

Systematic reviews of diagnostic test accuracy and the Cochrane collaboration

Gianni Virgili · Andrea A. Conti · Vittoria Murro ·
Gian Franco Gensini · Roberto Gusinu

Published online: 9 April 2009
© SIMI 2009

The methodologist's point of view

Gianni Virgili, Andrea A. Conti, Vittoria Murro

Introduction

Since Cochrane's Corner focuses on Cochrane systematic reviews, it is important that readers be informed about new publication types within the Cochrane Library (<http://www.thecochranelibrary.com>), for example systematic reviews of diagnostic test accuracy (DTA). This commentary serves to explain that research on diagnostic test efficacy and impact has recently begun to be based on an expanding set of complex methodological rules. These rules may still be unfamiliar to doctors who rely on and use tests for clinical decisions.

G. Virgili (✉) · V. Murro
Department of Oto-Neuro-Ophthalmological Surgical Sciences,
University of Florence, Florence, Italy
e-mail: gianni.virgili@unifi.it

A. A. Conti · G. F. Gensini
Department of Critical Care Medicine and Surgery,
University of Florence, and Don Carlo Gnocchi Foundation,
IRCCS, Florence, Italy

G. F. Gensini
e-mail: g.gensini@dac.unifi.it

R. Gusinu
DAI Cardiologico e dei Vasi,
Azienda Ospedaliero-Universitaria Careggi,
Florence, Italy

A historical perspective

A philosopher [1] wrote that even science relies on common sense, or on assumptions that are based on strong common beliefs. These beliefs are not only self-evident but necessarily unspoken.

For centuries a philosophical architecture based on individual observation was designed and taught by leading scientists and doctors, and served to explain and treat diseases.

Modern medicine was born when scientific methods for measuring physical phenomena began to be used by physicians and researchers. The observations of John Snow (1813–1858) on the relationship between water supply and cholera incidence in London in 1854–1855 (in 1855 the second edition of his classical text *On the Mode of Communication of Cholera* [2] was published) contributed to the collapse of the miasma theory of cholera. The miasma theory was dominant at the time, but Snow was not convinced that “bad air” could cause cholera. Filippo Pacini (1812–1883) in Italy had discovered the *Vibrio Cholerae* in the bowel of people affected by the disease in 1854, but although Snow had no knowledge of this he nonetheless accurately described the transmission pattern of cholera by connecting the disease with the presence of the water drunk in London. He observed in a systematic way that households receiving water from the Thames in upstream London were less subjected to cholera than those supplied with water in downstream London and identified the source of the epidemics in the public water pump of Broad Street. By many authors his text on cholera is referred to as one of the first great achievements of modern epidemiology.

Today causes of disease are studied using complex methods relying on rigorous and sophisticated epidemiological measurements of their occurrence, and evidence-based medicine serves to import this methodological body

of knowledge into the assessment of treatment efficacy. The cholera example provided shows how basic assumptions have changed through time. In recent years, these changes have particularly concerned therapy and now continue with diagnosis.

The investigation of diagnostic test efficacy

Deeks [3] has proposed the following classification of diagnostic studies:

- Studies of feasibility and reliability—which report on the consistency of diagnostic test results.
- Studies of diagnostic accuracy—which assess the performance of a diagnostic test in distinguishing between diseased and non-diseased patients.
- Studies of diagnostic and therapeutic process—which assess how the test may influence the diagnostic process.
- Studies of patient outcome—which assess the impact of a diagnostic test on patient outcomes.
- Studies of cost/effectiveness—which assess the cost/effectiveness of a diagnostic test in clinical practice.

Tatsioni et al. [4], in a technology assessment on Magnetic Resonance Spectroscopy for brain tumours, have suggested that most research has dealt with the feasibility and spectrum of disease, few studies have dealt with diagnostic accuracy, and none with the impact on patient outcome.

We may therefore pose ourselves the question why we are using complex rules for assessing treatment efficacy, such as costly randomised controlled trials, while discarding the use of hard, patient-centred end points to investigate the efficacy of diagnostic tests. We know that correct diagnosis, and the monitoring of disease and treatment effect, are essential in the proper use of modern drugs as also for the correct implementation of other health interventions, yet we continue to privilege therapy rather than diagnosis.

Diagnostic test accuracy studies

Diagnostic test accuracy studies aim at measuring the ability of a new—or simpler, cheaper, faster, less invasive—test, called index test, to detect the presence or absence of a specific disease or condition. The presence of this disease or condition is defined using a reference standard—the term “gold standard” is discouraged because it improperly implies perfection.

Two methodological tools should be considered regarding DTA studies:

- the STARD initiative [5]—a checklist adopted by major journal editors to report properly DTA studies;
- the QUADAS checklist [6]—a series of items used to score the methodological quality of a DTA study.

The QUADAS checklist is a user-friendly scale including fourteen quality items, among which the following may be cited:

- Representative spectrum. This refers to the fact that a consecutive series of patients in the study have a spectrum of disease sufficiently broad to inform practice.
- Acceptable reference standard. Here the reference is to whether the test used to define the presence or absence of the condition/disease is valid and reliable.
- Acceptable delay between tests. This requires that the time interval between index and reference test be short enough to avoid a change in the clinical condition.
- Partial verification avoided. Here is presented the need for no pre-selection of patients included in the study based on the index test results.
- Differential verification avoided. This requires that the same reference test be used for all patients in the study regardless of index test results.
- Incorporation avoided. In this case it is required that the index test be not a component of the reference test.
- Reference standard results blinded and index test results blinded. This requires that examinations by means of the index and reference tests be blinded from each other.
- Uninterpretable results reported. Here cases with indeterminate index test results are reported—number and characteristics—and shown to have limited potential for introducing bias.
- Withdrawals explained. People who were selected in the study but in fact did not complete the assessments should be reported and shown to have limited potential for introducing bias.

DTA systematic reviews and the Cochrane collaboration

The first Cochrane review of DTA studies was published in the *Cochrane Database of Systematic Reviews* (<http://www.cochrane.org>) Issue 4 in 2008 [7] and investigated the diagnostic performance of Galactomannan, compared to other tests, to diagnose invasive aspergillosis in immunocompromised patients. Useful suggestions are provided, even if the authors caution against the high heterogeneity of results across studies, which is typical of DTA systematic reviews.

Some key features of DTA reviews, among others, are:

- a clear definition of the clinical question—including who will use the index test, what tests come before or after the proposed index test or are combined with it [8];
- a broad bibliographic search—no pre-set strategy is recommended;
- title selection, study inclusion, data extraction and quality assessment independently by two reviewers, using pre-defined and validated methods;
- statistical analyses using appropriate techniques—including summary ROC curves that take into account heterogeneity when metanalysis is possible [9–11];
- conclusions incorporating study quality assessment.

What are the consequences of the paradigm shift regarding the assessment of diagnostic tests in medicine?

Beginning with what may concern doctors more directly and professionally, DTA may lead them to feel expropriated of their diagnostic role, given the shift of clinical investigation design towards methodologists and other professionals using costly resources. The answer to this is that the role of clinicians will continue to maintain its fundamental importance at a time when society requires a more complex doctor-patient relationship, and an ever increasing involvement on the part of doctors in public health issues and in the cost/effective use of diagnostic and treatment resources.

Incorporating new methodological tools regarding diagnosis is leading us to admit that we are in effect using some tests without a prior reliable investigation into their diagnostic performance and into their impact on patient outcome. In other words, we are bridging the two or three decade gap with respect to drug therapy investigation.

One expected advantage in investigating diagnostic tests will be the possibility of following more accurately the clinical path of patients through their disease. This will serve to clarify the purpose of the test itself in the logic of the modern concept of clinical governance which tends to optimize the efficiency of health care.

The cost sector is another area in which DTA may play a part by limiting the expenses related to purchasing devices and paying for their maintenance and use, even if there is the possibility that costs might increase if complex regulatory rules are set.

Finally DTA is also likely to improve a recent trend in medicine, namely, the standardisation of clinical practice.

A clinician's point of view

Gian Franco Gensini, Roberto Gusinu, Andrea A. Conti

In 1991 G. Guyatt published his first relevant paper on Evidence-based medicine [12], and since then many Authors consider 1991 the “year of birth” of Evidence Based Medicine. If this is accepted, today, in 2009, Evidence Based Medicine has reached its majority. From the beginning of its life, the methodological and clinical interest of Evidence Based Medicine has largely concentrated on therapeutics, and only in the last few years has the diagnostic area begun to be systematically evaluated. Not by chance, as the Authors of the previous paper point out, the first Cochrane review of DTA studies was published in the Cochrane Database of Systematic Reviews in 2008 [7].

The topic of DTA evaluation is a major one, especially if examined in the context of health technology assessment (HTA). HTA, an internationally acknowledged process, is the systematic analysis of the characteristics, consequences and impact of health care technology. Among its objectives there are those of providing high quality information targeted to policy making and of “objectively” supporting clinical decision making [13]. HTA consequently appears as an appropriate instrument for clinical governance, and clinicians are particularly interested in its correct implementation.

HTA, with its ample array of instruments and methods functional to promoting health, among them diagnostic tools, has already made a number of indications on novel technologies available. Clinicians therefore welcome the attention dedicated by Evidence Based Medicine and the Cochrane collaboration to DTA, given that the degree of “technologisation” in medicine is constantly on the increase and that the precise weight of modern diagnostic procedures benefit/cost is mandatory. At present, clinicians require that accurate and reliable information on the real performance of diagnostic tools be elaborated with regard to not only the most recent and sophisticated technologies but also to already routinely adopted diagnostic methods.

Conflict of interest statement The authors declare that they have no conflict of interest related to the publication of this manuscript.

References

1. Wittgenstein L (1969–1975) On certainty (Über Gewissheit). In: Anscombe GEM, von Wright GH (eds) Basil Blackwell, Oxford
2. Snow J (1855) On the mode of communication of cholera. John Churchill, London
3. Deeks JJ (1999) Using evaluation of diagnostic tests: understanding their limitation and making the most of available evidence. *Ann Oncol* 10:761–768

4. Tatsioni A, Zarin DA, Aronson N, Samson DJ, Flamm CR, Schmid C, Lau J (2005) Challenges in systematic reviews of diagnostic technologies. *Ann Intern Med* 142:1048–1055
5. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis GA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Drummond R, de Vet HCW (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 326:41–44
6. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 3:25
7. Leeflang MM, Debets-Ossenkopp YJ, Visser CE, Scholten RJ, Hooft L, Bijlmer HA, Reitsma JB, Bossuyt PMM, Vandenbroucke-Grauls CM. Galactomannan detection for invasive aspergillosis in immunocompromised patients. *Cochrane Database of Systematic Reviews* 2008, Issue 4. Art. No.: CD007394.
8. Bossuyt PM, Irwig L, Craig J, Glasziou P (2006) Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ* 332:1089–1092
9. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 58:982–990
10. Rutter CM, Gatsonis CA (2001) A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statist Med* 20:2865–2884
11. Leeflang MGM, Deeks JJ, Gatsonis C, Bossuyt PMM, on behalf of the Cochrane Diagnostic Test Accuracy Working Group (2008) Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 149:889–897
12. Guyatt G (1991) Evidence-based medicine. *ACP J Club (Ann Intern Med)* 14(suppl 2):A-16.
13. NIHR Health Technology Assessment programme. NHS National Institute for Health Research. Available at <http://www.ncchta.org/publicationspdfs/Infoleaflets/GeneralLeaflet.pdf>