E23. Statistical issues in the validation of biomarkers and surrogate endpoints

Marc Buyse*

International Drug Development Institute (IDDI), Louvain-la-Neuve, Belgium

Definitions

The definitions of biomarkers and endpoints used in this review are summarised in Table 1.[1] Biomarkers may be imaging-based or physiological indicators, but with the advent of the targeted therapy era, cellular, molecular and genetic biomarkers are becoming increasingly important. Biological considerations play a key role in the initial identification of prognostic and predictive biomarkers, but biological considerations must be interpreted with caution, and may in some cases prove misleading. For instance, recent studies suggest that HER2neu-directed therapies may be clinically beneficial even in the absence of HER2neu overexpression, a finding that would not be anticipated on the basis of HER2 biology as currently understood. Ultimately, biology cannot substitute for the clinical / statistical validation of biomarkers through randomised clinical trials and meta-analyses. The Early Breast Cancer Trialists' Collaborative Group (EBCTCG), for instance, currently holds the most extensive data on the hazard rates over time for hormone receptor positive and negative tumours.[2] The analyses conducted by this group have confirmed that receptor hormone status is both a prognostic marker for outcome in breast cancer, and a predictive marker for therapeutic response to endocrine therapies such as tamoxifen and aromatase inhibitors. It is essential that the EBCTCG continue to collect data on clinical outcomes but also on biomarkers to carry out large-scale validation of biomarkers in the future.

Table 1. Definitions

Term	Definition
Biomarker	A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention
Prognostic biomarker	Biomarker that forecasts the likely course of disease irrespective of treatment
Predictive biomarker	Biomarker that forecasts the likely response to a specific treatment
Clinical endpoint	Measurement providing systematic information on how a patient feels, functions or survives
Surrogate endpoint	Measurement providing early and accurate prediction of both a clinical endpoint, and the effects of treatment on this endpoint
Validation	Confirmation by robust statistical methods that a candidate prognostic biomarker, predictive biomarker or surrogate endpoint fulfils a set of conditions that are necessary and sufficient for its use in the clinic

Validating prognostic biomarkers

For a biomarker to be validated as prognostic, an association must be demonstrated between the presence and absence of the marker at baseline, or changes in the biomarker over time, and a clinical endpoint, independently of treatment (Table 2). This is a straightforward requirement that can be verified in small retrospective studies. However, the establishment of a statistical correlation may not be sufficient to make a prognostic biomarker useful in clinical practice.[3] For instance, patients with early breast cancer and a poor-prognosis by the 70-gene MammaPrint™ signature were found to have a large and highly significant odds ratio (around 15.0, $p < 0.0001$) for distant metastases within 5 years, when compared with patients with a good-prognosis signature. The negative predictive value was 0.9 (i.e. only one patient in 10 with a good-prognosis signature would be expected to develop metastases within 5 years), but the positive predictive value of the signature was only 0.63 (i.e. about two thirds of the patients with a poor-prognosis signature would be expected to develop metastases within 5 years). These findings, which were confirmed in an independent multicentre validation study, indicate that while the signature may be useful in avoiding aggressive chemotherapy in patients with a good prognosis, it is not a sufficiently accurate predictor of which patients will, or will not, develop metastases to provide the sole basis for a treatment decision.[4] The ultimate proof of usefulness in the clinic still requires randomised, prospective evidence in clinical trials.[5]

Validating predictive biomarkers

Predictive markers present even greater challenges, both with respect to initial demonstration of a correlation, and subsequent robust validation (Table 2). Statistical identification of predictive markers requires data from randomised trials which include patients with both high and low levels of the biomarker.[6] The highest level of evidence derives from trials with an "interaction" design, in which all patients are stratified by biomarker level and then randomised to the same two treatments. Large numbers of events are generally required to detect interactions reliably, which limits the number of interaction trials that can realistically be conducted. An

1359-6349/$ – see front matter

Table 2: Use, validation and examples of prognostic biomarkers, predictive biomarkers and surrogate endpoints in oncology

Type of biomarker	Uses in management and clinical trials	Identification	Validation	Examples
Prognostic biomarker	Treatment choice, patient selection and stratification in trials	Easy, but often flawed or biased	Frequent, but sometimes disappointing because of regression to the mean or flaws in the initial identification study	Nodal status and tumour stage in early breast cancer Performance status and the presence of visceral metastases in advanced breast cancer
Predictive biomarker	Treatment choice, patient selection and stratification in trials	Difficult, requires randomised trial	Uncommon, requires large randomised trial	Hormonal receptor status predictive of effect of tamoxifen and aromatase inhibitors in early and advanced breast cancer HER2neu amplification predictive of effect of trastuzumab and lapatinib in early and advanced breast cancer
Surrogate endpoint	Treatment choice, treatment evaluation in trials	Very difficult, requires meta-analysis or large randomised trial	Rare, requires meta-analysis or large randomised trial	DFS surrogate for OS for the efficacy of chemotherapy and endocrine therapy in early breast cancer

alternative approach consists of entering only marker-positive patients in the validation trial, but the presence of a benefit in these patients does not automatically imply that the marker is truly predictive.

Validating surrogate endpoints

At present, there are few accepted surrogate endpoints in clinical oncology, and none based on tumour response, molecular or genetic markers in solid tumours (in contrast, in non-solid tumours, haematological complete remission has long been considered a surrogate for time to disease progression and overall survival). While validation criteria for surrogate biomarkers are still an area of intense statistical research, there is an emerging consensus that validation can be based on a "correlation approach" involving demonstration in randomised trials or meta-analyses that the surrogate is prognostic for disease outcome ("individual-level surrogacy"), and that the effect of intervention on the surrogate is correlated with the effect on the true endpoint ("trial-level surrogacy") (Table 2). [7,8]

Conflict of interest statement

None declared.

References

[1] Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;69:89–95.

[2] Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Chemotherapy and hormonal therapy for early breast cancer: effects on recurrence and 15-year survival in an overview of the randomized trials. *Lancet* 2005;366:2087–106.

[3] Hayes DF, Trock B, Harris AL. Assessing the clinical impact of prognostic factors: when is "statistically significant" clinically useful? *Breast Cancer Res* 1998;52:305–19.

[4] Buyse M, Loi S, van't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for patients with node-negative breast cancer. *J Natl Cancer Instit* 2006;98:1183–92.

[5] Bogaerts J, Cardoso F, Buyse M, et al. Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat Clin Practice Oncol* 2006;3:540–51.

[6] Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol* 2005;23:2020–7.

[7] Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Stat Med* 2006;25:183–203.

[8] Burzykowski T, Molenberghs G, Buyse M (editors). *The Evaluation of Surrogate Endpoints*. New York: Springer; 2005.