

Editors' view

Biomarkers and surrogate endpoints

J. K. Aronson, Chairman of the Editorial Board, British Journal of Clinical Pharmacology

University Department of Clinical Pharmacology, Radcliffe Infirmary, Woodstock Road, Oxford OX2 6HE, UK

When David Beckham leaves the field towards the end of a match, the man who replaces him is a surrogate. and although I suspect that many footballers, if asked, would say that Surrogate is a town in Yorkshire, the word actually comes from the Latin word *subrogare*, to substitute.

The use of the term 'surrogate marker' in medicine dates from the late 1980s [1], but it had been preceded by some years by the term 'biomarker' [2] and was succeeded and replaced by yet another term, 'surrogate endpoint' [3]. To see where these terms stand in relation to each other we need to define them from the bottom up.

A surrogate endpoint has been defined as 'a biomarker intended to substitute for a clinical endpoint', the latter being 'a characteristic or variable that reflects how a patient feels, functions, or survives' [4]. So, what is a biomarker? Well, that has been defined as 'a characteristic that is objectively measured and evaluated as an indication of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention' [4]. Although what the marker marks is clearly defined here as being intrinsic, that is not true of the marker itself, which can be intrinsic or extrinsic. This leads me to propose a nosography of biomarkers, as listed in Table 1. The categories in the table could be further subdivided according to whether the markers are being used for diagnosis, staging, or monitoring of disease or for determining its response to therapy.

Advantages of biomarkers

Biomarkers are often cheaper and easier to measure than 'true' endpoints. For example, it is easier to measure a patient's blood pressure than to use echocardiography to measure left ventricular function, and it is much easier to do echocardiography than to measure morbidity and mortality from hypertension in the long term. Biomark-

ers can also be measured more quickly and earlier. Blood pressure can be measured today, whereas it takes several years to collect mortality data. In clinical trials the use of biomarkers leads to smaller sample sizes. For example, to determine the effect of a new drug on blood pressure a relatively small sample size of say 100–200 patients would be needed and the trial would be relatively quick (1–2 years). To study the prevention of deaths from strokes a much larger study group would be needed and the trial would take many years. There may also be ethical problems associated with measuring true endpoints. For example, in paracetamol overdose it is unethical to wait for evidence of liver damage before deciding whether or not to treat a patient; instead a pharmacological biomarker, the plasma paracetamol concentration, is used to predict whether treatment is required.

Criteria for useful biomarkers

There are many links in the chain of events that leads from the pathogenesis of a disease to its clinical manifestations; biomarkers can be used at any point in the chain, at the molecular, cellular, or organ levels. Likewise, a therapy might be developed to attack any one of these links, in order to try to manipulate the disease, symptomatically or therapeutically. Any measurement short of the actual outcome could be regarded as a surrogate endpoint biomarker. However, although all surrogate endpoints are biomarkers, not all biomarkers are useful surrogate endpoints.

The ideal biomarker is one through which the disease comes about or through which an intervention alters the disease. For example, the serum cholesterol concentration should be an excellent diagnostic marker for cardiovascular disease; however, there is no clear cut-off point, and only about 10% of those who are going to have a stroke or heart attack have a serum cholesterol

Table 1

Types of biomarkers that can be used in diagnosing, staging, and monitoring disease, and in determining its response to therapy

Type of biomarker	Example of a relevant surrogate endpoint	The relevant clinical endpoint
1. Extrinsic markers	Cigarette consumption Daily defined dose	Lung cancer Drug consumption
2. Intrinsic markers		
a. Physical		
i. Clinical	Lid lag	Hyperthyroidism
ii. Radiographic	White dots on MRI scan	Lesions of multiple sclerosis
b. Laboratory*		
i. Physiological	Blood pressure	Stroke
ii. Pharmacological		
1. Exogenous	Inhibition of CYP isozymes	Routes of drug metabolism
2. Endogenous	Docetaxel clearance	Febrile neutropenia
iii. Biochemical	Serum TSH	Hypothyroidism
iv. Haematological	INR with warfarin	Pulmonary embolism
v. Immunological	Autoantibodies	Autoimmune diseases
vi. Microbiological	<i>C. difficile</i> toxin	Pseudomembranous colitis
vii. Histological	Jejunal biopsy	Gluten-sensitive enteropathy
viii. Genetic	CYP2C19 isoforms	Warfarin dosage

*Including bedside tests requiring laboratory equipment.

concentration above the reference range. But even if cholesterol is not a good diagnostic marker, it can still be used as a marker of therapeutic response to cholesterol lowering drugs.

Other useful biomarkers are not directly related to the clinical endpoint, but are affected in parallel with the disease. In some cases they are good diagnostic markers but not good markers of progress (for example, prostate specific antigen in prostatic cancer), or conversely they may be good markers of progress but not helpful diagnostically (for example carcinoembryonic antigen in ovarian carcinoma).

In looking for criteria for deciding which biomarkers are good candidates for surrogate endpoints we can turn to the guidelines that Austin Bradford Hill propounded for helping to analyse association in determining causation (Table 2) [5, 6]. He propounded these guidelines in the context of environmental causes of disease, but they can be used in other spheres [7]. Whenever a biomarker conforms to these guidelines, it is more likely to be useful. Note that simply because a biomarker fulfills the guidelines it will not necessarily be useful; it merely makes it more likely to be useful.

Problems with surrogate endpoints

Surrogate endpoints are most likely to be useful when the pathophysiology of the disease and the mechanism

of action of the intervention are thoroughly understood. Otherwise, pitfalls await.

For instance, smoking causes lung cancer, and a trial of the benefit of education in preventing lung cancer might use smoking as a surrogate endpoint rather than the occurrence of the cancer itself. On the other hand, if chemotherapy is used as a measure for treating lung cancer, smoking could not be used as a surrogate endpoint. This is obvious, but alerts us to the possibility of similar but less obvious examples, in which the mechanisms are not understood.

Ventricular arrhythmias cause sudden death, and antiarrhythmic drugs prevent ventricular arrhythmias. It was therefore expected that antiarrhythmic drugs would prevent sudden death. In fact, in the Cardiac Arrhythmia Suppression Trial [8], Class I antiarrhythmic drugs increased sudden death significantly in patients with asymptomatic ventricular arrhythmias after a myocardial infarction, and the trial was stopped prematurely. The hypothesis was wrong.

Another good example is enalapril and vasodilators, such as hydralazine and isosorbide, whose haemodynamic effects and effects on mortality associated with heart failure are dissociated. Vasodilators improved exercise capacity and improved left ventricular function to a greater extent than enalapril. However, enalapril reduced mortality significantly more than vasodilators

Table 2

Austin Bradford Hill's guidelines that increase the likelihood that an association is causative

Guidelines	Characteristics of useful biomarkers
Strength	A strong association between marker and outcome, or between the effects of a treatment on each
Consistency	The association persists in different individuals, in different places, in different circumstances, and at different times.
Specificity	The marker is associated with a specific disease
Temporality	The time-courses of changes in the marker and outcome occur in parallel
Biological gradient (dose-responsiveness)	Increasing exposure to an intervention produces increasing effects on the marker and the disease
Plausibility	Credible mechanisms connect the marker, the pathogenesis of the disease, and the mode of action of the intervention
Coherence	The association is consistent with the natural history of the disease and the marker
Experimental evidence	An intervention gives results consistent with the association
Analogy	There is a similar result to which we can adduce a relationship

[9]. So in this case haemodynamic effects are not a good surrogate.

Patients with asthma feel breathless if they have a low peak expiratory flow rate (PEFR). However, in one study different drugs produced different relationships between PEFR and breathlessness [10]. Patients taking beclomethasone did not feel as breathless as those taking theophylline for a given PEFR. So what should the surrogate marker be – the 'hard' endpoint of peak flow or the 'soft' marker of how the patients felt? This also raises the question of whether more than one surrogate endpoint should be used in clinical trials.

Confounding factors can nullify the value of surrogate endpoints. For example, serum T₃ is used as a marker of the tissue damage that thyroid hormone causes in patients with hyperthyroidism. However, its usefulness is blunted in patients taking amiodarone, which interferes with the conversion of T₄ to T₃ without necessarily altering thyroid function.

Statistical problems with surrogate endpoints

A surrogate endpoint has been defined statistically as 'a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint' [11]. Often the surrogate endpoint is used as an entry criterion in clinical trials, and it is important to be aware that this can lead to statistical problems [12]. It introduces heterogeneous variance and the problem of regression to the mean. If someone is entered into a trial on the basis of an abnormal surrogate marker and then receives no treatment, the surrogate endpoint will still improve, simply because of the statistical variation in the measurement of variables. This reduces the power of a study. There is also a high

likelihood of missing data when surrogate endpoints are used. Using a small sample size when using a surrogate endpoint may also mean that a study is not big enough to detect adverse effects of drugs.

New biomarkers

We use biomarkers all the time. We measure the time to relapse in a patient with cancer as a surrogate endpoint for survival time. We measure ocular pressure instead of loss of vision in patients with glaucoma. We use biomarkers to stage disease (for example the number of lymph nodes affected by cancer), in diagnosis (for example serum T₃, electrocardiography, and autoantibodies), and to monitor the progress of a disease or its treatment (for example, blood glucose concentration, blood pressure, and FEV₁).

The search for useful biomarkers is a constant one. For example, in this month's issue of the *Journal* Dumont et al. have investigated neuropsychological, neurophysiological, and neuroendocrine tests and motor skills as biomarkers of the effects of SSRIs in healthy subjects [13]. They applied three of the Bradford Hill guidelines: consistency, dose-responsiveness in the therapeutic range, and biological plausibility. They concluded that there is no single marker of value, but that a combination of markers is best. Their search was impressively thorough – 171 different tests in 56 studies – but it is a little surprising that among the 79 references that they cited they did not include the citation classic by Hindmarch [14], which featured in our thirtieth anniversary special issue last year, and in which the same conclusion was reached about psychoactive drugs in general [15,16].

Another type of biomarker appears elsewhere in this issue of the *Journal*. If a drug is adsorbed by charcoal

and is excreted into the gut via the liver or secreted via enterocytes, activated charcoal will prevent its reabsorption. This property of activated charcoal has been put to good use in the treatment of self-poisoning. For example, in a large, randomized, placebo-controlled trial in cardiac glycoside poisoning in Sri Lanka, due to ingestion of yellow oleander seeds taken with suicidal intent, multiple-dose activated charcoal reduced mortality from 8% to 2.5%, a striking effect [17]. Stass *et al.* [18] have used this action of activated charcoal for a different purpose – as an exogenous pharmacological biomarker of the extent to which enterohepatic or enteroenteric recycling contributes to the systemic availability of moxifloxacin. Oral charcoal increased the clearance of a single intravenous dose of moxifloxacin by about 24%.

Other papers in this issue of the *Journal* imply the use of biomarkers, whether you realize it or not:

- a study of adherence of middle-aged patients to statins, given the premise that age and pre-existing cardiovascular disease are the best markers of the risk of stroke and heart attack [19];
- the potential use of the prevalence of khat chewers in populations in which the habit is common as a marker of the effectiveness of a heart attack prevention programme, since chewing khat is associated with a greatly increased risk of myocardial infarction [20];
- the use, in a Chinese population taking warfarin, of the INR as a marker of the risks of major bleeding or thromboembolism [21];
- the potential use of measurement of the N-terminal propeptide of type III procollagen as a marker for the haemodynamic effects of spirinolactone [22].

Conclusion

There are clear potential benefits in using biomarkers. Information can be obtained earlier, more quickly, and more cheaply. However, the chain of events in a disease process linking pathogenesis to outcome is fragile and the better we understand the nature of the path a disease takes and the pharmacology of a drug that affects it the better biomarkers we will be able to develop in diagnosing, staging, and monitoring disease and its response to therapy.

References

- 1 Brotman B, Prince AM. Gamma-glutamyltransferase as a potential surrogate marker for detection of the non-A, non-B carrier state. *Vox Sang* 1988; 54: 144–7.
- 2 Paone JF, Waalkes TP, Baker RR, Shaper JH. Serum UDP-galactosyl transferase as a potential biomarker for breast carcinoma. *J Surg Oncol* 1980; 15: 59–66.
- 3 Boone CW, Kelloff GJ. Intraepithelial neoplasia, surrogate endpoint biomarkers, and cancer chemoprevention. *J Cell Biochem Suppl* 1993; 17F: 37–48.
- 4 NIH Definitions Working Group. Biomarkers and surrogate endpoints in clinical research: definitions and conceptual model. In: Downing GJ (ed.) *Biomarkers and Surrogate Endpoints*. Amsterdam: Elsevier, 2000: 1–9.
- 5 Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965; 58: 295–300.
- 6 Legator MS, Morris DL. What did Sir Bradford Hill really say? *Arch Environ Health* 2003; 58: 718–20.
- 7 Shakir SA, Layton D. Causal association in pharmacovigilance and pharmacoepidemiology: thoughts on the application of the Austin Bradford-Hill criteria. *Drug Saf* 2002; 25: 467–71.
- 8 The Cardiac Arrhythmia Suppression Trial Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomised trial of arrhythmia suppression after myocardial infarction. *New Engl J Med* 1989; 321: 406–12.
- 9 Cohn J. Lessons from V-HeFT: questions for V-HeFT11 and the future therapy of heart failure. *Herz* 1991; 16: 267–71.
- 10 Higgs CMB, Laszlo G. Influence of treatment with beclomethasone, cromoglycate and theophylline on perception of bronchoconstriction in patients with bronchial asthma. *Clin Sci* 1996; 90: 227–34.
- 11 Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989; 8: 431–40.
- 12 Wittes J, Lakatos E, Probstfield J. Surrogate endpoints in clinical trials: cardiovascular diseases. *Stat Med* 1989; 8: 415–25.
- 13 Dumont GJH, De Visser SJ, Cohen AF, Van Gerven JMA. Biomarkers for the effects of selective serotonin reuptake inhibitors (SSRIs) in healthy subjects. *Br J Clin Pharmacol* 2005; 59: 495–510.
- 14 Hindmarch I. Psychomotor function and psychoactive drugs. *Br J Clin Pharmacol* 1980; 10: 189–209.
- 15 Hindmarch I. Psychomotor function and psychoactive drugs. Author's commentary. *Br J Clin Pharmacol* 2004; 58: S741–S742.
- 16 Cohen A. Psychomotor function and psychoactive drugs. Independent commentary. *Br J Clin Pharmacol* 2004; 58: S742–S743.
- 17 De Silva HA, Fonseka MM, Pathmeswaran A, Alahakone DG, Ratnatilake GA, Gunatilake SB, Ranasinha CD, Laloo DG, Aronson JK, de Silva HJ. Multiple-dose activated charcoal for treatment of yellow oleander poisoning: a single-blind, randomised, placebo-controlled trial. *Lancet* 2003; 361: 1935–8.
- 18 Stass H, Kubitzka D, Moller J-G, Delesen H. Influence of activated charcoal on the pharmacokinetics of moxifloxacin following intravenous and oral administration of a 400 mg single dose to healthy males. *Br J Clin Pharmacol* 2005; 59: 536–41.
- 19 Perreault S, Blais L, Lamarre D, Dragomir A, Berbiche D, Lalonde L, Laurier C, St-Maurice F, Collin J. Persistence and determinants of statin therapy among middle-aged patients for primary and secondary prevention. *Br J Clin Pharmacol* 2005; 59: 564–73.
- 20 Al-Motarreb A, Briancon S, Al-Jaber N, Al-Adhi B, Al-Jailani F, Salek MS, Broadley KJ. Khat chewing is a risk factor for acute myocardial infarction: a case-control study. *Br J Clin Pharmacol* 2005; 59: 574–81.
- 21 You JHS, Chan FWH, Wong RSM, Cheng G. Is INR between 2.0 and 3.0 the optimal level for Chinese patients on warfarin therapy for moderate-intensity anticoagulation? *Br J Clin Pharmacol* 2005; 59: 582–7.
- 22 Davies J, Gavin A, Band M, Morris A, Struthers A. Spirinolactone reduces brachial pulse wave velocity and PIIINP levels in hypertensive diabetic patients. *Br J Clin Pharmacol* 2005; 59: 520–3.