In this issue: Focusing on Guidelines for Laboratory Medicine and Biomarkers

The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine

eJIFCC

Evidence and cost effectiveness requirements for recommending new biomarkers

Paul Collinson

Departments of Clinical Blood Sciences and Cardiology, St George's Hospital and Medical School, London, UK

ARTICLE INFO

Corresponding author:
Paul Collinson
Departments of Clinical Blood Sciences and Cardiology
St George's Hospital and Medical School
London SW17 0QT, UK

Key words:
biomarkers, cost effectiveness, evidence-based medicine, test evaluation

Disclaimer:
This article is based on a lecture given at IFCC Worldlab 2014 in Istanbul.

IS THIS TEST 'APT'?

The literature is full of new biomarkers which are claimed to add to the laboratory repertoire in a range of conditions. The literature is often confusing and may be contradictory. The past 20 years is littered with publications claiming the next big thing in a biomarker, some of which have been implemented on high throughput laboratory platforms. The number of novel biomarkers which have reached widespread clinical acceptance and implementation is relatively small. How can the laboratory community realistically assess claims for new markers? There is, to date, no completely defined set of criteria which should be used. However, there are some common themes in biomarker assessment. The two major areas which need to be considered are evidence required to assess test performance and cost effectiveness.

Assessment of test performance can be broadly considered under three categories, **A**nalytical suitability, **P**lausibility and **T**reatment effectiveness; is the test APT. *Analytical suitability* means an assessment of the evidence-based analytical performance of the assay. This will include at least the following. Pre-analytical factors that will affect the test must be well understood before a test can be put into routine clinical

practice. This will include the collection conditions required, anticoagulant requirements, pre-analytical sample handling factors and stability in storage. A marker needs to be measurable in the routine clinical laboratory without the need for special handling conditions if it is to form part of the routine work-up of the patient. Tests requiring complex pre-analytical steps are tolerated by the laboratory, rather than embraced. Often there is no alternative; the test is confined to special circumstances and particular patient types which are usually rare. A test in the clinical routine which will be ordered in large numbers requires simplicity of laboratory handling. A recent example is the measurement of soluble CD40 ligand (sCD40l), a marker of platelet activation. Measurement of sCD40l was shown to be a powerful predictor of mortality in patients with unstable angina. In addition, it was shown to be a predictor of a successful therapeutic response to the anti-glycoprotein IIb/IIIa antagonist abciximab (1). These studies were done using serum as matrix. It was subsequently found that clotting releases significant but variable amounts of sCD40l. Studies demonstrated that the release of sCD40l was critically affected by sample handling and the assay utilised for measurement (2).Only EDTA plasma could be used and values were significantly affected by delay in sample processing (3,4). Finally, it was shown that sCD40l was primarily produced by *in vitro* platelet activation (5) and the first use of a commercial assay failed to confirm the promise of the initial publication (6). Analytical performance of the test needs to be also appropriate for clinical use. Bodies such as the Clinical Laboratory Standards Institute produce protocols for the routine assessment of limit of blank, limit of detection and imprecision profile. It is also important that these analytical performance measures are independently assessed and that laboratories do not rely on the manufacturers' datasheets as the sole source of this information. Assay imprecision has a profound influence on the ability to define the 99th percentile and the value of the relative change required between two consecutive measurements to be reliably different. It is an interesting observation that the redefinition of myocardial infarction (7-9) considers a 10% imprecision to be adequate at the 99th percentile but also recommended a 20% change in values. Unfortunately, if the data is modelled it is apparent that an imprecision rather less than 10% is required to reliably detect a 20% change (http://www.westgard.com/troponin-interpretations.htm). In addition to the ability to measure the biomarker with precision and accuracy, the analysis must be simple and have a rapid turnaround time. Ideally it should be implemented on existing laboratory equipment rather than requiring additional apparatus. In practice this means that a colorimetric or more likely an immunoassay for the marker is available. Population aspects of the test need to be understood in particular the influence of age, gender, ethnicity and comorbid conditions on the reference interval need to be considered. These can be quite subtle. Occult comorbid conditions profoundly influence the reference interval for cardiac troponin but can only be unmasked by the use of rigorous patient selection including cardiac imaging (10,11). The need for appropriate patient selection for troponin reference intervals has been the subject of discussion and recommendations made (12,13).

The *plausibility* of the biomarker for the putative clinical role needs also to be established. The pathobiology of the biomarker needs to be understood. This means an understanding of the genesis of the biomarker and of the relationship of the biomarker to the medical condition of interest. A good example of this is ischaemia modified albumin (IMA). The concept of a biomarker of ischaemia is very attractive.

Ischaemia would be detected prior to necrosis (we have excellent markers for this in the cardiac troponins) allowing intervention to abort the pathophysiology before irreversible cardiac injury occurs. The background concept of IMA was that the N terminus of albumin was altered during an ischaemic event resulting in the loss of the ability to bind transition metals. This was detectable by loss of the ability to bind cobalt, which could be determined by a simple colorimetric reaction (14). Preliminary studies using angioplasty as a model of human myocardial ischaemia showed that IMA increased after balloon inflation then returned rapidly to baseline levels, supporting the role as a biomarker of ischaemia (15,16). Subsequently, sequencing of the N terminus of IMA positive albumin showed that the N-terminal amino acid sequence was not removed (17). Physicochemical studies suggested that it was the binding of free fatty acids to albumin that induced a conformational change that reduced transition metal binding (18). A lack of fundamental understanding of the biomarker was therefore apparent and contributed to the lack of any clinical application (19). Plausibility also includes the clinical plausibility for the putative clinical role. This means that the biomarker must have appropriate sensitivity and specificity to detect the medical condition of interest in clinically appropriate populations where the test will actually be used in routine clinical practice. Many studies on biomarkers have evaluated them in clinical trial sample banks or alternatively in highly selected patient groups. This does not constitute an appropriate environment to evaluate test performance as disease prevalence is inappropriately high, often close to 100%. Such studies allow proof of concept that needs to be followed up by prospective evaluation in clinically representative populations. Comparison of a sensitive with a less sensitive troponin assay clearly shows earlier diagnostic sensitivity (20), as would be expected. Early studies of the new high sensitivity assays showed excellent analytical performance but compared them with the conventional assays and included patients with ST segment elevation in the evaluation (21,22), overstating the diagnostic performance of the assays.

Treatment effectiveness is the final and most important strand to assessment. This may be summed up as the "so what" factor. This is short for the question that should be asked by any clinician of a test "so what do I need to do differently with the result of this test". A new biomarker must offer either a significant proven diagnostic efficiency or result in a change in treatment. Ideally it should do both. The change in treatment may be a decision to give or withhold drug or other therapeutic

Table 1 **Key questions for evaluating the evidence base for clinical use**

Has this marker been measured with an appropriate method and been shown to be additive to or replace a contemporary test?
Have there been independent studies?
Has there been a multicentre study?
Is there meta-analysis of evidence?
Has there been an RCT?
Can I measure it in the routine lab without additional equipment and staff?

intervention or to change the management pathway such as more prompt hospital discharge or admission to an appropriate level of clinical care. The questions which should pass through the laboratory practitioners' mind are shown in Table 1 below.

An example of a randomised controlled trial of the diagnostic test is the Randomised Assessment of Treatment using Panel Assays of Cardiac markers (RATPAC) (23). This was a pragmatic randomised controlled trial which compared two treatment strategies, conventional management with measurement on admission and at 90 minutes of a panel of cardiac troponin I, creatine kinase MB and myoglobin by point of care testing. The outcome measure was a proportion of patients discharged or a decision to discharge within four hours of attendance with no adverse events during the following three months. Randomisation to the point of care arm of the study was reflected in increased successful discharge and no change in the frequency of adverse events. There was increased use of coronary care in the point of care arm. One of the most interesting aspects of this study was the significant differences between the six different sites with only two showing very large differences in length of stay in those randomised to the point of care arm (24). It highlights the importance of process within the utilisation of test results. Simple provision of rapid results will be ineffective unless it is accompanied by treatment decision.

IS THIS TEST COST EFFECTIVE?

Cost effectiveness considers the impact on health care resources utilisation and how we assess it. Cost effectiveness can be considered under four categories as shown in Table 2 below. It should be noted however that the terminology is often mixed.

Cost minimisation analysis is the most straightforward. It assumes that the consequences of the two interventions being compared are identical so the analysis reduces to the comparison of costs alone. An example would be the diagnosis of acute myocardial infarction using cardiac troponin (cTn) compared to the measurement of creatine kinase MB isoenzyme (CK-MB). If the assumption is that CK-MB costs 20 currency units (CU) and cTn 30 CU then a protocol involving three hourly CK-MB measurements for 12 hours (total cost 80 CU) will be more expensive than a protocol measuring cTn on admission and 12 hours from admission (total cost 60 CU). In cost effectiveness analysis differences can be expressed in terms of changes in one main parameter. The differences in costs are related to the main differences in events. An example of this type of analysis is the use of measurement of B type natriuretic peptide (BNP) in patients with suspected chronic heart failure. The basic

Table 2 Cost effectiveness categories

Type	Measurement and valuation of consequences
Cost minimisation analysis	No measurement. Consequences assumed or shown to be equivalent.
Cost effectiveness analysis	Natural units (Life years gained)
Cost utility analysis	Health state preference values (quality adjusted life years gained)
Cost benefit analysis	Monetary gains

premise is that two pathways are compared: direct referral for hospital assessment of patients with suspected heart failure and referral only of those with an elevated BNP. A simple analysis compares costs at the pathway level where the costs of echocardiography on all patients is compared with the combined cost of BNP measurement followed by echocardiography only in the those with BNP levels above a certain designated threshold. This is effectively a cost minimisation analysis and shows that the BNP based pathway is cheaper (25). A more sophisticated approach utilising a sequential testing strategy modelled on individual patient data meta-analysis was performed as part of a health technology assessment informing the National Institute of Clinical and health Excellence (NICE) guidelines on BNP testing. This modelling produced very similar results to the cost minimisation model. Cost effectiveness was driven by the prior probability of disease and favoured BNP measurement as the first test (as in strategy discussed above) unless the probability of heart failure was very high (26). Cost utility analysis typically utilises the quality adjusted life year (QALY). A QALY takes into account longevity and quality-of-life. The number of QALYs accrued by a patient is estimated by multiplying the years of survival by quality-of-life measured on a scale from zero (equivalent to death) to 1 (perfect health). States of health below zero are possible for a health state considered worse than death. QALYs have the advantage of allowing comparison between any healthcare intervention that can influence survival or quality-of-life. Analysis is based on willingness to pay (cost per QALY) with a typical threshold of £20,000 in the UK. An example would be comparison of the cost effectiveness of measurement of high sensitivity troponin on admission versus conventional troponin management at 10 hours (27). Such a study shows that high sensitivity troponin measurement on admission is superior to conventional troponin measurement and that measurement on admission and at three hours is the most sensitive approach. Measurement of conventional troponin at 10 hours is only cost effective if an immediate decision to discharge is made, highlighting again the importance of process in the application of laboratory testing. One problem with cost effectiveness analysis in diagnostics is that the data is often inadequate or even non-existent. Modelling approaches are typically used but the accuracy of the cost modelling is often challenging though mitigated by sensitivity analysis (changing the model parameter and looking at the impact, a large change suggests that the modelling is not robust). Very small differences in QALY's may be present.

A systematic attempt to evaluate the evidence for diagnostics including laboratory testing is used by the Diagnostics Assessment Committee of NICE. They utilise a systematic evidence-based review followed by cost economic modelling. The recommendations and their evidence base can be found on the NICE website (www.nice.org.uk) and in the publications of the UK health technology assessment programme. These are all available online. Examples are the recent recommendations for the use of faecal calprotectin (www.nice.org.uk/guidance/dg11) and the accompanying evidence report (28).

CONCLUSIONS

In conclusion, assessment of test suitability is a combination of the traditional laboratory attributes of the analytical performance of the test but combined with other features. The underlying scientific validity of the test needs to be understood and the diagnostic utility demonstrated in appropriate populations, to show the test is plausible. Finally, the test result must produce a treatment change. All of these, Analytical, Plausibility, Treatment will make a test APT. But an APT test clinically also

needs to be cost effective. Conversely, unless a test has been shown to be APT, the probability of demonstrating cost effectiveness is small. The challenge for the laboratory is to work together with clinicians to develop test evaluation strategies that will allow demonstration of all the attributes to show that the test is both APT and cost effective.

REFERENCES

1. Heeschen C, Dimmeler S, Hamm CW, van den Brand MJ, Boersma E, Zeiher AM, Simoons ML. Soluble CD40 ligand in acute coronary syndromes. N Engl J Med 2003;348:1104-11.

2. Halldorsdottir AM, Stoker J, Porche-Sorbet R, Eby CS. Soluble CD40 ligand measurement inaccuracies attributable to specimen type, processing time, and ELISA method. Clin Chem 2005;51:1054-7.

3. Weber M, Rabenau B, Stanisch M, Elsaesser A, Mitrovic V, Heeschen C, Hamm C. Influence of sample type and storage conditions on soluble CD40 ligand assessment. Clin Chem 2006;52:888-91.

4. Weber M, Rabenau B, Stanisch M, Nef HM, Mollmann H, Elsasser A et al. Influence of sample type on soluble CD40 ligand assessment in patients with acute coronary syndromes. Thromb Res 2007;120:811-4.

5. Ivandic BT, Spanuth E, Haase D, Lestin HG, Katus HA. Increased plasma concentrations of soluble CD40 ligand in acute coronary syndrome depend on in vitro platelet activation. Clin Chem 2007;53:1231-4.

6. Plaikner M, Peer A, Falkensammer G, Schmidauer C, Pechlaner C, Griesmacher A et al. Lack of association of soluble CD40 ligand with the presence of acute myocardial infarction or ischemic stroke in the emergency department. Clin Chem 2009;55:175-8.

7. Myocardial infarction redefined--a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. Eur Heart J 2000;21:1502-13.

8. Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD et al. Third universal definition of myocardial infarction. Eur Heart J 2012;33:2551-67.

9. Thygesen K, Alpert JS, White HD, Jaffe AS, Apple FS, Galvani M et al. Universal definition of myocardial infarction. Circulation 2007;116:2634-53.

10. Collinson PO, Heung YM, Gaze D, Boa F, Senior R, Christenson R, Apple FS. Influence of population selection on the 99th percentile reference value for cardiac troponin assays. Clin Chem 2012;58:219-25.

11. Koerbin G, Abhayaratna WP, Potter JM, Apple FS, Jaffe AS, Ravalico TH, Hickman PE. Effect of population selection on 99th percentile values for a high sensitivity cardiac troponin I and T assays. Clin Biochem 2013;46:1636-43.

12. Sandoval Y, Apple FS. The global need to define normality: the 99th percentile value of cardiac troponin. Clin Chem 2014;60:455-62.

13. Apple FS, Jaffe AS, Collinson P, Mockel M, Ordonez-Llanos J, Lindahl B et al. IFCC educational materials on selected analytical and clinical applications of high sensitivity cardiac troponin assays. Clin Biochem 2015;48(4-5):201-3.

14. Bar-Or D, Lau E, Winkler JV. A novel assay for cobalt-albumin binding and its potential as a marker for myocardial ischemia-a preliminary report. J Emerg Med 2000;19:311-5.

15. Bar-Or D, Winkler JV, Vanbenthuysen K, Harris L, Lau E, Hetzel FW. Reduced albumin-cobalt binding with transient myocardial ischemia after elective percutaneous transluminal coronary angioplasty: a preliminary comparison to creatine kinase-MB, myoglobin, and troponin I. Am Heart J 2001;141:985-91.

16. Sinha MK, Gaze DC, Tippins JR, Collinson PO, Kaski JC. Ischemia modified albumin is a sensitive marker of myocardial ischemia after percutaneous coronary intervention. Circulation 2003;107:2403-5.

17. Bhagavan NV, Lai EM, Rios PA, Yang J, Ortega-Lopez AM, Shinoda H et al. Evaluation of human serum albumin cobalt binding assay for the assessment of myocardial ischemia and myocardial infarction. Clin Chem 2003;49:581-5.

18. Mothes E, Faller P. Evidence that the principal CoII-binding site in human serum albumin is not at the N-terminus: implication on the albumin cobalt binding test for detecting myocardial ischemia. Biochemistry 2007;46:2267-74.

19. Collinson PO, Gaze DC. Ischaemia-modified albumin: clinical utility and pitfalls in measurement. J Clin Pathol 2008;61:1025-8.

20. Melanson SE, Morrow DA, Jarolim P. Earlier detection of myocardial injury in a preliminary evaluation using a new troponin I assay with improved sensitivity. Am J Clin Pathol 2007;128:282-6.

21. Keller T, Zeller T, Peetz D, Tzikas S, Roth A, Czyz E et al. Sensitive troponin I assay in early diagnosis of acute myocardial infarction. N Engl J Med 2009;361:868-77.

22. Reichlin T, Hochholzer W, Bassetti S, Steuer S, Stelzig C, Hartwiger S et al. Early diagnosis of myocardial

infarction with sensitive cardiac troponin assays. N Engl J Med 2009;361:858-67.

23. Goodacre SW, Bradburn M, Cross E, Collinson P, Gray A, Hall AS. The Randomised Assessment of Treatment using Panel Assay of Cardiac Markers (RATPAC) trial: a randomised controlled trial of point-of-care cardiac markers in the emergency department. Heart 2011;97:190-6.

24. Bradburn M, Goodacre SW, Fitzgerald P, Coats T, Gray A, Hassan T et al. Interhospital variation in the RATPAC Trial (Randomised Assessment of Treatment using Panel Assay of Cardiac markers). Emerg Med J 2012;29:233-8.

25. Collinson PO. The cost effectiveness of B-Type natriuretic peptide measurement in the primary care setting-a UK perspective. Congest Heart Fail 2006;12:103-7.

26. 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;13:1-207, iii.

27. Thokala P, Goodacre SW, Collinson PO, Stevens JW, Mills NL, Newby DE et al. Cost-effectiveness of presentation versus delayed troponin testing for acute myocardial infarction. Heart 2012;98:1498-503.

28. Waugh N, Cummins E, Royle P, Kandala NB, Shyangdan D, Arasaradnam R et al. Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel diseases: systematic review and economic evaluation. Health Technol Assess 2013;17:xv-211.