, R., et al. (2016). "Role of cell-free fetal DNA in the maternal plasma in the prenatal diagnosis of chromosomal abnormalities." International journal of clinical and experimental medicine 9(6): 11740-11747.	a) Exclude as only singleton pregnancies included. b) Exclude as NIPT method not described (possibly sequencing; performed by BGI Shenzhen Biotech Co., Ltd.).
53.	Mackie, F. L., et al. (2017). "Cell-free fetal DNA-based noninvasive prenatal testing of aneuploidy." Obstetrician & Gynaecologist 19(3): 211-218.	a & b) Exclude as non-systematic review.
54.	Manotaya, S., et al. (2016). "Clinical experience from Thailand: noninvasive prenatal testing as screening tests for trisomies 21, 18 and 13 in 4736 pregnancies." Prenatal Diagnosis 36(3): 224-231.	a) Exclude as no separate data for 167 twin pregnancies reported. b) Exclude as sequencing, no microarray-based NIPT.
55.	Martinez-Payo, C., et al. (2018). "Clinical results after the implementation of cell-free fetal DNA detection in maternal plasma." Journal of Obstetrics and Gynaecology Research.	a) Exclude as no test accuracy/failure data for multiple gestations reported. b) Exclude as NIPT method not described.

56.	McLennan, A., et al. (2016). "Noninvasive prenatal testing in routine clinical practice--an audit of NIPT and combined first-trimester screening in an unselected Australian population." Australian & New Zealand Journal of Obstetrics & Gynaecology 56(1): 22-28.	a) Exclude as only singleton pregnancies included. b) Harmony test was used, but unclear if sequencing or microarray-based. Study period: March 2013 and August 2014. Agreed to exclude as study period before November 2014.
57.	Meck, J. M., et al. (2015). "Noninvasive prenatal screening for aneuploidy: positive predictive values based on cytogenetic findings." American Journal of Obstetrics & Gynecology 213(2): 214.e211-215.	a) Exclude as no test accuracy/failure data for twins reported (only 2 FP mentioned). b) Exclude as no microarray-based NIPT (testing was performed by 4 different companies [Sequenom, Natera, Ariosa, Verinata]).
58.	Miltoft, C. B., et al. (2018). "Contingent first-trimester screening for aneuploidies with cell-free DNA in a Danish clinical setting." Ultrasound in Obstetrics & Gynecology 51(4): 470-479.	a) Exclude as only singleton pregnancies included. b) Agreed to exclude as Ariosa/Roche confirmed that the Harmony Prenatal Test with both sequencing and microarray approach was used for analysis. See Table 13 for details.
59.	Minarik, G., et al. (2015). "Utilization of Benchtop Next Generation Sequencing Platforms Ion Torrent PGM and MiSeq in Noninvasive Prenatal Testing for Chromosome 21 Trisomy and Testing of Impact of In Silico and Physical Size Selection on Its Analytical Performance." PLoS ONE [Electronic Resource] 10(12): e0144811.	a) Exclude as it is unclear if multiple gestations were included and if yes, no separate data are provided. b) Exclude as DNA sequencing, no microarray.
60.	Mnyani, C. N., et al. (2016). "The value and role of non-invasive prenatal testing in a select South African population." South African Medical Journal 106(10): 1047-1050.	a) Exclude as only singleton pregnancies included. b) Exclude as Natera Panorama test with SNP-sequencing was used.
61.	Neufeld-Kaiser, W. A., et al. (2015). "Positive predictive value of non-invasive prenatal screening for fetal chromosome disorders using cell-free DNA in maternal serum: independent clinical experience of a tertiary referral center." BMC Medicine 13: 129.	a) Exclude as unclear if multiple gestations were included; no separate data reported. b) Exclude as 92% of the NIPT tests were performed in one of the four major commercial laboratories offering testing during this timeframe. No information on methodology and no separate data for microarray-based methods.
62.	Neveling, K., et al. (2016). "Validation of two-channel sequencing-by-synthesis for noninvasive prenatal testing of fetal whole and partial chromosome aberrations." Prenatal Diagnosis 36(3): 216-223.	a) Exclude as it is unclear if multiple gestations were included and if yes, no separate data are provided. b) Exclude as DNA sequencing, no microarray.
63.	Norton, M. E., et al. (2016). "Cell-free DNA vs sequential screening for the detection of fetal chromosomal abnormalities." American Journal of Obstetrics & Gynecology 214(6): 727.e721-726.	a) Exclude as only singleton pregnancies and no NIPT performed. b) Exclude as no NIPT performed (just modelled performance), no microarray.

64.	Oepkes, D., et al. (2016). "Trial by Dutch laboratories for evaluation of non-invasive prenatal testing. Part I-clinical impact." Prenatal Diagnosis 36(12): 1083-1090.	a) Exclude as only singleton pregnancies included. b) Exclude as MPSS technology was used, not microarray.
65.	Palomaki, G. E., et al. (2015). "Evaluating first trimester maternal serum screening combinations for Down syndrome suitable for use with reflexive secondary screening via sequencing of cell free DNA: high detection with low rates of invasive procedures." Prenatal Diagnosis 35(8): 789-796.	a & b) Exclude as no NIPT performed.
66.	Palomaki, G. E., et al. (2015). "Circulating cell free DNA testing: are some test failures informative?" Prenatal Diagnosis 35(3): 289-293.	a) Exclude as only singleton pregnancies included. b) Exclude as DNA sequencing not microarray based NIPT.
67.	Palomaki, G. E., et al. (2017). "The clinical utility of DNA-based screening for fetal aneuploidy by primary obstetrical care providers in the general pregnancy population." Genetics in Medicine 19(7): 778-786.	a) Exclude as dizygotic twins were excluded. b) Exclude as SNP-based NIPT used (Natera), no microarray.
68.	Pantiukh, K. S., et al. (2016). "Report on noninvasive prenatal testing: Classical and alternative approaches [version 1; referees: 2 approved]." F1000Research 5 (no pagination)(722).	a) Exclude as unclear if multiple pregnancies were included; if so no separate outcome data reported. b) Exclude as whole-genome low coverage sequencing with GC correction, no microarray-based NIPT.
69.	Persico, N., et al. (2016). "Cell-free DNA testing in the maternal blood in high-risk pregnancies after first-trimester combined screening." Prenatal Diagnosis 36(3): 232-236.	a) Exclude as only singleton pregnancies included. b) Exclude as SNP sequencing was used (Natera), not microarray.
70.	Pertile, M. D., et al. (2017). "Rare autosomal trisomies, revealed by maternal plasma DNA sequencing, suggest increased risk of feto-placental disease." Science Translational Medicine 9(405): 30.	a) Exclude as it was unclear if multiple pregnancies were included; if yes, no separate data were reported. No test performance data for T21, T18 or T13 detection. b) Exclude as whole genome sequencing, not microarray.
71.	Pescia, G., et al. (2017). "Cell-free DNA testing of an extended range of chromosomal anomalies: clinical experience with 6,388 consecutive cases." Genetics in Medicine 19(2): 169-175.	a) Exclude as only singleton pregnancies included b) Exclude as analysis by shotgun sequencing on Illumina sequencers, not microarray-based NIPT.
72.	Petersen, A. K., et al. (2017). "Positive predictive value estimates for cell-free noninvasive prenatal screening from data of a large referral genetic diagnostic laboratory." American Journal of Obstetrics & Gynecology 217(6): 691.e691-691.e696.	a) Exclude as unclear if multiple gestations were included; no separate data reported. b) Exclude as NIPT performed by a variety of commercial laboratories including Ariosa Diagnostics, BGI, Natera, Sequenom, and Illumina, according to their specific methodologies. No separate data for microarray, if used.
73.	Poon, L. C., et al. (2016). "IONA test for first-trimester detection of trisomies 21, 18 and 13." Ultrasound in Obstetrics & Gynecology 47(2): 184-187.	a) Exclude as only singleton pregnancies included. b) Exclude as NIPT using Ion Proton™sequencing platform, not microarray-based.
74.	Qi, G., et al. (2016). "Noninvasive prenatal testing in routine clinical practice for a high-risk population: Experience from a center." Medicine 95(41): e5126.	a) Exclude as only singleton pregnancies included. b) Exclude as NIPT sequencing analysis, not microarray-based NIPT.

75.	Qian, Y. Q., et al. (2018). "Detection of fetal subchromosomal aberration with cell-free DNA screening led to diagnosis of parental translocation: Review of 11344 consecutive cases in a university hospital." European Journal of Medical Genetics.	a) Exclude as unclear if multiple pregnancies included; no separate outcome data reported. b) Exclude as NIPT sequencing analysis, not microarray-based NIPT
76.	Qiang, R., et al. (2017). "Detection of trisomies 13, 18 and 21 using non-invasive prenatal testing." Experimental and Therapeutic Medicine 13(5): 2304-2310.	a) Exclude as only singleton pregnancies included. b) Exclude as NIPT sequencing analysis, not microarray-based NIPT.
77.	Radoi, V. E., et al. (2015). "Cell free fetal DNA testing in maternal blood of Romanian pregnant women." Iranian Journal of Reproductive Medicine 13(10): 623-626.	a) Exclude as unclear if multiple gestations were included; no separate outcome data reported. b) Exclude as Panorama test (sequencing of SNPs), not microarray-based NIPT.
78.	Rao, R. R., et al. (2016). "The value of the first trimester ultrasound in the era of cell free DNA screening." Prenatal Diagnosis 36(13): 1192-1198.	a) Exclude as no separate data for multiple gestations reported. b) Exclude as NIPT was performed by Sequenom (Maternity 21), Verinata (Verify), Natera (Panomara), and Ariosa (Harmony); no separate data for microarray (if used).
79.	Revello, R., et al. (2016). "Screening for trisomies by cell-free DNA testing of maternal blood: consequences of a failed result." Ultrasound in Obstetrics & Gynecology 47(6): 698-704.	a) Exclude as only singleton pregnancies included. b) Exclude as Ariosa/Roche confirmed that the Harmony Prenatal Test with both sequencing and microarray approach was used for analysis. See Table 13 for details.
80.	Ryan, A., et al. (2016). "Validation of an Enhanced Version of a Single-Nucleotide Polymorphism-Based Noninvasive Prenatal Test for Detection of Fetal Aneuploidies." Fetal Diagnosis & Therapy 40(3): 219-223.	a) Exclude as only singleton pregnancies included. b) Exclude as sequencing of SNPs was used, not microarray-based NIPT.
81.	Saadati, N., et al. (2016). "Determining the role of mother race in neonatal outcome of trisomies 21, 18 and 13 using cell free DNA analysis." International Journal of Medical Research & Health Sciences 5(12): 365-369.	a) Exclude as only singleton pregnancies included. b) Exclude as no information on NIPT methodology.
82.	Samura, O., et al. (2017). "Current status of non-invasive prenatal testing in Japan." Journal of Obstetrics & Gynaecology Research 43(8): 1245-1255.	a) Exclude as no separate test performance/failure data for 73 twin pregnancies reported. b) Exclude as no separate test performance/failure data for microarray technology reported [94.7% of samples were sent to Sequenom and 5.3% were sent to four companies (Illumina; Ariosa Diagnostics; Labcorp; and Natera)].
83.	Santamaria, R., et al. (2018). "A National Referral Laboratory's Experience with the Implementation of SNP-Based Non-invasive Prenatal Screening for Fetal Aneuploidy and Select Microdeletion Syndromes." Journal of Fetal Medicine 5(1): 7-12.	a) Exclude as only singleton pregnancies included. b) Exclude as sequencing of SNPs, not microarray-based NIPT.
84.	Sbu (2015) Non-invasive prenatal test for Down's syndrome (Structured abstract). Health Technology Assessment Database	a & b) No full text available; possibly HTA report.

85.	Scott, F. P., et al. (2018). "Factors affecting cell-free DNA fetal fraction and the consequences for test accuracy." Journal of Maternal-Fetal & Neonatal Medicine 31(14): 1865-1872.	a) Exclude as only singleton pregnancies included. b) Harmony Prenatal test was used, but unclear if microarray was used for DNA quantification. Mention "sequencing bias" on page 1871, so probably targeted sequencing-based testing. Study period: March 2013 and August 2014. Agreed to exclude as study period before November 2014.
86.	Seyedoshohadaei, F., et al. (2017). "Evaluating the association between first trimester screening tests and adverse perinatal outcomes." Journal of Research in Medical and Dental Science 5(6): 14-19.	a) Exclude as only singleton pregnancies included. b) Exclude as NIPT methodology and test performance/failures not reported.
87.	Shani, H., et al. (2016). "Chromosomal abnormalities not currently detected by cell-free fetal DNA: a retrospective analysis at a single center." American Journal of Obstetrics & Gynecology 214(6): 729.e721-729.e711.	a & b) Exclude as no NIPT performed.
88.	Shi, W. L., et al. (2016). "Non-invasive prenatal testing (NIPT) detected chromosome aneuploidies and beyond in a clinical setting." International journal of clinical and experimental medicine 9(9): 18250-18254.	a) Exclude as only singleton pregnancies included. b) Exclude as sequencing-based NIPT, not microarray.
89.	Snyder, H. L., et al. (2016). "Follow-up of multiple aneuploidies and single monosomies detected by noninvasive prenatal testing: implications for management and counseling." Prenatal Diagnosis 36(3): 203-209.	a) Exclude as only singleton pregnancies included. b) Exclude as Verifi test with MPSS was used; no microarray.
90.	Srebniak, M. I., et al. (2017). "The influence of SNP-based chromosomal microarray and NIPT on the diagnostic yield in 10,000 fetuses with and without fetal ultrasound anomalies." Human Mutation 38(7): 880-888.	a) Exclude as no separate test performance or failure data reported for multiple gestations. b) Exclude as whole-genome sequencing was used, no microarray-based NIPT.
91.	Strah, D., et al. (2015). "Non-invasive prenatal cell-free fetal DNA testing for down syndrome and other chromosomal abnormalities." Zdravniski Vestnik 84(11): 727-733.	a) Exclude as no test accuracy/failure data can be calculated for the 3 twin pregnancies. b) Exclude as NIPT methodology not reported (Samples were analyzed at BGI Diagnostic Laboratories), most likely by sequencing-based NIPT.
92.	Sun, K., et al. (2017). "COFFEE: control-free noninvasive fetal chromosomal examination using maternal plasma DNA." Prenatal Diagnosis 37(4): 336-340.	a) Exclude as unclear if multiple pregnancies were included; no separate data reported. b) Exclude as sequencing-based NIPT, not microarray.
93.	Suo, F., et al. (2018). "Non-invasive prenatal testing in detecting sex chromosome aneuploidy: A large-scale study in Xuzhou area of China." Clinica Chimica Acta 481: 139-141.	a) Exclude as unclear if multiple pregnancies included; no separate data reported for multiples; no test accuracy data for trisomies 21, 18 or 13. b) Exclude as sequencing-based NIPT, not microarray.
94.	Suzumori, N., et al. (2016). "Fetal cell-free DNA fraction in maternal plasma is affected by fetal trisomy." Journal of Human Genetics 61(7): 647-652.	a) Exclude as only singleton pregnancies included. b) Exclude as MPS-based NIPT, not microarray.

95.	Taneja, P. A., et al. (2017). "Fetal aneuploidy screening with cell-free DNA in late gestation." Journal of Maternal-Fetal & Neonatal Medicine 30(3): 338-342.	a) Exclude as only singleton pregnancies included. b) Exclude as verifi Prenatal Test was used which analyses cfDNA using massively parallel next-generation whole-genome sequencing.
96.	Taneja, P. A., et al. (2016). "Noninvasive prenatal testing in the general obstetric population: clinical performance and counseling considerations in over 85000 cases." Prenatal Diagnosis 36(3): 237-243.	a) Exclude as only singleton pregnancies included. b) Exclude as verifi Prenatal Test was used which analyses cfDNA using massively parallel next-generation whole-genome sequencing.
97.	Tynan, J. A., et al. (2016). "Application of risk score analysis to low-coverage whole genome sequencing data for the noninvasive detection of trisomy 21, trisomy 18, and trisomy 13." Prenatal Diagnosis 36(1): 56-62.	a) Exclude as only singleton pregnancies included. b) Exclude as whole genome MPS based assay (VisibiliT™), not microarray.
98.	Van Opstal, D., et al. (2018). "Origin and clinical relevance of chromosomal aberrations other than the common trisomies detected by genome-wide NIPS: Results of the TRIDENT study." Genetics in Medicine 20(5): 480-485.	a) Exclude as only singleton pregnancies included (see Oepkes 2016). b) Exclude as MPSS technology was used, not microarray.
99.	Verma, I. C., et al. (2018). "Single Nucleotide Polymorphism-Based Noninvasive Prenatal Testing: Experience in India." Journal of Obstetrics and Gynecology of India: 1-9.	a) Exclude as only singleton pregnancies included. b) Exclude as SNP-based (Natera Inc) methodology used, not microarray.
100.	Vicic, A., et al. (2017). "Prenatal diagnosis of Down syndrome: A 13-year retrospective study." Taiwanese Journal of Obstetrics and Gynecology 56(6): 731-735.	a & b) No NIPT performance or test failure data reported.
101.	Wang, L., et al. (2015). "Maternal mosaicism of sex chromosome causes discordant sex chromosomal aneuploidies associated with noninvasive prenatal testing." Taiwanese Journal of Obstetrics & Gynecology 54(5): 527-531.	a) Exclude as unclear if multiple pregnancies were included; if yes there are no separate test performance / failure data. b) Exclude as whole genome sequencing was used.
102.	Wang, Y. J., et al. (2017). "PLAC4 mRNA SNP in non-invasive prenatal testing of Down syndrome." International Journal of Clinical and Experimental Pathology 10(7): 7962-7967.	a) Exclude as only singleton pregnancies included. b) Exclude as PLAC4 mRNA was measured using quantitative reverse transcription-PCR; not cfDNA, not microarray.
103.	Wax, J. R., et al. (2015). "Noninvasive prenatal testing: impact on genetic counseling, invasive prenatal diagnosis, and trisomy 21 detection." Journal of Clinical Ultrasound 43(1): 1-6.	a) Exclude as only singleton pregnancies included. b) Exclude as NIPT by MPSS was used; no microarray.
104.	Williams, J., 3rd, et al. (2015). "Utilization of noninvasive prenatal testing: impact on referrals for diagnostic testing." American Journal of Obstetrics & Gynecology 213(1): 102.e101-106.	a) Exclude as only singleton pregnancies included. b) Exclude as no information on NIPT methodology and no test performance/test failure data reported.
105.	Xi, Y., et al. (2017). "Noninvasive Prenatal Detection of Trisomy 21 by Targeted Semiconductor Sequencing: A Technical Feasibility Study." Fetal Diagnosis and Therapy 42(4): 302-310.	a) Exclude as unclear if multiple gestations included; no separate data. b) Exclude as targeted semiconductor sequencing, not microarray-based NIPT.

106.	Xie, M. J., et al. (2018). "Noninvasive Prenatal Testing of Rare Autosomal Aneuploidies by Semiconductor Sequencing." DNA & Cell Biology 37(3): 174-181.	a) Exclude as unclear if multiple gestations included; no separate data. b) Exclude as semiconductor sequencing, not microarray-based NIPT.
107.	Xu, C., et al. (2017). "Noninvasive Prenatal Screening of Fetal Aneuploidy without Massively Parallel Sequencing." Clinical Chemistry 63(4): 861-869.	a) Exclude as only singleton pregnancies included. b) Exclude as high-throughput ligation-dependent probe amplification (HLPA) assay (multiplex PCR), not microarray-based NIPT.
108.	Xu, X. P., et al. (2016). "A Method to Quantify Cell-Free Fetal DNA Fraction in Maternal Plasma Using Next Generation Sequencing: Its Application in Non-Invasive Prenatal Chromosomal Aneuploidy Detection." PLoS ONE [Electronic Resource] 11(1): e0146997.	a) Exclude as no separate data for multiple gestations reported. b) Exclude as NIPT on Ion Proton, a semiconductor sequencing platform, was used; not microarray.
109.	Yamada, T., et al. (2018). "Maternal age-specific risk for trisomy 21 based on the clinical performance of NIPT and empirically derived NIPT age-specific positive and negative predictive values in Japan." Journal of Human Genetics: 1-5.	a) Exclude as unclear if multiple pregnancies included; no separate data reported. b) Exclude as NIPT based on massively parallel sequencing: MateriT21 Plus® and GeneTech NIP; no microarray.
110.	Yang, S. F., et al. (2018). "Diagnostic differences between patients opting for non-invasive prenatal testing and patients having traditional prenatal diagnosis." International Journal of Clinical and Experimental Pathology 11(5): 2831-2838.	a) Exclude as unclear if multiple pregnancies included; no separate data. b) Exclude as NIPT method not described.
111.	Yared, E., et al. (2016). "Obesity increases the risk of failure of noninvasive prenatal screening regardless of gestational age." American Journal of Obstetrics & Gynecology 215(3): 370.e371-376.	a) Exclude as unclear if multiple pregnancies included; no separate data reported. b) Exclude as NIPT based on sequencing of SNPs (Panorama test) was used, not microarray.
112.	Yaron, Y., et al. (2016). "Current controversies in prenatal diagnosis 2: for those women screened by NIPT using cell free DNA, maternal serum markers are obsolete." Prenatal Diagnosis 36(13): 1167-1171.	a & b) Exclude as no primary research article.
113.	Yaron, Y., et al. (2017). "Current controversies in prenatal diagnosis 2: For those women screened by NIPT using cell-free DNA, maternal serum markers are obsolete." Obstetrical and Gynecological Survey 72(4): 216-217.	a & b) Exclude as note.
114.	Yu, B., et al. (2018). "Clinical evaluation of NIPS for women at advanced maternal age: a multicenter retrospective study." Journal of Maternal-Fetal and Neonatal Medicine: 1-6.	a) Exclude as unclear if multiple pregnancies included; no separate data reported. b) Exclude as NIPT based on sequencing, not microarray.
115.	Yu, B., et al. (2017). "Overall evaluation of the clinical value of prenatal screening for fetal-free DNA in maternal blood." Medicine 96(27): e7114.	a) Exclude as no separate test accuracy or failure data for the 119 included twin pregnancies. b) Exclude as NIPT based on sequencing, not microarray.
116.	Zhang, H., et al. (2015). "Statistical Approach to Decreasing the Error Rate of Noninvasive Prenatal Aneuploid Detection caused by Maternal Copy Number Variation." Scientific Reports 5: 16106.	a) Exclude as unclear if multiple pregnancies included; no separate data reported. b) Exclude as shotgun MPS-based NIPT, not microarray.

117.	Zhang, J. and B. Zhang (2016). "Second-generation non-invasive high-throughput DNA sequencing technology in the screening of down's syndrome in advanced maternal age women." Biomedical Reports 4(6): 715-718.	a) Exclude as only singleton pregnancies were included. b) Exclude as sequencing, not microarray technology.
118.	Zhang, L., et al. (2017). "Count-based size-correction analysis of maternal plasma DNA for improved noninvasive prenatal detection of fetal trisomies 13, 18, and 21." American Journal of Translational Research 9(7): 3469-3473.	a) Exclude as unclear if multiple gestations included; no separate data reported. b) Exclude as massively parallel DNA sequencing with an Ion Proton™ Sequencer, not microarray-based NIPT.
119.	Zhou, X., et al. (2017). "Contribution of maternal copy number variations to false-positive fetal trisomies detected by noninvasive prenatal testing." Prenatal Diagnosis 37(4): 318-322.	a) Exclude as unclear if multiple pregnancies included; no separate data. b) Exclude as massively parallel sequencing-based NIPT, not microarray.

Appendix 3 — Summary and appraisal of individual studies

Data Extraction

Table 12. Studies relevant to criterion 4 (The test)

a) Twin/multiple pregnancies (15 articles reporting on 16 studies)

Full citation	**Benachi, A., et al. (2015). "Cell-free DNA analysis in maternal plasma in cases of fetal abnormalities detected on ultrasound examination." Obstetrics & Gynecology 125(6): 1330-1337.**
Key questions	1a) Test accuracy of cfDNA testing in twin pregnancies at higher chance of fetal trisomies. 4a) Test failure rate of cfDNA testing in twin pregnancies.
Study characteristics	
Study design / Setting / Study period	Multicenter, prospective study (SEHDA study). 29 French fetal medicine centres. December 2012 to October 2013.
Population	Pregnant women, with or without fetal ultrasound findings (structural or “soft markers” whenever fetal karyotyping was thought necessary), who were considered at higher chance for fetal aneuploidies based on maternal age alone (older than 38 years, standard French maternal age cutoff), maternal serum screening (first-trimester combined test or second trimester), or a history of pregnancy with trisomy and who were willing to undergo invasive procedures, at least 18 years old, more than 10 weeks of gestation and singleton or twin pregnancy. N=900 included in the study. Exclusion from the study NR. 35.1% first trimester. 57% higher chance but no ultrasound anomalies, 43% very high chance with ultrasound anomalies. Chorionicity for 7 twin pregnancies NR.

Full citation	**Benachi, A., et al. (2015). "Cell-free DNA analysis in maternal plasma in cases of fetal abnormalities detected on ultrasound examination." Obstetrics & Gynecology 125(6): 1330-1337.**
	N=886 included in analysis (7 twin pregnancies and 879 singleton pregnancies). 14 women excluded from the analysis (8 without karyotype results, 6 with nonreportable cfDNA assay).
Index test / Comparator / Reference standard	Index test: Analysis of stored frozen samples. 12-plex random whole-genome sequencing on HiSeg1500 (Illumina). The classification was based on a standard normal transformed cutoff value of Z = 3 for chromosome 21 and Z = 3.95 for chromosomes 18 and 13. Comparator: None. Reference standard: Invasive testing.
Outcomes	1a) Test accuracy. 4a) Test failure rate.
Funding source or sponsor of the study	Funding source not reported. Jean-Marc Costa is an employee of CERBA, in which he is also a shareholder. The other authors did not report any potential conflicts of interest.
Information about the authors contacted	No contact needed.
Information about other contacts	No contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: Unclear Applicability concerns: High	
DOMAIN II: Index test Risk of bias: Low Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Low Applicability concerns: Low	
DOMAIN IV: Flow & timing	

Full citation	**Benachi, A., et al. (2015). "Cell-free DNA analysis in maternal plasma in cases of fetal abnormalities detected on ultrasound examination." Obstetrics & Gynecology 125(6): 1330-1337.**
Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: Unclear	

Full citation	**Beulen, L., et al. (2017). "Clinical utility of non-invasive prenatal testing in pregnancies with ultrasound anomalies." Ultrasound in Obstetrics & Gynecology 49(6): 721-728.**
Key questions	1a) Test accuracy of cfDNA testing in twin pregnancies at higher chance of fetal trisomies. 4a) Test failure rate of cfDNA testing in twin pregnancies.
Study characteristics	
Study design / Setting / Study period	Cohort study, retrospective analysis. 5 prenatal diagnostic centres linked to the Network for Prenatal Diagnosis Nijmegen (Netherlands). April 2014 and November 2015.
Population	All pregnant women at higher chance for fetal chromosomal abnormality based on findings at sonographic examination, who underwent NIPT as an alternative to fetal genotyping by QF-PCR and microarray analysis at prenatal diagnostic centres linked to the Network for Prenatal Diagnosis Nijmegen, between April 2014 and November 2015. N=251 included in the study (21 twin pregnancies, 230 singleton pregnancies). 339 women who did not choose cfDNA were excluded from the study. Median gestational age 20 weeks (range 10-34 weeks). All higher chance of fetal chromosomal abnormalities based on findings at sonographic examination. 17/21 (81%) dichorionic, 4/21 (19%) monochorionic. N=232 included in analysis (17 twin pregnancies, 215 singleton pregnancies). 4 twin pregnancies without reference standard excluded from the analysis. 14 singleton pregnancies without reference standard and 1 singleton pregnancy with cfDNA test failure excluded from the analysis.
Index test / Comparator / Reference standard	Index test: All samples were processed and analysed at the Department of Human Genetics of the Radbound University Medical Center in Nijmegen. Random whole-genome sequencing on SOLiD 5500 XL Genetic Analyzer (Life Technologies, Foster City, CA, USA) or on a NextSeq 500 desktop sequencer (Illumina, San Diego, CA, USA). Threshold NR. Comparator: None. Reference standard: Prenatal or postnatal diagnostic testing, newborn examination.
Outcomes	1a) Test accuracy in singleton and twin pregnancies at higher chance of fetal trisomies.

Full citation	**Beulen, L., et al. (2017). "Clinical utility of non-invasive prenatal testing in pregnancies with ultrasound anomalies." Ultrasound in Obstetrics & Gynecology 49(6): 721-728.**
	4a) Test failure rate in singleton and twin pregnancies.
Funding source or sponsor of the study	This study was supported financially by the Foundation for Prenatal Screening in the Nijmegen Region. Article does not state role of sponsor or conflicts of interest.
Information about the authors contacted	No contact needed.
Information about other contacts	No contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: Low Applicability concerns: High	
DOMAIN II: Index test Risk of bias: Low Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Low Applicability concerns: Low	
DOMAIN IV: Flow & timing Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: Unclear	

Full citation	**Brison, N., et al. (2018). "Predicting fetoplacental chromosomal mosaicism during non-invasive prenatal testing." Prenatal Diagnosis 38(4): 258-266.**
Key questions	a) Test accuracy in twin pregnancies.
Study characteristics	
Study design / Setting / Study period	Cohort study, consecutive non-selected series of 19,735 pregnant women. Retrospective analysis of the NIPT sequencing profiles of the study cohort using the novel analysis pipeline. 1 University-based lab (Centre for Human Genetics, KU Leuven, Leuven) in Belgium. Study period NR.
Population	Consecutive non-selected series of 19,735 pregnant women from 10 weeks' gestation onwards tested for common fetal aneuploidies by cfDNA testing. No exclusions from the study. Mean gestational age 13.2 (SD 2.32) weeks. 24% advanced maternal age (> 36 years), 12% higher chance of foetal trisomy as indicated by the combined test, 2% family history of congenital or hereditary diseases, 62% maternal comfort. 131 with predicted full trisomy (1 twin pregnancy) and 34 with predicted mosaic chromosomal trisomy (2 twin pregnancies) included in the analysis. Chorionicity reported for 1 false negative T18 case (dichorionic).
Index test / Comparator / Reference standard	Index test: cfDNA testing based on random whole-genome sequencing performed at the Centre for Human Genetics, KU Leuven, Leuven, Belgium. Threshold: Z > 3: trisomy call. Full, non-mosaic chromosomal trisomies (100% of the placenta) are predicted when Z > 3 and TriZ ≥ −3 are labelled as TriZ-high/normal. Cases with a Z > 3 and TriZ < −3 are labelled as TriZ-low and are predicted to be mosaic chromosomal trisomies. Comparator: None. Reference standard: When cfDNA profiling indicated the presence of a chromosomal abnormality, the women were offered follow-up by

Full citation	**Brison, N., et al. (2018). "Predicting fetoplacental chromosomal mosaicism during non-invasive prenatal testing." Prenatal Diagnosis 38(4): 258-266.**
	standard invasive prenatal diagnosis based on DNA extracted from chorionic villus sampling or amniotic fluid or postnatal genetic testing. Subsequently, when a discrepancy between the NIPT and the invasive genetic test result was detected, women were proposed to donate the placenta upon delivery.
Outcomes	a) Test accuracy for full trisomies and fetoplacental mosaicism (follow-up of positive results only).
Funding source or sponsor of the study	This work was made possible by grants from the University of Leuven (KU Leuven): GOA (GOA/12/015 to J.R.V., K.D., and H.V.E.) and SymBio-Sys (PFV/10/016 to J.R.V.). Prof Vermeesch's laboratory receives license fees from Agilent for the analysis pipeline used in this study. There are no other conflicts of interest.
Information about the authors contacted	No contact needed.
Information about other contacts	No contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: Low Applicability concerns: High	
DOMAIN II: Index test Risk of bias: High Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Low Applicability concerns: Low	
DOMAIN IV: Flow & timing Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: Low	

Full citation	**Du, E., et al. (2017). "Massively Parallel Sequencing (MPS) of Cell-Free Fetal DNA (cffDNA) for Trisomies 21, 18, and 13 in Twin Pregnancies." Twin Research & Human Genetics: the Official Journal of the International Society for Twin Studies 20(3): 242-249.**
Key questions	a) Overall test accuracy of cfDNA testing in twin pregnancies (mixed chances for fetal trisomies or general obstetric population). 4a) Test failure rate of cfDNA testing in twin pregnancies.
Study characteristics	
Study design / Setting / Study period	Prospective cohort study. NR (possibly 1 hospital in China: Zhongnan Hospital of Wuhan University, Wuhan, China). 1st January 2013 to 1st October 2016.
Population	Twin pregnancies that required invasive prenatal diagnosis using amniocentesis, or with clinical examination and follow-up of the neonates. Women with intrauterine fetal demise at the time of sampling or without fetal karyotype results or clinical examination and follow-up were excluded from this study. N=92 included in the study. Number of exclusions from the study NR. 100% second trimester. Women had prior ultrasound and blood test but that was not used to calculate a combined test chance result. Possibly general chanceof fetal trisomies. 39/92 (42%) monochorionic and 53/92 (58%) dichorionic. N=92 included in analysis.
Index test / Comparator / Reference standard	Index test: cfDNA tesing based on random whole-genome sequencing performed in China according to a previously reported workflow [61]. Threshold: A-value > 3 and T-value > 3: high-risk zone → sample considered affected. If either A-value > 3 or T value > 3: warning zone. If in Warning Zone 1: sample considered affected by mosaicism or partial trisomy. Such cases were reported as high risk but were accompanied by appropriate comments. Samples in Warning zone 3 were likely affected by inadequate fetal DNA concentrations, If clinically permitted, blood sampling and sequencing were repeated. Otherwise, a high risk result was issued.

Full citation	**Du, E., et al. (2017). "Massively Parallel Sequencing (MPS) of Cell-Free Fetal DNA (cffDNA) for Trisomies 21, 18, and 13 in Twin Pregnancies." Twin Research & Human Genetics: the Official Journal of the International Society for Twin Studies 20(3): 242-249.**
	If A-value ≤ 3 and T-value ≤ 3: low-risk zone.
Outcomes	a) Test accuracy. 4a) Test failure rate.
Funding source or sponsor of the study	This work was supported by Department of Obstetrics and Gynecology, Zhongnan Hospital of Wuhan University, Department of Obstetrics and Gynecology, Renmin Hospital, Hubei University ofMedical, Shiyan, China. This work was also supported by National Natural Science Foundation of China (81370707) to W.-H. Z, Science and technology support. The authors declare that they do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.
Information about the authors contacted	Contacted corresponding author by email to clarify reference standard for 2 cases. No reply.
Information about other contacts	No other contacts needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: High Applicability concerns: High	
DOMAIN II: Index test Risk of bias: Low Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Low Applicability concerns: Low	
DOMAIN IV: Flow & timing Risk of bias: Unclear	
DOMAIN V: Role of sponsor Risk of bias: Low	

Full citation	**Fosler, L., et al. (2017). "Aneuploidy screening by non-invasive prenatal testing in twin pregnancy." Ultrasound in Obstetrics & Gynecology 49(4): 470-477.**
Key questions	1a) Test accuracy of cfDNA testing in twin pregnancies (mostly) at higher chance of fetal trisomies. 4a) Test failure rate of cfDNA testing in twin pregnancies.
Study characteristics	
Study design / Setting / Study period	Clinical study B: Prospectively collected samples for which outcomes were requested from providers. Laboratory experience report with verifi prenatal test. College of American Pathologists-accredited and Clinical Laboratory Improvement Act-certified Illumina Laboratory. Providers in the USA requesting the commercially available verify ® Prenatal Test (Illumina, Inc, San Diego, CA). Study period NR.
Population	Maternal blood samples received during the study period at the Illumina lab, indicated as twin gestation on the test requisition forms from providers in the USA. Samples received from distributor laboratories and/or heath systems located in the USA were excluded due to the inability to obtain clinical follow-up. N=487 included in study. Number of exclusions from the study NR. 68.4% first trimester. >90% higher chance of fetal trisomies due to advanced maternal age (63%), abnormal ultrasound findings (17%), previous affected pregnancy (5%), positive serum screen (3%), multiple indications (7.5%). Chorionicity information not available to the laboratory for all samples. N=169 included in analysis. 318 excluded from the analysis (8 tests were cancelled, 308 without karyotype outcome, 2 with only ultrasound findings available).
Index test / Comparator / Reference standard	Index test: Verifi prenatal test (Illumina, Inc, San Diego, CA) performed at College of American Pathologists-accredited and Clinical Laboratory Improvement Act-certified Illumina Laboratory. Random whole-genome sequencing performed on IlluminaHiSeq 2000 sequencers (Illumina, Inc.). Report NIPT result as “aneuploidy detected”, “no aneuploidy detected” or “aneuploidy suspected” for each trisomy as described previously [62]. Threshold from Bianchi 2012 [62]: NCV > 4.0 aneuploid NCV < 2.5 euploid

Full citation	**Fosler, L., et al. (2017). "Aneuploidy screening by non-invasive prenatal testing in twin pregnancy." Ultrasound in Obstetrics & Gynecology 49(4): 470-477.**
	2.5 ≤ NCV ≤ 4.0 unclassified. Comparator: None. Reference standard: Invasive diagnostic procedure, newborn testing/physical examination or ultrasound evaluation (2 samples). Data on karyotype outcome collected retrospectively after the estimated delivery date had passed for all samples. From Futch et al. (2013) [(63)]: Data collected via phone call to the ordering provider.
Outcomes	1a) Test accuracy. 4a) Test failure rate.
Funding source or sponsor of the study	This study was funded by Illumina. L. Fosler, P. Winters, K. W. Jones, K. J. Curnow, A. J. Sehnert and S. Bhatt are, or were, employees of, and hold equity in, Illumina. L. D. Platt is a paid consultant for Illumina.
Information about the authors contacted	Contacted L. Fosler by email to clarify the number of cases with reference standard. No reply received.
Information about other contacts	No other contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: High Applicability concerns: High	
DOMAIN II: Index test Risk of bias: Unclear Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Unclear Applicability concerns: Unclear	
DOMAIN IV: Flow & timing Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: High	

Full citation	**Lee, M.-Y., et al. (2015). "Performance of Momguard, a new non-invasive prenatal testing protocol developed in Korea." Obstetrics & Gynecology Science 58(5): 340-345.**
Key questions	1a) Test accuracy of cfDNA testing in multiple pregnancies at higher chance of fetal trisomies.
Study characteristics	
Study design / Setting / Study period	Preliminary study as part of a large prospective cohort study. Asan Medical Center, Seoul, Korea. August 2014 and February 2015.
Population	Pregnant women who were >18 years old, gestational age >8 weeks, and who met at least one of the following additional criteria: advanced maternal age (≥35 years), a positive serum biochemical screening test, the presence of fetal anomalies detected by ultrasound, or a personal/family history of fetal aneuploidy. To evaluate the performance of Momguard in twin pregnancies, multiple gestations were also included. For this preliminary study, part of the clinical data collected between August 2014 and February 2015 was analysed. This analysis included all pregnant women who underwent either CVS, amniocentesis, or cordocentesis for confirming fetal karyotype. Fetuses with karyotypes that were confirmed using peripheral blood after birth or conceptual tissues in cases of missed abortion were also included. N=92 included in study (1 case with cfDNA test failure excluded from the study; other exclusions NR). 8.7% first trimester. 29% advanced maternal age, 26% positive serum screening, 80% presence of ultrasonic markers, 2% personal/family history of aneuploidies. Chorionicity for 2 multiple gestations NR. N=92 included in the analysis (2 multiple pregnancies and 90 singleton pregnancies).
Index test / Comparator / Reference standard	Index test: All cfDNA analyses were performed in the LabGenomics Clinical Laboratory (Seongnam, Korea), where a multi-platform NGS-based noninvasive test was implemented for fetal aneuploidy screening (Momguard). Analysis of stored (frozen) samples. Random whole-genome sequencing; up to 12 samples pooled for sequencing on MiSeq (Illumina, San Diego,CA, USA) or 96 samples pooled for sequencing on NextSeq (Illumina). For T21 and T18: Z-score > 4 high risk of aneuploidy. Z-scores between 2.5 and 4 intermediate risk for T21 and T18.

Full citation	**Lee, M.-Y., et al. (2015). "Performance of Momguard, a new non-invasive prenatal testing protocol developed in Korea." Obstetrics & Gynecology Science 58(5): 340-345.**
	For T13: Z-score > 2.8 high risk, Z-scores between 1.9 and 2.8 intermediate risk. Comparator: None. Reference standard: CVS, amniocentesis, or cordocentesis for confirming fetal karyotype. Karyotypes that were confirmed using peripheral blood after birth or conceptual tissues in cases of missed abortion.
Outcomes	a) Test accuracy
Funding source or sponsor of the study	This study was supported by a grant from the LabGenomics Clinical Research Institute. No potential conflict of interest relevant to this article was reported. 3/7 authors are affiliated to LabGenomic Clinical Research Institute, LabGenomics (Seongnam, Korea), the provider of the Momguard cfDNA test.
Information about the authors contacted	No contact needed.
Information about other contacts	No contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: Unclear Applicability concerns: High	
DOMAIN II: Index test Risk of bias: Unclear Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Low Applicability concerns: Low	
DOMAIN IV: Flow & timing Risk of bias: Low	
DOMAIN V: Role of sponsor Risk of bias: High	

Full citation	**Li, W. H., et al. (2015). "Noninvasive prenatal testing for fetal trisomy in a mixed risk factors pregnancy population." Taiwanese Journal of Obstetrics & Gynecology 54(2): 122-125.**
Key questions	a) Overall test accuracy in multiple pregnancies (mixed risk factors). 4a) Test failure rate in multiple pregnancies.
Study characteristics	
Study design / Setting / Study period	Prospective cohort study. 1 tertiary medical centre in Taiwan. July 2012 to June 2014
Population	Pregnant women undergoing prenatal aneuploidy screening in a single tertiary medical centre was conducted. Indications included maternal anxiety, advanced maternal age, abnormal nuchal translucency, and high/moderate chance result of first trimester Down's syndrome screening. Multifetal pregnancies and patients receiving in vitro fertilization were also enrolled for analysis. N=169 included in study (12 twin pregnancies, 1 triplet pregnancy and 156 singleton pregnancies). Number of exclusions from the study NR. Median gestational age 13.45 weeks (7-31 weeks). "Mixed risk factors pregnancy population": 4% advanced maternal age (AMA, ≥ 34 years) and higher-chance result at maternal serum screening, 4% AMA and intermediate-chance result at maternal serum screening, 63% AMA only, 6% higher-chance result at maternal serum screening and <34 years, 6% intermediate-chance result at maternal serum screening and <34 years, 16% purely anxiety. 2.4% history or family history of aneuploidies. Chorionicity for 13 multiple gestations NR. N=169 included in analysis (12 twin pregnancies, 1 triplet pregnancy and 156 singleton pregnancies).
Index test / Comparator / Reference standard	Index test: NR Comparator: None. Reference standard: Amniocentesis in 2 women with NIPT positive results. NR for screen-negative women ("confirmed correct after birth").
Outcomes	a) Test accuracy. 4a) Test failure rate.

Full citation	**Li, W. H., et al. (2015). "Noninvasive prenatal testing for fetal trisomy in a mixed risk factors pregnancy population." Taiwanese Journal of Obstetrics & Gynecology 54(2): 122-125.**
Funding source or sponsor of the study	Funding source not reported. All the authors declare no conflict of interest.
Information about the authors contacted	Contacted the corresponding author via email to clarify patient selection, patient applicability, the cfDNA test methodology, reference standard in screen-negative cases and funding. No reply received.
Information about other contacts	No contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: Unclear Applicability concerns: Unclear	
DOMAIN II: Index test Risk of bias: Low Applicability concerns: Unclear	
DOMAIN III: Reference standard Risk of bias: Unclear Applicability concerns: Unclear	
DOMAIN IV: Flow & timing Risk of bias: Unclear	
DOMAIN V: Role of sponsor Risk of bias: Unclear	

Full citation	**Livergood, M. C., et al. (2017). "Obesity and cell-free DNA "no calls": is there an optimal gestational age at time of sampling?" American Journal of Obstetrics & Gynecology 216(4): 413.e411-413.e419.**
Key questions	4a) Test failure rate in multiple pregnancies.
Study characteristics	
Study design / Setting / Study period	Retrospective cohort study of prospectively collected data from the perinatal genetics database at Mercy Hospital St Louis. Tertiary referral centre Mercy Hospital St Louis in St Louis (MO, USA). November 30, 2011, through March 15, 2016 (~4.5 years).
Population	All women who underwent cfDNA testing through the perinatal genetics department at Mercy Hospital St Louis from Nov. 30, 2011, through March 15, 2016. Women with missing data on either height and/or weight were excluded from the study. N=2,385 included in the study and analysis (5 with missing data on height or weight and 3,212 without cfDNA testing were excluded from the study). Mean gestational age 12-13 weeks. 41% advanced maternal age, 9% ultrasound findings, 20% higher-chance result at serum screen, 3% family history, 36% lower-chance result at serum screen. 72/2,385 (3%) multiple pregnancies. No information on chorionicity.
Index test / Comparator / Reference standard	Index test: The cfDNA testing included the vast majority of modality options commercially available: Panorama (SNPs), Harmony (DANSR with targeted sequencing or microarray), MaterniT21 & Verifi (random whole-genome sequencing) Comparator: None. Reference standard: Not applicable (this study only assessed the cfDNA test failure rate).
Outcomes	4a) Test failure rate in multiple pregnancies.
Funding source or sponsor of the study	Funding source not reported. The authors report no conflict of interest.

Full citation	**Livergood, M. C., et al. (2017). "Obesity and cell-free DNA "no calls": is there an optimal gestational age at time of sampling?" American Journal of Obstetrics & Gynecology 216(4): 413.e411-413.e419.**
Information about the authors contacted	No contact needed.
Information about other contacts	No contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: High Applicability concerns: High	
DOMAIN II: Index test Risk of bias: Unclear Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: NA Applicability concerns: NA	
DOMAIN IV: Flow & timing Risk of bias: Low	
DOMAIN V: Role of sponsor Risk of bias: Low	

Full citation	**Meck, J. M., et al. (2015). "Noninvasive prenatal screening for aneuploidy: positive predictive values based on cytogenetic findings." American Journal of Obstetrics & Gynecology 213(2): 214.e211-215.**
Key questions	a) Test accuracy in twin pregnancies (mostly higher chance of fetal trisomies).
Study characteristics	
Study design / Setting / Study period	Cohort study; retrospective analysis 4 cytogenetics laboratories (GeneDx, Gaithersburg, MD; Stanford Hospital and Clinics, Palo Alto, CA; GenPath, Elmwood Park, NJ; the Genetics Center GenPath, Smithtown, NY) in USA; no information on number of centres. November 2011 through October 2014 (3 years).
Population	All of the cases received for follow-up cytogenetic testing during an approximate 3-year period (November 2011 through October 2014) in 4 cytogenetics laboratories (GeneDx, Gaithersburg, MD; Stanford Hospital and Clinics, Palo Alto, CA; GenPath, Elmwood Park, NJ; the Genetics Center GenPath, Smithtown, NY) in which the referring physician noted on the test request form that there had been prior NIPS performed on that pregnancy. N=216 included in the study. No exclusions from the study but only included cases with cytogenetic testing performed at the 4 study labs. Gestational age at time of referral for cytogenetic testing 10 - 28 5/7 weeks. 90% higher chance for fetal trisomies either by virtue of a fetal ultrasound abnormality or advanced maternal age. N=212 included in the analysis (4 with inconclusive or failed cfDNA test result excluded). Mention 2 false positive results for T21 in twin pregnancies. Chorionicity NR.
Index test / Comparator / Reference standard	Index test: cfDNA test provider was reported by the clients: Ariosa (Harmony) 54; Sequenom (MaterniT21) 48; Natera (Panorama) 36; and Verinata (Verifi) 23. In 55 cases, no information provided. Comparator: None. Reference standard: Cytogenetic analysis was performed on amniotic fluid in 137 cases, on CVS in 69 cases, on products of conception in 4 cases, and on 6 neonatal blood samples.
Outcomes	a) Mention 2 false positive cfDNA test results in twin pregnancies.

Full citation	**Meck, J. M., et al. (2015). "Noninvasive prenatal screening for aneuploidy: positive predictive values based on cytogenetic findings." American Journal of Obstetrics & Gynecology 213(2): 214.e211-215.**
Funding source or sponsor of the study	Funding source not reported. All authors work (or were working at the time of inception of this study) for a company (J.M.M., E.K.D., L.M., A.A., C.T., D.P.-A., S.A., R.T.K.) or academic institution (A.M.C.) that performs prenatal cytogenetic testing.
Information about the authors contacted	No contact needed.
Information about other contacts	No contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: High Applicability concerns: Unclear	
DOMAIN II: Index test Risk of bias: Low Applicability concerns: Unclear	
DOMAIN III: Reference standard Risk of bias: Low Applicability concerns: Low	
DOMAIN IV: Flow & timing Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: Unclear	

Full citation	**Papageorghiou, A. T., et al. (2016). "Clinical evaluation of the IONA test: a non-invasive prenatal screening test for trisomies 21, 18 and 13." Ultrasound in Obstetrics & Gynecology 47(2): 188-193.**
Key questions	1a) Test accuracy in twin pregnancies at higher chance of fetal trisomies. 4a) Test failure rate in twin pregnancies.
Study characteristics	
Study design / Setting / Study period	Multicenter blinded case-control study. The study included all trisomy 21, 18 and 13 samples available that met the eligibility criteria, plus unaffected samples that were selected to reflect the prevalence of trisomy 21 observed during sample collection in a hogher-chance population (i.e. 1:9). The unaffected samples were selected using random sampling techniques from all eligible samples. 6 hospitals in England. April 2008 to November 2014.
Population	Women ≥ 18 years of age, with a singleton or twin pregnancy of at least 10 weeks' gestation and a clinical indication for an invasive procedure (screen-positive result from conventional aneuploidy screening such as the combined test, quadruple test, fetal structural anomaly on ultrasound examination or advanced maternal age). Exclusion criteria included higher-order multiple pregnancy (triplets or more), known mosaicism, partial trisomy or translocations, fetal demise, disappearing twin, malignancy or known aneuploidy in the pregnancy. N=442 included in study. Number of unaffected samples not included in the study NR. Gestational age NR. Trisomy prevalence in this case-control study reflected an obstetric population at higher chance of fetal trisomies (1:9). N=437 included in analysis (11 twin pregnancies and 426 singleton pregnancies). 5 samples excluded from the analysis (did not meet validity criteria applied by the IONA software). Chorionicity for 11 twin pregnancies NR.
Index test / Comparator / Reference standard	Index test: cfDNA testing based on semiconductor whole-genome sequencing was performed using Ion Chef and Ion Proton systems (Thermo Fisher Scientific) as a multiplex of eight samples (Iona test, Premaitha Health). For the purposes of the IONA test, a 'trisomy test result' is considered to be 'positive' if the likelihood ratio result is > 1; this threshold value follows implicitly from the conventional statistical interpretation of a likelihood ratio value.

Full citation	**Papageorghiou, A. T., et al. (2016). "Clinical evaluation of the IONA test: a non-invasive prenatal screening test for trisomies 21, 18 and 13." Ultrasound in Obstetrics & Gynecology 47(2): 188-193.**
	With regards to age-adjusted probability (chance) of trisomy, a probability ≥ 1 in 150 was considered to be a screen-positive (higher-chance) result. Explored cut-offs of 1 in 50, 1 in 250 and 1 in 500. Comparator: None. Reference standard: Amniocentesis, chorionic villus sampling (CVS) or birth outcome.
Outcomes	1a) Test accuracy. 4a) Test failure rate.
Funding source or sponsor of the study	Funding source not reported. M.F., R.H., R.M. and W.D. are employees of Premaitha Health plc; the other authors are Principal Investigators for the protocol under which samples were collected.
Information about the authors contacted	No contact needed.
Information about other contacts	No contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: High Applicability concerns: High	
DOMAIN II: Index test Risk of bias: High Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Low Applicability concerns: Low	
DOMAIN IV: Flow & timing Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: High	

Full citation	**Sarno, L., et al. (2016). "Prospective first-trimester screening for trisomies by cell-free DNA testing of maternal blood in twin pregnancy." Ultrasound in Obstetrics & Gynecology 47(6): 705-711.**
Key questions	a) Overall test accuracy in twin pregnancies. 4a) Test failure rate in twin pregnancies.
Study characteristics	
Study design / Setting / Study period	Prospective cohort study. 2 National Health Service (NHS) hospitals and one private clinic in England. October 2012 to August 2015.
Population	Women with singleton or twin pregnancies attending routine care at 11+0 to 13+6 weeks' gestation in one of 2 National Health Service (NHS) hospitals in England or from cfDNA testing as part of routine screening in pregnancies at 10+0 to 13+6 weeks' gestation attending the Fetal Medicine Centre in London, which is a private clinic. Cases with no known karyotype or no known pregnancy outcome, chromosomal abnormalities other than trisomies 21, 18 and 13 were excluded from the study. 438 twin pregnancies and 10,698 singleton pregnancies included in the study. Data from singleton pregnancies also reported in the publication by Revello et al. (2016) [49]. Number of exclusions NR for the singleton pregnancies. For the twin pregnancies, 29 were excluded from the study (23 termination of pregnancy, miscarriage or stillbirth with no known karyotype; 4 lost to follow-up, 2 with chromosomal abnormalities other than trisomies 21, 18 and 13). 100% first trimester pregnancies. cfDNA testing following (or in addition to?) first-trimester combined screening. 373/438 (85%) dichorionic and 65/438 (15%) monochorionic. 417 twin pregnancies and 10,530 singleton pregnancies included in the analysis. 21 twin pregnancies and 168 singleton pregnancies with cfDNA test failure excluded from the analysis.
Index test / Comparator / Reference standard	Index test: Harmony Prenatal Test (Ariosa Diagnostics, Inc., San Jose, CA) performed by Ariosa Diagnostics, Inc., San Jose, CA. DANSR product quantitation method unclear from the paper; Ariosa/Roche confirmed use of DNA microarrays.* In twin pregnancies, the FORTE algorithm used for singletons was modified so that the smallest contribution of fetal fraction from the two foetuses was considered. Comparator: None.

Full citation	**Sarno, L., et al. (2016). "Prospective first-trimester screening for trisomies by cell-free DNA testing of maternal blood in twin pregnancy." Ultrasound in Obstetrics & Gynecology 47(6): 705-711.**
	Reference standard: Karyotype of chorionic villi, amniotic fluid or neonatal blood or phenotypic newborn assessment.
Outcomes	a) Test accuracy. 4a) Test failure rate.
Funding source or sponsor of the study	The study was supported by a grant from The Fetal Medicine Foundation (UK Charity No: 1037116). The cost of collection and analysis of some of the samples was covered by Ariosa Diagnostics, Inc. San Jose, CA, USA.
Information about the authors contacted	Contacted the corresponding author and one of his co-workers (Maria del Mar Gil) to provide false positive data separately for the 3 trisomies. Reply received from Maria del Mar Gil who provided data for the twin pregnancies only.
Information about other contacts	* Contacted Ariosa Diagnostics Inc. (San Jose, CA) to clarify cfDNA test methodology for study samples.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: Unclear Applicability concerns: Low	
DOMAIN II: Index test Risk of bias: Low Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Low Applicability concerns: Low	
DOMAIN IV: Flow & timing Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: Unclear	

Full citation	**Strom, C. M., et al. (2017). "Improving the Positive Predictive Value of Non-Invasive Prenatal Screening (NIPS)." PLoS ONE [Electronic Resource] 12(3): e0167130.**
Key questions	a) Overall test accuracy in twin pregnancies.
Study characteristics	
Study design / Setting / Study period	Validation study: Case-control study, stored twin samples provided by Sequenom. Clinical implementation: Prospective cohort study of the first 10,713 samples tested. Quest Diagnostics (Quest Diagnostics provides non-invasive prenatal screening and other testing services): 3 labs? (Quest Diagnostics, Athena Diagnostics and Celera Diagnostics). Study periods not reported.
Population	Validation study (Twins): NR, stored twin samples provided by Sequenom. 115 twin samples included in study and analysis. Exclusions from the study NR. Gestational age NR. No information on chorionicity. Clinical implementation study: First 10,713 samples received included in study. Accepted samples beginning at the 10th gestational week. Exclusions from the study NR. 90% between 10th and 15th gestation week. Prior chance of fetal trisomy NR. 4 twin pregnancies with positive cfDNA test result included in analysis. No information on chorionicity.
Index test / Comparator / Reference standard	Index test: Second-generation NIPS test: QNatal Advanced (Quest Diagnostics) based on random whole-genome sequencing. 12-plex sequencing on HiSeq2500 system (Illumina). Threshold: Validation study: No threshold, unblended analysis ("*All samples with autosomal trisomies had Z scores >11 and all unaffected pregnancies had Z scores <4.*") Clinical implementation:

Full citation	**Strom, C. M., et al. (2017). "Improving the Positive Predictive Value of Non-Invasive Prenatal Screening (NIPS)." PLoS ONE [Electronic Resource] 12(3): e0167130.**
	Z score cutoff of ≤4 for unaffected pregnancies and >8 for affected pregnancies. Z scores >4 but < 8 prompted further examination. Comparator: None. Reference standard: Validation study: NR Clinical implementation: Confirmation of positive NIPT result by karyotype, ultrasound (not used for T21 confirmation due to lack of specificity of “soft” findings) or physical exam. (3 twins: invasive testing or physical examination at delivery.)
Outcomes	Validation study: a) Test accuracy. Clinical implementation study: a) Only true positive and false positive cfDNA test results for 4 twin pregnancies mentioned.
Funding source or sponsor of the study	This study and preparation of this manuscript were done as part of routine work at Quest Diagnostics. Athena Diagnostics and Celera Diagnostics are wholly owned subsidiaries of Quest Diagnostics. Quest Diagnostics provided support in the form of salaries for CMS, BA, DT, KZ, YL, KL, QN, PK, MM, JW, DR, JC, RO and WS, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Athena/Quest Diagnostics provided support in the form of salaries for CB, ME, and CE, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Celera/Quest Diagnostics provided support in the form of salary for DW, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the `author contributions' section. This study and preparation of this manuscript were done as part of routine work at Quest Diagnostics. Charles M Strom, Ben Anderson, David Tsao, Ke Zhang, Yan Liu, Kayla Livingston, Quoclinh Nguyen, Paula Kolacki, Megan Maxwell, Jia-Chi Wang, Douglas Rabin, Joseph Catanese, Renius Owen, and Weimin Sun are employees of Quest Diagnostics.

Full citation	**Strom, C. M., et al. (2017). "Improving the Positive Predictive Value of Non-Invasive Prenatal Screening (NIPS)." PLoS ONE [Electronic Resource] 12(3): e0167130.**
	Cory Braastad, Matthew Evans, and Christopher Elzinga are employees of Athena/Quest Diagnostics. David Wolfson is an employee of Celera/Quest Diagnostics. Quest Diagnostics provides non- invasive prenatal screening and other testing services. Quest diagnostics is neither a manufacturing company nor marketing company. QNatal Advanced is a laboratory developed test by a US CLIA approved clinical laboratory.
Information about the authors contacted	No contact needed.
Information about other contacts	No contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: Validation study: High Clinical implementation: High	 Applicability concerns: Validation study: High Clinical implementation: Unclear
DOMAIN II: Index test Risk of bias: Validation study: High Clinical implementation: Low	 Applicability concerns: Validation study: Low Clinical implementation: Low
DOMAIN III: Reference standard Risk of bias: Validation study: Unclear Clinical implementation: Unclear	 Applicability concerns: Validation study: Unclear Clinical implementation: High
DOMAIN IV: Flow & timing Risk of bias: Validation study: Unclear Clinical implementation: High	
DOMAIN V: Role of sponsor Risk of bias:	

Full citation	**Strom, C. M., et al. (2017). "Improving the Positive Predictive Value of Non-Invasive Prenatal Screening (NIPS)." PLoS ONE [Electronic Resource] 12(3): e0167130.**
Validation study: High Clinical implementation: High	

Full citation	**Tan, Y., et al. (2016). "Noninvasive prenatal testing (NIPT) in twin pregnancies with treatment of assisted reproductive techniques (ART) in a single center." Prenatal Diagnosis 36(7): 672-679.**
Key questions	a) Overall test accuracy in twin pregnancies. 4a) Test failure rate in twin pregnancies.
Study characteristics	
Study design / Setting / Study period	Prospective cohort study 1 hospital (Reproductive and Genetic Hospital of CITIC-Xiangya) in China. January 2012 to December 2013 (2 years).
Population	1) Pregnant women with twin pregnancies after assisted reproductive technologies (ART) treatment from January 2012 to December 2013; (2) over 18 years old; (3) for one- to two-embryo transfer, confirmation of live twin pregnancy by ultrasound scan before enrolling in the study; for three-embryo transfer, confirmation of live twin pregnancy and no demise fetus by ultrasound scan before enrolling in the study; (4) voluntarily received NIPT screening for fetal trisomy 21 (T21), trisomy 18 (T18), and trisomy 13 (T13), with or without prior Down's syndrome screening result; and (5) gestational age (GA) of >10 weeks. Participants were treated with ART at the study hospital, and once twin pregnancy was confirmed, pregnant women were offered the choice of receiving NIPT. Only pregnant women who chose to undergo NIPT were included in this study. N=565 included in the study (of 8,136 twin pregnancies through ART at the study hospital). 81.4% first trimester. General chance of fetal trisomies. 544/565 (96%) dichorionic, 18 (3%) monochorionic, 3 (paper says 4?) (<1%) other. N=510 included in the analysis. 55 excluded from the analysis (5 test failures; 17 lost to follow-up; 33 birth defect, stillbirth, or miscarriage with unconfirmed reasons).
Index test / Comparator / Reference standard	Index test: cfDNA testing based on random whole-genome sequencing performed at the Minister of Health accredited and ISO/IEC17025-certified clinical laboratories of BGI-Shenzhen, China. Fetal aneuploidy risk was evaluated using a binary hypothesis t-test and logarithmic likelihood ratio L-score. If t-score was >2.5 and L-score was >1, a 'high risk' result was given.

Full citation	**Tan, Y., et al. (2016). "Noninvasive prenatal testing (NIPT) in twin pregnancies with treatment of assisted reproductive techniques (ART) in a single center." Prenatal Diagnosis 36(7): 672-679.**
	If t-score was <2.5 and L-score was <1, a 'low risk' result was given. If either t-score was >2.5 or L-score was >1, the sample was at risk of mosaicism or low fetal fraction, and re-sampling was recommended. Comparator: None. Reference standard: For NIPT positive results, amniocentesis followed by karyotyping. For NIPT negative results, standard healthcare procedures were provided: one month after the date of expected confinement, pregnant outcome was surveyed by telephone interview. To encourage reporting of NIPT false positive and false negative results, an insurance policy was provided to each participant. Briefly, for NIPT positive results, the insurance policy reimbursed the costs for prenatal diagnosis, and for each confirmed NIPT false negative results the insurance policy compensated CNY200 000 to the patient.
Outcomes	a) Test accuracy. 4a) Test failure rate.
Funding source or sponsor of the study	This work was supported by grants from the Major State Basic Research Development Program of China (No. 2012CB944901) and the National Natural Science Foundation of China (Nos.81222007 and 81471432) and the Program for New Century Excellent Talents in University. The work is also supported by Key Laboratory of Cooperation Project in Guangdong Province (No. 2011A060906007), Shenzhen Birth Defect Screening Project Lab (No. [2011] 861), and Shenzhen Municipal Government of China (Nos. CXZZ20130517144604091 and CXZZ20140808170655268). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Ya Gao, Xuyang Yin, Fang Chen, and Wei Wang are the employees of BGI-Shenzhen. Jing Li and Huanhuan Peng are the employees of Clinical laboratory of BGI Health, BGI-Wuhan. Meili Fu, Yuying Yuan, Fuman Jiang, and Hongyun Zhang are the employees of Clinical laboratory of BGI Health, BGI-Shenzhen. Fei Gong, Yueqiu Tan, Xihong Li, Juan Du, Wen Li, Guangxiu Lu, and Ge Lin are the employees of Reproductive and Genetic Hospital of CITICXiangya and have no financial relationship with BGI-Shenzhen, or Clinical laboratory of BGI Health.
Information about the	No contact needed.

Full citation	**Tan, Y., et al. (2016). "Noninvasive prenatal testing (NIPT) in twin pregnancies with treatment of assisted reproductive techniques (ART) in a single center." Prenatal Diagnosis 36(7): 672-679.**
authors contacted	
Information about other contacts	No contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: High Applicability concerns: High	
DOMAIN II: Index test Risk of bias: Low Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Unclear Applicability concerns: Unclear	
DOMAIN IV: Flow & timing Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: High	

Full citation	**Valderramos, S. G., et al. (2016). "Cell-free DNA screening in clinical practice: abnormal autosomal aneuploidy and microdeletion results." American Journal of Obstetrics & Gynecology 215(5): 626.e621-626.e610.**
Key questions	a) Overall test accuracy in twin pregnancies.
Study characteristics	
Study design / Setting / Study period	Retrospective cohort study. 1 referral maternal-fetal medicine practice in Los Angeles (CA, USA). March 2013 through July 2015.
Population	All patients with abnormal cfDNA results from a referral maternal-fetal medicine practice in Los Angeles, CA, from March 2013 through July 2015. N=121 included in the study (10 twin pregnancies, 111 singleton pregnancies). No exclusions from the study. Gestational age and prior chance of fetal trisomy NR. N=83 included in analysis (8 twin pregnancies, 75 singleton pregnancies). 38 excluded from the analysis due to non-reportable result (n=13), microdeletion only n=16), triploidy (n=2), more than 1 abnormal parameter (n=3), or no confirmatory testing (n=4). Chorionicity of the twin pregnancies NR.
Index test / Comparator / Reference standard	Index test: 4 commercially available laboratories used. cfDNA testing based on SNPs (1 lab, *Panorama test by Natera?*), based on random whole-genome sequencing (2 labs) or targeted sequencing (1 lab, *Harmony Prenatal Test by Ariosa?*). Comparator: None. Reference standard: Confirmatory diagnostic testing (by chorionic villus sampling or amniocentesis) or postnatal genetic evaluation. Neonatal outcomes were reported by the patient's primary obstetric provider in follow-up.
Outcomes	a) Test accuracy in twin pregnancies (PPV only)
Funding source or sponsor of the study	Funding source not reported. L.D.P. and N.S.S. are on the speaker's bureau for Illumina and L.D.P. served on their medical advisory board. L.D.P. is treasurer of the Perinatal Quality Foundation. The remaining authors report no conflict of interest.

Full citation	**Valderramos, S. G., et al. (2016). "Cell-free DNA screening in clinical practice: abnormal autosomal aneuploidy and microdeletion results." American Journal of Obstetrics & Gynecology 215(5): 626.e621-626.e610.**
Information about the authors contacted	No contact needed.
Information about other contacts	No contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: High Applicability concerns: Unclear	
DOMAIN II: Index test Risk of bias: Low Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Low Applicability concerns: Low	
DOMAIN IV: Flow & timing Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: Unclear	

Full citation	**Yang, J., et al. (2018). "Performance of non-invasive prenatal testing for trisomies 21 and 18 in twin pregnancies." Mol Cytogenet 11: 47.**
Key questions	a) Overall test accuracy in twin pregnancies; 4a) Test failure rate in twin pregnancies.
Study characteristics	
Study design / Setting / Study period	Prospective cohort study. NR (all authors affiliated to Prenatal Diagnosis Centre, Guangdong Women and Children Hospital, China). January 2015 to December 2016 (2 years).
Population	Twin pregnancies undergoing cfDNA testing. N=432 included in the study. Number excluded from the study NR. 14.1% first trimester. Prior chance of fetal trisomy NR. 337/432 (78%) dichorionic and 95/432 (22%) monochorionic. N=373 included in the analysis (59 without karyotype or live birth feedback excluded from the analysis).
Index test / Comparator / Reference standard	Index test: JingXin Fetal Chromosome Aneuploidy (T21, T18, T13) Testing Kits (CFDA registration permit No. 0153400300). Possibly based on random whole-genome semiconductor sequencing (see Liao 2014 [64]). Z score range from − 3 to 3 was considered to indicate a low risk for a trisomy chromosome. Z score were > 3, high-risk zone. Comparator: None. Reference standard: Invasive sampling was performed for cases at higher chance of fetal trisomies. Otherwise “live birth feedbacks”.
Outcomes	a) Test accuracy. 4a) Test failure rate.
Funding source or sponsor of the study	National Key Research and Development Program of China, 2016YFC1000700, 2016YFC1000703. Guangdong Medical Science and Technology Research Project, 2016118171659322.

Full citation	**Yang, J., et al. (2018). "Performance of non-invasive prenatal testing for trisomies 21 and 18 in twin pregnancies." Mol Cytogenet 11: 47.**
	The authors declare that they have no competing interests.
Information about the authors contacted	No contact needed.
Information about other contacts	No contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: Unclear Applicability concerns: High	
DOMAIN II: Index test Risk of bias: Low Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Unclear Applicability concerns: Unclear	
DOMAIN IV: Flow & timing Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: Low	

b) DNA microarray-based cfDNA testing (5 studies)

Full citation	**Gil, M. M., et al. (2017). "Screening for trisomies 21 and 18 in a Spanish public hospital: from the combined test to the cell-free DNA test." Journal of Maternal-Fetal & Neonatal Medicine 30(20): 2476-2482.**
Key questions	b) Overall test accuracy of DNA microarray-based cfDNA testing. 1b) cfDNA testing as follow-on test in pregnant women with higher-chance result from FTCS (>1 in 250 and normal ultrasound). 4b) Test failure rate.
Study characteristics	
Study design	Uncontrolled before-after study. 1 centre in Spain. Before NIPT: November 2011 – December 2014; After NIPT: January 2015 – January 2016.
Population	6,011 women with singleton pregnancies attending Torrejon University Hospital in Madrid, Spain, from 11/2011-01/2016 at 11-13 weeks for first-trimester combined screening (FTCS). After NIPT introduction: Women with singleton pregnancies screened from 01/2015-01/2016. All 1st trimester (11-13 weeks' gestation for FTCS, 12-14 weeks for cfDNA testing). cfDNA testing offered to women with FTCS result > 1 in 250 without ultrasound abnormalities (Nuchal translucency thickness <3.5 mm and no fetal defects). 54/72 chose cfDNA testing.
Index test / Comparator / Reference standard	Index test: Harmony Prenatal Test (Ariosa Diagnostics, Inc., San Jose, CA) performed by Ariosa Diagnostics, Inc., San Jose, CA. DANSR product quantitation method unclear from the paper; Ariosa/Roche confirmed use of DNA microarrays.* FORTE risk score ≥ 1%: high risk. Reference standard: Pregnancy outcome was ascertained at least two months after the expected due date to optimise accuracy by three methods: firstly, prenatal or postnatal karyotyping; secondly, neonatal examination and all paediatrics medical records available for the baby from Madrid region database; and thirdly, by contacting the patients' general practitioners when

Full citation	**Gil, M. M., et al. (2017). "Screening for trisomies 21 and 18 in a Spanish public hospital: from the combined test to the cell-free DNA test." Journal of Maternal-Fetal & Neonatal Medicine 30(20): 2476-2482.**
	the previous sources were insufficient or unavailable
Outcomes	b) Overall test accuracy of DNA microarray-based cfDNA testing (54/72 included in analysis). 1b) cfDNA test accuracy (54/72 included in analysis). 4b) Indeterminate results.
Funding source or sponsor of the study	Funding source not reported. No conflicts of interest to declare.
Information about the authors contacted	Contacted the corresponding author via email to clarify cfDNA test methodology and missing data on true negatives and false negatives.
Information about other contacts	* Contacted Ariosa Diagnostics Inc. (San Jose, CA) via email to clarify cfDNA test methodology for study samples.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: High Applicability concerns: Low	
DOMAIN II: Index test Risk of bias: Low Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Unclear Applicability concerns: Unclear	
DOMAIN IV: Flow & timing Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: Unclear	

Full citation	**Juneau, K., et al. (2014). "Microarray-based cell-free DNA analysis improves noninvasive prenatal testing." Fetal Diagnosis & Therapy 36(4): 282-286.**
Key questions	b) Overall test accuracy of DNA microarray-based cfDNA testing. Direct comparison of test accuracy in DNA microarray- and sequencing- based Harmony Prenatal Test.
Study characteristics	
Study design	Retrospective study of frozen maternal plasma samples (Cohort study? High prevalence of trisomies in the study population.) Study period NR.
Population	Singleton pregnancies in women at least 18 years old. Prior chance of fetal trisomy NR. Gestational age: Mean 14.8 weeks, SD 4.2 weeks, Range 10-34 weeks. 392/878 with appropriate reference standard included in analyses. The remaining 486 samples "*were originally tested using the Harmony Prenatal Test from Ariosa Diagnostics Inc. (San Jose, Calif., USA)*…."
Index test / Comparator / Reference standard	Index tests: Harmony Prenatal Test (Ariosa Diagnostics, Inc., San Jose, CA) performed by Ariosa Diagnostics, Inc., San Jose, CA. DANSR products from each sample were divided and analysed by next generation sequencing on Illumina HiSeq 2500 and by custom DNA microarrays from Affymetric Inc. imaged on an Affymetrix GeneTitan MultiChannel Instrument, respectively. For both quantitation methods: FORTE risk score ≥ 1%: high risk. Reference standard: Invasive genetic testing or postnatal newborn examination followed by detailed genetic analysis, when trisomy was suspected.
Outcomes	b) DNA microarray-based cfDNA test accuracy (392/878 included in analysis)* Sequencing-based cfDNA test accuracy (392/878 included in analysis)*
Funding source or sponsor of the study	Study designed, performed, interpreted and published by employees of Ariosa Diagnostics Inc. (San Jose, CA). All 10 authors employees of Ariosa Diagnostics Inc. (San Jose, CA).
Information about the authors contacted	Contacted the corresponding author via email to obtain test accuracy data for the subgroup of women with appropriate reference standard.
Information about other contacts	* Contacted Ariosa Diagnostics Inc. (San Jose, CA) via email to obtain test accuracy data for the subgroup of women with appropriate reference standard.

Full citation	**Juneau, K., et al. (2014). "Microarray-based cell-free DNA analysis improves noninvasive prenatal testing." Fetal Diagnosis & Therapy 36(4): 282-286.**
Methodological quality	
DOMAIN I: Patient selection	
Risk of bias: High	Applicability concerns: High
DOMAIN II: Index test	
Risk of bias: Unclear	Applicability concerns: Low
DOMAIN III: Reference standard	
Risk of bias: Low	Applicability concerns: Low
DOMAIN IV: Flow & timing	
Risk of bias: High	
DOMAIN V: Role of sponsor	
Risk of bias: High	

Full citation	**Kagan, K. O., et al. (2018). "First-trimester risk assessment based on ultrasound and cell-free DNA vs combined screening: a randomized controlled trial." Ultrasound in Obstetrics & Gynecology 51(4): 437-444.**
Key questions	b) Overall test accuracy of DNA microarray-based cfDNA testing. 2b) cfDNA testing as replacement test in lower-chance pregnant women with normal first trimester ultrasound (US). Head-to-head comparison with first-trimester combined screening (FTCS). 4b) Test failure rate.
Study characteristics	
Study design	Randomised controlled trial (1:1); FTCS versus ultrasound (US) & cfDNA testing (reflex approach). 1 centre in Germany. October 2015 - December 2016
Population	Women with singleton pregnancy undergoing first-trimester screening, performed at the prenatal medicine department of the University of Tuebingen, Germany, between October 2015 and December 2016 with normal ultrasound examination (Nuchal translucency [NT] thickness ≤3.5mm and no fetal defects) at 11-13 weeks' gestation. All 1st trimester, low chance of fetal trisomy. 1,400 included in study: FTCS: n=699; US & cfDNA: n=701.
Index test / Comparator / Reference standard	Index test: Harmony Prenatal Test (Ariosa Diagnostics, Inc., San Jose, CA) performed by Cenata GmbH (Tuebingen, Germany). DANSR product quantitation using DNA microarrays. FORTE risk score > 1%: high risk Comparator: FTCS at 11-13 weeks (maternal and gestational age, fetal NT thickness, and maternal levels of serum PAPP-A and free β-hCG). Combined chance result for T21 computed based on the most recent Fetal Medicine Foundation (FMF) algorithm; cutoff: 1 in 100. Reference standard: Newborn examination or genetic testing (pre- or postnatal).
Outcomes	b) Overall test accuracy of DNA microarray-based cfDNA testing. 2b) Test accuracy of cfDNA testing (678/701 included in analysis); test accuracy of the FTCS test (688/699 included in

Full citation	**Kagan, K. O., et al. (2018). "First-trimester risk assessment based on ultrasound and cell-free DNA vs combined screening: a randomized controlled trial." Ultrasound in Obstetrics & Gynecology 51(4): 437-444.**
	analysis). 4b) Indeterminate results.
Funding source or sponsor of the study	Roche/Ariosa Diagnostics, Inc. (San Jose, CA, USA) provided the kits for the Harmony® Prenatal Test. Cenata GmbH (Tuebingen, Germany) performed the cfDNA analysis One author is an employee of Roche Sequencing Solutions Inc.; another of the authors is an employee of Cenata GmbH.
Information about the authors contacted	Contacted the corresponding author via email to clarify cfDNA test methodology.
Information about other contacts	No further contact needed.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: Low Applicability concerns: High	
DOMAIN II: Index test Risk of bias: Low Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Low Applicability concerns: Low	
DOMAIN IV: Flow & timing Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: Unclear	

Full citation	**Langlois, S., et al. (2017). "Comparison of first-tier cell-free DNA screening for common aneuploidies with conventional publically funded screening." Prenatal Diagnosis 37(12): 1238-1244.**
Key questions	b) Overall test accuracy of DNA microarray-based cfDNA testing. 2b) cfDNA testing as replacement test in the general obstetric population. Head-to-head comparison with standard screening. 4b) Test failure rate.
Study characteristics	
Study design	Prospective cohort study. Substudy of PEGASUS (clinicaltrials.gov identifier NCT01925742). 3/5 centres from PEGASUS study selected for this substudy: Vancouver, Calgary, Quebec (Canada). Study period NR.
Population	Women needed to be 19 years or older, have a singleton gestation, be recruited before 14 weeks gestation, have decided to undertake the provincially funded screening test. All 1st trimester (10 weeks – 13 weeks 6 days); general obstetric population. cfDNA test: 1,159/1,198 women included in analysis. First-trimester combined screening (FTCS): 287/300 women from Calgary centre included in analysis. The centres in Vancouver and Quebec did not use FTCS (see below).
Index test / Comparator / Reference standard	Index test: Harmony Prenatal Test (Ariosa Diagnostics, Inc., San Jose, CA) performed by Ariosa Diagnostics, Inc., San Jose, CA. DANSR product quantitation method unclear from the paper; Ariosa/Roche confirmed use of DNA microarrays.* Cutoff NR. Comparator: FTCS: 1st trimester PAPP-A, free βhCG, and nuchal translucency thickness (Calgary centre). Vancouver and Quebec centres offered serum integrated prenatal screening (SIPS) or quadruple screening to women < 35, and integrated prenatal screening (IPS, first-trimester ultrasound plus SIPS) for women ≥ 35 years. Reference standard:

Full citation	**Langlois, S., et al. (2017). "Comparison of first-tier cell-free DNA screening for common aneuploidies with conventional publically funded screening." Prenatal Diagnosis 37(12): 1238-1244.**
	Prenatal or postnatal cytogenetic analysis, newborn and follow-up outcome at age 6 weeks.
Outcomes	b) Overall test accuracy of DNA microarray-based cfDNA testing. 2b) Test accuracy of cfDNA testing (1,159/1,165 included in analysis); test accuracy of the FTCS test (287/300 included in analysis). 4b) Indeterminate results.
Funding source or sponsor of the study	Genome Canada, the Canadian Institutes for Health Research, Genome Québec, Genome BC, Genome Alberta, the Québec Ministère de l'enseignement supérieur, de la recherche, de la science et de la technologie. Arms' length in-kind co-funding for this study was also provided by Roche/Ariosa Diagnostics Inc (San Jose, CA) in the form of cell-free DNA (cfDNA) testing (Harmony Prenatal Test) free of charge for the women enrolled in the present study. Roche/Ariosa Diagnostics Inc (San Jose, CA) had no role in the design of the study, interpretation of the results, or approval of the manuscript.
Information about the authors contacted	Contacted the corresponding author via email to clarify cfDNA test methodology.
Information about other contacts	* Contacted Ariosa Diagnostics Inc. (San Jose, CA) to clarify cfDNA test methodology for study samples.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: Unclear Applicability concerns: High	
DOMAIN II: Index test Risk of bias: Low Applicability concerns: High (for comparator as <80% of women had FTCS).	
DOMAIN III: Reference standard Risk of bias: Low Applicability concerns: Low	
DOMAIN IV: Flow & timing Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: Low	

Full citation	**Stokowski, R., et al. (2015). "Clinical performance of non-invasive prenatal testing (NIPT) using targeted cell-free DNA analysis in maternal plasma with microarrays or next generation sequencing (NGS) is consistent across multiple controlled clinical studies." Prenatal Diagnosis 35(12): 1243-1246.**
Key questions	b) Overall test accuracy of DNA microarray-based cfDNA testing. 4b) Test failure rate.
Study characteristics	
Study design	Multicentre cohort study; retrospective analysis. Stored (frozen) blood samples from Sweden, UK and USA. Study period NR.
Population	799 pregnant women (759 singleton pregnancies, 40 twin pregnancies). Prior chance of fetal trisomies NR. High prevalence of trisomies in study population. Gestational age at blood sampling: Median 16 weeks, IQR 13-19 weeks.
Index test / Comparator / Reference standard	Index test: Analysis of frozen samples. Harmony Prenatal Test (Ariosa Diagnostics, Inc., San Jose, CA) based on DNA microarrays; performed by Ariosa Diagnostics, Inc., San Jose, CA. FORTE risk score ≥ 1%: high risk. Reference standard: Diagnostic testing (amniocentesis and/or chorionic villi sampling) or newborn examination with any suspected aneuploidies at birth confirmed with karyotyping.
Outcomes	b) Overall test accuracy of microarray-based cfDNA testing (791/799 included in analysis). 4b) Indeterminate results.
Funding source or sponsor of the study	This study was supported by Ariosa Diagnostics, Inc. (San Jose, CA). 8/11 authors are paid employees of Ariosa Diagnostics.
Information about the authors contacted	No need for further contact.

Full citation	**Stokowski, R., et al. (2015). "Clinical performance of non-invasive prenatal testing (NIPT) using targeted cell-free DNA analysis in maternal plasma with microarrays or next generation sequencing (NGS) is consistent across multiple controlled clinical studies." Prenatal Diagnosis 35(12): 1243-1246.**
Information about other contacts	No need for further contact.
Methodological quality	
DOMAIN I: Patient selection Risk of bias: Unclear Applicability concerns: High	
DOMAIN II: Index test Risk of bias: Unclear Applicability concerns: Low	
DOMAIN III: Reference standard Risk of bias: Low Applicability concerns: Low	
DOMAIN IV: Flow & timing Risk of bias: High	
DOMAIN V: Role of sponsor Risk of bias: High	

Table 13. Studies using the Harmony Prenatal Test with both sequencing and DNA microarray technologies (7 studies published in 8 articles)

Reference		Study characteristics	Findings relevant to the review questions
1.	Bevilacqua, E., et al. (2018). "Cell-Free DNA Analysis in Maternal Blood: Differences in Estimates between Laboratories with Different Methodologies Using a Propensity Score Approach." Fetal Diagnosis and Therapy: 1-10.	Prospective cohort study comparing two different cfDNA tests. Propensity score analysis to match patients between the 2 groups. "Harmony Prenatal Test": Department of Obstetrics and Gynecology, University Hospital Brugmann, Brussels, Belgium; January 2013 and October 2016. "Cerba test": Various French fetal medicine centers and private practitioners; November 2014 and February 2016. Singleton pregnancies with cfDNA testing performed after 10 weeks of gestational age and with known pregnancy outcomes. Included: 5,505/7,121 Harmony Prenatal Test: 2,870/2,932 Cerba test: 2,635/4,189 Significant differences between the 2 groups in maternal age, maternal weight, % smokers, % higher-chance pregnancies (17% vs 61%), and gestational age. Index tests: Harmony Prenatal Test (Ariosa Diagnostics, San Jose, CA, USA) using sequencing- as well as DNA microarray-based technologies.* "Cerba test" (Laboratoire Cerba, SaintOuen-l'Aumône, France) using genome-wide massively parallel sequencing. Reference standard: "Known pregnancy outcome".	**b) Test accuracy (mixed chances of fetal trisomy)** **Harmony Prenatal Test** / **Cerba Test** **T21** 41/41 (100%) detected / 93/93 (100%) detected **T18** 11/12 (91.7%) detected / 7/7 (100%) detected **T13** 5/6 (83.3%) detected / 5/5 (100%) detected **FPR** 0.1% / 0.2% 1 FP for T21 / 1 FP for T21 1 FP for T18 / 3 FP for T18 1 FP for T13 / 1 FP for T13 Test performances for the detection of the major fetal trisomies 21, 18, and 13 were comparable, mainly regarding trisomy 21. The FPR was higher with the Cerba test (0.2 vs. 0.1%). **4b) Test failures** **Harmony Prenatal Test** / **Cerba Test** **Initial test** 46/2,811 *(1.6%)* / 20/2,530 *(0.8%)* **Repeat test** 13/41 *(31.7%)* / 2/13 *(15.4%)* **Overall no-result** 18/2,811 (0.6%) / 9/1,530 (0.4%) After matching, the data indicate a higher initial no-result rate in the Harmony group (1.30%) than in the Cerba group (0.75%; p = 0.039).

Reference		Study characteristics	Findings relevant to the review questions
2.	Bjerregaard, L., et al. (2017). "The rate of invasive testing for trisomy 21 is reduced after implementation of NIPT." Danish Medical Journal 64(4).	Before-after study without concurrent control group; Aalborg University Hospital, Denmark; Before NIPT: 1 March 2011 to 1 February 2013 After NIPT: 1 March 2013 to 1 February 2015. All singleton higher-chance pregnancies (first trimester combined test chance of T21 ≥ 1:300) Before: n=253 After: n=302 (132/302 chose cfDNA testing). Index test: Harmony Prenatal Test (Ariosa Diagnostics, San Jose, CA, USA) using sequencing- as well as DNA microarray-based technologies.* Reference standard: Pre- or postnatal karyotyping or phenotype at birth.	**1b) Test accuracy in higher-chance pregnant women (FTCS ≥ 1:300)** (132/302 opted to have NIPT and were included in the analysis) **T21** **TP** 4 **TN** *128* **FP** 0 **FN** 0 **4b) Test failures** **Initial test** 1/132 (0.8%) **Repeat test** 0/1 **Overall no-result** 0/132
3.	Chan, N., et al. (2018). "Implications of failure to achieve a result from prenatal maternal serum cell-free DNA testing: a historical cohort study." BJOG: An International Journal of Obstetrics and Gynaecology 125(7): 848-855.	Historical cohort study. Private specialist, multi-site prenatal screening service (Sydney Ultrasound For Women) in Sydney, Australia. June 2013 and March 2016. Women who failed to obtain a result from cfDNA testing (n=131), no exclusions from the study. cfDNA test as first-tier test? A total of 12,033 women had cfDNA testing. Harmony Prenatal Test: n=6,375 GeneSyte Test: n=5,658 Index test: Initially: Harmony Prenatal Test (Ariosa Diagnostics, San Jose, CA, USA) using sequencing- as well as DNA microarray-based technologies.* Availability of cfDNA testing during study period led to change in provider: Genea (Sydney, Australia) for analysis by GeneSyte (based on sequencing). Reference standard: NA as only data on test failures reported.	**4b) Test failures (prior chance for fetal trisomies NR)** **Harmony Prenatal Test** / **GeneSyte Test** 12/5,658 (0.2%) **Initial test** 119/6,375 (1.9%) **Repeat test** 13/46 (28.3%) **Overall no-result** *86/6,375 (1.3%)* *P* < 0.0001 for initial test failure rate; binomial test.

Reference		Study characteristics	Findings relevant to the review questions
4.	Gil, M. M., et al. (2016). "Clinical implementation of routine screening for fetal trisomies in the UK NHS: cell-free DNA test contingent on results from first-trimester combined test." Ultrasound in Obstetrics & Gynecology 47(1): 45-52.	Prospective cohort study. 2 NHS hospitals in England (King's College Hospital, London, and Medway Maritime Hospital, Gillingham, Kent). October 2013 and February 2015 12,134 singleton pregnancies were offered FTCS. 11,692 with known outcome included in analysis. cfDNA testing offered to women with chance of ≥ 1 in 100 (higher-chance result) and chance between 1 in 101 and 1 in 2,500 (intermediate-chance result). 3,698/4,012 (92%) chose cfDNA testing. 449/460 (97.6%) higher-chance women chose cfDNA testing. 3,249/3,552 (91.5%) intermediate-chance women chose cfDNA testing. Index test: Harmony Prenatal Test (Ariosa Diagnostics, San Jose, CA, USA) using sequencing- as well as DNA microarray-based technologies.* Reference standard: Karyotype of chorionic villi, amniotic fluid or neonatal blood or phenotype examination at birth.	**1b) Test accuracy in intermediate (FTCS between 1:101 and 1:2,500) and higher chance (FTCS ≥ 1:100) pregnant women** (3,633/3,698 included in analysis) **T21** / **T18** / **T13** **TP** 43 / 21 / 2 **TN** *3,588* / *3,608* / *3,625* **FP** 1 / 4 / 4 **FN** 1 / 0 / 2 **4b) Test failures** **Initial test** 99/3,698 *(2.7%)* **Repeat test** 20/54 *(37%)* **Overall no-result** 65/3,698 *(1.8%)* 3 T18

<table>
<tr><th colspan="2">Reference</th><th>Study characteristics</th><th>Findings relevant to the review questions</th></tr>
<tr><td>5.</td><td>Lee, T. J., et al. (2018). "Cell-free fetal DNA testing in singleton IVF conceptions." Human Reproduction 33(4): 572-578.</td><td>Retrospective cohort study.
Single private obstetric and gynaecological ultrasound clinic in Melbourne, Australia.
April 2013 and November 2016.

5,625 singleton pregnancies after 10 weeks' gestation had cfDNA testing performed, consecutive sampling. cfDNA testing as primary screening test before 12 weeks' gestation or as follow-on test after high-chance first or second trimester screening result.
>93% first trimester.
4,633 spontaneously conceived
992 IVF.

Index test:
Harmony Prenatal Test (Ariosa Diagnostics, San Jose, CA, USA) using sequencing- as well as DNA microarray-based technologies.*

Reference standard:
Pre- or postnatal karyotype and/or phenotype at birth.</td><td>b) Test accuracy
(mostly general obstetric population)
(5,569/5,625 included in analysis)

<table>
<tr><th></th><th>Spontaneous</th><th>IVF</th><th>Total</th></tr>
<tr><td>T21</td><td></td><td></td><td></td></tr>
<tr><td>PPV</td><td>40/40 (100%)</td><td>3/3 (100%)</td><td>43/43 (100%)</td></tr>
<tr><td>TP</td><td>40</td><td>3</td><td>43</td></tr>
<tr><td>FP</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>T18</td><td></td><td></td><td></td></tr>
<tr><td>PPV</td><td>10/13 (76.9%)</td><td>1/2 (50%)</td><td>11/15 (73.3%)</td></tr>
<tr><td>TP</td><td>10</td><td>1</td><td>11</td></tr>
<tr><td>FP</td><td>3</td><td>1</td><td>4</td></tr>
<tr><td>T13</td><td></td><td></td><td></td></tr>
<tr><td>PPV</td><td>1/4 (25%)</td><td>0/5 (0%)</td><td>1/9 (11.1%)</td></tr>
<tr><td>TP</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>FP</td><td>3</td><td>5</td><td>8</td></tr>
</table>

4b) Test failures
<table>
<tr><th></th><th>Spontaneous</th><th>IVF</th><th>Total</th></tr>
<tr><td>Initial test</td><td>2.2%</td><td>5.2%</td><td>NR</td></tr>
<tr><td>Overall no-result</td><td>0.7%</td><td>2.4%</td><td>NR</td></tr>
</table>
</td></tr>
</table>

Reference		Study characteristics	Findings relevant to the review questions
6.	Miltoft, C. B., et al. (2018). "Contingent first-trimester screening for aneuploidies with cell-free DNA in a Danish clinical setting." Ultrasound in Obstetrics & Gynecology 51(4): 470-479.	Prospective cohort study. 2 hospitals in in Copenhagen (Copenhagen University Hospitals, Rigshospitalet and Herlev and Gentofte Hospital), Denmark. August 2014 and May 2015. 6,449 women aged ≥18 years with a singleton pregnancy undergoing FTCS. 597/869 with FTCS chance for T21 of ≥1 in 1,000 had cfDNA testing. Index test: Harmony Prenatal Test (Ariosa Diagnostics, San Jose, CA, USA) using sequencing- as well as DNA microarray-based technologies.* Reference standard: Pre- or postnatal karyotypes or newborn examination.	**1b) Test accuracy in higher-chance (FTCS ≥ 1 in 1,000) pregnant women** (581/597 included in analysis) **T21** / **T18** / **T13** **TP** 13 / 1 / 2 **TN** *567* / *580* / *579* **FP** 0 / 0 / 0 **FN** 1 / 0 / 0 **4b) Test failures** **Initial test** 19/597 (3.2%) **2nd sample** 5/7 (71.4%) **3rd sample** 2/3 (66.7%) **Overall no-result** 16/597 *(2.7%)*

<table>
<tr><th colspan="2">Reference</th><th>Study characteristics</th><th>Findings relevant to the review questions</th></tr>
<tr><td>7.</td><td>Revello, R., et al. (2016). "Screening for trisomies by cell-free DNA testing of maternal blood: consequences of a failed result." Ultrasound in Obstetrics & Gynecology 47(6): 698-704.

Sarno, L., et al. (2016). "Prospective first-trimester screening for trisomies by cell-free DNA testing of maternal blood in twin pregnancy." Ultrasound in Obstetrics & Gynecology 47(6): 705-711.</td><td>Prospective cohort study;
2 National Health Service (NHS) hospitals in England (King’s College Hospital, London, and Medway Maritime Hospital, Gillingham, Kent);
1 private clinic (Fetal Medicine Centre in London).
October 2002 to August 2015.

10,963 singleton and 467 twin pregnancies had cfDNA testing and FTCS at 10-14 weeks’ gestation.
10,698/10,963 singleton and 438/467 twin pregnancies with pregnancy outcome and excluding chromosomal abnormalities other than T21, T18, and T13 included for further analysis.
General obstetric population, 100% first trimester.

Index test:
Harmony Prenatal Test (Ariosa Diagnostics, San Jose, CA, USA) using sequencing- as well as DNA microarray-based technologies.*

Reference standard:
Pre- or postnatal karyotypes or newborn examination.</td><td>b) Test accuracy (general obstetric population)
(10,530/10,698 singleton and 417/438 twin pregnancies included in analyses)

Detection rate | Singleton | Twin | Total
T21 | 156/158 (98.7%) | 8/8 (100%) | 164/166 (98.8%)
T18 | 41/46 (89.1%) | 3/4 (75%) | 44/50 (88.0%)
T13 | 8/15 (53.3%) | 0/1 (0%) | 8/16 (50%)
FPR | 23/10,311 (0.22%) | 1/404 (0.25%) | 24/10,715 (0.22%)

4b) Test failures
Singleton | Twin | Total
Initial test | 316/10,698 (3.0 %) | 41/438 (9.4%) | 357/11,136 (3.2%)
Repeat test | 87/235 (37.0%) | 19/39 (48.7%) | 106/274 (38.7%)
Overall no-result | 168/10,698 (1.6%) | 21/438 (4.8%) | 189/11,136 (1.7%)</td></tr>
</table>

cfDNA, cell-free deoxyribonucleic acid; FN, false negative; FP, false positive; FPR, false positive rate = FP/(FP+TN) = 1 – specificity; FTCS, First trimester combined screening; NA, not applicable; NIPT, non-invasive prenatal testing (here: cfDNA testing); NR, not reported; PPV, positive predictive value; T13, trisomy 13; T18, trisomy 18; T21, trisomy 21; TN, true negative; TP, true positive.

* Information received via personal communication with Ariosa Diagnostics Inc. (San Jose, CA).

Numbers in italics were calculated based on information given in the paper.

Appraisal for quality and risk of bias

Quality assessments of included studies are reported below.

Table 14. Quality assessment of all included studies (n=21)

Study	Risk of bias					Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Role of sponsor	Patient selection	Index test	Reference standard
a) Studies on cfDNA testing in twin / multiple gestations (n=16)								
Benachi 2015 [24]	Unclear	Low	Low	High	Unclear	High	Low	Low
Beulen 2017 [25]	Low	Low	Low	High	Unclear	High	Low	Low
Brison 2018 [26]	Low	High	Low	High	Low	High	Low	Low
Du 2017 [27]	High	Low	Low	Unclear	Low	High	Low	Low
Fosler 2017 [28]	High	Unclear	Unclear	High	High	High	Low	Unclear
Lee 2015 [22]	Unclear	Unclear	Low	Low	High	High	Low	Low
Li 2015 [21]	Unclear	Low	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear
Livergood 2017 [29]	High	Unclear	NA	Low	Low	High	Low	NA
Meck 2015 [30]	High	Low	Low	High	Unclear	Unclear	Unclear	Low
Papageorghiou 2016 [35]	High	High	Low	High	High	High	Low	Low
Sarno 2016 [31]	Unclear	Low	Low	High	Unclear	Low	Low	Low
Strom 2017 [32] Validation study	High	High	Unclear	Unclear	High	High	Low	Unclear
Strom 2017 [32] Clinical implementation	High	Low	Unclear	High	High	Unclear	Low	High

Study	Risk of bias					Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Role of sponsor	Patient selection	Index test	Reference standard
Tan 2016 [33]	High	Low	Unclear	High	High	High	Low	Unclear
Valderramos 2016 [34]	High	Low	Low	High	Unclear	Unclear	Low	Low
Yang 2018 [23]	Unclear	Low	Unclear	High	Low	High	Low	Unclear
b) Studies on DNA microarray-based cfDNA testing (n=5)								
Gil 2017 [36]	High	Low	Unclear	High	Unclear	Low	Low*	Unclear
Juneau 2014 [15]	High	Unclear	Low**	High	High	High	Low	Low**
Kagan 2018 [50]	Low	Low	Low	High	Unclear	High	Low	Low
Langlois 2017 [51]	Unclear	Low	Low	High	Low	High	High*	Low
Stokowski 2015 [52]	Unclear	Unclear	Low	High	High	High	Low	Low

* Confirmed as DNA microarray-based cfDNA test by personal communication with Ariosa Diagnostics Inc. (San Jose, CA).

** Rating for subgroup of 392 samples with suitable reference standard that were included in the review.

Appendix 4 – Test performance outcomes

Test accuracy outcomes for all included studies are reported below. For twin/multiple pregnancies, cases were categorised as (1) 'true positive' (TP) if cfDNA test result was positive and matched the karyotype or birth outcome of at least one fetus/baby; (2) 'false positive' (FP) if cfDNA test result was positive and did not match the karyotype or birth outcome of either fetus/baby; (3) 'true negative' (TN) if the cfDNA test result was negative and all fetuses/babies were determined to be unaffected by karyotyping or birth outcome; and (4) 'false negative' (FN) if at least one fetus/baby were determined to be affected by karyotyping or birth outcome.

Table 15. Test accuracy outcomes in individual studies (n=21).

Reference	Fetal Fraction	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						
a) cfDNA testing in twin/multiple pregnancies (n=16)												
Benachi 2015 [24] Twins	NR	T21	*2*	*5*	*0*	*0*	*100 (19.8-100)*	*100 (46.3-100)*	*100 (19.8-100)*	*100 (46.3-100)*	NR	0 test failures / 0 inconclusive results / Exclusions from analysis not reported separately for singleton and twin pregnancies (8 without fetal karyotype results)
Singletons	NR	T21	*74*	*804*	*1*	*0*	*100 (93.9-100)*	*99.88 (99.2-99.99)*	*98.7 (91.8-99.9)*	*100 (99.4-100)*	NR	6 test failures / 0 inconclusive results / Exclusions from analysis not reported separately for singleton and twin pregnancies (8 without fetal karyotype results)

Reference	Fetal Fraction	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						
Beulen 2017 [25] Twins	NR	T21	*1*	*16*	*0*	*0*	*100 (5.5-100)*	*100 (75.9-100)*	*100 (5.5-100)*	*100 (75.9-100)*	NR	0 test failures / 0 inconclusive results / 4 without reference standard.
		T18	*0*	*17*	*0*	*0*	NA	*100 (77.1-100)*	NA	*100 (77.1-100)*	NR	
		T13	*0*	*16*	*0*	*1*	*0 (0-94.5)*	*100 (75.9-100)*	NA	*94.1 (69.2-99.7)*	NR	
Singletons	NR	T21	*11*	*204*	*0*	*0*	*100 (67.9-100)*	*100 (97.7-100)*	*100 (67.9-100)*	*100 (97.7-100)*	NR	1 test failure / 0 inconclusive result / 14 without reference standard.
		T18	*5*	*210*	*0*	*0*	*100 (46.3-100)*	*100 (97.8-100)*	*100 (46.3-100)*	*100 (97.8-100)*	NR	
		T13	*3*	*211*	*0*	*1*	*75.0 (21.9-98.7)*	*100 (97.8-100)*	*100 (31.0-100)*	*99.5 (97.0-99.98)*	NR	
Brison 2018 [26]	NR	T21	1	NR	NR	NR	NA	NA	NA	NA	NR	Test failures NR / Inconclusive results NR / *19,732* excluded from the analysis.
		T18	NR	NR	NR	1	NA	NA	NA	NA	NR	
		T13	1	NR	NR	NR	NA	NA	NA	NA	NR	
Du 2017 [27]	NR	T21	2	*90*	0	0	*100 (19.8-100)*	*100 (94.9-100)*	*100 (19.8-100)*	*100 (94.9-100)*	NR	0 test failures / 0 inconclusive results / 0 exclusions from analysis (reference standard unclear for 2 cases).
		T18	0	*92*	0	0	NA	*100 (95.0-100)*	NA	*100 (95.0-100)*	NR	
		T13	0	*91*	1	0	NA	*98.9 (93.2-99.9)*	*0 (0-94.5)*	*100 (95.0-100)*	NR	
Fosler 2017 [28]	NR	T21	6	*162*	1	0	*100 (51.7-100)*	*99.4 (96.1-99.97)*	*85.7 (42.0-99.2)*	*100 (97.1-100)*	NR	8 tests were cancelled / 2 had “trisomy suspected” / 308 without reference standard. 2 with ultrasound finding as reference standard.
		T18	0	*169*	0	0	NA	*100 (97.2-100)*	NA	*100 (97.2-100)*	NR	
		T13	0	*169*	0	0	NA	*100 (97.2-100)*	NA	*100 (97.2-100)*	NR	

Reference	Fetal Fraction	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						
Lee 2015 [22] Multiples	NR	T21	*0*	*2*	*0*	*0*	NA	*100 (19.8-100)*	NA	*100 (19.8-100)*	NR	0 test failures / 0 inconclusive results / 0 exclusions from analysis.
		T18	*0*	*2*	*0*	*0*	NA	*100 (19.8-100)*	NA	*100 (19.8-100)*	NR	
		T13	*0*	*2*	*0*	*0*	NA	*100 (19.8-100)*	NA	*100 (19.8-100)*	NR	
Singletons	NR	T21	*5*	*85*	*0*	*0*	*100 (46.3-100)*	*100 (94.6-100)*	*100 (46.3-100)*	*100 (94.6-100)*	NR	0 test failures / 0 inconclusive results / 0 exclusions from analysis.
		T18	*2*	*88*	*0*	*0*	*100 (19.8-100)*	*100 (94.8-100)*	*100 (19.8-100)*	*100 (94.8-100)*	NR	
		T13	*1*	*89*	*0*	*0*	*100 (5.5-100)*	*100 (94.8-100)*	*100 (5.5-100)*	*100 (94.8-100)*	NR	
Li 2015 [21] Multiples (12 twins, 1 triplet)	NR	T21	0	*13*	0	0	NA	*100 (71.7-100)*	NA	*100 (71.7-100)*	NR	0 test failures / 0 inconclusive results / 0 exclusions from analysis.
		T18	0	13	0	0	NA	*100 (71.7-100)*	NA	*100 (71.7-100)*	NR	
		T13	0	13	0	0	NA	*100 (71.7-100)*	NA	*100 (71.7-100)*	NR	
Singletons	NR	T21	*0*	*156*	*0*	*0*	NA	*100 (97.0-100)*	NA	*100 (97.0-100)*	NR	0 test failures / 0 inconclusive results / 0 exclusions from analysis.
		T18	*1*	*155*	*0*	*0*	*100 (5.5-100)*	*100 (97.0-100)*	*100 (5.5-100)*	*100 (97.0-100)*	NR	
		T13	*0*	*156*	*0*	*0*	NA	*100 (97.0-100)*	NA	*100 (97.0-100)*	NR	
Livergood 2017 [29]	NR	This study did not assess test accuracy, only test failure rate.					NA	NA	NA	NA	NR	In total 105/2,385 (4.4%) "no call". Multiple gestations: 2/72 (2.8%) with "no call". / No exclusions from analysis.

Reference	Fetal Fraction	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						
Meck 2015 [30]	NR	Mention 2 FP cases for T21 in twin pregnancies. 1 case involved a demise of 1 fetus at 10 weeks' gestation with no karyotype available on the deceased twin.					NA	*NA*	NA	NA	NR	In total (singleton and twin pregnancies), 4 with inconclusive or failed NIPT / All with reference standard (inclusion criteria).
Papageorghiou 2016 [35] Twins	NR	T21	1	*10*	0	0	*100 (5.5-100)*	*100 (65.5-100)*	*100 (5.5-100)*	*100 (65.5-100)*	NR	5 test failures in total / 0 inconclusive results / 0 without reference standard.
		T18	1	*10*	0	0	*100 (5.5-100)*	*100 (65.5-100)*	*100 (5.5-100)*	*100 (65.5-100)*	NR	
		T13	0	*11*	0	0	NA	*100 (67.9-100)*	NA	*100 (67.9-100)*	NR	
Singletons	NR	T21	42	*384*	0	0	*100 (89.6-100)*	*100 (98.8-100)*	*100 (89.6-100)*	*100 (98.8-100)*	NR	
		T18	9	*417*	0	0	*100 (62.9-100)*	*100 (98.9-100)*	*100 (62.9-100)*	*100 (98.9-100)*	NR	
		T13	5	*421*	0	0	*100 (46.3-100)*	*100 (98.9-100)*	*100 (46.3-100)*	*100 (98.9-100)*	NR	
Sarno 2016 [31] Twins	Median 8.0%	T21	8	*409*	*0*	0	100 *(59.8-100)*	*100 (98.8-100)*	*100 (59.8-100)*	*100 (98.8-100)*	Total FP rate 1/404 (0.25%)	21 test failures / 0 inconclusive results / 0 without reference standard.
		T18	3	*413**	*0**	1	75.0 *(21.9-98.7)*	*100 (98.9-100)*	*100 (31.0-100)*	*99.76 (98.4-99.99)*		
		T13	0	*415**	*1**	1	0 *(0.0-94.5)*	*99.76 (98.5-99.99)*	0 *(0.0-94.5)*	*99.76 (98.5-99.99)*		
Singletons	Median 11.0%	T21	156	NR	NR	2	98.7 *(95.0-99.7)*	NA	NA	NA	Total FP rate 23/10,311 (0.22%)	168 test failures / 0 inconclusive results / 0 without reference standard.
		T18	41	NR	NR	5	89.1 *(75.6-95.9)*	NA	NA	NA		
		T13	8	NR	NR	7	53.5 *(27.4-77.7)*	NA	NA	NA		

Reference	Fetal Fraction	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						
Strom 2017 [32] Validation study	NR	T21	10	*105*	0	0	*100 (65.5-100)*	*100 (95.6-100)*	*100 (65.5-100)*	*100 (95.6-100)*	NR	0 test failures / 0 inconclusive results / 0 without reference standard.
		T18	4	*111*	0	0	*100 (39.6-100)*	*100 (95.8-100)*	*100 (39.6-100)*	*100 (95.8-100)*	NR	
		T13	13	*102*	0	0	*100 (71.7-100)*	*100 (95.5-100)*	*100 (71.7-100)*	*100 (95.5-100)*	NR	
Clinical implementation: Twins	NR	T21	*3*	NR	*0*	NR	NA	NA	*100 (31.0-100)*	NA	NR	In total (twin and singleton pregnancies) 92 test failures / 0 inconclusive results / *10,529* without reference standard.
		T18	*0*	NR	*1*	NR	NA	NA	*0 (0.0-94.5)*	NA	NR	
		T13	*0*	NR	*0*	NR	NA	NA	NA	NA	NR	
Clinical implementation: Singletons	NR	T21	*38*	NR	*1*	NR	NA	NA	*97.4 (84.9-99.9)*	NA	NR	
		T18	*23*	NR	*1*	NR	NA	NA	*95.8 (76.9-99.8)*	NA	NR	
		T13	*9*	NR	*4*	NR	NA	NA	*69.2 (38.9-89.6)*	NA	NR	
Tan 2016 [33]	Mean 8.9% SD 4.2%	T21	4	*506*	0	0	*100 (39.6-100)*	*100 (99.1-100)*	*100 (39.8-100)*	*100 (99.1-100)*	NR	5 test failures / 0 inconclusive results / 50 without reference standard.
		T18	0	*510*	0	0	NA	*100 (99.1-100)*	NA	*100 (99.1-100)*	NR	
		T13	0	*510*	0	0	NA	*100 (99.1-100)*	NA	*100 (99.1-100)*	NR	
Valderramos 2016 [34]	NR	T21	4	NR	0	NR	NA	NA	*100 (39.6-100)*	NA	NR	13 non-reportable results / In total 38 cases excluded from the analysis.
		T18	1	NR	0	NR	NA	NA	*100 (5.5-100)*	NA	NR	
		T13	1	NR	2	NR	NA	NA	*33.3 (1.8-87.5)*	NA	NR	
Yang 2018 [23]	NR	T21	4	*369*	0	0	*100 (39.6-100)*	*100 (98.7-100)*	*100 (39.6-100)*	*100 (98.7-100)*	NR	0 test failures / 0 inconclusive results / 59 without reference standard.
		T18	1	*372*	0	0	*100 (5.5-100)*	*100 (98.7-100)*	*100 (5.5-100)*	*100 (98.7-100)*	NR	
		T13	*0*	*373*	*0*	*0*	NA	*100 (98.7-100)*	NA	*100 (98.7-100)*	NR	

Reference	Fetal Fraction	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						
b) DNA microarray-based cfDNA testing (n=5)												
Gil 2017 [36; 65]	NR	T21	1	53	0	0	*100 (5.4-100)*	*100 (91.6-100)*	*100 (5.4-100)*	*100 (91.6-100)*	NR	0 test failures after repeat test / 0 inconclusive results / 18 women with higher-chance FTCS result did not choose cfDNA testing.
		T18	0	*54***	0	*0***	NA	*100 (91.7-100)*	NA	*100 (91.7-100)*		
		T13	0	*54***	0	*0***	NA	*100 (91.7-100)*	NA	*100 (91.7-100)*		
Juneau 2014 [15] DNA microarray-based test	NR	T21	72**	*320***	0**	0**	*100 (93.7-100)*	*100 (98.5-100)*	*100 (93.7-100)*	*100 (98.5-100)*	NR	Test failures excluded from study / 0 inconclusive results / 486 with unclear reference standard.
		T18	13**	*379***	0**	0**	*100 (71.7-100)*	*100 (98.7-100)*	*100 (71.7-100)*	*100 (98.7-100)*		
		T13	7**	*385***	0**	0**	*100 (56.1-100)*	*100 (98.8-100)*	*100 (56.1-100)*	*100 (98.8-100)*		
Sequencing-based test	NR	T21	72**	*320***	0**	0**	*100 (93.7-100)*	*100 (98.5-100)*	*100 (93.7-100)*	*100 (98.5-100)*	NR	Test failures excluded from study / 0 inconclusive results / 486 with unclear reference standard.
		T18	13**	*379***	0**	0**	*100 (71.7-100)*	*100 (98.7-100)*	*100 (71.7-100)*	*100 (98.7-100)*		
		T13	7**	*385***	0**	0**	*100 (56.1-100)*	*100 (98.8-100)*	*100 (56.1-100)*	*100 (98.8-100)*		
Kagan 2018 [50] cfDNA testing	Median 12.5%	T21	*0*	*678*	*0*	*0*	NA	*100 (99.3-100)*	NA	*100 (99.3-100)*	FP rate T21: 0%	10 test failures / 0 inconclusive results / 13 without reference standard.
		T18	*0*	*678*	*0*	*0*	NA	*100 (99.3-100)*	NA	*100 (99.3-100)*		
		T13	*0*	*678*	*0*	*0*	NA	*100 (99.3-100)*	NA	*100 (99.3-100)*		
FTCS	NA	T21	0	671	17	0	NA	*97.5 (96.0-98.5)*	*0 (0-22.9)*	*100 (99.3-100)*	FP rate T21: 2.5%	0 test failures / 0 inconclusive results / 11 without reference standard.

Reference	Fetal Fraction	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						
Langlois 2017 (51) cfDNA testing	NR	T21	6	*1,153*	0	0	*100 (51.7-100)*	*100 (99.6-100)*	*100 (51.7-100)*	*100 (99.6-100)*	FP rate T21: 0% FP rate T18: 0.09% FP rate T13: 0.09%	6 test failures after repeat test / 0 inconclusive results / 33 without reference standard.
		T18	0	*1,158*	1	0	NA	*99.9 (99.4-100)*	*0 (0-94.5)*	*100 (99.6-100)*		
		T13	0	*1,158*	1	0	NA	*99.9 (99.4-100)*	*0 (0-94.5)*	*100 (99.6-100)*		
FTCS	NA	T21	5	263	19	0	*100 (46.3-100)*	*93.3 (89.5-95.8)*	*20.8 (7.9-42.7)*	*100 (98.2-100)*	NR	0 test failures / 0 inconclusive results / 33 without reference standard / 878 other standard screening test than FTCS.
Stokowski 2015 (52)	Median 13.8% IQR 10.7-16.9%	T21	107	*683*	0	1	99.1 *(94.2-99.95)*	100 *(99.3-100)*	*100 (95.7-100)*	*99.85 (99.1-99.99)*	NR	8 test failures / 0 inconclusive results.
		T18	29	*761*	0	1	96.7 *(80.9-99.8)*	100 *(99.4-100)*	*100 (85.4-100)*	*99.87 (99.2-99.99)*		
		T13	12	*779*	0	0	100 *(69.9-100)*	100 *(99.4-100)*	*100 (69.9-100)*	*100 (99.4-100)*		

cfDNA, cell-free deoxyribonucleic acid; CI, confidence interval; FTCS, First trimester combined screening; FP, false positive; FP rate = FP / (FP+TN) = 1 – Specificity; FN, false negative; FN rate = FN / (FN+TP) = 1 – Sensitivity; IQR, interquartile range; NA, not applicable; NR, not reported; NPV, negative predictive value; PPV, positive predictive value; 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.

* Unpublished data received by email from a co-worker (Dr Maria del Mar Gil) of the corresponding author.
** Unpublished data received from Ariosa Diagnostics Inc. (San Jose, CA) on 12th October 2018.

Table 16. Initial test failure rates, reasons for failure and success of repeat tests (research question 4)

Reference / Population	Initial cfDNA test failure rate					Repeat tests successful	Causes of cfDNA test failures	Failure rate of combined test
	8 weeks	10 weeks	12 weeks	14 weeks	Total			
a) In twin/multiple pregnancies (10 studies)								
Benachi 2015 [24]	NR	NR	NR	NR	*Twins: 0/7* *Singletons: 6/885 (0.7%)*	42/42 for whom the test had to be run twice because of technical issues during the first assay (ie, failure in library preparation).	Fetal fraction < 4% in 4/6 or result appeared atypical (ie, positive Z-score for more than one chromosome).	NA
Beulen 2017 [25]	NR	NR	NR	NR	*Twins: 0/21** *Singletons: 2/230 (0.4%)**	NR	NIPT did not meet quality criteria.	NA
Du 2017 [27]	NR	NR	NR	NR	0/92*	NR	NA	NA
Fosler 2017 [28]	NR	NR	NR	NR	8/487 *(1.6%)*	Not performed.	8 tests were cancelled (e.g. testing before 10 weeks gestation, insufficient sample quantity, patient or physician request).	NA
Li 2015 [21]	NR	NR	NR	NR	*Multiples: 0/13** *Singletons: 0/156**	NR	NA	NA
Livergood [29]	NR	NR	NR	NR	Multiples: *2/72 (2.8%)**	NR	NR	NA
Papageorghiou 2016 [35]	NR	NR	NR	NR	5/442 (1.1%)	Not performed.	Low fetal fraction in 3/442 (0.7%) and low counts in 2/442 (0.4%).	NA
Sarno 2016 [31]	NR	NR	NR	NR	Twins: 41/438 (9.4%) Singletons: 316/10,698 (2.9%) (p<0.0001)	Twins: 20/39 (51.3%) Singletons: 148/235 (63.0%)	NR	NR
Tan [33]	NR	NR	NR	NR	18/565 (3.2%)	*13/15 (87%)*	Failing quality control (low fetal fraction).	NR
Yang 2018 [23]	NR	NR	NR	NR	0/432*	NR	NR (NA?)	NA

Reference / Population	Initial cfDNA test failure rate					Repeat tests successful	Causes of cfDNA test failures	Failure rate of combined test
	8 weeks	10 weeks	12 weeks	14 weeks	Total			
b) Using DNA microarray-based approach (4 studies)								
Gil 2017 [36] Singletons *12-14 weeks*	NR	NR	NR	NR	1/54 *(1.9%)*	2nd blood draw: 1/1	NR	NR
Kagan 2018 [50] Singletons 11-13 weeks	NR	NR	NR	NR	10/688 (1.5%)	Not performed	NR	0/688
Langlois 2017 [51] Singletons 10-14 weeks	NR	NR	NR	NR	11/1,165 (0.9%, 95% CI, 0.47-1.7%)	2nd blood draw: 5/11 (45.5%)	1st blood draw: 10/11 low fetal fraction; 1/11 unusually high variance in cfDNA count.	0/287
Stokowski 2015 [52] Singletons & twins Median 16 weeks	NR	NR	NR	NR	8/799* *(1.0%)*	NR	8/8 insufficient fetal DNA	NA

cf, cell-free; CI, confidence interval; DNA, deoxyribonucleic acid; NA, not applicable; NR, not reported.

* Unclear if the reported failure rate is after initial testing or includes repeat testing.

Numbers in italics calculated by reviewers from information given in the paper.

Appendix 5 – UK NSC reporting checklist for evidence summaries

All items on the UK NSC Reporting Checklist for Evidence Summaries have been addressed in this report. A summary of the checklist, along with the page or pages where each item can be found in this report, is presented in Table 17.

Table 17. UK NSC reporting checklist for evidence summaries

	Section	Item	Page no.
1.	TITLE AND SUMMARIES		
1.1	Title sheet	Identify the review as a UK NSC evidence summary.	Title page
1.2	Plain English summary	Plain English description of the executive summary.	5-6
1.3	Executive summary	Structured overview of the whole report. To include: the purpose/aim of the review; background; previous recommendations; findings and gaps in the evidence; recommendations on the screening that can or cannot be made on the basis of the review.	7-19
2.	INTRODUCTION AND APPROACH		
2.1	Background and objectives	Background – Current policy context and rationale for the current review – for example, reference to details of previous reviews, basis for current recommendation, recommendations made, gaps identified, drivers for new reviews Objectives – What are the questions the current evidence summary intends to answer? – statement of the key questions for the current evidence summary, criteria they address, and number of studies included per question, description of the overall results of the literature search. Method – briefly outline the rapid review methods	20-26

		used.	
2.2	Eligibility for inclusion in the review	State all criteria for inclusion and exclusion of studies to the review clearly (PICO, dates, language, study type, publication type, publication status etc.) To be decided *a priori*.	26-28
2.3	Appraisal for quality/risk of bias tool	Details of tool/checklist used to assess quality, e.g. QUADAS 2, CASP, SIGN, AMSTAR.	29-30
3.	SEARCH STRATEGY AND STUDY SELECTION (FOR EACH KEY QUESTION)		
3.1	Databases/ sources searched	Give details of all databases searched (including platform/interface and coverage dates) and date of final search.	26
3.2	Search strategy and results	Present the full search strategy for at least one database (usually a version of Medline), including limits and search filters if used. Provide details of the total number of (results from each database searched), number of duplicates removed, and the final number of unique records to consider for inclusion.	Appendix 1 (73-78)
3.3	Study selection	State the process for selecting studies – inclusion and exclusion criteria, number of studies screened by title/abstract and full text, number of reviewers, any cross checking carried out.	Appendix 2 (79-96)
4.	STUDY LEVEL REPORTING OF RESULTS (FOR EACH KEY QUESTION)		
4.1	Study level reporting, results and risk of bias assessment	For each study, produce a table that includes the full citation and a summary of the data relevant to the question (for example, study size, PICO, follow-up period, outcomes reported, statistical analyses etc.). Provide a simple summary of key measures, effect estimates and confidence intervals for each study where available. For each study, present the results of any assessment of quality/risk of bias.	Appendix 3 97-146 Appendix 4 149-157 147-148

4.2	Additional analyses	Describe additional analyses (for example, sensitivity, specificity, PPV, etc.) carried out by the reviewer.	30-32
5.	QUESTION LEVEL SYNTHESIS		
5.1	Description of the evidence	For each question, give numbers of studies screened, assessed for eligibility, and included in the review, with summary reasons for exclusion.	a) 34 b) 49
5.2	Combining and presenting the findings	Provide a balanced discussion of the body of evidence which avoids over reliance on one study or set of studies. Consideration of four components should inform the reviewer's judgement on whether the criterion is 'met', 'not met' or 'uncertain': quantity; quality; applicability and consistency.	a) 34-47 b) 49-62
5.3	Summary of findings	Provide a description of the evidence reviewed and included for each question, with reference to their eligibility for inclusion. Summarise the main findings including the quality/risk of bias issues for each question. Have the criteria addressed been 'met', 'not met' or 'uncertain'?	63-65
6.	REVIEW SUMMARY		
6.1	Conclusions and implications for policy	Do findings indicate whether screening should be recommended? Is further work warranted? Are there gaps in the evidence highlighted by the review?	66-69
6.2	Limitations	Discuss limitations of the available evidence and of the review methodology if relevant.	69-72

References

1. Taylor-Phillips S, Freeman K, Geppert J *et al.* (2015) *Systematic review and cost-consequence assessment of cell-free DNA testing for T21, T18 and T13 in the UK – Final report*. https://legacyscreening.phe.org.uk/fetalanomalies (Last external review - Systematic review: cfDNA).

2. UK National Screening Committee (2016) UK NSC non-invasive prenatal testing (NIPT) recommendation. https://legacyscreening.phe.org.uk/fetalanomalies (Recommendation statement).

3. Coppus AM, Evenhuis HM, Verberne GJ *et al.* (2008) Survival in elderly persons with Down syndrome. *J Am Geriatr Soc* **56**, 2311-2316.

4. NHS (2017) Edwards' syndrome (trisomy 18) 08/10/2018)

5. Morris JK, Springett A (2014) *The National Down Syndrome Cytogenetic Register for England and Wales: 2013 Annual Report*. http://www.binocar.org/content/annrep2013_FINAL.pdf.

6. Wu J, Springett A, Morris JK (2013) Survival of trisomy 18 (Edwards syndrome) and trisomy 13 (Patau Syndrome) in England and Wales: 2004-2011. *Am J Med Genet A* **161A**, 2512-2518.

7. Tidy C (2016) Patau's Syndrome - Trisomy 13. (accessed 08/10/2018

8. Carothers AD, Hecht CA, Hook EB (1999) International variation in reported livebirth prevalence rates of Down syndrome, adjusted for maternal age. *J Med Genet* **36**, 386-393.

9. Morris JK, Mutton DE, Alberman E (2002) Revised estimates of the maternal age specific live birth prevalence of Down's syndrome. *J Med Screen* **9**, 2-6.

10. Kagan KO, Sonek J, Wagner P *et al.* (2017) Principles of first trimester screening in the age of non-invasive prenatal diagnosis: screening for chromosomal abnormalities. *Arch Gynecol Obstet* **296**, 645-651.

11. Beta J, Lesmes-Heredia C, Bedetti C *et al.* (2018) Risk of miscarriage following amniocentesis and chorionic villus sampling: a systematic review of the literature. *Minerva Ginecol* **70**, 215-219.

12. Sehnert AJ, Rhees B, Comstock D *et al.* (2011) Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free fetal DNA from maternal blood. *Clin Chem* **57**, 1042-1049.

13. Sparks AB, Struble CA, Wang ET *et al.* (2012) Noninvasive prenatal detection and selective analysis of cell-free DNA obtained from maternal blood: evaluation for trisomy 21 and trisomy 18. *Am J Obstet Gynecol* **206**, 319 e311-319.

14. Zimmermann B, Hill M, Gemelos G *et al.* (2012) Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci. *Prenat Diagn* **32**, 1233-1241.

15. Juneau K, Bogard PE, Huang S *et al.* (2014) Microarray-based cell-free DNA analysis improves noninvasive prenatal testing. *Fetal Diagn Ther* **36**, 282-286.

16. Bumgarner R (2013) Overview of DNA microarrays: types, applications, and their future. *Curr Protoc Mol Biol* **Chapter 22**, Unit 22 21.

17. Hartwig TS, Ambye L, Sorensen S *et al.* (2017) Discordant non-invasive prenatal testing (NIPT) - a systematic review. *Prenat Diagn* **37**, 527-539.

18. Gil MM, Accurti V, Santacruz B *et al.* (2017) Analysis of cell-free DNA in maternal blood in screening for aneuploidies: updated meta-analysis. *Ultrasound in Obstetrics & Gynecology* **50**, 302-314.

19. Whiting PF, Rutjes AW, Westwood ME *et al.* (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine* **155**, 529-536.

20. Badeau M, Lindsay C, Blais J *et al.* (2017) Genomics-based non-invasive prenatal testing for detection of fetal chromosomal aneuploidy in pregnant women. *Cochrane Database Syst Rev* **11**, CD011767.

21. Li WH, Wang PH, Chuang CM *et al.* (2015) Noninvasive prenatal testing for fetal trisomy in a mixed risk factors pregnancy population. *Taiwan* **54**, 122-125.

22. Lee M-Y, Cho D-Y, Won H-S *et al.* (2015) Performance of Momguard, a new non-invasive prenatal testing protocol developed in Korea. *Obstet* **58**, 340-345.

23. Yang J, Qi Y, Hou Y *et al.* (2018) Performance of non-invasive prenatal testing for trisomies 21 and 18 in twin pregnancies. *Mol Cytogenet* **11**, 47.

24. Benachi A, Letourneau A, Kleinfinger P *et al.* (2015) Cell-free DNA analysis in maternal plasma in cases of fetal abnormalities detected on ultrasound examination. *Obstet Gynecol* **125**, 1330-1337.

25. Beulen L, Faas BHW, Feenstra I *et al.* (2017) Clinical utility of non-invasive prenatal testing in pregnancies with ultrasound anomalies. *Ultrasound in Obstetrics & Gynecology* **49**, 721-728.

26. Brison N, Neofytou M, Dehaspe L *et al.* (2018) Predicting fetoplacental chromosomal mosaicism during non-invasive prenatal testing. *Prenat Diagn* **38**, 258-266.

27. Du E, Feng C, Cao Y *et al.* (2017) Massively Parallel Sequencing (MPS) of Cell-Free Fetal DNA (cffDNA) for Trisomies 21, 18, and 13 in Twin Pregnancies. *Twin Res Hum Genet* **20**, 242-249.

28. Fosler L, Winters P, Jones KW *et al.* (2017) Aneuploidy screening by non-invasive prenatal testing in twin pregnancy. *Ultrasound in Obstetrics & Gynecology* **49**, 470-477.

29. Livergood MC, LeChien KA, Trudell AS (2017) Obesity and cell-free DNA "no calls": is there an optimal gestational age at time of sampling? *Am J Obstet Gynecol* **216**, 413.e411-413.e419.

30. Meck JM, Kramer Dugan E, Matyakhina L *et al.* (2015) Noninvasive prenatal screening for aneuploidy: positive predictive values based on cytogenetic findings. *Am J Obstet Gynecol* **213**, 214.e211-215.

31. Sarno L, Revello R, Hanson E *et al.* (2016) Prospective first-trimester screening for trisomies by cell-free DNA testing of maternal blood in twin pregnancy. *Ultrasound in Obstetrics & Gynecology* **47**, 705-711.

32. Strom CM, Anderson B, Tsao D *et al.* (2017) Improving the Positive Predictive Value of Non-Invasive Prenatal Screening (NIPS). *PLoS ONE* **12**, e0167130.

33. Tan Y, Gao Y, Lin G *et al.* (2016) Noninvasive prenatal testing (NIPT) in twin pregnancies with treatment of assisted reproductive techniques (ART) in a single center. *Prenat Diagn* **36**, 672-679.

34. Valderramos SG, Rao RR, Scibetta EW *et al.* (2016) Cell-free DNA screening in clinical practice: abnormal autosomal aneuploidy and microdeletion results. *Am J Obstet Gynecol* **215**, 626.e621-626.e610.

35. Papageorghiou AT, Khalil A, Forman M *et al.* (2016) Clinical evaluation of the IONA test: a non-invasive prenatal screening test for trisomies 21, 18 and 13. *Ultrasound in Obstetrics & Gynecology* **47**, 188-193.

36. Gil MM, Brik M, Casanova C *et al.* (2017) Screening for trisomies 21 and 18 in a Spanish public hospital: from the combined test to the cell-free DNA test. *Journal of Maternal-Fetal & Neonatal Medicine* **30**, 2476-2482.

37. del Mar Gil M, Quezada MS, Bregant B *et al.* (2014) Cell-free DNA analysis for trisomy risk assessment in first-trimester twin pregnancies. *Fetal Diagn Ther* **35**, 204-211.

38. Huang X, Zheng J, Chen M *et al.* (2014) Noninvasive prenatal testing of trisomies 21 and 18 by massively parallel sequencing of maternal plasma DNA in twin pregnancies. *Prenat Diagn* **34**, 335-340.

39. Shaw SW, Hsiao CH, Chen CY *et al.* (2014) Noninvasive prenatal testing for whole fetal chromosomal aneuploidies: a multicenter prospective cohort trial in Taiwan. *Fetal Diagn Ther* **35**, 13-17.

40. Beamon CJ, Hardisty EE, Harris SC *et al.* (2014) A single center's experience with noninvasive prenatal testing. *Genet Med* **16**, 681-687.

41. Bevilacqua E, Gil MM, Nicolaides KH *et al.* (2015) Performance of screening for aneuploidies by cell-free DNA analysis of maternal blood in twin pregnancies. *Ultrasound Obstet Gynecol* **45**, 61-66.

42. Zhang H, Gao Y, Jiang F *et al.* (2015) Non-invasive prenatal testing for trisomies 21, 18 and 13: clinical experience from 146,958 pregnancies. *Ultrasound Obstet Gynecol* **45**, 530-538.

43. Bevilacqua E, Jani JC, Letourneau A *et al.* (2018) Cell-Free DNA Analysis in Maternal Blood: Differences in Estimates between Laboratories with Different Methodologies Using a Propensity Score Approach. *Fetal Diagn Ther*, 1-10.

44. Bjerregaard L, Stenbakken AB, Andersen CS *et al.* (2017) The rate of invasive testing for trisomy 21 is reduced after implementation of NIPT. *Dan Med J* **64**.

45. Chan N, Smet ME, Sandow R *et al.* (2018) Implications of failure to achieve a result from prenatal maternal serum cell-free DNA testing: a historical cohort study. *BJOG* **125**, 848-855.

46. Gil MM, Revello R, Poon LC *et al.* (2016) Clinical implementation of routine screening for fetal trisomies in the UK NHS: cell-free DNA test contingent on results from first-trimester combined test. *Ultrasound in Obstetrics & Gynecology* **47**, 45-52.

47. Lee TJ, Rolnik DL, Menezes MA *et al.* (2018) Cell-free fetal DNA testing in singleton IVF conceptions. *Human Reproduction* **33**, 572-578.

48. Miltoft CB, Rode L, Ekelund CK *et al.* (2018) Contingent first-trimester screening for aneuploidies with cell-free DNA in a Danish clinical setting. *Ultrasound in Obstetrics & Gynecology* **51**, 470-479.

49. Revello R, Sarno L, Ispas A *et al.* (2016) Screening for trisomies by cell-free DNA testing of maternal blood: consequences of a failed result. *Ultrasound Obstet Gynecol* **47**, 698-704.

50. Kagan KO, Sroka F, Sonek J *et al.* (2018) First-trimester risk assessment based on ultrasound and cell-free DNA vs combined screening: a randomized controlled trial. *Ultrasound in Obstetrics & Gynecology* **51**, 437-444.

51. Langlois S, Johnson JA, Audibert F *et al.* (2017) Comparison of first-tier cell-free DNA screening for common aneuploidies with conventional publically funded screening. *Prenat Diagn* **37**, 1238-1244.

52. Stokowski R, Wang E, White K *et al.* (2015) Clinical performance of non-invasive prenatal testing (NIPT) using targeted cell-free DNA analysis in maternal plasma with microarrays or next generation sequencing (NGS) is consistent across multiple controlled clinical studies. *Prenat Diagn* **35**, 1243-1246.

53. Sparks AB, Wang ET, Struble CA *et al.* (2012) Selective analysis of cell-free DNA in maternal blood for evaluation of fetal trisomy. *Prenat Diagn* **32**, 3-9.

54. Norton ME, Jacobsson B, Swamy GK *et al.* (2015) Cell-free DNA analysis for noninvasive examination of trisomy. *N Engl J Med* **372**, 1589-1597.

55. Willems PJ, Dierickx H, Vandenakker E *et al.* (2014) The first 3,000 Non-Invasive Prenatal Tests (NIPT) with the Harmony test in Belgium and the Netherlands. *Facts Views Vis Obgyn* **6**, 7-12.

56. Ashoor G, Syngelaki A, Wagner M *et al.* (2012) Chromosome-selective sequencing of maternal plasma cell-free DNA for first-trimester detection of trisomy 21 and trisomy 18. *Am J Obstet Gynecol* **206**, 322 e321-325.

57. Nicolaides KH, Syngelaki A, Ashoor G *et al.* (2012) Noninvasive prenatal testing for fetal trisomies in a routinely screened first-trimester population. *Am J Obstet Gynecol* **207**, 374 e371-376.

58. Norton ME, Brar H, Weiss J *et al.* (2012) Non-Invasive Chromosomal Evaluation (NICE) Study: results of a multicenter prospective cohort study for detection of fetal trisomy 21 and trisomy 18. *Am J Obstet Gynecol* **207**, 137 e131-138.

59. Verweij EJ, Jacobsson B, van Scheltema PA *et al.* (2013) European non-invasive trisomy evaluation (EU-NITE) study: a multicenter prospective cohort study for non-invasive fetal trisomy 21 testing. *Prenat Diagn* **33**, 996-1001.

60. Quezada MS, Gil MM, Francisco C *et al.* (2015) 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* **45**, 36-41.

61. Dan S, Wang W, Ren J *et al.* (2012) Clinical application of massively parallel sequencing-based prenatal noninvasive fetal trisomy test for trisomies 21 and 18 in 11,105 pregnancies with mixed risk factors. *Prenat Diagn* **32**, 1225-1232.

62. Bianchi DW, Platt LD, Goldberg JD *et al.* (2012) Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. *Obstet Gynecol* **119**, 890-901.

63. Futch T, Spinosa J, Bhatt S *et al.* (2013) Initial clinical laboratory experience in noninvasive prenatal testing for fetal aneuploidy from maternal plasma DNA samples. *Prenat Diagn* **33**, 569-574.

64. Liao C, Yin AH, Peng CF *et al.* (2014) Noninvasive prenatal diagnosis of common aneuploidies by semiconductor sequencing. *Proc Natl Acad Sci U S A* **111**, 7415-7420.

65. Gil MM, Accurti V, Santacruz B *et al.* (2017) Analysis of cell-free DNA in maternal blood in screening for aneuploidies: updated meta-analysis. *Ultrasound Obstet Gynecol* **50**, 302-314.