Systematic reviews of diagnostic test accuracy
Leeflang, M.M.G.

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.
Introduction
Introduction

General introduction

Diagnostic tests aim to reduce uncertainty about an individual’s condition. A plethora of tests is available for almost every condition imaginable. Examples include physical examination to rule out ankle fractures, mammograms to screen for breast cancer, magnetic resonance imaging for detecting herniated discs, portable chemical tests for blood glucose monitoring, nucleic acid amplification assays to detect infectious agents, and over the counter pregnancy tests.

A perfect test would identify all patients with the target condition, without making mistakes. This target condition may refer to a disease, or a disease stage, such as, for example, the healing phase of a fracture. Because perfect tests rarely exist, the users of a test may wish to know how well the test discriminates between individuals who have the target condition and those who have not. This is called diagnostic test accuracy.

The accuracy of a diagnostic test is studied by comparing the results of the test (or tests) under evaluation (also called index test) with the results of a reference standard. The reference standard is regarded as the best available method to establish the presence or absence of the target condition. The participants of a diagnostic accuracy study ideally undergo both the index test and the reference standard after which the results of both tests are compared (see Table 1). With dichotomous tests and a single target condition diagnostic test accuracy is often expressed as the proportion of people with the target condition who have indeed a positive test result (the test’s sensitivity, or true positive fraction) and the proportion of people without the target condition who have a negative test result (the test’s specificity, or true negative fraction)\(^1\).

\[
\begin{array}{|c|c|c|}
\hline
 & \text{Reference standard} & \\
 & + & - \\
\hline
\text{Index test} & + & TP & FP \\
 & - & FN & TN \\
\hline
TP+FN & FP+TN & \\
\hline
\end{array}
\]

In a 2 by 2 table, the results of the index test are compared with the results of the reference standard. TP=true positive; FP=false positive; FN=false negative; TN=true negative. Sensitivity = TP/(TP+FN); Specificity = TN/(TN+FP).
Central theme of the thesis

Health care professionals who are looking for evidence about how good a diagnostic test is in discriminating between patients with and without the target condition of interest, rely increasingly on systematic reviews of diagnostic test accuracy studies. Systematic reviews examine whether scientific findings are consistent and can be generalised across populations, settings, and treatment variations, or whether findings vary significantly by particular subsets\textsuperscript{2,3}.

Like any other research, the methodology of systematic reviews should be transparent and explicit, in order to minimise bias and maximise the informativeness in all parts of the review process. The methodology of systematic reviews involves the following steps

(1) formulating a research question;
(2) searching for the available evidence regarding the research question;
(3) assessing the quality of the available evidence;
(4) analysing the data;
(5) interpreting the results.

The objective of this thesis is to provide empirical evidence to improve and guide the further development of the methodology behind systematic reviews of diagnostic test accuracy. The focus is specifically on the search process, incorporation of study quality, and analysis of the data.

Outline of the thesis

Chapter 1 gives an overview of the development of the methodology for diagnostic test accuracy systematic reviews over the last decade. The steps involved in a review are introduced. The following chapters offer a more detailed discussion of some of the features of a systematic review diagnostic test accuracy.

In Chapter 2 we look at the usefulness of search strategies for retrieving diagnostic test accuracy studies in electronic bibliographical databases. We present the fraction of relevant studies that will be missed if search filters are used and we determine whether the search filters decrease the number of articles that one needs to screen to find one relevant article.

After retrieving the studies that are relevant for the systematic review, the quality of these studies needs to be assessed. The results of this quality assessment can be incorporated in the meta-analysis in many ways. In Chapter 3 we compare three different strategies for incorporating quality, to test the hypothesis that adjustment for quality produces less optimistic estimates of diagnostic accuracy and narrower confidence intervals.
Introduction

Chapter 4 addresses a possible source of bias when evaluating a test that produces a continuous result: post-hoc determination of an optimal cut-off value. We aim to determine the magnitude of bias in sensitivity and specificity associated with data-driven selection of cut-off values and to examine potential solutions to reduce this bias.

Chapter 5 addresses a possible source of heterogeneity between studies: differences in the prevalence of the target condition across studies. Although it is sometimes claimed that sensitivity and specificity do not depend on disease prevalence, we provide a number of real life examples in which accuracy varied with prevalence.

Chapters 6 and 7 are examples of systematic reviews of diagnostic test accuracy. In Chapter 6 we report a review of the accuracy of fibronectin tests for the prediction of pre-eclampsia, one of the most important causes of maternal and fetal mortality and morbidity worldwide.

Chapter 7 presents a systematic review about the diagnostic accuracy of a commercially available galactomannan test for the diagnosis of invasive aspergillosis in immunocompromized patients. This systematic review served as a pilot review for the Cochrane Diagnostic Test Accuracy Working Group that was constituted to develop and test methods for the inclusion of diagnostic test accuracy reviews in The Cochrane Library.

Chapter 8 summarizes the main findings of the research presented in this thesis and discusses a number of options for future research.
References


