



Published in final edited form as:

Head Neck. 2014 July ; 36(7): 1069–1075.

Designing biomarker studies for head and neck cancer

Kelly Y. Kim, Ph.D.,

Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD

Lisa M. McShane, Ph.D., and

Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD

Barbara A. Conley, M.D.

Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD

Abstract

While there is ample literature reporting on the identification of molecular biomarkers for head and neck squamous cell carcinoma, none is currently recommended for routine clinical use. A major reason for this lack of progress is the difficulty in designing studies in head and neck cancer to clearly establish the clinical utility of biomarkers. Consequently, biomarker studies frequently stall at the initial discovery phase. In this paper, we focus on biomarkers for use in clinical management, including selection of therapy. Using several contemporary examples, we identify some of the common deficiencies in study design that hinder success in biomarker development for this disease area, and we suggest some potential solutions. The goal of this article is to provide guidance that can assist investigators to more efficiently move promising biomarkers in head and neck cancer from discovery to clinical practice.

Keywords

prognostic biomarkers; molecular markers; clinical utility; head and neck cancer; HNSCC

INTRODUCTION

Despite a long list of prognostic and predictive biomarker candidates that can be found in the literature, *not one molecular marker has been widely accepted for routine use in managing patients with head and neck squamous cell carcinoma (HNSCC)*. The National Comprehensive Cancer Network (NCCN) recognizes human papilloma virus (HPV) and p16INK4A (p16) as prognostic markers for oropharyngeal cancer, whereby tumors positive for HPV infection or p16 overexpression have better prognosis than HPV negative tumors; and NCCN recognizes the association of Epstein-Barr virus (EBV) infection with nasopharyngeal cancer.¹ How we should apply HPV or EBV positive status in treatment decision-making is under investigation at various institutions, and right now these markers are mostly used in experimental, clinical trial settings to determine patient eligibility and to stratify patients as part of trial design.

Corresponding Author: Kelly Y. Kim, Division of Cancer Treatment and Diagnosis, National Cancer Institute, 9609 Medical Center Dr., Room 4W430 [MSC 9730], Bethesda, MD 20892, 240-276-5938 (phone), 240-276-7889 (fax), kimke@mail.nih.gov.

A review article by Lothaire et al. compared the published findings on the most extensively studied molecular biomarkers for head and neck squamous cell carcinoma (HNSCC), including the epidermal growth factor receptor (EGFR), cyclin D1 (*CCND1*), Bcl-2 (B-cell lymphoma 2), cyclin-dependent kinase inhibitor p27 (Kip1), vascular endothelial growth factor (VEGF), and p53; and found that the reported conclusions about their prognostic or predictive value by the different research groups were not always consistent.² More recent studies on the association of biomarker expression with clinical outcomes have reported new results for Bcl-2^{3,4} that seem to contradict the findings in some older papers;^{5,6} and produced additional reports that contradict each other for EGFR^{7,8} as well as for p27^{9,10}. A number of factors could contribute to different results being produced by different laboratories, including true clinical variability in the patient cohorts studied and variations in the assay (e.g., technological platforms for detection and measurement, sources of reagents, whether the tissue was fresh or fixed, scoring procedures, and cut points). These sources of variability could consequently cause one laboratory to find that the overexpression of a biomarker is associated with certain clinical outcomes and another laboratory to see no such association. Biomarker expression, for example, could be measured in terms of mRNA levels by one laboratory and protein levels by another. The measurement of mRNA expression could be accomplished via the reverse transcriptase-polymerase chain reaction (RT-PCR) method by one laboratory and by using Affymetrix expression arrays by another. Antibodies from different vendors and even different lots from the same vendor may differ in their specificity and binding affinity to the same protein biomarker, and so on.

Variations in the study design and model development (e.g., specimen selection, patient population, clinical endpoints, analysis methods) could result in differing prognostic and predictive models. The endpoint used to represent clinical outcome in the analysis could have a significant impact on the findings regarding the prognostic or predictive strength of a biomarker. That is, a biomarker associated with better local control rate or disease free survival may not be prognostic of better overall survival. Use of response rate as an endpoint may not be appropriate if the study is evaluating a predictive biomarker for a cytostatic agent that may not cause tumor shrinkage. For evaluation of the clinical utility of a biomarker, it is first necessary to have an analytically and clinically validated assay. In this paper, we do not delve into the technical and analytical validity issues, which have been covered by other publications.¹¹⁻¹⁵ Rather, we discuss in greater detail the common pitfalls related to study design that could produce discrepancies and raise concerns about the clinical validity and utility of the findings.

The basic framework for biomarker development is comprised of the following processes: 1) discovery of initial correlations using retrospective specimens, 2) defining the intended use for the biomarker, 3) performing analytical validation and refining the test to settle on the final form (i.e., “locked down” version), 4) clinical validation of the test in appropriate study populations, and 5) performing confirmatory studies that address clinical utility. In the following section, we discuss the common pitfalls related to study design that weaken the validity of findings, and suggest ways to approach biomarker study design so that appropriate interpretations of marker performance and clinical utility can be made. Inherent

hurdles to sound biomarker study design due to the complexity of head and neck tumors' genetic and molecular underpinnings and the rarity of HNSCC are also discussed.

DEFINING CLINICAL UTILITY

There are relatively few markers in oncology that have demonstrated clinical utility, and this is particularly true for head and neck cancer. To establish clinical utility for a biomarker, investigators must show that use of the biomarker in guiding patient care results in an overall benefit to the patient (e.g., improved survival and/or quality of life) when weighed against the risks associated with use of the biomarker. The two types of markers that are most often considered for potential clinical utility in the care of patients with cancer are *predictive* and *prognostic* markers. Predictive markers are those that identify patients who benefit from a *particular* therapy (relative to other available therapies). Prognostic markers are those markers that are associated with a clinical outcome in the absence of therapy, or sometimes in the context of standard therapy that all patients are likely to receive. Potential for clinical utility is clear for predictive markers, but it can be a greater challenge to identify prognostic markers that achieve clinical utility rather than the more limited claim of clinical validity. Clinical validity is demonstrated by establishing a suitably strong association between the prognostic marker and a clinical outcome of interest, but this does not guarantee that the marker will be useful for clinical decision making.

Simply demonstrating that a prognostic marker can distinguish two groups of patients with different survival outcome is not sufficient to establish its clinical utility. If survival in two groups of patients defined on the basis of a marker is different, but survival in both groups is poor and no treatment that will improve survival in either group is available, then the marker does not have clinical utility. For example, if we are dealing with locally advanced oral squamous cell carcinoma (OSCC) patients with resectable tumor for whom the standard of care is surgery plus radiation therapy (RT) and the goal of the biomarker test is to identify low-risk patients who may safely receive lower dose adjuvant RT, then being able to achieve a good separation is not useful if the low-risk group still has significant risk of poor outcome such that reducing RT cannot be justified. In other words, results can show differences that are statistically significant but may not be clinically significant. Such a marker might be used to stratify patients in a clinical trial to reduce noise and increase statistical efficiency of the trial, but that does not establish clinical utility of the marker for guiding treatment decisions for individual patients. Also, if studies are carried out using archived specimens with incomplete clinical annotation (e.g., no information about treatment), the type of conclusions that can be drawn from such studies is limited. It may be possible to perform discovery studies looking for *prognostic* markers by examining the associations between markers and clinical outcome, but treatments received may confound those associations.

A situation in which a prognostic marker could have clinical utility is the setting in which a biomarker-defined patient group has such good prognosis in the absence of further therapy that the toxicities of more aggressive treatment are not justified. One example of a prognostic test with clinical utility of this type is the OncotypeDX test (Genomic Health, Inc., Redwood City, CA) for women with early-stage, hormone receptor-positive breast cancer. The test is comprised of a panel of 21 genes whose expression is used to generate a

recurrence score that can identify patients who can forgo adjuvant chemotherapy. An example of a prognostic marker in head and neck cancer with potential for clinical utility by identifying a group of patients with good prognosis who might benefit more from less aggressive therapies is the favorable risk marker, HPV or p16, for oropharyngeal cancers. Clinical trials evaluating HPV/p16 for this clinical use are mentioned in greater detail later in this paper.

Evaluation of clinical utility requires a systematic approach considering many factors. An important first step is to identify the intended clinical use. The *specific clinical context* in which the marker could inform and improve treatment decision-making should be clear, and discussed in the context of the standard of care or standard practice not only with respect to the specific therapy administered but also how clinicopathological information is currently used to select among the treatment options available. For example, postoperative RT but not concomitant chemoradiation (CRT) is generally administered as adjuvant therapy to surgery in patients with the following risk features: multiple positive nodes, perivascular or perineural invasion, advanced primary T classification, or nodal involvement at levels IV or V (for oral cavity and oropharyngeal cancers). However, when the risk features include *extracapsular nodal extension or positive resection margins*, postoperative CRT is recommended over RT because it has been shown that patients with these adverse risk features benefited from the addition of cisplatin as a radiosensitizer to postoperative RT, while no survival advantage was observed in patients with multiple involved regional nodes without extracapsular spread.¹⁶ Therefore, investigators must consider and concurrently evaluate the established clinicopathologic factors that have already been rigorously shown to be associated with outcome or with sensitivity to a particular treatment, in order to establish the utility of a new prognostic or predictive molecular marker, respectively. In other words, investigators need to show that the treatment decision guided by use of the marker is different from standard of care and results in an overall positive balance of benefits to risks. Below, we elaborate on a minimal set of criteria for approaching molecular marker research with a view towards defining and demonstrating clinical utility (Table 1).

The intended use for the biomarker test should be identified based on medical need

Identifying whether the marker will be used for determining prognosis, treatment response, risk of recurrence, or risk of toxicity should be the first step. Investigators should describe the area of medical need by explaining how the current approaches or factors that guide clinical decision-making are inadequate, problematic, or controversial. These could involve cases where there is uncertainty or no uniform recommendation on treatment selection, or for the subset of patients who do not benefit from the standard treatment as expected. For example, there is a need for better prognostic markers and assays when the clinicopathologic features currently being used to determine the patient's risk and prognosis are not highly accurate; or there is lack of consensus among clinicians regarding the treatment and management of patient subgroups characterized by certain risk features, which might be resolved by further refinement or stratifications of the subgroups. In order for the test results to be actionable and change practice, alternative treatment options need to be available or developed, should the results from better assays indicate that the standard treatment approaches are not optimal.

The target patient population for that intended use should be precisely defined early in the biomarker test development process

The intended use that has been identified will dictate which patient population should be selected for the study. Investigators need to define the characteristics of the patient population for which the molecular marker is being developed in sufficient detail, vis-a-vis tumor subsite, stage, and other factors that would make the test relevant for that particular context of medical application. It may not be sufficient to characterize the target population as “advanced stage” patients, for example, if the treatment decisions are going to differ depending on whether the tumors are locally advanced, regionally advanced, or distantly metastatic.

Specimens used in the study must represent the tumors that will be sampled in the clinical setting or context relevant for the intended use of the biomarker

Early in the biomarker development process, due consideration needs to be given to the appropriateness of specimens or datasets, particularly convenience samples, being used to address the specific clinical questions regarding treatment decision-making. Specimens obtained by biopsy may have different characteristics than surgical excision specimens. It may be tempting to study prognostic markers for locoregionally advanced stage cancers using surgical specimens from stage I and II disease because those specimens are easier to obtain, but the association of a marker with a clinical outcome of interest or its usefulness for clinical management may differ appreciably depending on stage. Investigators may be tempted to answer prognostic questions about recurrent tumors by studying specimens from primary tumors. This could be problematic because the biomarker levels and the behavior of tumors may not be comparable between recurrent and primary tumors. Prior treatment could affect the biomarker profile and behavior of the recurrent tumor, and confound the associations between prognostic markers and clinical outcome. Chemotherapy that was received, for example, could have resulted in the selection of a clonal group, whose molecular characteristics may be different from those of treatment-naïve tumor. Feasibility of collecting the necessary specimens also plays a role in determination of clinical utility of a marker test, which is another reason that biomarker studies should be conducted using specimens in the format that will be required to perform the test in routine clinical practice. Also, consideration should be given to choosing the appropriate assay platform that will have robust performance characteristics for the types of samples expected to be available once the test is in clinical use. Restricting to the use of specimens typical of the target clinical setting as early as possible in the biomarker development process will be most efficient in the evaluation of a test’s potential clinical utility.

Data analysis should account and adjust for clinical, pathological, and molecular confounders

The biological differences among HNSCCs due to the tumors’ genetic and molecular underpinnings, and variable tumor behavior in terms of tumor progression and treatment responsiveness that are observed from tumors at different stages and different anatomic subsites require that investigators remain vigilant to ensure that the interpretation of studies is free of influences from *confounding variables*. Factors such as primary tumor site, disease

stage, and different treatment regimens affect prognosis and could produce differential treatment response. For instance, tumors originating from different sites can exhibit varying behavior that is not predictable by histopathology (e.g., the distant metastasis rate is much greater for nasopharyngeal carcinomas than laryngeal or oropharyngeal carcinomas).^{17–20} *Resection margin status* is also an important prognostic indicator, since incomplete surgical resection of the tumor would increase the chances for disease recurrence. The presence of *cervical lymph node metastasis* is another powerful indicator of poor prognosis, and it is independent of T stage, as small tumors can be highly metastatic while some large tumors are not aggressive.

Investigators also need to be aware of and account for molecular confounders that could alter the expression or function of the biomarker, and interfere with our interpretation of a biomarker's prognostic or predictive significance. For example, overexpression of a mutant p53 could down-regulate Bcl-2 at both the protein and mRNA levels,²¹ such that studies looking at Bcl-2 as a potential biomarker may need to account for the p53 status of the tumors. Proper interpretation of what might appear to be a straightforward analysis of a marker's prognostic value can be extremely challenging in light of this complexity.

While multivariable analyses can be used to adjust for the effects of standard variables, typical biomarker studies are not sufficiently large to reliably detect the presence of interactions between standard factors and the biomarker of interest. An example of *marker by subsite interaction* is represented by HPV, which is a prognostic marker for oropharyngeal tumors but generally does not exhibit prognostic value in head and neck cancers in other subsites. If a study includes multiple sites, then even if “subsite” is adjusted for in a multivariate model, the subsites represented in largest proportion could drive the overall estimate of the effect of a marker in the study, unless the model also incorporated a term for the interaction between subsite and marker.

An alternative approach to addressing the diversity in HNSCC biology is to conduct studies in focused subgroups (e.g., focused on a single anatomic site, etc.), but the rarity of head and neck cancer presents a significant inherent hurdle. Limited numbers of patients available for prospective tissue collection, the difficulty in obtaining tissue samples from some subsites, and the extremely limited numbers of head and neck cancer specimens available from retrospective tissue collections make it extraordinarily challenging to conduct marker studies in focused subgroups of head and neck cancer. The presence of real predictive or prognostic effect may be missed due to random variability of the observed effect and lack of statistical power inherent in small studies. Conducting multiple subgroup analyses within large studies of patients with heterogeneous characteristics also runs the risks of generating false positive findings due to exploratory analyses and statistical testing in multiple subgroups.

Decisions about which HNSCC subtypes can be meaningfully combined, and under what circumstances, rely heavily on biological and medical rationales that support the assumptions about the behaviors (e.g., aggressiveness, treatment responsiveness, etc.) of the tumors to justify this approach. If sufficient numbers of comparable independent small studies are available, then use of meta-analysis techniques can also be considered to formally combine results across studies.

The molecular marker test should be more efficient (e.g., more cost-effective, less invasive) or contribute clinically important information beyond that provided by other biomarkers and clinicopathologic factors that are already part of existing treatment guidelines

Clinicians may not find a new marker test to be useful if it provides little added value compared to using clinicopathological risk factors (e.g., extracapsular nodal spread, perineural or perivascular invasion, and, for oral cavity and oropharyngeal cancers, nodal involvement at levels IV or V) or other existing diagnostic methods that are already established as a part of standard clinical practice. Studies should adjust for the influences of these confounding variables in order to provide reliable evidence that the marker adds clinically relevant information beyond what is already known from standard clinical and pathologic factors. Markers that are merely correlated with tumor stage or lymph node metastasis are not useful because no additional insight would be provided. Another way to evaluate whether a prognostic marker performs better than established prognostic variables is to first generate a risk score, for example, as could be obtained by Cox regression modeling of the association between the clinicopathological variables and survival. This first step establishes the baseline for comparison with the prognostic markers under investigation. Techniques for comparing the performance of two prognostic models include analyzing the change in the concordance index, analyzing the difference in the area under the time-dependent receiver operating characteristic (ROC) curve between the two prognostic models, and analyzing the differences in the positive predictive values and negative predictive values for predicting failure time (PPV(t) and NPV(t), respectively) of the two prognostic models.²²

Include specimens from patients that have not received the treatment in question (i.e., control specimens) for studies aimed at distinguishing predictive versus prognostic role of a marker

Although the distinction between prognostic versus predictive markers was explained earlier in the paper, it is reiterated here because we have noticed on many occasions that investigators make the error of trying to evaluate putative predictive markers by using study designs that can only assess the prognostic value of a biomarker. When the specimens used in the study are from patients who were all treated with cisplatin, for example, it would be erroneous to conclude that the biomarker can identify patients who are sensitive to cisplatin based on differential outcome in low versus high expressers of the biomarker are observed. The biomarker could be a prognostic marker that tells us who will have better outcome regardless of treatment. In order to establish that a biomarker is predictive, investigators need to demonstrate *that relative to a non-cisplatin treatment option*, cisplatin provides a clinically meaningful improvement in outcome in the biomarker-positive but does not provide such a benefit in the biomarker-negative subgroup.

AREAS OF NEED FOR MOLECULAR MARKERS IN HNSCC

In this section, we describe a few examples of areas that have been the focus of HNSCC research activities in recent years. We use the first example to go over how the guidelines for study design and clinical utility assessment that have been outlined in the previous

sections could be applied. One area of active research is the development of molecular biomarkers that could help predict occult lymph node metastasis for OSCC.^{23–25} This was identified to be an area of need by a number of different research groups because of a potential overuse of elective neck dissections in current clinical practice as a preventive measure against undetected metastatic disease at initial diagnosis. This practice stems from the fact that the status of cervical lymph nodes is considered the most important prognostic factor for OSCC,²⁶ and current methods to assess risk (e.g., techniques for measuring tumor thickness) have problems with respect to accuracy and uniformity in measurement techniques.²⁷ For example, the recommended tumor thickness cutoff for prescribing elective neck dissection varies greatly in the literature, ranging from 1.5 to 8 mm; and there is also variability in how tumor thickness is measured, with some measuring the entire tumor thickness from the surface, others measuring from a line that approximates the boundary of where the normal mucosa would be to the deepest extent of the tumor, and still others only measuring from the basement membrane to the deepest extent of tumor.²⁷ Neck dissection carries the risk of serious morbidities, which must be weighed against the risk of poor outcome with an observation approach. The risk of regional nodal involvement at presentation for OSCC ranges from 20% to 45%²⁷ although it varies greatly according to the subsite: primaries of the alveolar ridge and hard palate, for example, infrequently involve the neck, but the incidence of occult neck metastasis is 50% to 60% in patients with anterior tongue cancer.¹

Here the investigators need to *identify the intended use* for their biomarker test. The most appropriate intended use would be to help clinicians identify which oral cancer patients do not need to have their cervical lymph nodes removed and can be spared the risk of morbidity and cost associated with a neck dissection. The *appropriate target population* for such a study would be patients who were the clinically node-negative at diagnosis and who have been followed for an adequate period of time to accurately determine who was truly negative versus who had occult disease that later manifested itself clinically. The specimens assessed would have to be those collected at diagnosis, regardless of whether the patients were later determined to have involved lymph nodes at time of surgery. The development of the prognostic model should *account for factors that could confound the results*, such as treatment that could eradicate microscopic neck metastases. According to the NCCN guidelines, definitive RT is a treatment option for T1-2, N0 disease, where at least 44–64 Gy is given to the neck, and postoperative CRT is recommended for all patients with resected oral cavity cancers that have positive margins.¹

The prognostic model must also perform better in terms of sensitivity, specificity, negative predictive value, and positive predictive value than the risk assessment methods currently used in practice. However, it would be challenging to quantitate the performance of established clinicopathologic risk factors associated with occult lymphatic metastasis, such as tumor thickness and depth of invasion, for comparison with the performance of new molecular markers, considering the variable practices by surgeons for measuring these risk factors and including them in their decision-making process. Also, if the specimens are from a convenience sample, not all of the information on which the neck management decision was based might be available from a retrospective specimen set. Some investigators

incorporate the clinicopathologic factors along with the molecular markers into the prognostic model that is being developed. Because many in the head and neck community consider up to 20% probability of occult metastasis as acceptable risk, some investigators set this value as the bar that a biomarker test needs to exceed. This risk level was established by Weiss et al. through computer-assisted mathematical modeling of three different treatment decisions (neck dissection, RT, and observation) and the outcomes associated with each decision (cure, death, cure with surgery, cure with RT, and cure with salvage), to determine the threshold at which the benefits outweigh the costs to patients of prophylactically treating the N0 neck.²⁸ However, given that the adverse risks associated with having cervical lymph node metastasis (e.g., risk for tumor recurrence, low success of salvage surgery, and death) are much more threatening than the risks associated with neck surgery, it is not certain whether surgeons and patients would be willing to accept a test results that has 20%, or even a 10%, false negative rate as good enough for them to forgo elective neck dissection. With the use of *selective* neck dissection as the preferred elective treatment currently, neck treatment is much less likely to be associated with the detriment to quality of life that could result from radical neck dissection (e.g., long-term damage to shoulder function and chronic pain).

While the previous example was used to demonstrate how the criteria for productive biomarker study design could be applied, we now present an example of a biomarker whose performance as a favorable prognostic marker has become established and the need now is to integrate the biomarker testing into prospective studies to produce the level of evidence needed for clinical application. Claims that prognostic markers are clinically useful because they can be used to identify which patients should be monitored more closely for disease recurrence or progression or who could receive less aggressive therapy because their prognosis is good are generic statements that can be ascribed indiscriminately to almost all prognostic markers for all diseases. These claims are meaningful only when clinical utility can be confirmed in the setting of treatment trials or from appropriate prospective-retrospective studies.²⁹ While this paper has not delved into clinical trial design issues, the final step of biomarker development in which a definitive assessment of clinical utility is made generally involves large phase II or phase III clinical trials, although sometimes convincing findings from phase II studies or large retrospective studies may also be acceptable. Generally, results of a retrospective study would need to be confirmed with corroborating evidence from at least one additional similar study. These prospective studies to confirm the findings from retrospective studies also need to be rigorously defined and executed in terms of selecting the right endpoint, making sure that there is adequate power, randomization as appropriate, and so on as described by Freidlin et al.³⁰

Although many studies have shown that patients with HPV-positive oropharyngeal cancer have improved survival when compared to patients with HPV-negative oropharyngeal cancer,^{31–33} NCCN guidelines state that “HPV testing should not change management decisions.”¹ The reason for this is that clinical utility and how the information should be used in routine clinical decision-making are still under investigation. The following clinical trials (found in the National Institutes of Health registry of clinical trials: www.clinicaltrials.gov) show the head and neck oncology community’s efforts toward

defining the clinical utility of HPV testing through a better understanding of the effects of HPV status on responsiveness to various treatment intensities and also the patterns of failure. University of Michigan Cancer Center is currently recruiting participants for their phase II clinical trial (NCT01663259) of reduced-intensity therapy for locally advanced oropharyngeal cancer in non-smoking HPV-positive patients. The “reduced-intensity therapy” under investigation is the replacement of concurrent chemotherapy with cetuximab, and the goal is to see if there will be a reduction in long-term toxicity without an increase in the tumor recurrence rate. Similarly, Radiation Therapy Oncology Group is conducting a phase III trial (NCT01302834) to see if cetuximab could replace cisplatin in treating p16-positive, locally advanced oropharyngeal cancer patients. University of North Carolina Lineberger Comprehensive Cancer Center is currently recruiting participants for its phase II study (NCT01530997) of de-intensification of radiation and chemotherapy for HPV-related OPSCC. Their standard CRT regimen for OPSCC consists of 7 weeks of radiation with high doses of cisplatin. The goal of this study is to evaluate whether a shorter, less intensive regimen (6 weeks of 54–60 Gy (total doses) of intensity modulated radiotherapy (IMRT)) with a lower weekly dose of cisplatin (30 mg/m²) could provide a similar complete pathological response rate. There are several other trials examining reduced doses of IMRT as well as different schedules for cisplatin or replacing cisplatin with different drugs or biologics.

DISCUSSION

While clinical validity of a biomarker involves demonstrating that the test result correlates with a clinical outcome of interest, there is no assurance that the marker will be useful for clinical decision-making. In order to demonstrate clinical utility, investigators need to show that the biomarker test result would be actionable and that the change in patient treatment as directed by the test result would result in a significant improvement in patient survival or quality of life that would outweigh the risks and costs of testing. However, assessment of clinical utility is not possible when hampered by deficiencies early in the development process. These deficiencies might include lack of focus on a specific target patient population (in terms of disease stage, anatomic site, primary or recurrent, etc.), use of specimens that are not representative of the correct patient population or target tissue (including site and timing of specimen collection), or failure to account for factors that could confound the results. Also, in order for the test results to be actionable and change practice, alternative treatment options need to be available should the results from better predictive or prognostic assays indicate that the standard treatment approaches are not optimal. A test used to direct patient treatment must be performed in a CLIA certified laboratory and, if used in a clinical trial, may also need an Investigational Device Exemption from the Centers for Diagnostics and Radiologic Health, U.S. Food and Drug Administration (Code of Federal Regulations, Title 21, Part 812; <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?cfrpart=812>)

Many biomarker research proposals tend to straddle basic and clinical research pursuits. By focusing on establishing a rigorous experimental approach required for mechanistic studies to elucidate a marker’s biological role in tumor aggressiveness or metastatic potential and on finding a correlation between the biomarker with tumor stage, grade, or metastasis, the

investigators can demonstrate clinical validity but fall short of addressing the clinical context in which the biomarker could have utility. Scrutiny of these proposals does not proceed very far before flaws are found at the fundamental level of having the appropriate clinical specimens or having an acceptable statistical design, including adequate study sample size. With the current promise of predictive medicine, patients and clinicians are eager to see new predictive and prognostic biomarkers implemented in the clinic. Peer review panels evaluating biomarker research proposals now pay closer attention to the potential clinical utility of biomarker tests as well. Thus, it is important to understand how to develop a robust and useful test that can improve medical practice.

Acknowledgments

The authors would like to thank Dr. Bhupinder Mann in the Cancer Therapy Evaluation Program of the National Cancer Institute for his invaluable comments on the manuscript.

References

1. NCCN. National Comprehensive Cancer Network (NCCN). Clinical Practice Guidelines in Oncology: Head and Neck Cancers. 2011. Version 2.2011:http://www.nccn.org/professionals/physician_gls/f_guidelines.asp
2. Lothaire P, de Azambuja E, Dequanter D, et al. Molecular markers of head and neck squamous cell carcinoma: Promising signs in need of prospective evaluation. *Head and Neck-Journal for the Sciences and Specialties of the Head and Neck*. 2006; 28:256–69.
3. Michaud WA, Nichols AC, Mroz EA, Faquin WC, Clark JR, Begum S. Bcl-2 Blocks Cisplatin-Induced Apoptosis and Predicts Poor Outcome Following Chemoradiation Treatment in Advanced Oropharyngeal Squamous Cell Carcinoma. *Clinical cancer research*. 2009; 15:1645–54. [PubMed: 19240170]
4. Nichols AC, Finkelstein DM, Faquin WC, Westra WH, Mroz EA, Kneuert P. Bcl2 and Human Papilloma Virus 16 as Predictors of Outcome following Concurrent Chemoradiation for Advanced Oropharyngeal Cancer. *Clinical cancer research*. 2010; 16:2138–46. [PubMed: 20233885]
5. Wilson GD, Grover R, Richman PI, Daley FM, Saunders MI, Dische S. BCL-2 expression correlates with favourable outcome in head and neck cancer treated by accelerated radiotherapy. *Anticancer Research*. 1996; 16:2403–8. [PubMed: 8816842]
6. Wilson GD, Saunders MI, Dische S, Richman PI, Daley FM, Bentzen SM. bcl-2 expression in head and neck cancer: An enigmatic prognostic marker. *International Journal of Radiation Oncology Biology Physics*. 2001; 49:435–41.
7. Chung CH, Zhang Q, Hammond EM, et al. Integrating Epidermal Growth Factor Receptor Assay with Clinical Parameters Improves Risk Classification for Relapse and Survival in Head-and-Neck Squamous Cell Carcinoma. *International Journal of Radiation Oncology Biology Physics*. 2011; 81:331–8.
8. Reimers N, Kasper HU, Weissenborn SJ, et al. Combined analysis of HPV-DNA, p16 and EGFR expression to predict prognosis in oropharyngeal cancer. *International Journal of Cancer*. 2007; 120:1731–8.
9. Mineta H, Miura K, Suzuki I, et al. Low p27 expression correlates with poor prognosis for patients with oral tongue squamous cell carcinoma. *Cancer*. 1999; 85:1011–7. [PubMed: 10091782]
10. Kapranos N, Stathopoulos GP, Manolopoulos L, et al. p53, p21 and p27 protein expression in head and neck cancer and their prognostic value. *Anticancer Research*. 2001; 21:521–8. [PubMed: 11299798]
11. Chau CH, Rixe O, McLeod H, Figg WD. Validation of analytic methods for biomarkers used in drug development. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2008; 14:5967–76. [PubMed: 18829475]

12. Lee JW, Devanarayan V, Barrett YC, et al. Fit-for-purpose method development and validation for successful biomarker measurement. *Pharmaceutical research*. 2006; 23:312–28. [PubMed: 16397743]
13. Lee JW, Weiner RS, Sailstad JM, et al. Method validation and measurement of biomarkers in nonclinical and clinical samples in drug development: a conference report. *Pharmaceutical research*. 2005; 22:499–511. [PubMed: 15846456]
14. Valentin MA, Ma S, Zhao A, Legay F, Avrameas A. Validation of immunoassay for protein biomarkers: bioanalytical study plan implementation to support pre-clinical and clinical studies. *Journal of pharmaceutical and biomedical analysis*. 2011; 55:869–77. [PubMed: 21530130]
15. Williams PM, Lively TG, Jessup JM, Conley BA. Bridging the gap: moving predictive and prognostic assays from research to clinical use. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2012; 18:1531–9. [PubMed: 22422405]
16. Bernier J, Cooper JS, Pajak TF, van Glabbeke M, Bourhis J, Forastiere A. Defining risk levels in locally advanced head and neck cancers: A comparative analysis of concurrent postoperative radiation plus chemotherapy trials of the EORTC (#22931) and RTOG (#9501). *Head & neck*. 2005; 27:843–50. [PubMed: 16161069]
17. Belbin TJ, Singh B, Barber I, et al. Molecular classification of head and neck squamous cell carcinoma using cDNA microarrays. *Cancer Research*. 2002; 62:1184–90. [PubMed: 11861402]
18. Ahmad A, Stefani S. Distant Metastases of Nasopharyngeal Carcinoma - a Study of 256 Male-Patients. *Journal of Surgical Oncology*. 1986; 33:194–7. [PubMed: 3773537]
19. Geara FB, Sanguineti G, Tucker SL, et al. Carcinoma of the nasopharynx treated by radiotherapy alone: Determinants of distant metastasis and survival. *Radiotherapy and Oncology*. 1997; 43:53–61. [PubMed: 9165137]
20. Lee AWM, Law SCK, Foo W, et al. Retrospective Analysis of Patients with Nasopharyngeal Carcinoma Treated during 1976–1985 - Survival after Local Recurrence. *International Journal of Radiation Oncology Biology Physics*. 1993; 26:773–82.
21. Basu A, Haldar S. The relationship between Bcl2, Bax and p53: consequences for cell cycle progression and cell death. *Molecular Human Reproduction*. 1998; 4:1099–109. [PubMed: 9872359]
22. Subramanian J, Simon R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *Journal of the National Cancer Institute*. 2010; 102:464–74. [PubMed: 20233996]
23. Bhattacharya A, Roy R, Snijders AM, et al. Two distinct routes to oral cancer differing in genome instability and risk for cervical node metastasis. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2012; 17:7024–34. [PubMed: 22068658]
24. Morita Y, Hata K, Nakanishi M, Nishisho T, Yura Y, Yoneda T. Cyclooxygenase-2 promotes tumor lymphangiogenesis and lymph node metastasis in oral squamous cell carcinoma. *Int J Oncol*. 2012; 41:885–92. [PubMed: 22735610]
25. Wang C, Liu X, Chen Z, et al. Polycomb group protein EZH2-mediated E-cadherin repression promotes metastasis of oral tongue squamous cell carcinoma. *Mol Carcinog*. 2012
26. Jalisi S. Management of the clinically negative neck in early squamous cell carcinoma of the oral cavity. *Otolaryngol Clin North Am*. 2005; 38:37–46. viii. [PubMed: 15649497]
27. Cheng A, Schmidt BL. Management of the N0 neck in oral squamous cell carcinoma. *Oral and maxillofacial surgery clinics of North America*. 2008; 20:477–97. [PubMed: 18603204]
28. Weiss MH, Harrison LB, Isaacs RS. Use of decision analysis in planning a management strategy for the stage N0 neck. *Arch Otolaryngol Head Neck Surg*. 1994; 120:699–702. [PubMed: 8018319]
29. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *Journal of the National Cancer Institute*. 2009; 101:1446–52. [PubMed: 19815849]
30. Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. *Journal of the National Cancer Institute*. 2010; 102:152–60. [PubMed: 20075367]
31. Ang KK, Harris J, Wheeler R, et al. Human Papillomavirus and Survival of Patients with Oropharyngeal Cancer. *New England Journal of Medicine*. 2010; 363:24–35. [PubMed: 20530316]

32. Gillison ML, Harris J, Westra W, et al. Survival outcomes by tumor human papillomavirus (HPV) status in stage III–IV oropharyngeal cancer (OPC) in RTOG 0129. *Journal of Clinical Oncology*. 2009;27.
33. Rischin D, Young RJ, Fisher R, et al. Prognostic Significance of p16(INK4A) and Human Papillomavirus in Patients With Oropharyngeal Cancer Treated on TROG 02. 02 Phase III Trial. *Journal of Clinical Oncology*. 2010; 28:4142–8. [PubMed: 20697079]

Table 1

Criteria for designing molecular marker studies to demonstrate clinical utility.

Identify the intended use for the biomarker.
Precisely define the target patient population for that intended use.
Use specimens representing the tumors that will be sampled in the clinical setting relevant for the intended use of the biomarker.
Account for clinical, pathological, and molecular confounders.
Compare the marker's performance to that of other biomarkers and clinicopathologic factors that are currently in use to guide treatment.
Include in the study patients from appropriate control groups in order to distinguish predictive versus prognostic role of a marker.