Evaluation of diagnostic tests: from accuracy to outcome
Lijmer, J.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
Chapter 1

What is a diagnostic test? A definition of a diagnostic test could be any method for obtaining information on a patient’s health status. This includes information generated by laboratory tests, imaging procedures or histological examinations, but also information that can be obtained from history of physical examination. As such, diagnostic tests have always been part of western medicine, in which presenting symptoms are linked to possible underlying diseases.

It goes without saying, that diagnostic tests yield information that is gathered to establish a diagnosis. However data from tests can also be collected for other, related purposes: to judge the severity of the disease, for making statements about prognosis, to support therapy decisions, and to monitor the effect of treatments. In general terms, the results from the test are used in all of these cases to reduce the uncertainty whether or not the patient has a specific target condition. The target condition refers to any identifiable condition that is linked with a particular action in health care, such as communicating the likely cause for the complaints, or the initiation or modification of treatment.

Over the last decades many diagnostic technologies have emerged such as magnetic resonance imaging, DNA amplification techniques and immunohistochemical methods. The diagnostic test arsenal is still increasing, while their application extends from the hospital to the home situation. Physicians or patients may find it difficult to resist the temptation to call on those new and attractive diagnostic technologies often presented in glossy brochures or stylised newspaper adds. More information is always better, is it not?

Despite the intuitive appeal of the information they produce, diagnostic tests, including self-tests, are not as innocent as some of the manufacturing companies would like to present them. Erroneous or unwanted information can lead to regrettable actions with negative consequences. An inaccurate test can lead to false diagnoses and the initiation of unwarranted, sometimes even harmful therapy or, alternatively, give false reassurance and lead to unnecessary, possible dangerous, delays in starting appropriate therapy. Examples of premature dissemination and inappropriate use of at first promising new tests have been the dexamethasone suppression test for depression,\(^2\) 125I-fibrinogen leg scan for the diagnosis of deep venous thrombosis,\(^3\) and the carcinoembryonic antigen (CEA) for colon cancer.\(^4\) A striking feature of the latter example was that inappropriate use continued for a long time after the first overoptimistic reports were overtaken by less positive results in later evaluations.\(^4\)\(^5\)

In addition, the increasing costs of health care have put a pressure on available budgets, calling for the elimination of ineffective medical technologies. Nowadays, many feel that there should be sound evidence that the use of a diagnostic test does more good than harm before it is introduced in daily practice. For this reason, new
diagnostic technologies should, like therapeutic interventions, be subjected to a thorough evaluation. This thesis deals with the evaluation of diagnostic tests, presenting both methodological contributions as well as applications of existing and new methodology to clinical problems.

A thorough evaluation of diagnostic tests can in itself be a time-consuming and costly process. An efficient use of resources available for research calls for a well-planned evaluation strategy. In such a strategy, more elaborate and therefore expensive phases should only be performed if sufficient evidence has been obtained in previous steps of the evaluation process. Several comparable models have been proposed for the evaluation of diagnostic tests. These models and their merits are discussed in detail in chapter 2. Existing evaluation models contain two crucial steps: first, an assessment of the diagnostic accuracy and, second, an evaluation of the effect of the test's results on patient outcome.

The diagnostic accuracy of a test is assessed by performing that test and a reference test in a series of patients. The reference test is the best available test to establish the presence of the target condition of interest. Subsequently the information from the test under evaluation is compared with the results from the reference test in the same patients. Diagnostic accuracy can then be characterised in terms of sensitivity and specificity, predictive values, likelihood ratios, or receiver operating characteristic curves and derived measures.

It is valuable to know whether a result from a test corresponds to the truth. Can this value be trusted? Is the question that comes first to the mind in the evaluation of medical tests. From a patient perspective, mere knowledge about the true state of things is not enough. An improved estimate of the probability of disease being present or absent is by itself of little use. To answer the question "Does this test do more good than harm?" it is necessary to take the downstream consequences of the test into account. One would like to know if the information generated by that test could be correctly used in subsequent decisions, including actions undertaken to the improve of patient outcome. Randomised controlled trials of diagnostic strategies have been forward to answer this question in a systematic, valid way.

In practice there are several problems, both with the assessment of diagnostic accuracy and the evaluation of the effect on patient outcome. Estimates of diagnostic accuracy can vary largely between different evaluations of the same test. Part of this variability can be attributed to the poor design of these studies. A survey of the diagnostic articles published in five major journals (1990-1993) showed that only 18% of the studies satisfied 5 of 7 methodological standards examined. Variations in the study populations, the tests and the reference tests among the studies can also result in genuine variation in diagnostic accuracy.
In Chapters 3 to 9 potential sources of variation in the estimation of diagnostic accuracy are discussed. The objective of the study outlined in Chapter 3 was to estimate the bias induced by flaws in study design in characterising diagnostic accuracy. Knowledge on the quantitative effect of these biases can help readers and reviewers of the medical literature in appraising diagnostic studies. The analysis, reported in Chapter 4, focused on the delay between onset of symptoms and test performance as a possible source for variation. The optimal timing of sampling of myoglobin for early diagnosis of acute myocardial infarction within the first 24 hours was evaluated.

Decision analyses of diagnostic strategies often use test characteristics reported for a group of consecutive patients suspected of the target condition. However, it is likely that the diagnostic accuracy of a test varies between subgroups as defined by tests preceding the test of interest. This may affect the results and recommendations from such decision analyses. This hypothesis was examined in a study reported in Chapter 5 with data from a large prospective study on the diagnostic accuracy of several non-invasive tests for pulmonary embolism. Chapter 6 is an application of the incorporation of variation in diagnostic accuracy in a decision analysis. Different diagnostic strategies for patients suspected of pulmonary embolism were compared using subgroup estimates of diagnostic accuracy where possible rather than overall estimates.

The large variation in the results and the poor quality of the primary studies poses a challenge to meta-analyses of diagnostic tests, where one wants to summarise the available evidence into more precise estimates of test performance. In the light of the heterogeneity in the primary studies it is sometimes questionable whether the results can actually be pooled into a single summary estimate. Chapter 7 describes the potential sources for variations in study results and the methods to explore them in meta-analyses of diagnostic tests. In Chapters 8 and 9, these methods are applied to explore the sources of heterogeneity in study results in a meta-analysis of D-Dimer assays for the diagnosis of venous thromboembolism and a meta-analysis of magnetic resonance angiography for the assessment of peripheral vascular disease respectively.

The next chapters are dedicated to the second step: evaluation of the effect of a test on patient outcome. Evaluations of the harms and benefits of diagnostic tests, as proposed by the existing evaluation models, are infrequent as is shown in overviews of diagnostic test studies. Such evaluations are indeed quite complex. The relation between test and outcome is almost always indirect. These studies require a close collaboration of radiologists and other diagnostic specialists with the health care professionals that take the subsequent decisions. The methodology for clinical research in this field is still under development. Current textbooks on clinical
epidemiology do not go beyond basic discussions of diagnostic accuracy. In Chapters 10 and 11 we present some methodological contributions to this field and discuss possible designs and their pitfalls to evaluate the effect of diagnostic strategies on patient outcome.

References


