Tumour marker development: Towards validation of clinically useful markers

Daniel F. Hayes

Breast Oncology Program, University of Michigan Comprehensive Cancer Center, Ann Arbor, MI, USA

Introduction

The overall goal in caring for a patient with any disease is to provide benefit and minimise toxicity. Given the nature of anti-neoplastic therapies, this goal is especially important in the practice of oncology, where benefits may include saving lives or inducing remarkable palliation but risks can involve extremely problematic side effects and occasional life-threatening toxicities. There has been great hope that the ongoing expansion of our understanding of the molecular biology of cancer would lead to both novel targeted therapies as well as markers that might direct established standard therapies. However, the reality of this hope has been relatively disappointing. Indeed, the Tumour Markers Guidelines Panel of the American Society of Clinical Oncology (ASCO), composed of a group of experts in the field, has reviewed many putative markers (including both tissue-based and circulating markers), but its ultimate recommendations have been surprisingly sparse (Table 1) [1–4].

Why are these guidelines so conservative? A major reason is that most tumour marker investigations have been studies of convenience in which an interesting new assay is applied to archived samples that happen to be available. However, for a marker to be clinically useful, the following three conditions must pertain: (1) the use of the marker must be clearly articulated and the studies to define whether outcomes are improved by use of that marker must be carefully planned; (2) the differences in point estimates of outcomes for patients who are 'positive' versus those who are 'negative' must be sufficiently different that a clinician would treat those patients differently; and (3) the estimate of these differences must be reliable, based on both technical and reproducible factors and on the quality of the trial design, conduct and analysis [5].

Table 1
American Society of Clinical Oncology clinical practice guidelines for use of tumour markers in breast cancer *(tissue factors only)* [a]

Factor	Use	Guideline
Oestrogen and progesterone receptors	Predictive factors for endocrine therapy	Measure on every primary breast cancer and on metastatic lesions if results influence treatment planning
DNA flow cytometrically derived parameters	Prognosis or prediction	Data are insufficient to recommend obtaining results
erbB-2 (HER-2/neu)	Prognosis	Data are insufficient to recommend obtaining results for this use
	Prediction for: trastuzumab CMF-like regimens doxorubicin taxanes endocrine Rx	*erbB-2* should be evaluated on every primary breast cancer at time of diagnosis or at time of recurrence for use as predictive factor for trastuzumab; Committee could not make definitive recommendations regarding CMF-like regimens. *erbB-2* may identify patients who particularly benefit from anthracycline-based therapy but should not be used to exclude anthracycline treatment. *erbB-2* should not be used to prescribe taxane-based therapy or endocrine therapy.
p53	Prognosis or prediction	Data are insufficient to recommend use of p53
Cathepsin-D	Prognosis	Data are insufficient to recommend use of cathepsin-D

[a] Modified from Bast et al. (2000) [3].

Table 2
Levels of evidence for grading clinical utility of tumour markers[a]

Level	Type of evidence
I	Evidence from a single high-powered prospective study that is specifically designed to test marker or evidence from meta-analysis and/or overview of Level II or III studies. In the former case, the study must be designed so that therapy and follow-up are dictated by protocol. Ideally, the study is a prospective randomised trial in which diagnostic and/or therapeutic clinical decisions in one arm are determined based at least in part on marker results, and diagnostic and/or therapeutic clinical decisions in control arm are made independently of marker results. However, may also include prospective but not randomised trials with marker data and clinical outcome as primary objective.
II	Evidence from study in which marker data are determined in relationship to prospective therapeutic trial that is performed to test therapeutic hypothesis but not specifically designed to test marker utility (i.e. marker study is secondary objective of protocol). However, specimen collection for marker study and statistical analysis are prospectively determined in protocol as secondary objectives.
III	Evidence from large but retrospective studies from which variable numbers of samples are available or selected. Therapeutic aspects and follow-up of patient population may or may not have been prospectively dictated. Statistical analysis for tumor marker was not dictated prospectively at time of therapeutic trial design.
IV	Evidence from small retrospective studies which do not have prospectively dictated therapy, follow-up, specimen selection, or statistical analysis. May be matched case controls, etc.
V	Evidence from small pilot studies designed to determine or estimate distribution of marker levels in sample population. May include 'correlation' with other known or investigational markers of outcome, but not designed to determine clinical utility.

[a] From Hayes et al. (1996) [6].

Several years ago, members of the ASCO Tumour Markers Guidelines Panel developed a proposal for a framework in which previously published tumour marker studies might be critically evaluated [6]. This framework can also be used by investigators to plan future studies in a fashion that leads to more rapid acceptance, or refutation, of a given marker in the clinical arena. A key component of TMUGS is a scale of Levels of Evidence, designed to help place tumour marker studies into a context of validity (Table 2). Overall, this framework and suggestions by others have led to an increasing understanding that tumour marker studies should be designed, conducted, and analysed and reported with the same rigor as classic laboratory or clinical therapeutic studies [7,8]. In concert with this effort, an ad hoc multi-disciplinary committee convened by the National Cancer Institute-European and the Organisation for Research and Treatment of Cancer (NCI-EORTC) has suggested criteria for reporting recommendations for tumour marker prognostic studies (REMARK) (Box 1). [9–11].

In summary, tumour marker studies should be designed, conducted, and analysed with the same rigor as laboratory research or therapeutic clinical trials.

What is the question?

The first critical component to determine the clinical utility of a tumour marker is to ask how it might be used. Tumour markers can be helpful in one of many situations, including assessment of risk of developing a cancer, screening for a new cancer, differential diagnosis, determination of prognosis, or monitoring clinical course, either to detect recurrence or progression. Each of these uses is different and requires a carefully conducted study to address it. Most tumour markers are used to determine prognosis. In this regard, it is critical to understand the difference between prognosis and prediction [5]. The value of a prognostic factor is to determine if, after some preceding therapeutic action, the patient has a sufficiently favourable prognosis that he/she would not need subsequent therapy. A predictive factor provides an estimate of whether a given therapy is likely to work, or more precisely, whether the therapy is so unlikely to work that the patient would forego therapy to avoid the risks. Many markers can be both prognostic and predictive. Moreover, a marker can be favourably prognostic yet unfavourably predictive for a specific therapy, or vice verse. Even more confusing, a marker can even predict sensitivity to one type of therapy and for resistance to another. Nonetheless, careful consideration of prognosis and prediction can be used to calculate the absolute odds of a patient's benefit, and then he/she can decide if that outweighs the absolute toxicity and costs of the therapy of interest. Computerised on-line programs are available to make this calculation for breast and colon cancer [12].

It is critical to understand that a tumour marker study is highly confounded by the type of therapy

Reporting recommendations for tumour marker prognostic studies (REMARK)

Introduction: State the marker examined, the study objectives, and any pre-specified hypotheses.

Materials and Methods: *Patients*: Describe the characteristics (e.g. disease stage or co-morbidities) of the study patients, including their source and inclusion and exclusion criteria. Describe treatments received and how chosen (e.g. randomised or rule-based).

Specimen characteristics: Describe type of biological material used (including control samples) and methods of preservation and storage.

Assay methods: Specify the assay method used and provide (or reference) a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. Specify whether and how assays were performed blinded to the study endpoint.

Study design:

- State the method of case selection, including whether prospective or retrospective and whether stratification or matching (e.g. by stage of disease or age) was used. Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.
- Precisely define all clinical endpoints examined.
- List all candidate variables initially examined or considered for inclusion in models.
- Give rationale for sample size; if the study was designed to detect a specified effect size, give the target power and effect size.

Statistical analysis methods:

- Specify all statistical methods, including details of any variable selection procedures and other model-building issues, how model assumptions were verified, and how missing data were handled.
- Clarify how marker values were handled in the analyses; if relevant, describe methods used for cutpoint determination.

Results:

Data:

- Describe the flow of patients through the study, including the number of patients included in each stage of the analysis (a diagram may be helpful) and reasons for dropout. Specifically, both overall and for each subgroup extensively examined report the numbers of patients and the number of events.
- Report distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumour marker, including numbers of missing values.

Analysis and presentation:

- Show the relation of the marker to standard prognostic variables.
- Present univariate analyses showing the relation between the marker and outcome, with the estimated effect (e.g. hazard ratio and survival probability). Preferably provide similar analyses for all other variables being analysed. For the effect of a tumour marker on a time-to-event outcome, a Kaplan–Meier plot is recommended.
- For key multivariable analyses, report estimated effects (e.g. hazard ratio) with confidence intervals for the marker and, at least for the final model, all other variables in the model.
- Among reported results, provide estimated effects with confidence intervals from an analysis in which the marker and standard prognostic variables are included, regardless of their statistical significance.

If done, report results of further investigations, such as checking assumptions, sensitivity analyses, and internal validation.

Discussion:

- Interpret the results in the context of the pre-specified hypotheses and other relevant studies; include a discussion of limitations of the study.
- Discuss implications for future research and clinical value.

After McShane et al. (2005) [9–11].

that the patients in the study received. Depending on the relative mix of patients and how they were treated, results from a study of convenience might be interpreted to indicate that a marker is a positive factor, a negative factor, or has no prognostic significance at all. Too often, interesting new technology is applied to available archived samples without thought to how the assay might be used in real clinical practice. In the Tumour Marker Utility Grading System, such studies are considered Level of Evidence III at best. These studies are extraordinarily vulnerable to a variety of obvious or hidden biases that make their results nearly incomprehensible [13,14]. Over the last several years, it has become increasingly obvious that adoption of a new marker into routine clinical practice requires that the studies be designed prospectively with a specific use in mind. In that regard, it is important to understand what that use might be, and how results of a given study may be adversely affected to produce false positive and negative findings. A major reason that so many promising tumour markers never achieve widespread clinical utility is that the investigators fail to understand these concepts and, therefore, fail to design their studies appropriately.

How strong is the marker?

The more a given therapy is effective and/or has little toxicity or monetary costs, the more likely that the patient will accept its use and society will agree to pay for it [15–17]. To be useful, a marker has to be exquisitely accurate to dictate that a patient should not receive the therapy. With lower efficacy and/or higher toxicities and costs, the marker can be less accurate since more patients are willing to forego the therapy anyway. Although difficult to perform, several studies have addressed patient willingness to accept systemic therapies in different settings, and as expected have demonstrated substantial variation [15–17]. Thus, simply observing that outcomes, such as response, recurrence, progression, or survival rates, differ statistically is not an indication that the marker is useful. One has to consider what the 'next' treatment would be, and whether a patient would forego it to avoid toxicity. Such a decision varies according to the treatment, the endpoint, and the toxicities and costs.

In summary, for a tumour marker to be accepted for routine clinical application, the intended use must be considered in the initial study design, with an understanding of the three components that make the results of the study worthwhile. The study should then provide an estimate of the magnitude of differences between marker 'positive' and 'negative' patients. Once these two critical issues are considered, then the validity of this magnitude must be ensured in regards to technical aspects of the assay, statistical design and analysis of the study, and an understanding of bias and over-fitting of the results that lead to the investigators' conclusions.

Conflict of interest statement

The author had the following financial affiliations during the preceding twelve months:
Consultant/Advisory Panel Participant or Lecture/Honorarium During Last 12 months: Abraxis, American Biosciences, AviaraDx, Cytogen Corp., Monogram Bioscience, Pfizer, Precision Therapeutics, Inc., Siemens Medical Solutions Diagnostics, StemCapture.
Sponsored Clinical Research – Principal or co-Investigator During Last 12 months: Glaxo Smith Kline, Wyeth Ayerst-Genetics Institute, Pfizer, Novartis.

Acknowledgements

Supported in part by NIH grant CA64057 and by the Fashion Footwear Association of New York (FFANY)/QVC Presents/Shoes on Sale®.

References

1 ASCO Expert Panel. Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer: Report of the American Society of Clinical Oncology Expert Panel. *J Clin Oncol* 1996, **14**, 2843–2877.
2 ASCO Expert Panel. 1997 update of recommendations for the use of tumor markers in breast and colorectal cancer. *J Clin Oncol* 1998, **16**, 793–795.
3 Bast RC, Jr., Ravdin P, Hayes DF, *et al.* 2000 Update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol* 2001, **19**, 1865–1878.
4 Bast RC, Jr, Ravdin P, Hayes DF, *et al.* Errata: 2000 Update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol* 2001, **19**, 4185–4188.
5 Henry NL, Hayes DF. Uses and abuses of tumor markers in the diagnosis, monitoring, and treatment of primary and metastatic breast cancer. *Oncologist* 2006, **11**, 541–552.
6 Hayes DF, Bast R, Desch CE, *et al.* A tumor marker utility grading system (TMUGS): a framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst* 1996, **88**, 1456–1466.
7 McGuire WL. Breast cancer prognostic factors: Evaluation guidelines. *J Natl Cancer Inst* 1991, **83**, 154–155.

8 Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994, **69**, 979–985.
9 McShane LM, Altman DG, Sauerbrei W, *et al.* Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 2005, **97**, 1180–118.
10 McShane LM, Altman DG, Sauerbrei W, *et al*. Reporting recommendations for tumor marker prognostic studies. *J Clin Oncol* 2005, **23**, 9067–9072.
11 McShane LM, Altman DG, Sauerbrei W, *et al*. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005, **93**, 387–391.
12 Ravdin PM, Siminoff LA, Davis GJ, *et al.* Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol* 2001, **19**, 980–991.
13 Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer* 2004, **4**, 309–314.
14 Ransohoff DF. Cancer. Developing molecular biomarkers for cancer. *Science* 2003, **299**, 1679–1680.
15 Coates AS, Simes RJ. *Patient Assessment of Adjuvant Treatment in Operable Breast Cancer*. New York, NY, John Wiley & Sons Ltd, 1992.
16 Ravdin P, Siminoff I, Harvey J. Survey of breast cancer patients concerning their knowledge and expectations of adjuvant therapy. *J Clin Oncol* 1998, **16**, 515–521.
17 Lindley C, Vasa S, Sawyer T, *et al.* Quality of life and preferences for treatment following systemic adjuvant therapy for early stage breast cancer. *J Clin Oncol* 1998, **16**, 1380–1387.