



Qualification versus validation of biomarkers

Jenny Doust

To cite this article: Jenny Doust (2010) Qualification versus validation of biomarkers, Scandinavian Journal of Clinical and Laboratory Investigation, 70:sup242, 40-43, DOI: [10.3109/00365513.2010.493380](https://doi.org/10.3109/00365513.2010.493380)

To link to this article: <https://doi.org/10.3109/00365513.2010.493380>



Published online: 01 Jun 2010.



Submit your article to this journal [↗](#)



Article views: 1033



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)

ORIGINAL ARTICLE

Qualification versus validation of biomarkers

JENNY DOUST

Faculty of Health Sciences and Medicine, Bond University, Robina, Queensland, Australia

Abstract

The phases of research used to evaluate new drugs provide a useful reference point for determining the studies that need to be conducted to evaluate new biomarkers. However, biomarkers do not have a single pathway for changing health outcomes and may be used for a variety of purposes, such as improving diagnostic criteria, improving prognosis, improving the monitoring of disease or as a measurement of health outcomes. The impact on health outcomes is also less direct and is dependent on the sequence of actions taken as a consequence of the test results. The different purposes of biomarkers and the less direct effect on health outcomes require different study designs to those used for the evaluation of pharmaceutical products and a more careful interpretation of results. Greater collaboration between researchers designing laboratory-based qualification studies and researchers designing clinical validation studies could achieve a process of evaluation for biomarkers that is both reliable and efficient.

Introduction

Over the past 50 years or so, there has been considerable progress in defining the methods needed to evaluate the effectiveness of health care interventions, particularly pharmaceutical products. Much of this increased understanding also applies to the evaluation of biomarkers. However, there are important differences that researchers, regulators and clinicians need to consider in the design and interpretation of studies to evaluate biomarkers.

The pathway between the use of a pharmaceutical intervention and the change in clinical outcome is reasonably direct. We can generally assume that the effects observed in a trial can be transferred to patients in other settings. The pathway between the use of a biomarker and a change in clinical outcome is both less direct and can occur at multiple points in the clinical pathway. Biomarkers can be used to diagnose the patient, measure disease severity, measure the response to treatment, monitor patients over time or to predict the prognosis of patients. In each case, the potential impact of the biomarker on health outcomes depends on the sequence of actions taken as a consequence of the test result. The potential multiple uses of biomarkers and their indirect impact on health outcomes means that researchers need to choose different study designs and a different sequence of study designs for their evaluation. As a previous author stated “the

complexity of these choices makes drug evaluation research seem simple and almost pedestrian”[1].

The pathway for evaluating biomarkers

As with pharmaceuticals, the pathway for establishing clinical applicability begins with the synthesis and purification of the product, the “discovery phase”. The proportion of biomarkers that go on to reach commercialisation after this phase is probably even smaller than the proportion of pharmaceutical products, in part because of the complexity of the evaluation process.

In general, the discovery phase is followed by the “qualification” or analytical validity studies. In these studies, the analytical procedures for the detection of the biomarker are standardised and factors which may impact on measurement are identified. Various guidelines identify this stage of the research cycle as being pre-clinical, Phase 1 or Phase 2 studies [1,2]. However it is labelled, some of the questions that may be addressed in these studies are:

- a) The optimal procedures for collecting and performing the assay
- b) The intra-individual and inter-individual reproducibility of the assay
- c) The relationship between the biomarker and severity of disease;

- d) The effect of factors such as age and sex on the biomarker and the establishment of normal reference ranges.

Qualification studies can be differentiated from validation studies to establish the clinical validity of the biomarker. Depending upon the proposed purpose of the biomarker, these studies address questions such as:

- a) How well does the biomarker differentiate between those who do or do not have a disease (diagnostic accuracy studies)?
- b) How well does a biomarker predict prognosis in patients with a disease?
- c) Does monitoring patients using the biomarker improve health outcomes?
- d) Can the biomarker be used as an intermediate outcome to measure response to treatment, for example in pharmaceutical evaluations?

The final phase of assessment of a biomarker is establishing the clinical utility of the biomarker – how much does it improve health outcomes and is it cost-effective?

In the evaluation of pharmaceutical products, the pathway from discovery to clinical utility is essentially a linear process, with each phase of studies following sequentially. In research on biomarkers, however, it may not be desirable or efficient to follow this sequence.

For example, a study of biomarkers in stored specimens is relatively quick and easy to produce. This type of case-control study can quickly establish differences in biomarker concentrations between those who do and do not have a disease, or those who do or do not progress to develop a disease and therefore can generate early hypotheses concerning potential diagnostic and prognostic markers. Because such studies are relatively cheap to produce, it is more efficient for these studies to be conducted early in the research pathway. Although these studies cannot establish the clinical validity of the biomarker, such as estimate the diagnostic accuracy of a biomarker, if these studies do not show a clear relationship between the target disorder and the biomarker, then further research on the analytical validity of the biomarker is unnecessary.

Analytical validity studies can also be informed by clinical validity studies. For example, there are two main forms of the B-type natriuretic peptide test: the biologically active BNP, and the biologically inactive amino-terminal peptide that is cleaved off after the release of the peptide from the myocardium, NT-proBNP. For both tests, there are several assays available including a point of care test. It is important to know if any of these assays is superior to other assays and if diagnostic accuracy varies by the type of assay. One way to do this is to look at the results of diagnostic accuracy studies using the BNP test and the results of tests using NT-proBNP and to compare the

results of the two sets of studies. However, it is important to avoid what is termed an “ecological fallacy” in epidemiology. For example, there have been 20 studies examining the accuracy of BNP for the diagnosis of clinically defined heart failure in primary care settings and 16 studies examining the accuracy of NT-proBNP for same purpose and in the same setting. We could compare the results of the two sets of studies. However, any difference in accuracy between the two sets of studies may be due to other differences between the two sets of studies, such as the severity of disease or the age of the patients. Such comparisons are likely to give rise to inaccurate conclusions. The most appropriate way to compare the two forms of the test would be to combine the results of studies which have directly compared the two forms of the test in the same patients, thus ensuring that the disease profile is the same for both tests.

Establishing the clinical validity of diagnostic biomarkers

One of the great advances in clinical medicine in the twentieth century was the development of the randomised controlled trial. Using the results of clinical trials, doctors and funding bodies are able to make decisions about the potential harms and benefits of health care interventions. We can do this because we can assume that the effects observed in a clinical trial are similar to the effect that we will see in the patients we treat. Of course, there are likely to be some differences. For example, we may see a patient who is older than patients included in the clinical trials and we may therefore adjust the dose of the drug, monitor more closely for side effects or not use the drug at all. But overall, the results in clinical trials are assumed to be the best evidence available for clinical decision making about treatments.

The same assumption is much more difficult, however, when we evaluate a diagnostic test. Much will depend on the sequence of events that occurs after diagnosis, such as the package of treatment options available for patients diagnosed with the target disorder and even the diagnostic and treatment options for those determined to have an alternative diagnosis. Because the treatment packages available can differ widely between clinical settings, the effect of a diagnostic test on health outcomes can be substantially different in different clinical settings.

Even more profoundly, and more difficult to predict, is the impact of a new diagnostic test on the spectrum of patients being tested. As doctors become more familiar with the use of a biomarker, it can change the spectrum of patients being tested for a disorder or the way that the biomarker is used in a diagnostic algorithm. Consider the example of troponins. The introduction of this biomarker has resulted in the spectrum of patients being tested for myocardial infarction

to change substantially over that time. In this case, even the definition of the target disorder has been changed, so that the spectrum of patients who are defined as having a myocardial infarction is considerably different now to twenty years ago. The change in the spectrum of patients being tested over time magnifies the difficulties in evaluating the effect of a biomarker on health outcomes.

The way that clinicians use diagnostic tests also makes it difficult to assess the effect of an individual clinical test on clinical outcomes. Clinicians use tests in a variety of ways. If you have studied some epidemiology you may have learnt about the Clinical epidemiologists emphasise the process of probability revision, that is combining an estimate of the pre-test probability of disease with the sensitivity and specificity of the test to estimate a post-test probability of disease. However, clinicians do not use tests in such a strictly quantitative way. The most common method used by clinicians for diagnosis is pattern recognition. The clinician recognises an overall pattern combining information from a variety of sources. The reason why experienced doctors are able to diagnose more accurately and more quickly than junior doctors is because they are able to recognise the pattern of a diagnosis in patients [3]. Although the human mind is superbly good at this process, it can be difficult to determine how one individual piece of information informs the understanding of the entire pattern.

There are also logistical difficulties in conducting clinical trials to evaluate the effect of diagnostic biomarkers. For example, in a trial of a pharmaceutical drug, the power of the trial is determined by the number of events that occur in the treatment and the control arms of the trial. In a trial of a diagnostic test, the power of the trial is driven by the proportion of patients who have a change in clinical outcomes as a result of the change in diagnosis. This can only occur in patients who have discordant results between the old method of diagnosis and the new method and therefore any observed difference in health outcomes is driven by only a small proportion of the patients enrolled in the trial. Therefore, the size of a trial of a diagnostic test needs to be several magnitudes larger than a trial of a therapeutic intervention. Clinical trials of diagnostic tests are highly subject to type II errors and this needs to be considered in their interpretation [4].

Because of these factors, it is not always reliable or efficient to test diagnostic biomarkers in randomised controlled trials [5,6] and studies of diagnostic accuracy may be sufficient. However, even in these studies it is important to clarify the proposed role of the biomarker in clinical practice and for this to be considered in the design of the clinical validity studies.

Traditional diagnostic accuracy studies examine the ability of a diagnostic test to differentiate between those who have and do not have a target disorder in a cohort of patients presenting with a suspected disease in a spe-

cific clinical setting. These studies provide the familiar estimates of sensitivity and specificity as measured against a reference standard test. However, such studies are not that helpful in establishing the clinical utility of a diagnostic test and helping to determine whether a clinician should or should not use a diagnostic test in a particular clinical setting or whether a funding body should or should not fund it. We cannot determine from these studies how much the test adds to other clinical information that is available or compare the diagnostic accuracy of two potential diagnostic tests.

It also difficult to generalise the results of diagnostic accuracy studies conducted in one clinical setting to another clinical setting or to patients with a different spectrum of disease. For example, there have been nearly 100 studies to evaluate the diagnostic accuracy of B-type natriuretic peptide, including randomised controlled trials of its use in various settings. However, it has not been studied in patients presenting with suspected heart failure in the primary care setting prior to referral to a specialist clinic, making it difficult to determine its effectiveness and cost-effectiveness in that setting [7].

Establishing clinical validity of other biomarkers

Diagnosis is only one of the potential uses of a biomarker. Other potential uses are monitoring, screening and prognosis. Each of these areas requires its own types of study designs. There has been surprising little research into the use of tests for monitoring disease, even though this is a common clinical activity and increasingly so with the shift towards greater care of chronic diseases [8]. For example, a recent randomised controlled trial of blood glucose monitoring in patients with type 2 diabetes showed that there was no difference in diabetic control and worse quality of life after 12 months in patients who used home blood glucose monitoring than in those monitored by a clinic test each 3 months [9]. This overturns results of previous trials which may be due to differences in treatments other than the blood glucose monitoring between the two arms of the trials and puts into questions a health care intervention which is estimated to cost more than £100,000,000 a year in the UK alone.

As for diagnostic biomarkers, it is not always necessary to use randomised controlled trials to establish if a new screening test improves health outcomes. For example, it has been shown in a systematic review of randomised controlled trials that faecal occult blood testing reduces mortality from colorectal cancer [10]. It would not be necessary to conduct a randomised controlled trial of a new biomarker for this disease if a new biomarker were to become available. The most appropriate study design in this case is a paired comparison of the diagnostic accuracy of the two tests.

Conclusion

Clinicians, funders and regulators need to be able to answer the following questions:

- Should clinicians use this test?
- Should clinicians use this test in this patient? and
- How do I use the results of the test in this patient?

Given the time and resources that are required to conduct studies of biomarkers, and the difficulties with both conducting and interpreting such studies, researchers need to look for more efficient and cost-effective methods for answering clinically relevant questions. There is a clear need for greater collaboration between researchers involved in analytical studies and those involved in clinical validation studies to ensure that research in biomarkers is both efficient and reliable. Even more than in other areas of clinical decision making, there is a large gap between the information that is needed by clinicians and policy makers to inform health care decisions and the information that is available from research. As more clinicians and researchers begin to understand the complexity of research in this area, it is to be hoped that this gap will be bridged more successfully.

Key points

Studies in biomarker research can be broadly divided into those used in the discovery phase, qualification studies, clinical validation studies and clinical utility studies.

Biomarker research does not necessarily follow a sequential order of research as for pharmaceutical products.

The purpose of the biomarker in clinical practice needs to be considered before the design of validation studies.

Randomised controlled trials of diagnostic biomarkers may not provide results that are transferable or generalisable and are often underpowered to detect a difference in health outcomes.

In some circumstances other study designs, such as diagnostic accuracy studies, can be sufficient to determine the evidence needed for clinical decision making.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

- [1] Ransohoff DF. How to improve reliability and efficiency of research about molecular markers: roles of phases, guidelines, and study design. *J Clin Epidemiol*. 2007 Dec;60(12):1205–19.
- [2] Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ*. 2002 Mar 2;324(7336):539–41.
- [3] Kassirer JP, Kopelman RI. Cognitive errors in diagnosis: instantiation, classification, and consequences. *Am J Med*. 1989 Apr;86(4):433–41.
- [4] Doust JA, Craig JC. Evaluating diagnostic tests—should the same methods apply? *Am Heart J*. 2008 Jul;156(1):4–6.
- [5] Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ*. 2006 May 6;332(7549):1089–92.
- [6] Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med*. 2006 Jun 6;144(11):850–5.
- [7] Mant J, Doust J, Roalfe A, Barton P, Cowie MR, Glasziou P, et al. Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care. *Health Technol Assess*. 2009 Jul;13(32):1–207.
- [8] Glasziou P, Irwig L, Mant D. Monitoring in chronic disease: a rational approach. *BMJ*. 2005 Mar 19;330(7492):644–8.
- [9] Farmer A, Wade A, Goyder E, Yudkin P, French D, Craven A, et al. Impact of self monitoring of blood glucose in the management of patients with non-insulin treated diabetes: open parallel group randomised trial. *BMJ*. 2007 Jul 21;335(7611):132.
- [10] Hewitson P, Glasziou P, Irwig L, Towler B, Watson E. Screening for colorectal cancer using the faecal occult blood test, Hemoccult. *Cochrane Database Syst Rev*. 2007 (1):CD001216.