Chapter 14
Research Methodology for Studies of Diagnostic Tests

Stephen P. Glasser

Abstract Much of clinical research is aimed at assessing causality. However, clinical research can also address the value of new medical tests, which will ultimately be used for screening for risk factors, to diagnose a disease, or to assess prognosis. In order to be able to construct research questions and designs involving these concepts, one must have a working knowledge of this field. In other words, although traditional clinical research designs can be used to assess some of these questions, most of the studies assessing the value of diagnostic testing are more akin to descriptive observational designs, but with the twist that these designs are not aimed to assess causality, but are rather aimed at determining whether a diagnostic test will be useful in clinical practice. This chapter will introduce the various ways of assessing the accuracy of diagnostic tests, which will include discussions of sensitivity, specificity, predictive value, likelihood ratio, and receiver operator characteristic curves.

Introduction

Up to this point in the book, we have been discussing clinical research predominantly from the standpoint of causality. Clinical research can also address the value of new medical tests, which will ultimately be used for screening for risk factors, to diagnose a disease, or to assess prognosis. The types of research questions one might formulate for this type of research include: “How does one know how good a test is in giving you the answers that you seek?” or “What are the rules of evidence against which new tests should be judged?” In order to be able to construct research questions and designs involving these concepts, one must have a working knowledge of this field. In other words, although traditional clinical research designs can be used to assess some of these questions, most of the studies assessing the value of diagnostic testing are more akin to descriptive observational designs, but with the twist that these designs are not aimed to assess causality, but are rather aimed at determining whether a diagnostic test will be useful in clinical practice.
Bayes Theorem

Thomas Bayes was an English theologian and mathematician who lived from 1702–1761. In an essay published posthumously in 1863 (by Richard Price), Bayes’ offers a solution to the problem “…to find the chance of probability of its happening (a disease in the current context) should be somewhere between any two named degrees of probability.” Bayes’ Theorem provides a way to apply quantitative reasoning to the scientific method. That is, if a hypothesis predicts that something should occur and it does, it strengthens our belief in that hypothesis; and, conversely if it does not occur, it weakens our belief. Since most predictions involve probabilities i.e. a hypothesis predicts that an outcome has a certain percentage chance of occurring, this approach has also been referred to as probabilistic reasoning. Bayes’ Theorem is a way of calculating the degree of belief one has about a hypothesis. Said in another way, the degree of belief in an uncertain event is conditional on a body of knowledge. Suppose we’re screening people for a disease (D) with a test which gives either a positive or a negative result (A and B, or T+ and T– respectively). Suppose further that the test is quite accurate, in the sense that, for example, it will give a positive result 95% of the time when the disease is present (D+), i.e. \( p(T+ \mid D+) = 0.95 \) (this formula asks what is the probability of the disease being present GIVEN a positive test?), or said another way, what is the probability that a person who tests positive has disease? The naive answer is 95%; but this is wrong. What we really want to know is \( p(D+ \mid T+) \), that is, what is the probability of testing positive if one has the disease; and, Bayes’s theorem (or predictive value) tells us that.

In modern medicine the first useful application of Bayes’ theorem was reported in 1959. Ledley and Lusted demonstrated a method to determine the likelihood that a patient had a given disease when various combinations of symptoms known to be associated with that disease were present. Redwood et al. utilized Bayesian logic to reconcile seemingly discordant results of treadmill exercise testing and coronary angiography. In 1977, Rifkin and Hood pioneered the routine application of Bayesian probability in the non-invasive detection of coronary artery disease (CAD). This was followed by other investigative uses of Bayesian analysis, an approach which has now become one of the common ways of evaluating all diagnostic testing.

As noted above, diagnostic data can be sought for a number of reasons beside just the presence or absence of disease. For example, the interest may be the severity of the disease, the ability to predict the clinical course of a disease, or to predict a therapy response. For a test to be clinically meaningful one has to determine how the test results will affect clinical decisions, what are its cost, risks, and what is the acceptability of the test; in other words, how much more likely will one be about this patients problem after a test has been performed than one was before the test; and, is it worth the risk and the cost? Recall, that the goal of studies of diagnostic testing seeks to determine whether a test is useful in clinical practice. To derive the latter we need to determine whether the test is reproducible, how accurate it is, whether the test affects clinical decisions, etc. One way to statistically assess test reproducibility (i.e. inter and intra-variability of test interpretation), is with a kappa statistic. Note that reproducibility does not require a gold standard, while accuracy