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

Domain 1: Patient selection

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

A. Risk of bias

Guidance:

Was a consecutive or random sample of patients enrolled?

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

Was a case-control design avoided?

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

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

Did the study avoid inappropriate exclusions?

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

B. Concerns regarding applicability

Guidance:

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

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

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

Domain 2: Index test

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

The QUADAS 2 tool requires a threshold to be pre-specified in the methods in order to avoid adjustment of the threshold according to the test outcome. However, the testing strategies considered in this review present a further level of concern. While an explicit threshold can be reported by studies (e.g. z- 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. This was demonstrated by one study that needed to adjust a pre-specified threshold value that a previous study had determined.[2] Since the population mean and standard deviation are not known, studies will have to determine their own threshold values. This review will, therefore, consider independent samples of participants to determine the threshold value as aiming to reduce bias.

A. Risk of bias

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

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

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

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

Was a threshold explicitly pre-specified?

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

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

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

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

B. Concerns about applicability

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

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

Domain 3: Reference standard

Due to the nature of the reference standards there is little concern about bias introduced by the choice of reference standard. We accepted prenatal or postnatal karyotyping or phenotypic newborn assessment as appropriate reference standard. They all display a detection rate of over 99% and are routine procedures in prenatal diagnosis [3]. If the index test is carried out before the reference standard, blinding to the results of the index tests is important.

A. Risk of bias

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

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

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

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

B. Concerns about applicability

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

Domain 4: Flow and Timing

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

A. Risk of bias

Did all patients receive a reference standard?

This question can be answer with ‘yes’ if the participants are recruited on the basis of their karyotype results.

Did all patients receive the same reference standard?

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

Were all patients included in the analysis?

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

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

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

Domain 5: Role of sponsor

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

A. Risk of bias

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

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

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

References

1. Lo YM, Lau TK, Zhang J, Leung TN, Chang AM, Hjelm NM, et al. Increased fetal DNA concentrations in the plasma of pregnant women carrying fetuses with trisomy 21. *Clin Chem* 1999;45(10):1747-51.
2. Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM, Ehrich M, et al. DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. *Genetics in medicine : official journal of the American College of Medical Genetics* 2011;13(11):913-20.
3. Dick P. Periodic health examination, 1996 update: 1. Prenatal screening for and diagnosis of Down syndrome. Canadian Task Force on the Periodic Health Examination. *Canadian Medical Association Journal* 1996;154(4):465-79.