Prognostic and predictive markers in cancer

Barbara A. Conley[a,*] and Sheila E. Taube[b]
[a]*Medical Oncology Clinical Research Unit, Center for Cancer Research, NCI, NIH, DHHS, Bethesda, MD, USA*
[b]*Cancer Diagnosis Program, Division of Cancer Treatment and Diagnosis, NCI, NIH, DHHS, Bethesda, MD, USA*

Abstract. The elucidation of the human genome and advances in knowledge about molecular abnormalities, signaling pathways, influence of the local tissue milieu and the relevance of genetic polymorphisms offer hope of designing more effective, individualized cancer treatment plans. Although the scientific and medical literature is replete with reports of putative prognostic or predictive markers for cancer, few new diagnostics have been incorporated into routine clinical practice. Criteria are needed to a) identify markers that have the promise to be clinically useful; b) assess the best methodology for clinical evaluation of the markers in question and c) confirm or validate that using the marker adds useful information compared to using standard prognostic factors alone. This review presents a methodology for the clinical evaluation of putative prognostic and predictive markers in cancer, with considerations of pitfalls in the early evaluation, rationale for development and optimization of assay methodology, and examples of possible clinical trials for assessing the clinical utility of putative markers.

1. Introduction

In 2003, over 1,300,000 cases of cancer will be diagnosed and over 550,000 people will die of cancer [18]. The elucidation of the human genome and fifty years of biological studies have laid the groundwork for a more informed method for treating cancer with the prospect of realizing improved survival. Advances in knowledge about the molecular abnormalities, signaling pathways, influence of the local tissue milieu and the relevance of genetic polymorphisms offer hope of designing effective therapies tailored for a given cancer in a particular individual, as well as the possibility of avoiding unnecessary toxicity.

The challenge is to develop methodology that will allow us to use the available knowledge and the available tools to evaluate the usefulness of molecular abnormalities for improving individual and population outcomes for cancer. The scientific and medical literature is replete with reports of putative prognostic or predictive

markers for cancer. However, few new diagnostics have been added to the oncologist's armamentarium, particularly for solid tumors. The translation of exploratory reports to clinically useful diagnostic tools will require a systematic process for clinical development of prognostic and predictive markers.

2. Definitions

"Clinical development" of markers refers to the series of steps from the identification of a potentially useful (or promising) marker, marker panel, or signature, through the systematic evaluation of its clinical utility. This process involves discrete decision points at which a choice is made to either pursue development further or not. These decisions are based on assessment of the likelihood that knowledge of the marker will add value to current diagnostic and treatment methods.

"Prognostic" and "predictive" are terms that are generally used to define markers with different functions. The term "prognostic" usually refers to a marker that provides information about the natural history of the cancer in an individual [36]. For example, a prognostic marker measured at diagnosis may predict a shorter survival compared to that observed in patients without the marker, both in the absence of as well as in the presence

---
*Corresponding author: Barbara A. Conley, M.D., Head, Aerodigestive Cancer Clinical Research Section, Medical Oncology Clinical Research Unit, National Cancer Institute, Center for Cancer Research, 10 Center Drive, Room 12C103A, Bethesda, MD 20892, USA. Tel.: +1 301 435 4004; Fax: +1 301 402 0172; E-mail: conleyb@mail.nih.gov.

of systemic treatment. Retention of heterozygosity of chromosome 18q in microsatellite stable cancers and mutation of the gene for TGF beta 1 receptor in cancers with high microsatellite instability among patients with resectable colorectal cancer have been reported as examples of favorable prognostic markers [39].

A "predictive" marker, on the other hand, identifies how likely an individual is to respond to a particular therapy. If such a marker is purely predictive, then the outcome will be associated only with a particular therapy, and not with all therapies. One example of this type of marker is the prediction of response to tamoxifen therapy by estrogen receptor (ER) expression in breast cancer. If estrogen receptor expression is present, then the likelihood of benefit from tamoxifen for patients with metastatic breast cancer is about 50%, while if it is absent, the likelihood of benefit is 5% [30]. However, estrogen receptor expression is not purely a predictive marker; studies have shown that ER positive patients may have a more indolent disease compared to ER negative patients regardless of what therapy is used, even chemotherapy [27].

This discussion suggests that the distinction between prognostic and predictive is somewhat artificial. Markers may be, and often are, both prognostic and predictive, to different degrees and in different contexts. This issue has implications that will be discussed for the design of studies evaluating the utility of a given marker.

3. Prognostic markers

There are at least two reasons to develop a purely prognostic marker. The first reason is to separate good and poor prognosis patients at the time of diagnosis. If expression of the marker clearly separates patients with an excellent prognosis from those with a poor prognosis, then the marker can be used to aid the decision about how aggressive the therapy needs to be. The poor prognosis patients might be considered for clinical trials of novel therapies that will, hopefully, be more effective. For example, a prognostic marker for early stage breast cancer that identified the 95% of patients treated with surgery alone whose survival will be the same as women who have not had cancer would help patients and their physicians decide whether it is necessary to expose them to the toxicity, inconvenience and expense of adjuvant chemotherapy. If studies in patients with early stage disease who did not have the marker status evaluated suggest that adjuvant therapy is

effective, it will be necessary to evaluate the usefulness of the therapy in the good prognosis patients identified by the marker to assess whether there is benefit even in this good prognostic group. The decision of whether or not to prescribe or to accept adjuvant therapy could then be made on a firm basis. If patients whose tumor lacked the marker did poorly with surgery alone, these patients could be given standard adjuvant chemotherapy or be offered a trial that randomizes patients to standard therapy or novel therapy with the potential of greater efficacy. Such markers could also provide information about the likelihood for success of localized treatments (e.g. surgery, radiation), but would likely not be able to determine which patients would receive more benefit from surgery versus radiation, in cases where both are used as primary curative therapy (e.g. early prostate adenocarcinomas or head/neck squamous cancers).

The second reason to develop prognostic markers is that they may provide clues to the possible mechanism(s) responsible for the poor prognosis. This can lead to the identification of new targets for treatment as well as new effective therapeutics. For example, several independent studies suggested that measures of angiogenesis such as vascular endothelial growth factor (VEGF) in plasma and/or tumor portended a worse prognosis in colorectal cancer [6,21,22]. These studies, although exploratory, led to a randomized phase II trial comparing 5-fluorouracil and leucovorin with and without bevacizumab [20], an antibody to VEGF, in patients with metastatic disease that hinted at a better response rate, longer median time to disease progression, and longer median survival, as well as defined better the expected toxicities of combining the bevacizumab with the chemotherapy. A recent randomized phase III clinical trial in patients with metastatic colorectal cancer who had not received any prior therapy reported a statistically significant survival benefit for the use of bevacizumab along with systemic chemotherapy (Irinotecan, 5-fluorouracil and leucovorin) [29], compared to systemic chemotherapy alone. This example also illustrates the iterative nature of prognostic and predictive marker development. A targeted agent has been shown to have a clinical benefit, and now development of a predictive marker or markers that would identify patients likely to benefit from the targeted therapy, as well as a reliable analytic technique with which to detect expression of the predictive marker(s) is warranted.

4. Predictive markers

It is helpful to know before initiation of treatment whether or not a patient is likely to respond to a par-

ticular therapy. If there were a marker that could accurately predict the success of a given therapy, then that therapy could be recommended with more confidence. Again, estrogen-receptor positivity in breast cancer patients provides a good example. A patient with an estrogen receptor positive tumor has a much better chance of response to tamoxifen and aromatase inhibitors than patients without ER expression in the tumor, while a patient with a tumor negative for the estrogen receptor should be offered chemotherapy, not hormonal therapy.

Identifying markers to predict response to systemic treatments can be quite complicated. The ideal marker, or possibly panel of markers, would predict that a drug would interact with its target in such a way as to benefit the patient. This has been demonstrated for some "targeted" drugs such as receptor antibodies (e.g., trastuzumab, rituximab) [1,3,4,26,32], and receptor tyrosine kinase inhibitors (e.g., imatinib mesylate) [5, 19]. Traditional cytotoxic drugs, such as those targeting tubulin, thymidylate synthase, topoisomerase I or other specific targets, may interact with their targets effectively, but we do not have many reliable markers to predict which tumors will respond. This lack of predictive markers affects the assessment of the intervention since the effect of the agent will not be recognized if only a small percent of the patients randomized have the necessary alteration of the target that will result in response to the agent.

The existence of the putative target may not be sufficient to predict response. Increasing numbers of studies are being published on mechanisms of resistance that may need to be accounted for, as well as mechanisms of relative resistance related to genetic polymorphisms of the target, for example, polymorphisms of the thymidylate synthase promoter [23,24,38]. In addition, polymorphisms of enzymes involved in antineoplastic drug metabolism that alter drug exposure (or area under the concentration x time curve), and thus drug toxicity, such as polymorphisms of UGT1A1 for irinotecan [2] are being increasingly defined. Designing appropriate studies to evaluate markers to predict both response and possible resistance is a growing challenge.

5. Considerations for clinical development of markers

Regardless of whether a putative prognostic or predictive indicator is a single marker, a panel of a few defined markers, or a complex gene expression or proteomic signature, it is desirable to have criteria to a) identify markers that have the promise to be clinically useful; b) assess the best methodology for clinical evaluation of the markers in question and c) confirm or validate that additional clinical utility is gained by using the marker compared to using standard prognostic factors. While no uniform methodology appears to exist as yet, investigators have begun to address these issues in the clinical development of prognostic and predictive markers for cancer.

6. Identification of promising markers

The underlying assumption in the development of new markers is that there is a clinical decision that will be facilitated by knowledge of the marker status. Markers are potentially useful if they have a biologic rationale or if they demonstrate a correlation with an outcome of interest such as survival or response to therapy, even in the absence of a defined biologic rationale [10]. Studies must have sufficient statistical power to ensure that the observed correlation could not occur by chance alone. A relatively large sample will also give a better estimate of the prevalence of the putative marker in the population of interest. Confidence in a marker can be increased by the accumulation of reports of the same correlation by different laboratories using the same or different methods. The promise of or interest in a new marker also depends on whether the new marker adds to information obtainable by using standard prognostic characteristics, such as stage or grade.

The criteria for whether a marker is promising enough to warrant further clinical evaluation also include the expected clinical impact of the marker. This assessment will depend on the intended use(s) of the marker, its prevalence in the target population, the feasibility of measuring the marker, and the hazard ratio attributable to the marker. If the marker is to be used to predict response to therapy, the availability of a specific therapy relevant to the marker will be critical. A prognostic marker with 50% prevalence in the target population, as well as associated with a 50% improved outcome would be very promising if this marker added to the ability of the standard prognostic variables to predict the behavior of the cancer. A marker that predicted an 80% response rate compared to a 30% response rate would also generate interest. The specificity of the marker for the organ and disease under study may be more important for circulating markers than for tissue based markers. However, the tissue distribution of the marker may be crucial for a predictive marker, espe-

cially if expression is related to the likelihood of severe toxicity in normal as well as tumor tissue.

If exploratory evaluation reveals that a marker either has a biologic rationale or a robust association with an outcome of interest, is fairly specific for a particular cancer or group of cancers, has a favorable tissue distribution for detection, has a prevalence among the population of interest that is conducive to clinical evaluation (ideally between 30–70%), has an independent and strong relationship to the outcome of interest, then the marker may be promising for development into a clinical tool.

7. Assay methodology

Once it is decided that a marker is promising, and a decision is made to evaluate clinical utility, a reliable assay needs to be developed that can be used in the matrix or tissue that will be available from patients. The assay's range of useful values, scoring system if any, performance, techniques, accuracy and precision should be defined prior to any clinical trial designed to assess the utility of the marker. If several assays for the marker are available, then comparison of these assays should be considered in order to choose the most appropriate assay methodology to complete the clinical evaluation of the marker. The chosen assay should be standardized; if possible an internal or quality standard should be employed. If multiple laboratories will be performing the assay, then comparability of the assay results between laboratory sites needs to be evaluated and shown to be satisfactory; the assay may need to be further optimized to obtain acceptable agreement level between laboratories. A quality assurance plan should be in place to ensure the assay's fidelity throughout the clinical trial.

Methods for obtaining, storing and handling, as well as preserving tissue should be evaluated for the effect on the assay. If needed, these methods should be optimized as well.

8. Evaluation of clinical utility

Assessment of the clinical utility is the most important aspect of marker development. It can proceed only after successfully completing the previous steps, i.e., determining whether the marker can be used as an independent predictor of outcome, deliberating about which patient population is likely to benefit most from use of the marker, ascertaining the prevalence of the marker and the postulated magnitude of effect on outcome, and developing an assay that can be used in a large trial in an appropriate surrogate tissue or actual biopsy. In the recent past, many phase III clinical trials of therapeutic regimens have also assessed several markers to evaluate correlations between marker expression and outcome. These types of trials are very useful in generating hypotheses, but may generate spurious results since they do not have sufficient power to test the relevant marker hypotheses. These trials also usually test multiple comparisons, with the increased risk of finding a "significant" but spurious correlation by chance alone. These trials generally cannot be relied on for definitive validation of useful clinical tools. They can and do, however, estimate whether the marker is an independent predictor of outcome, assess prevalence and magnitude of effect, and serve as excellent preliminary work to a more definitive clinical assessment of marker utility.

In order to justify using a marker to choose or to adjust therapy, it is important to demonstrate that such use improves upon the outcome obtained without use of the marker. Very few clinical studies have been designed to answer this question. It may be possible to perform such a trial retrospectively, using archived samples and data collected previously, if appropriate samples can be found from a uniformly treated population. However, it is often difficult to obtain retrospectively a population in which most of the likely confounding factors, such as treatment regimen, performance status, uniform staging and uniform follow-up are available. If the marker is tested retrospectively in a population of patients treated on a uniform clinical treatment trial, then factors such as uniform staging, treatment or follow-up are less of a problem. However, it is important that samples have been obtained from all the clinical trial participants. There are often biases in retrospective collections, particularly toward larger tumors [25]. Other biases may also be present. For example, Pajak and colleagues showed a significant survival difference for prostate cancer patients expressing p53, based on assessing p53 in samples from 129 of the 456 assessable patients in a randomized trial. Though a statistical comparison of clinical characteristics of those patients for whom tissue was available and those for whom tissue was not available for assessment of p53 expression showed no significant differences, there was a statistically significant difference in the survival of those patients who did have p53 evaluated in their tumor (whether or not expression was positive) com-

pared to those who did not, with the patients in whom p53 expression was evaluated faring worse [8,31]. This suggests that in most cases, the evaluation of clinical utility for a marker or marker panel will have to involve a prospective clinical trial.

The clinical evaluation of marker usefulness must be based on a testable hypothesis. The study design to test the hypothesis should specify endpoints, cut-off values for the marker if dichotomized values will be used, and the subset of patients, types of treatment and standard prognostic variables relevant to the hypothesis. It has been suggested that clinical trials to evaluate prognostic or predictive markers should have at least 90% power to detect a clinically relevant difference between marker groups [31]. The power calculation is based on the number of events needed to observe the defined effect, not the number of patients. An "event" is the outcome of interest: for evaluation of prognostic and predictive markers, "event" is likely to be survival (death), disease-recurrence, or response to therapy. If a multivariate analysis is to be part of the study, one "rule of thumb" is to have at least 5–10 outcome events for each potential predictor. The initial model, particularly if based on retrospective sets of specimens, often does not perform as well when tested on a separate (independent) population of patients [7,14,37]. If tests of interaction of the marker value with treatment are to be done, the number of events needed for an adequately powered study may be greater. The sample size needed for adequate statistical power of a clinical marker study should be calculated based on the hypothesized marker effect, the prevalence of the marker in the specified population, and the expected number of events in the population under study.

Two types of trial design have been proposed to evaluate the utility of a marker in the context of a phase III therapeutic clinical trial, that is, a trial testing a new treatment against a standard treatment. We shall call these the Marker x Treatment Interaction Design and the Marker Strategy Design [36].

In the first type of trial, (Marker x Treatment Interaction Design) the marker is measured in all patients, and patients are stratified into well separated marker status groups. For a predictive marker question, the design assesses if the outcome of one treatment depends on the marker status. Two rather than more groups are desirable for simplicity, but the trial has to reflect perceived clinical reality. Patients in both groups are randomized to receive one or the other treatment regimen (Fig. 1). This is the equivalent of doing two randomized trials, one in each marker status group. In statis-
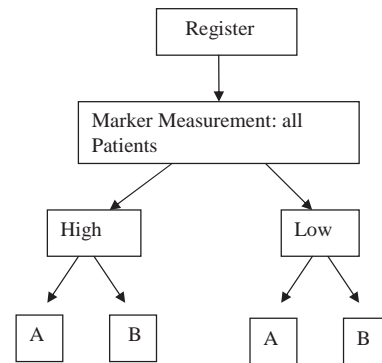

Fig. 1. Marker x Treatment Interaction Design for assessment of the Clinical Utility of Predictive Marker. The trial is the equivalent of 2 randomized treatment trials done in each marker status group. The design can assess whether the treatment effect depends on the marker status.

tical terms, this is called a marker x treatment interaction. For a marker with prevalence close to 50%, and at least a 30–40% expected effect on outcome, such a trial should not require a sample size much larger than that used for a typical therapeutic randomized clinical trial. The more distant the prevalence from 50% and the less the expected marker effect, the higher the accrual goal will have to be. This design could be used to evaluate whether improved outcome can be expected in one vs. another marker value group by using a particular treatment. For example, suppose we wished to evaluate whether thymidylate synthase (TS) expression could be used to choose a successful treatment or to avoid an unsuccessful one. Low TS expression correlates with better response to 5-fluorouracil (5-FU) in exploratory clinical trials. Using the marker x treatment interaction trial design, the TS level in all tumors would be measured, a cut-point would be determined, and patients with low TS values would be randomized to a treatment containing 5FU or not; likewise, patients with high TS values would be randomized to the same 2 treatments. Outcomes would be observed in the low and high TS expression groups. If the outcome for the low expression group were better for the 5FU containing treatment than for the treatment not containing 5FU, then we could say that low TS expression was predictive of a better outcome for patients receiving 5FU containing therapy. It would be necessary to know that the non-5FU containing treatment was not an inferior treatment in general, in order to make this assumption. In the high TS group, the outcome of the patients who received the 5FU-containing regimen should be the same or inferior to the outcome of the high TS patients who received the non-5FU containing treatment. If both treatments
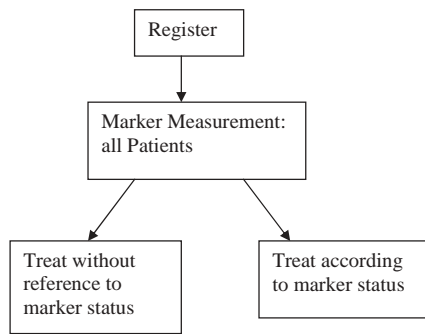
Fig. 2. Marker Strategy Design: This design assesses whether there is benefit to using a marker when choosing a treatment for a patient, compared to not using the marker (treating as is commonly done now, without considering predictive marker).

were inferior in the high TS marker group compared to the low TS marker group, this effect could be due to a prognostic effect of high TS expression, or to other factors not evaluated.

To test the advantage associated with using a marker rather than not using it, the most direct trial design would be to measure the marker, then randomize patients to either have their treatment determined by the marker, or not (referred to as the Marker Strategy Design) [13,36] (Fig. 2). Ideally, outcomes of the patients treated by marker status will be at least 30% better than those obtained by treatment that does not take into account marker status (standard therapy or physician choice).

For a marker to have value, the likelihood that one treatment (based on marker status) will work better than another should be relatively high. The marker should be able to identify the group in which the treatment works and clearly separate this group from the group in which the treatment will not work. The ideal situation would have confidence intervals that do not overlap.

One example of a clinical need that could be addressed by identification of appropriate prognostic markers is the adjuvant treatment of localized resected colorectal cancer. Current recommendations in the United States call for adjuvant 5FU based chemotherapy in all stage III patients. There is no consensus on whether or which stage II patients, who have a survival rate of 75–80% after surgery alone, should receive adjuvant therapy. When colorectal cancer metastasizes, even good performance status patients with the best available chemotherapy have a median survival of less than 2 years. Therefore, adjuvant treatment is very desirable if it is effective in preventing the development of metastatic colon cancer, without inflicting undesirable harms or death on the patient who may well be cured

even without the adjuvant therapy. The initial adjuvant regimens, based on 5-fluorouracil, increased the survival of the study population by about 30% in stage III disease [28]. This information is now being extrapolated by some oncologists and patients to stage II patients, who have a 70–80% chance of normal lifespan after surgery alone. Twenty to thirty percent of these patients will develop recurrent or metastatic disease (which is usually fatal). It may seem reasonable to take or give the same adjuvant therapy recommended for stage III patients to maximize the chance of cure. However, the probability of recurrence for a given patient is not 20–30%; it is either 100% or 0. The incidence of severe (grade 3 or higher) toxicity of 5-fluorouracil and leucovorin given for adjuvant treatment can be about 25% in the population of treated patients, depending on regimen, but it is 100% in an individual who develops it [24]. The chance of improved survival with adjuvant chemotherapy of any type in stage II has not been defined. Even if it is the same as in stage III (30%), then only a small number of stage II patients have the possibility of benefit from the drugs, while all are at risk for at least some toxicity, cost, and inconvenience. In such a situation, it seems rational to try to develop some way of discerning which patient will relapse and which will not. In addition, now that there is a choice of therapy for colorectal cancer (5-fluorouracil with or without leucovorin, given by one of several schedules, with or without irinotecan, as well as consideration of other agents active in metastatic disease, such as oxaliplatin, capecitabine, bevacizumab . . .) predictive markers should be developed to help choose the best regimen for that patient.

An organized approach to this problem can be taken. A review of the literature initially can determine whether markers exist that have been evaluated for their ability to identify which patients with stage II colon or colorectal cancer are likely to recur. Several studies indicate that a high level of microsatellite instability is associated with a better outcome than microsatellite low or stable tumors [9,11,39]. Such studies include mostly stage III cancers, but also some stage II cancers. Current evidence is not conclusive, but suggests that patients with tumors having high microsatellite instability may not benefit from 5-fluorouracil-based adjuvant therapy [35]. Also, loss of heterozygosity LOH on chromosome 18q may be an indicator of a poorer prognosis. It has been proposed that a combination of factors, including MSI "high" status and lack of LOH 18q may define a group of stage II (and perhaps stage III) colorectal cancers that will have a better prog-

nosis [39]. Currently there is no estimate of how much better the prognosis will be in such patients. In such a case, it may be reasonable to collect a large number of archived stage II colorectal cancers with adequate surgery and good follow-up (and hopefully no adjuvant chemotherapy) to test retrospectively whether these two or a group of markers define a population with survival so good that adjuvant chemotherapy should not be recommended. The definition of how good that prognosis has to be depends on many things, including how much risk (recurrence and toxicity) patients are willing to tolerate. Deliberation may be well informed by patient advocates who can actually query the patients throughout the country and who are generally well informed as to the current state of the art and science.

It is not likely that all parameters can be explored in a single trial. It may be more reasonable first to define retrospectively a group of patients that is likely not to need further treatment and one that is likely to recur. The benefit of adjuvant therapy can then be evaluated in one or both of these populations. One could measure the prognostic factors up front, then assign patients to either a low risk or high risk group, and randomize patients in the low risk group to adjuvant therapy or none and treat all the high risk patients with adjuvant therapy. During the study, putative predictive factors for toxicity and/or response could also be measured. After these steps, the benefit of adjuvant therapy for low risk patients, and the survival and recurrence rate of high risk patients treated with adjuvant therapy could be assessed. In addition, exploratory data on other possibly relevant factors concerning drug sensitivity and/or metabolism could be obtained. In order to engage the support of patients for such a trial, it would be very helpful to have done a very careful, statistically designed "retrospective" trial in archived samples from patients treated with surgery alone to assess the prognostic strength of the markers to be used. Reliable assays will also need to be available. The accrual goal of such a trial will depend on the prevalence of the marker or panel of markers in the population, the hazard ratio associated with the marker or panel of markers, the event rate, and the size of the difference in outcome that the investigators wish to detect in the randomized group. Such a trial is a serious undertaking, and careful deliberation on the details is warranted. Other trial designs could be considered: Good prognosis patients could be assigned to no further treatment beyond surgery, and poor prognosis patients could be randomized to no further treatment or to adjuvant therapy, or to standard adjuvant therapy vs. a new therapy. Each type of design will have its advantages and drawbacks.

Such considerations for adjuvant therapy could be applied to a variety of neoplasms, each of which will have its own ethical and practical dilemmas: prostate cancer, early stage lung cancer, bladder cancers of various stages, early stage breast cancer, etc.

One carefully considered example of the clinical development of a prognostic marker can be found in the studies of urokinase plasminogen activator and its inhibitor (uPA/PAI-1) as prognostic markers for node negative breast cancer. The initial clinical problem was identified as a need to distinguish which node-negative breast cancer patients do not benefit from (or need) adjuvant treatment. Traditional prognostic indicators, such as tumor size, grade, age, steroid hormone receptor status or menopausal status were not able to identify the 25–30% of patients who would relapse from those who would be cured by surgery alone. Early stage breast cancer patients represent over 50% of all breast cancer patients and this percentage is increasing. Current methods are inadequate to select the patients who need adjuvant therapy, and this situation results in up to 90% of lymph node negative patients being treated.

Investigators in the European Organization for Research and Treatment of Cancer (EORTC) had been studying the plasminogen activator system, which plays a role in metastasis and invasion. With this biologic rationale, they studied retrospective specimens from populations of breast cancer patients and noted that tumor levels of uPA and its inhibitor seemed to be strong, independent, prognostic factors in both node positive and node negative breast cancer. Patients with high levels of these markers in their tumors had poorer survival [15, 16]. A model of risk strata was developed [12,33]. The outcome of node negative patients after surgery alone was determined for both risk strata. The assay was developed and validated, as were methods of tumor procurement, handling and preservation. Finally, a multinational trial was designed with the hypotheses that low tumor levels of uPA and PAI-1 identify patients with low risk of relapse and that patients with high levels of uPA and PAI-1 would benefit from adjuvant chemotherapy. For this trial, patients with node negative breast cancer of 1–5 cm in size were stratified by marker level. High risk patients were randomized to either adjuvant chemotherapy or to observation (the European standard) and low risk patients were observed only. Those patients who refused randomization but were identified by the markers as high risk were allowed to choose whether they wanted adjuvant therapy or not and were included in the final assessment, though analyzed separately. The trial was able to confirm the

low risk of relapse for low tumor levels of the marker, as well as a benefit for high risk patients from adjuvant chemotherapy [17]. The marker must be assessed in fresh tumor tissue, and this requirement has hindered its acceptance, particularly in the US. Other assay methods are under development. However, this 10 year sequence is an excellent example of the rational development of a marker from determination of whether or not it is promising, to evaluation of the strength of the association with outcome, development of a reliable assay, the determination of cut-points, the prevalence in high and low risk groups, and evaluation in a separate independent group of patients in a prospective clinical trial.

For metastatic cancers, the goal is to find predictive markers for effective therapy, defined as therapy that results in a clinical benefit of some kind. For this purpose, response rate and survival are key outcomes of interest. Complications for systematic clinical development of markers arise in this setting except when a patient presents de novo with metastatic disease since an additional biopsy will not be necessary to diagnose the disease. On the other hand, if metastases present after an initial treatment, there is no assurance and sometimes evidence to the contrary, that the metastatic disease will be similar to the initial (and presumably archived) tissue sample with respect to expression of the markers of interest. In this case, if the marker seems to be promising in other ways, such as biologically rational, correlated with survival in archived samples with known clinical outcome, etc., then clinical development may require a search for a feasible and ethical way to obtain a measurement of the marker from patients with metastatic disease. Development of a serum test, or proteomic profile, and correlation of that determination with the outcome of interest, or the use of a surrogate, non-malignant but easily and safely obtained tissue or fluid may be possible. The correlation between results on the surrogate and that on the tumor would need to be assessed during development of the surrogate marker.

9. Conclusion

While we have come a long way in generating new molecular knowledge of cancer and the human genome, we have not yet reached the goal of being able to use this knowledge to prescribe the treatment that will be most beneficial and least harmful for the patient. Knowledge of molecular attributes that denote subsets of patients with differing prognosis and response to therapies will help individual patients avoid therapies that will not be beneficial, will streamline developmental therapeutic trials, and will lead to even more discoveries of effective treatments. Progress toward this goal will be greatly facilitated by a methodological approach that discovers potential new markers, assesses them for the potential of clinical usefulness, and then proves that clinical usefulness with well designed clinical trials. Such an endeavor will require the cooperation of the medical profession, basic scientists, industry, and most of all patients.

References

[1] J. Albanell, J. Codony, A. Rovira, B. Mellado and P. Gascon, Mechanism of action of anti-HER2 monoclonal antibodies: Scientific update on trastuzumab and 2C4, *Adv Exp Med Biol* **532** (2003), 253–268.

[2] Y. Ando, H. Saka, M. Ando, T. Sawa, K. Muro, H. Ueoka, A. Yokoyama, S. Saitoh, K. Shimokata and Y. Hasegawa, Polymorphisms of UDP-glucuronosyltransferase gene and irinotecan toxicity: A pharmacogenetic analysis, *Cancer Res* **60** (2000), 6921–6926.

[3] C.L. Arteaga, Trastuzumab, an appropriate first-line single-agent therapy for HER2-overexpressing metastatic breast cancer, *Breast Cancer Res* **5**(2) (2003), 96–100.

[4] J. Boye, T. Elter and A. Engert, An overview of the current clinical use of the anti-CD20 monoclonal antibody rituximab, *Ann Oncol* **14** (2003), 520–535.

[5] R. Capdeville and S. Silberman, Imatinib: A targeted clinical drug development, *Semin Hematol* **40**(Supple 3) (2003), 15–20.

[6] S. Cascinu, M.P. Staccioli, G. Gasparini, P. Giordani, V. Catalano, R. Ghiselli, C. Rossi, A.M. Baldelli, F. Graziano, V. Saba, P. Muretto and G. Catalano, Expression of vascular endothelial growth factor can predict event-free survival in stage II colon cancer, *Clin Cancer Res* **6** (2000), 2803–2807.

[7] M.E. Charlson, K.L. Ales, R. Simon and C.R. MacKenzie, Why predictive indexes perform less well in validation studies, *Arch Intern Med* **147** (1987), 2155–2161.

[8] D.J. Grignon, R. Caplan, F.H Sarkar, C.A. Lawton, E.H. Hammond, M.V. Pilepich, J.D. Forman, J. Mesic, K.K. Fu, R.A. Abrams, T.F. Pajak, W.U. Shipley and J.D. Cox, P53 status and prognosis of locally advanced prostatic adenocarcinoma: a study based on RTOG 8610, *J Natl Cancer Inst* **89**(2) (1997), 158–165.

[9] R. Gryfe, H. Kim, E.T.K. Hsieh, M.D. Aronson, E.J. Holowaty, S.B. Bull, M. Redston and S. Gallinger, Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer, *N Engl J Med* **342** (2000), 59–77.

[10] M.E.H. Hammond and S. Taube, Issues and barriers to development of clinically useful tumor markers: A developmental pathway proposal, *Sem Oncol* **29**(3) (2002), 213–221.

[11] K.C. Halling, A.J. French, S.K. McDonnell, L.J. Burgart, D.J. Schaid, B.J. Peterson, L. Moon-Tasson, M.R. Mahoney, D.J. Sargent, M.J. O'Connell, T.E. Witzig, G.H. Farr Jr, R.M. Goldberg and S.N. Thibodeau, Microsatellite instability and 8p allelic imbalance in stage B2 and C colorectal cancers, *J Natl Cancer Inst* **91**(15) (1999), 1295–1303.

[12] N. Harbeck, P. Dettmar, C. Thomssen, U. Berger, K. Ulm, R. Kates, H. Hofler, F. Janicke, H. Graeff and M. Schmitt, Risk-group discrimination in node negative breast cancer using invasion and proliferation markers: 6-year median follow-up, *Br J Cancer* **80**(3–4) (1999), 419–426.

[13] D.F. Hayes, B. Trock and A.L. Harris, Assessing the clinical impact of prognostic factors: when is "statistically significant" clinically useful? *Breast Cancer Res Treat* **52** (1998), 305–319.

[14] S.G. Hilsenbeck, G.M. Clark and W.L. McGuire, Why do so many prognostic factors fail to pan out? *Breast Cancer Res Treat* **22**(3) (1992), 197–206.

[15] F. Janicke, M. Schmitt and H. Graeff, Clinical relevance of the urokinase-type and tissue-type plasminogen activators and of their type 1 inhibitor in breast cancer, *Semin Thromb Hemost* **17** (1991), 303–312.

[16] F. Janicke, M. Schmitt, L. Pache, K. Ulm, N. Harbeck, H. Hofler and H. Graeff, Urokinase (uPA) ad its inhibitor PAI-1 are strong and independent prognostic factors in node-negative breast cancer, *Breast Cancer Res Treat* **24**(3) (1993), 195–208.

[17] F. Janicke, A. Prechtl, C. Thomssen, N. Harbeck, C. Meisner, M. Untch, C.G. Sweep, H.-K. Selbmann, H. Graeff and M. Schmitt, Randomized adjuvant chemotherapy trial in high-risk, lymph node-negative breast cancer patients identified by urokinase-type plasminogen activator and plasminogen activator inhibitor type 1, *J Natl Cancer Inst* **93**(12) (2001), 913–920.

[18] A. Jemal, T. Murray, A. Samuels, A. Ghafoor, E. Ward and M.J. Thun, Cancer Statistics, 2003, *CA Cancer J Clin* **53** (2003), 5–26.

[19] J.R. Johnson, P. Bross, M. Cohen, M. Rothmann, G. Chen, A. Zajicek, J. Gobburu, A. Rahman, A. Staten and R. Pazdur, Approval Summary: Imatinib mesylate capsules for treatment of adult patients with newly diagnosed Philadelphia chromosome-positive chronic myelogenous leukemia in chronic phase, *Clin Cancer Res* **9** (2003), 1972–1979.

[20] F. Kabbinavar, H.I. Hurwitz, L. Fehrenbacher, N.J. Meropol, W.F. Novotny, G. Lieberman, S. Griffing and E. Bergsland, Phase II, randomized trial comparing bevacizumab plus fluorouracil (FU)/leucovorin (LV) with FU/LV alone in patients with metastatic colorectal cancer, *J Clin Oncol* **21**(1) (2003), 60–65.

[21] A.J. Karayiannakis, K.N. Syrigos, A. Zbar, N. Baibas, A. Polychronidis, C. Simopoulos and G. Karatzas, Clinical significance of preoperative serum vascular endothelial growth factor levels in patients with colorectal cancer and the effect of tumor surgery, *Surgery* **131** (2002), 548–555.

[22] J.-C. Lee, N.-H. Chow, S.-T. Wang and S.-M. Huang, Prognostic value of vascular endothelial growth factor expression in colorectal cancer patients, *Eur J Cancer* **36** (2000), 748–753.

[23] S. Marsh and H.L. McLeod, Thymidylate synthase pharmacogenetics in colorectal cancer, *Clin Colorectal Cancer* **1**(3) (2001), 175–178.

[24] A.D. McCollum, P.J. Catalano, D.G. Haller, R.J. Mayer, J.S. Macdonald, A.B. Benson and C.S. Fuchs, Outcomes and toxicity in African-American and Caucasian patients in a randomized adjuvant chemotherapy trial for colon cancer, *J Natl*

[25] W.L. McGuire, Breast cancer prognostic factors: evaluation guidelines, *J Natl Cancer Inst* **83**(3) (1991), 154–155.

[26] J.M. Nabholtz, D.M. Reese, M.A. Lindsay and A. Riva, HER2-positive breast cancer: update on Breast Cancer International Research Group Trials, *Clin Breast Cancer* **suppl2** (2002), S75–S79.

[27] National Institutes of Health Consensus Development Panel. National Institutes of Health Consensus Development Conference Statement: Adjuvant therapy for breast cancer, November 1–3, 2000, *J Natl Cancer Inst* **93**(13) (2000), 979–989.

[28] NIH Consensus Conference: Adjuvant therapy for patients with colon and rectal cancer, *JAMA* **264**(11) (1990), 1444–1450.

[29] Oral Presentation: American Society of Clinical Oncology (ASCO) annual meeting, Chicago, June 1, 2003.

[30] C.K. Osborne, Tamoxifen in the treatment of breast cancer, *N Engl J Med* **339**(22) (1998), 1609–1618.

[31] T.F. Pajak, G.M. Clark, D.J. Sargent, L.M. McShane and E.H. Hammond, Statistical issues in tumor marker studies, *Arch Pathol Lab Med* **124** (2000), 1011–1015.

[32] R. Pettengell and D. Linch, Position paper on the therapeutic use of rituximab in CD20-positive diffuse large B-cell non-Hodgkin's lymphoma, *Br J Haematol* **121** (2003), 44–48.

[33] A. Prechtl, N. Harbeck, C. Thomssen, C. Meisner, M. Braun, M. Untch, M. Wieland, B. Lisboa, T. Cufer, H. Graeff, K. Selbmann, M. Schmitt and F. Janicke, Tumor-biological factors uPA and PAI-1 as stratification criteria of a multicenter adjuvant chemotherapy trial in node-negative breast cancer, *Int J Biol Markers* **15** (2000), 73–78.

[34] S.T. Pullarkat, J. Shoehlmacher, V. Ghaderi, Y.-P. Xiong, S.A. Ingles, A. Sherrod, R. Warren, D. Tsao-Wei, S. Groshen and H.-J. Lenz, Thymidylate synthase gene polymorphism determines response and toxicity of 5-FU chemotherapy, *The Pharmacogenomics Journal* **1** (2001), 65–70.

[35] C.M. Ribic, D.J. Sargent, M.J. Moore, S.N. Thibodeau, A.J. French, R.M. Goldberg, S.R. Hamilton, P. Laurent-Puig, R. Gryfe, L.E. Shepherd, D. Tu, M. Redston and S. Gallinger, Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer, *N Engl J Med* **349**(3) (2003), 247–257.

[36] D. Sargent and C. Allegra, Issues in clinical trial design for tumor marker studies, *Sem Oncol* **29**(3) (2002), 222–230.

[37] R. Simon and D.G. Altman, Statistical aspects of prognostic factors studies in oncology, *Br J Cancer* **69** (1994), 979–985.

[38] E. Villafranca, Y. Okruzhnov, M.A. Dominguez, J. Garcia-Foncillas, I. Azinovic, E. Martiez, J.J. Illarramendi, F. Arias, R. Martinez Monge, E. Salgado, S. Angeletti and A. Brugarolas, Polymorphisms of the repeated sequences in the enhancer region of the thymidylate synthase gene promoter may predict downstaging after preoperative chemoradiation in rectal cancer, *J Clin Oncol* **19**(6) (2001), 1779–1786.

[39] T. Watanabe, T.-T. Wu, P.J. Catalano, T. Ueki, R. Satriano, D.G. Haller, A.B. Benson III and S.R. Hamilton, Molecular predictors of survival after adjuvant chemotherapy for colon cancer, *N Engl J Med* **344**(16) (2001), 1196–1206.