Autoimmunity Reviews 12 (2012) 89–96

Contents lists available at SciVerse ScienceDirect

Autoimmunity Reviews

journal homepage: www.elsevier.com/locate/autrev

Review

Tools for critical appraisal of evidence in studies of diagnostic accuracy

Davide Giavarina *

Clinical Pathology Dept., St. Bortolo Hospital, Vicenza, Italy

ARTICLE INFO

Available online 8 July 2012

Keywords:
Sensitivity
Specificity
Meta-analysis
Forest plot
Summary ROC curve
Diagnostic test accuracy

ABSTRACT

Studies of accuracy are often more complex to understand than clinical trials, since there can be more than one outcome and scope (screening, diagnosis, and prognosis) and because results have to be reported in more than one way, than in clinical trials (relative risk or odds ratio). Sensitivity and specificity are common terms for practitioners, but to remember that sensitivity is the "ratio between true positive rate and true positive rate plus false negative rate" may sometime cause some frustration. Moreover, likelihood ratio, predictive values, diagnostic odds ratio, and pre- and post-test probability complicate the framework. To summarize these indexes from multiple studies can be also a little more difficult. However, understanding diagnostic test accuracy from different study results and how to interpret systematic reviews and meta-analysis can help every practitioner improve critical appraisal of evidence about the best use of diagnostic tests. Avoiding complicated mathematical formulas, this paper attempts to explain the meaning of the most important diagnostic indexes and how to read a Forest plot and a summary Receiver Operative Characteristic curve.

Contents

1. Introduction

"Do the right thing". In a nutshell, this is what the Evidence-Based Medicine (EBM) movement would like to teach; that is: "At every opportunity, use the most accurate diagnostic test, the most effective treatment, for the right patient, at the right time and without wasting resources".

The EBM movement was officially launched in 1992 in the Journal of the American Medical Association [1] (actually, the definition was used for the first time in an editorial of the American College of Pathologists

Abbreviations: AUC, Area under curve; CER, Control Event Rate; CI, Confidence intervals; DOR, Diagnostic odds ratio; DTA, Diagnostic test accuracy; EBM, Evidence-Based Medicine; EBP, Evidence-based practice; EER, Experimental Event Rate; FN, False negative; FP, False positive; LR, Likelihood ratio; OR, Odds ratio; PV−, Negative predictive value; PV+, Positive predictive value; RCT, Randomized controlled trial; ROC, Receiver Operating Characteristic; RR, Relative risk; Se, Sensitivity; Sp, Specificity; sROC, Summary Receiver Operating Characteristic; TN, True negative; TP, True positive.

* Clinical Chemistry and Haematology Laboratory, ULSS 6 Vicenza, Viale Rodolfi 37, I-36100 Vicenza, Italy. Tel.: +39 0444 753372; fax: +39 0444 752501.
E-mail address: davide.giavarina@ulssvicenza.it.

1568-9972/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.autrev.2012.07.004

Club Journal by Guyatt, in 1991 [2]) and the statement said: "Use of strongest available data to make informed, unbiased decisions about the diagnosis and treatment of patients". Deriving from the works of Archibald Cochrane, the main aim of EBM was to demonstrate the efficacy of health treatment, since there cannot be efficiency without effectiveness [3]. The quality improvement of primary clinical studies and a critical appraisal of the available evidence became the goal for EBM researchers.

Today, after 20 years, the EBM movement is more oriented to the clinical practice, to apply the principles in operational contexts of health, integrating the best evidence with the practitioners' experience and patients' expectations. While finding and appraisal of evidence are matters of study for methodologists and researchers working for systematic reviews, for most clinicians the challenge is to apply the best evidence for decision-making. However, an EBM process and some tools are needed also for good decision-making. The application of EBM to daily practice comprises five steps, referred to as the EBM A-5 cycle: Ask — structure the question that states the problem; Acquire — find the evidence; Appraise — determine whether the evidence is reliable and usable (since critical appraisal is the use of explicit, transparent methods to assess the data in published research, applying the rules of evidence to factors such as internal validity, adherence to reporting standards, conclusions and generalizability); Apply — take the evidence and apply it to your problem; Audit — assess the process you have used for effectiveness.

Systematic reviews remain the best publications giving complete summaries of all the available evidence, pooled and weighted. This type of publication responds to a defined question, reviews all the literature on the topic and summarizes the evidence using meta-analysis statistics to sum up the results. Although results of systematic reviews are usually explained and commented upon, understanding meta-analyses is a required skill for the practitioner so that he can make his own appraisal and make his own decisions for the care of his patients. More skills are needed to understand and evaluate studies of diagnostic accuracy than of clinical trials, because the statistics are more numerous and pooling statistics are more complex.

The aim of this paper is to present a review of the major indexes of accuracy and then to consider the opportunities and issues related to the synthesis of evidence of diagnostic accuracy.

2. Evaluating the effectiveness of a diagnostic test

2.1. Differences between clinical effectiveness and diagnostic efficacy

Beginning with available evidence, it is easier to evaluate the effectiveness of a given treatment than to evaluate a diagnostic test for accuracy. Two reasons can be suggested. Firstly, there are few randomized controlled trials (RCTs) on the diagnostic accuracy, due to less economic interest and less "culture" in comparing different tests for the same diagnosis Secondly, the outcomes of a diagnostic test are not well defined but they depend on the actions arising from and following the diagnosis or lack thereof. Patient outcome rarely depends only on the efficacy of the diagnostic test; the action that follows the diagnosis is critical to the outcome (Fig. 1).

But even if we evaluate the efficacy of the test based only on discriminative diagnostic ability, the evaluation shows important differences compared to studies of therapeutic efficiency.

Randomized control studies provide the strongest evidence for (relative) efficacy of treatment. In such studies, and generally in all the evaluation studies of medical and surgical therapy, a single measure of effectiveness is assessed: the Relative Risk (RR), or alternatively and more commonly, the Odds Ratio (OR). To calculate these indexes, we evaluate the presence of the desired (or adverse) outcome in the two arms of the study, where in the first it is applied to the treatment under study, and in the other, to placebo, current therapy or no treatment.

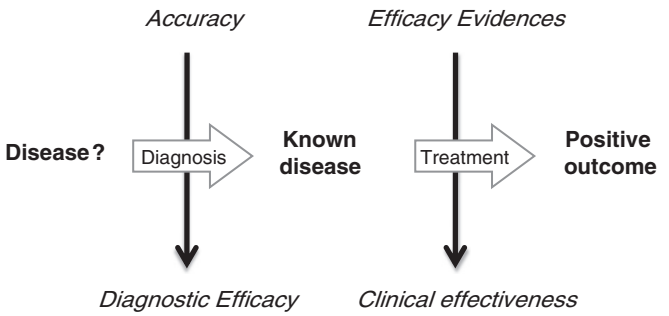

Fig. 1. Interaction between diagnosis and treatment to determine the outcome.

The percentage of outcomes in the treated subjects defines the Experimental Event Rate (EER); the percentage in the controls, defines the Control Event Rate (CER). The ratio of EER and CER is the RR.

The OR of a treatment is the ratio of the frequency of the outcome in the treated group and the frequency of the same event in the control group (Table 1). One single index, derived from many different studies, can be easily summarized in the Forest plot of meta-analysis.

In studies of diagnostic test accuracy (DTA) there are at least two, but frequently more than two, efficacy indexes: sensitivity (Se) and specificity (Sp), but also positive predictive value (PV+) and negative predictive value (PV−), positive and negative likelihood ratio (LR), diagnostic odds ratio (DOR), Receiver Operating Characteristic (ROC) curve. It can be a little more difficult to summarize these indexes; to do so, we first have to understand the meaning and the power of each index.

2.2. Indexes of diagnostic efficacy

2.2.1. Sensitivity and specificity

Every doctor has encountered Se and Sp many times during his working life, and sometimes everyone experiences frustration trying to remember the definitions: that Se = TP/TP + FN, where TP is the number of true positives and FN is the number of false negatives; and similarly, that Sp = TN/TN + FP, where TN is the number of true negatives and FP is the number of false positives.

Se and Sp were firstly used as indexes of the accuracy of a diagnostic test in 1947 [4]. Se is the ability to recognize the "positive", the ill, while Sp is the ability to recognize the "negative", the healthy people. If the population of healthy people and ill people were completely separated with respect to the concentration of a biomarker or a test, a correct decision level might be able to divide the healthy from the diseased. In reality, as shown in Fig. 2, the distribution of values in healthy and in ill people frequently overlaps. Any chosen cutoff level will cause a number of false positive and/or false negative results. In the example given in Fig. 2, among 100 healthy people, the chosen cutoff correctly classifies 93 people (TN), while seven are classified as positive (FP); this is the Sp of the test at this cutoff. The ability to recognize the negative in this case is 93/100 = 0.93. The same cutoff correctly classifies 97 people with the disease (TP), while 3 are misclassified (FN). This is the Se of the test at this cutoff. The ability to recognize the positive in this case is 97/100 = 0.97. The dashed squares represent the area of calculation: Sp is calculated in the healthy people, which are expected to be

Table 1
A simple way to calculate the relative risk and the odds ratio.

	Outcome present	Outcome absent
Diseased patients	A	B
Control patients	C	D

Relative risk: RR = EER/CER.
Odds ratio: OR = (A/B)/(C/D).
Experimental Event Rate (EER) = A/(A + B).
Control Event Rate (CER) = C/(C + D).