Strengthening methods of diagnostic accuracy studies
Ochodo, Eleanor

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Chapter 10

Summary and Discussion
Summary

To treat or not to treat? The decision to administer or withhold treatment based on test results could have far reaching effects on the health of a patient. The evidence supporting the accuracy of medical tests introduced in practice should therefore be credible and generated with valid and robust methods. In this thesis we assessed how recommended methods of diagnostic accuracy have been applied in reports of diagnostic test accuracy. We specifically focused on the methods of reporting of primary studies and on methods for assessing risk of bias, performing meta-analyses and investigating publication bias in systematic reviews of diagnostic test accuracy.

Our study in Chapter 1 shows that about three out of ten studies of the diagnostic accuracy of biomarkers or other medical tests published in journals with an impact factor of 4 or higher overinterpret their results, while 99% of studies contain practices that facilitate such overinterpretation. The most frequent form of overinterpretation is an overoptimistic abstract and the most common practices facilitating overinterpretation are not reporting a sample size calculation and not stating a study hypothesis a priori. Diagnostic accuracy studies with optimistic conclusions may be highly cited, leading to a cascade of inflated and questionable evidence in the literature. This may translate to the premature adoption of tests into clinical practice.

In Chapter 2, we found that in a cohort of test accuracy studies registered in ClinicalTrials.gov, only slightly more than half of the studies that had been completed 18 months or longer before our analyses were published in a peer reviewed biomedical journal. Although publication rates increased over time, about one-third of the studies completed before 2009 was still unpublished halfway 2013. The non-reporting of research results may lead to unnecessary duplication of research efforts and the absence of vital results limits the evidence base on which clinical decisions are made.
Our opinion article on the STARD reporting guideline in Chapter 3 highlights that 10 years after the introduction of STARD, the reporting of diagnostic accuracy studies is improving, albeit slowly. We advocate for the revision of the STARD guidelines to guide the reporting of abstracts and protocols of diagnostic accuracy studies. Just like clinical trials, we call for the formal requirement of registration for prospectively designed accuracy studies. We also propose for the extension of the guidelines to guide the reporting of other types of accuracy studies such as prognostic tests.

In Chapter 4 we found that in a sample of 53 recently published diagnostic accuracy reviews with a meta-analysis, almost all (92%) had assessed the methodological quality of included studies. Yet only 2 reviews (4%) considered results of quality assessment when drawing conclusions. Simply relying on the test results without considering the quality of the underlying research could lead to the uptake of poor performing tests in practice and could consequently lead to suboptimal patient management.

Chapter 5 revealed a lack of clarity about recommended methods for meta-analysis of diagnostic accuracy data in an online survey conducted among authors of diagnostic accuracy reviews. Most authors who used traditional methods reported that they felt that these methods are recommended. Most authors who used advanced methods also reported that the methods they used are currently recommended. Bearing in mind the variability of results of test accuracy studies, it is essential that meta-analyses account for and if possible explain this variability. This will help physicians or policy makers make an objective assessment about the suitability of the tests to their settings and hopefully diminish the uptake of tests based on inadequate evidence. Advanced models of meta-analyses of test accuracy in this case have an upper hand, as they are better capable of dealing with variability than the more traditional methods.

In Chapter 6, we found that in a sample of 114 reviews, most authors (65%) mentioned or investigated the potential impact of publication bias. However, most (90%) used statistical methods not recommended for diagnostic accuracy
reviews to investigate publication bias. In addition, our comparison of the statistical tests recommended for intervention reviews (Begg and Egger tests) and those recommended for diagnostic reviews (Deeks’ tests) indicated that these tests do give different results and thus are not interchangeable. The use of wrong methods leads to an overestimation of the impact of publication bias on results of diagnostic accuracy reviews, thereby misleading readers.

Chapter 7 contains a systematic review of the diagnostic accuracy of rapid diagnostic tests and PCR for the detection of malaria in pregnant women. Our findings suggest that RDTs and PCR may have performance characteristics good enough to serve as alternatives for microscopy for the diagnosis of malaria in pregnancy. One challenge with evaluating the accuracy of the index tests was the limited use of the reference test, placental histology. Because histology is not always available or practical, few studies used it as a reference standard. To counter this we evaluated the index tests against alternative reference standards such as the microscopy of placental blood and microscopy of peripheral blood.

In Chapter 8, we systematically reviewed the diagnostic accuracy of circulating antigen tests and urine reagent strips for the detection of active schistosomiasis in endemic areas. There is currently no gold standard or recommended clinical reference standard for the detection of active schistosomiasis. However, because in practice, microscopy is the most commonly used test and often used as the reference test in studies, we selected it for use as the reference standard to detect *S. haematobium* and *S. mansoni* within this review. Among the evaluated tests for *S. haematobium* infection, microhaematuria detected the largest proportion of infections and non-infections identified by microscopy. For *S. mansoni*, the CCA POC test detects a very large proportion of infections identified by microscopy but misclassifies a large proportion of microscopy negatives as positives in endemic areas with a moderate to high prevalence of infection. Poor and inconsistent reporting limited our investigations of sources of heterogeneity and risk of bias assessment.

We conclude this thesis by urging for more frequent use of systematic reviews of test accuracy in Africa in Chapter 9. In a continent saddled with many diseases
amidst limited resources, evidence based practice is needed to effectively prioritize the use of the limited resources. We discuss ways of increasing the number and uptake of diagnostic reviews relevant to the African setting.

Way forward
To strengthen methods for diagnostic accuracy studies, we recommend the following actions and areas for future research.

To curb overinterpretation of reports of diagnostic accuracy studies
- Journals need to continuously emphasize that manuscripts submitted adhere to STARD reporting guidelines
- Extension of STARD guidelines to guide reporting of abstracts
- Revision of the language used in STARD to be in tandem to that used in the CONSORT guidelines. This may make STARD easier to follow.
- Research into the potential drivers of overinterpretation in diagnostic accuracy studies.
- More research comparing the number of citations accrued for overinterpreted reports of diagnostic accuracy studies versus those that are not overinterpreted.
- Research into the frequency of overinterpretation in reports of diagnostic accuracy studies used to inform guidelines or policy on diagnostic tests or strategies.

To curb and investigate publication bias
- We advocate formal registration of test accuracy studies
- Extension of STARD to guide reporting of protocols of diagnostic accuracy studies.
- Research to identify drivers of publication/outcome reporting bias in test accuracy studies.
- Research is needed among a cohort of registered studies showing the extent of discrepancies between original protocols and final publications.
• More research is needed to identify new statistical tests or improve on the existing Deek’s test to investigate publication bias in diagnostic accuracy reviews.

**To improve the incorporation of assessments of quality in diagnostic reviews**

• The Cochrane diagnostic handbook needs to be explicit in advising authors to incorporate quality in the conclusions of diagnostic reviews.
• A survey among authors of diagnostic reviews could reveal the challenges encountered in assessing and interpreting results of methodological quality.

**To improve the use of recommended methods of meta-analysis of test accuracy studies**

• Research is needed to highlight how the results obtained from both traditional and advanced methods of meta-analysis influence test results in clinical practice.
• Clear guidance is needed to guide authors on suitable alternative methods of meta-analysis should the recommended advanced methods of meta-analysis fail to work.

**To improve testing on Malaria and Schistosomiasis,**

• Studies are needed to identify a suitable and practical clinical reference standard for malaria in pregnancy and schistosomiasis.
• For malaria in pregnancy, more evaluations on the accuracy of RDTs and PCRs in different prevalence settings of plasmodium falciparum and plasmodium vivax species.
• For schistosomiasis, more evaluations of circulating antigen tests in different prevalence settings and in low intensity settings in light of the current control programs with the drug praziquantel.
• Authors conducting studies of these tests in developing countries need to be sensitized on the use of STARD reporting guidelines.
Concluding remarks

Preferred methods for reporting, assessing risk of bias and meta-analysing data in diagnostic accuracy studies and reviews have been brought forward. This thesis however shows that the uptake or correct use of these methods is suboptimal. This could be in part due to the complexity of diagnostic research and limited understanding of its methodology. Instead of focusing on the development of more advanced methods while the current ones are still poorly understood, I believe the emphasis in diagnostic research should be on more qualitative research into innovative knowledge translation methods. In other words, how can these methods be communicated in a simpler manner? How can these methods be made more relatable to clinical or public health practice? How can methods of reporting such as STARD be communicated better so that its uptake can be increased? More research into the usability or comprehension of these methods would come in handy in increasing their uptake. Preferably this research should be done in context. For example, the reasons for poor uptake or comprehension of the methodology for diagnostic accuracy research may vary according to specialties (infectious versus non-infectious diseases) or type of test (imaging tests versus laboratory tests).

In addition, in order to mitigate publication bias in diagnostic research, calls are currently being made to advocate for the formal prospective registration of diagnostic accuracy studies. However, methodologists in the field of test evaluation should not lose sight of the ‘why’ behind this phenomenon. Qualitative research into the drivers of non-reporting or non-publication of diagnostic accuracy studies should also be carried out in order to come up effective recommendations.