Version 1.0: 31 January 2022

Financial conflicts and manufacturer involvement in the evidence base evaluating the performance of prenatal cfDNA testing for fetal trisomies 21, 18 and 13 – Short protocol

Authors: Julia Geppert
Karoline Munro
Rosanna Fforde
Chris Stinton
Sian Taylor-Phillips

Correspondence to: Dr Julia Geppert
Populations, Evidence and Technologies
Division of Health Sciences
Warwick Medical School
University of Warwick
Coventry CV4 7AL
Email: J.Geppert@warwick.ac.uk

Date completed: 31 January 2022

Aim

This study aims to assess financial conflicts and manufacturer involvement in the evidence base evaluating the performance of prenatal cfDNA testing for fetal trisomies 21, 18 and 13 by

1. Analysing the degree of financial conflicts and involvement of commercial test manufacturers and their change over time;
2. Exploring any associations between the conflicts of interest or manufacturer involvement and the methodological quality of the studies; and
3. Exploring any associations between financial conflicts or manufacturer involvement and test accuracy outcomes.

Methods

Study identification

We will identify the evidence base on the accuracy of prenatal cfDNA testing for fetal trisomy 21 (T21), trisomy 18 (T18) and trisomy 13 (T13) detection using the 2017 Cochrane review on non-invasive prenatal testing (NIPT) for fetal chromosomal aneuploidies.[1] For this systematic review, 13 electronic databases (including MEDLINE, Embase and Web of Science) were searched from 1 January 2007 to 12 July 2016 without any language, search filter or publication type restrictions. Badeau et al. also screened reference lists of relevant full-text articles, websites of private prenatal diagnosis companies and conference abstracts. Eligible were studies that included pregnant women of any age, ethnicity and gestational age with singleton or multifetal pregnancy. The women must have had a screening test for fetal aneuploidy by random whole genome sequencing or targeted sequencing of cfDNA in maternal blood and a reference standard such as fetal karyotype or medical records from birth. More details on the methods of the 2017 Cochrane review[1] can be found in **Appendix 1**.

From the 63 articles that were included in the 2017 Cochrane review,[1] the present analysis will include those that investigated the performance of maternal cfDNA testing for fetal trisomies 21, 18 or 13 detection. Articles reporting on the test performance for the detection of other trisomies or fetal sex chromosome aneuploidies (namely 45,X; 47,XXY; 47,XXX and 47,XYY) only will be excluded from our analysis. One reviewer (JG) will assess the eligibility using information given in the Cochrane review.[1]

Data extraction

Two reviewers (JG, KM, RF) will independently extract information on authors' affiliations, study funding and conflicts of interest declarations from the published original articles or supporting online material provided by the journals. Two reviewers (JG, KM, RF) will independently extract QUADAS ratings and 2x2 tables from the Cochrane review.[1] Disagreements will be resolved by consensus or discussion with a third reviewer (CS).

Definitions

Authors' conflicts of interest

We will use the following definitions to class authors' interests as 'any conflict', 'no conflict' or 'uncertain conflicts' based on the International Committee of Medical Journal Editors (ICMJE)[2] quoted sources of financial conflict that were adapted to the area of prenatal cfDNA testing as follows:

Any conflict:

- Financial relationships of the authors over the 36 months prior to submission of the work with entities in the area of prenatal cfDNA testing that could be perceived to be affected financially by the published work (such as test manufacturers, laboratories/companies processing the test or holding a licence for test processing) or foundations supported by entities that could be perceived to have a financial stake in the outcome.
- These relationships include, but are not limited to, study funding, personal grants or departmental funding outside the published work, personal fees (honoria, royalties, or fees for consulting, lectures, speakers bureaus, paid expert testimony, employment, or other affiliations), stock ownership or options, non-financial support (e.g. free test processing, technical support, travel paid by the entity, writing assistance, administrative support), and patents relevant to cfDNA testing (held or applied for).

Conflict-free: If the authors actively declared they had no conflicts of interest and the study was completely funded by public funding sources, such as government agencies, charitable foundations or academic institutions, that are not perceived to have a financial stake in the outcome.

Uncertain: If there was lack of reporting on conflicts of interest and study funding in the published article and/or online supporting material, and no clear bias was identifiable from the

affiliations of the authors or if an individual author declares no conflicts of interest but there was manufacturer funding (e.g. employment) for other individual authors.

We will then divide the identified conflicts of interest as *higher* or *lower* using the following definitions:

Higher conflict: Patents relevant to the published work (planned, pending or issued), company employment or ownerships, stock holdings or options, consultancies, honoraria, speaking or advisory positions. Conflicted companies include commercial test manufacturers and laboratories/companies processing a cfDNA test or holding a license for cfDNA test processing (e.g. Aria Diagnostics, Ariosa Diagnostics, Berry Genomics, BGI group, Bionet Corp, CERBA, GENOMA group, Genome Care, Genome Ltd., LabGenomics Clinical Research Institute, LifeCodexx, Illumina, Natera, Premaitha Health plc, Sequenom, TheragenEtex Bio Institute, Verinata Health).

Lower conflict: Study grants, departmental funding, payment in kind (for example free test processing, technical support), employment by a laboratory or clinical facility offering (i.e. taking the blood sample and sending it of for processing) cfDNA testing for-profit (e.g. GENDIA, Labco Diagnostics), employment by a laboratory performing whole genome sequencing for other conditions (Clinomics, GATC Biotech AG, The Genomics Institute [TGI] of the Ulsan National Institute of Science and Technology [UNIST]).

The body of literature will be considered collectively, so if one author is clearly conflicted towards ONE manufacturer (e.g. HIGHER conflicts due to patents or company employment) that conflict will be taken to exist for other publications including those where the publication did not offer a declaration if the articles were published within 36 months of each other.

Two reviewers (JG, RF) will independently rate the prevalence and degree of authors' conflicts of interest by using the declared conflicts of interest from the published articles or supporting information (CS). Disagreements will be resolved by consensus or discussion with a third reviewer.

Sources of funding for the study

We will use the following definitions:

Manufacturer funded: Any salaries paid from a commercial test manufacturer to the authors and / or any other resources that the authors received from a commercial test manufacturer either directly or indirectly (via the authors' organisations) for any aspect of the published work, e.g. funding/grants, payment in kind (like provision of free test processing, technical support or personnel for the study). The time frame is that of the work itself, from the initial conception and planning of the study to the submission of the manuscript.

Not manufacturer funded: If no financial support (e.g. salaries, grants, funding or payments in kind) was received from a commercial test manufacturer, neither directly nor indirectly, during any time from the initial conception and planning of the study to the submission of the manuscript.

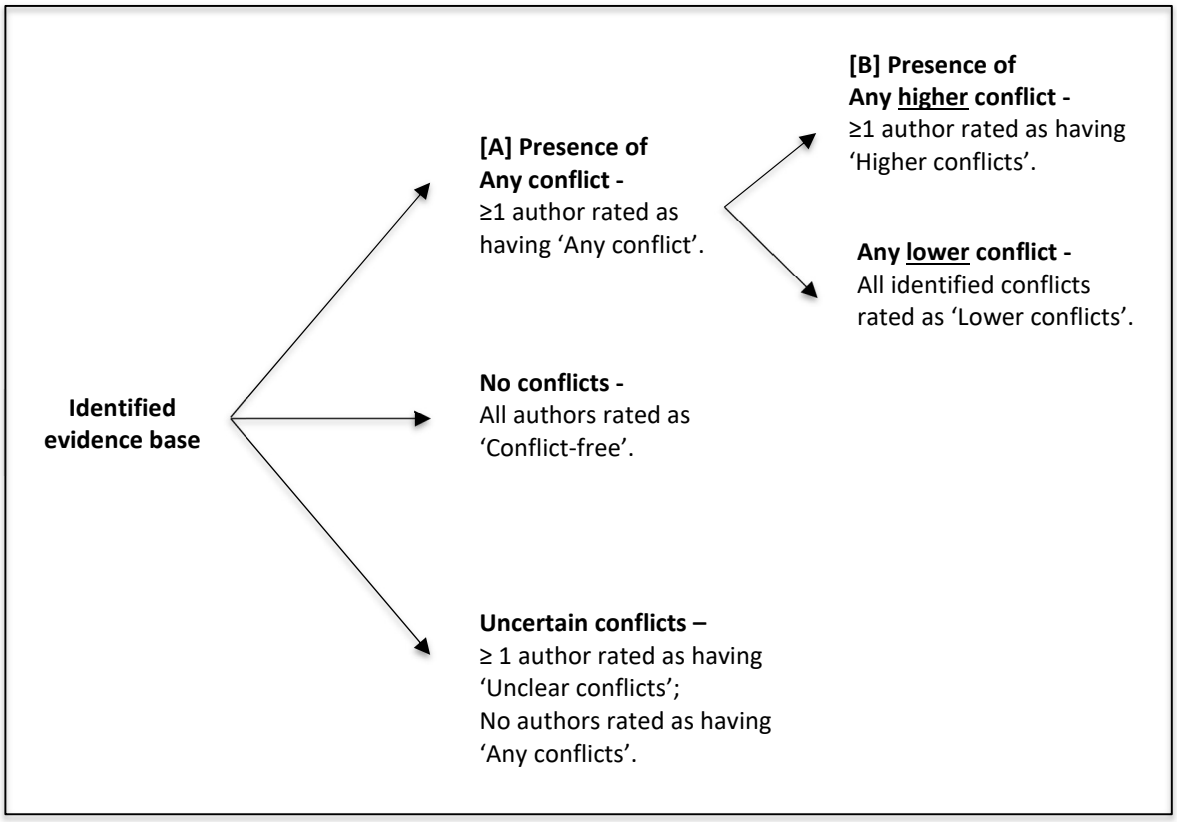
Unclear: Lack of reporting in the published article and/or online supporting material to allow a judgement.

Two reviewers (JG, RF) will independently rate manufacturer involvement by using the statements in the published articles or supporting information. Disagreements will be resolved by consensus or discussion with a third reviewer (CS).

Analysis / synthesis plan

Overall approach

The studies will be categorised based in the number and proportion of conflicted authors as follows:



Three comparisons considering different subsets of the identified evidence base will be explored:

[A] Studies with at least one author having 'any conflicts' versus studies with all authors being rated having 'no conflicts' or 'uncertain conflicts';

[B] Studies with at least one author having 'higher conflicts' versus studies with all authors being rated as having no 'higher conflicts' (i.e. 'lower conflicts', 'uncertain conflicts' or 'no conflicts');

[C] Studies that received 'manufacturer funding' versus studies with 'no manufacturer funding' or 'unclear manufacturer funding'.

Degree of financial conflicts and manufacturer funding in the evidence base

First, we will perform a descriptive analysis of the degree of financial conflicts and involvement of commercial test manufacturers in the evidence base. We will map studies by publication year to assess visually if there seems to be a change in the presence of financial conflicts (comparisons [A] 'Any conflicts' and [B] 'Higher conflicts', respectively) or in the presence of manufacturer involvement (comparison [C]) over time.

Associations with methodological quality

We will use the overall risk of bias ratings (low/high/unclear) as well as the answers to the individual signalling questions (yes/no/unclear) for the four domains from the QUADAS-2 tool[3] as classified by Badeau et al.[1]

In addition, we will dichotomise the study quality in the following way:
Higher quality studies = Cohort studies with either consecutive women included or a random sample of consecutive women;
Lower quality studies = All other study designs.

We will evaluate if the proportion of studies with 'high' risk of bias ratings for each of the four QUADAS-2 domains, the proportion of studies with a 'no' answer (flagging the potential for bias) for each of the 10 corresponding signalling questions and with a 'lower quality' overall quality rating, respectively, differs between the above mentioned subsets of the evidence base (comparisons [A], [B] and [C]).

Frequencies of 'high' risk of bias ratings, individual signalling questions answered with 'no' and 'lower quality' overall rating between the subsets of studies (comparisons [A], [B] and [C]) will be compared using the chi-square test; in cases of expected values smaller than 5, a Fishers exact test will be used. A p value below 0.05 will used to determine whether there is a statistically significant difference in the quality of the studies between groups. Statistical analyses will be performed in Stata software package (version 17; StataCorp, College Station, Texas 77845, USA).

Associations with test accuracy outcomes

We will use the extracted data for the primary studies from the 2017 Cochrane review[1] to obtain the four cell values of a diagnostic 2x2 table in order to calculate test accuracy measures:

sensitivity, specificity and corresponding 95% CI. We will stratify test accuracy measures according to condition (T21, T18 and T13).

We will match the meta-analysis methods from the 2017 Cochrane review.[1] According to Badeau et al., there was a limited or absent threshold effect and therefore no requirement to account for the correlation between sensitivity and specificity across studies in the meta-analyses. Badeau et al. therefore removed the correlation parameter from the bivariate model and simplified it to two univariate random effects logistic models for separate meta-analysis of sensitivity and specificity.

If the random-effects logistic regression models fail to converge, we will follow the methods of the 2017 Cochrane review[1] to simply sum up the counts of true positives, false positives, false negatives and true negatives across 2x2 tables and calculate confidence intervals using the Wilson method.[4 5]

The association between author conflicts or manufacturer involvement and the study's test accuracy estimates sensitivity and specificity, respectively, will be compared indirectly by adding covariates (as defined by the three comparisons ([A], [B] and [C]) to the univariate random-effects regression models. In addition to the three comparisons, we will explore study design (1-gate studies vs 2-gate studies) as covariate to confirm findings from previous studies that 2-gate studies might overestimate test accuracy.[6]

Meta-analyses will be performed using the melogit command in the Stata software package (version 17; StataCorp, College Station, Texas 77845, USA).

References

1. Badeau M, Lindsay C, Blais J, et al. Genomics-based non-invasive prenatal testing for detection of fetal chromosomal aneuploidy in pregnant women. *Cochrane Database of Systematic Reviews* 2017;11:CD011767. doi: https://dx.doi.org/10.1002/14651858.CD011767.pub2

2. International Committee of Medical Journal Editors (ICMJE). Disclosure of Financial and Non-Financial Relationships and Activities, and Conflicts of Interest 2021 [Available from: http://www.icmje.org/recommendations/browse/roles-and-responsibilities/author-responsibilities--conflicts-of-interest.html accessed 15 September 2021.

3. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine* 2011;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009

4. Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med* 1998;17(8):873-90. [published Online First: 1998/05/22]

5. Wilson EB. Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association* 1927;22(158):209-12. doi: 10.2307/2276774

6. Taylor-Phillips S, Freeman K, Geppert J, et al. Accuracy of non-invasive prenatal testing using cell-free DNA for detection of Down, Edwards and Patau syndromes: a systematic review and meta-analysis. *BMJ Open* 2016;6(1):e010002. doi: 10.1136/bmjopen-2015-010002 [published Online First: 2016/01/20]

Appendix

Appendix 1. Methods section from 2017 Cochrane review (copied from Badeau et al.[1])

nant women at increased risk of fetal aneuploidy should be offered further testing, that is, after a first-tier screening, but before a diagnostic test.

To assess the screening performance of MPSS and TMPS as a first-tier test in pregnant women without prior risk (i.e. in unselected pregnant women or general population) as a replacement for current offered first-tier tests (biochemical, ultrasound or both).

To assess the diagnostic performance of MPSS and TMPS as a second-tier test as potential diagnostic tests to replace current invasive diagnostic tests.

Secondary objectives

To investigate potential sources of heterogeneity that may influence the diagnostic accuracy of MPSS and TMPS such as gestational age at the time of blood collection and type of reference standard used.

METHODS

Criteria for considering studies for this review

Types of studies

We included studies that met the following inclusion criteria:

- randomised studies where pregnant women were randomised to receive one gNIPT (MPSS or TMPS) as well as the reference standard;
- retrospective and prospective cohort studies where all pregnant women were tested with one or more gNIPT methods and the reference standard (including head-to-head studies); and
- retrospective and prospective case-control studies comparing one or more of the gNIPT methods with the reference standard.

Although studies with a retrospective or case-control design are prone to biases, we included such studies because we anticipated a paucity of other study designs. When data were sufficient, we explored the effect of excluding case-control studies in sensitivity analyses.

We excluded studies for which it was not possible to extract or derive the number of true positives, false positives, false negatives and true negatives.

Participants

We included women of any age, ethnicity and gestational age with a singleton or multifetal (monochorionic and dichorionic) pregnancy.

Index tests

Genomics-based non-invasive prenatal tests based on plasma cffDNA in maternal blood, analysis by either MPSS or TMPS methods.

Target conditions

We considered seven fetal aneuploidies, namely T21, T18, T13, 45,X, 47,XXY, 47,XXX and 47,XYY.

Reference standards

We considered the following test as reference standard: fetal karyotyping performed on cells obtained from chorionic villi sampling (CVS), amniotic fluid, placental tissue, a fetus lost by miscarriage or other equivalent and recognised methods on the same materials. By "fetal karyotyping" we mean traditional banding techniques, spectral karyotyping, fluorescence in situ hybridisation (FISH), array comparative genomic hybridisation (aCGH) or quantitative fluorescence polymerase chain reaction (QF-PCR). If fetal karyotyping was not performed, we used neonatal clinical examination or medical records from birth as a secondary reference standard for T21, T18 or T13. For sex chromosome aneuploidies (SCA), only fetal karyotype was an appropriate reference standard because newborns usually have a normal phenotype.

Search methods for identification of studies

Electronic searches

We used a sensitive search strategy that included the following three sets of search terms and synonyms:

- index test (e.g. cell-free DNA, sequencing, non-invasive and genetic diagnosis);
- participants' description (e.g. pregnant women, fetus and prenatal); and
- target condition (e.g. aneuploidy and chromosome anomalies).

We combined free-text words and subject headings used within each set with the Boolean operator OR and then combined the three sets using AND. We reviewed publications from 1st January 2007 because MPSS and TMPS were introduced in the literature in 2008 (Chiu 2008; Fan 2008). We did not limit our search by language, search filter or publication type (e.g. journal article, clinical trial, validation study, review and comment).

We applied a comparable search strategy (Appendix 4) with adaptations for each of the following databases:

- MEDLINE (Ovid) (January 2007 to July 2016);
- Embase (January 2007 to July 2016);
- Web of Science (ISI) (January 2007 to July 2016);

- Cochrane Register of Diagnostic Test Accuracy Studies, Cochrane Library (January 2007 to October 2016);
- ClinicalTrials.gov (January 2007 to September 2016);
- European Clinical Trials Register (January 2007 to September 2016);
- WHO ICTRP (January 2007 to September 2016);
- The National Technical Information Service (NTIS) (January 2007 to September 2016);
- OpenGrey (January 2007 to October 2016); and
- National Guideline Clearing House (January 2007 to September 2016).

Searching other resources

We examined references cited in potentially relevant full-text papers and those cited in previous reviews by cross-checking bibliographies. We examined grey literature by searching data available on the websites of private prenatal diagnosis companies (Ariosa Diagnostics 2016; BGI 2016; Berry Genomics 2016; Genoma 2016; Genome Care 2016; Illumina 2016; LabGenomics 2016; LifeCodexx 2016; Natera 2016; Genesupport 2016; Premaitha Health plc 2016; Sequenom 2016) using gNIPT technologies (January 2007 to December 2016). We also searched for conference abstracts and theses in appropriate sources (e.g. TheseNet, Theses Canada Portal) (January 2007 to October 2016).

Data collection and analysis

We used the methods suggested by the Cochrane Diagnostic Test Accuracy Working Group (Deeks 2013). For selection of studies, data extraction and assessment of methodological quality, we conducted a pilot using 20 randomly selected articles to trial our forms in order to ensure criteria were applied consistently.
None of the review authors involved in conducting a gNIPT primary study (FL, FR, SL and YG) took part in the selection of studies, nor in any decisions/analyses related to their own studies. Furthermore, by the final date of data collection, these authors had not published a primary gNIPT study.

Selection of studies

Two review authors (MB and CL) independently identified relevant studies by screening the titles and abstracts of all studies identified by the search strategy. We obtained the full-text version of all potentially relevant studies and assessed them for inclusion by using a study eligibility table based on prespecified inclusion criteria. The data collection form (Excel® format) for classifying studies during the full-text assessment is presented in Appendix 5. We considered all comments, statements or errata related to included studies. We excluded studies that did not match the inclusion criteria and we recorded the reason(s) for exclusion. If results from the same study cohort were reported in multiple publications, we considered all the publications and included results from the most relevant and comprehensive publications. We excluded papers with preliminary results whose full published results were available. We resolved any disagreement between assessors (MB and CL) by iteration, discussion and consensus. If required, we consulted a third review author (JB or LN).

Data extraction and management

Two review authors (MB and CL, JB or LN) independently extracted information and data from each included study by using a data extraction form that we developed in Excel® format. We included the following items:

- study characteristics (e.g. reference details allowing identification of the publication, language and study design);
- population characteristics (e.g. gestational age, maternal age, ethnicity, total number of pregnant women, number of aneuploid cases, number of euploid cases, recruitment location (country, geographic locations or regions), recruitment period and other relevant tests carried out prior to index test (e.g. ultrasonography, biochemical screening));
- features of the reference standard (e.g. fetal karyotyping, chromosome analysis or clinical examination);
- features of the index test (e.g. technical details, commercial or in-house gNIPT, cutpoint, failure rate, blood sample collection time (before or after reference standard) and first-tier test or second-tier test); and
- data for constructing two-by-two tables (number of true positives, false positives, false negatives and true negatives) or summary statistics from which the data were derived. In the two-by-two tables, the true negative cases were patients with any other aneuploidy than the one under analysis and all euploid cases were considered unaffected. When data were presented in three-by-two tables due to unclassified index test results (defined as grey zone between positive and negative test results), we constructed two-by-two tables by considering all unclassified gNIPT results as test positives. This is because in practice such results will lead to further testing and investigation to ensure a case of fetal aneuploidy is not missed.

We cross-checked all extracted and recorded data and we resolved any disagreement by iteration, discussion and consensus between two review authors (MB and CL, JB or LN). If required, we consulted a third author (JB, LN or CL). We wrote to the study contact author if information was missing or unclear or to clarify potential overlap between publications based on the same dataset to avoid including the same women more than once. If an article presented results including other aneuploidies than the ones under review, we considered only the subset of the cohort with the aneuploidies of interest.

Assessment of methodological quality

We used the revised QUality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool for assessment of methodological quality of included studies (Whiting 2011). We tailored the tool to this review question using the operational criteria detailed in Appendix 6 to answer signalling questions and make the overall judgment of risk of bias and applicability concerns for each domain of the tool. We answered each signalling question with a 'yes', 'no' or 'unclear' response for each included study and we recorded the reason for the judgment made. If a study was recorded as 'yes' on all signalling questions related to risk of bias, then it was deemed appropriate to have an overall judgment of 'low risk of bias'. If a study is recorded 'no' or 'unclear' on one or more signalling questions in a domain, then it was judged as having 'high or unclear risk of bias'. Judgments about applicability concern were rated as 'low', 'high' or 'unclear' in relation to our review question. 'Unclear concern' was used only if insufficient information was available. Two review authors (MB and CL, JB or LN) independently applied the QUADAS-2 tool to each included study and we resolved any disagreement by iteration, discussion and consensus. If required, we consulted a third review author (JB, LN or CL).