Statistical consideration for clinical biomarker research in bladder cancer

Shahrokh F. Shariat[1,2], **Yair Lotan**[3], **Andrew Vickers**[1], **Pierre I. Karakiewicz**[4], **Bernd J. Schmitz-Dräger**[5], **Peter J. Goebell**[6], and **Nuria Malats**[7]

[1]Department of Surgery (Urology), Memorial Sloan-Kettering Cancer Center, New York, NY, USA

[2]Department of Urology, New York Presbyterian Hospital, Weill Medical College of Cornell University, New York, NY, USA

[3]Department of Urology, UT Southwestern Medical Center, Dallas, TX, USA

[4]University of Montreal, Montreal, Quebec, Canada

[5]Department of Urology, University of Erlangen, Erlangen, Germany

[6]Department of Urology, Euromed Clinic, Fürth, Germany

[7]Spanish National Cancer Research Centre, Madrid, Spain

Abstract

Purpose—To critically review and illustrate current methodologic and statistical considerations for bladder cancer biomarker discovery and evaluation.

Methods—Original, review, and methodological articles, and editorials were reviewed and summarized.

Results—Biomarkers may be useful at multiple stages of bladder cancer management: early detection, diagnosis, staging, prognosis, and treatment; however, few novel biomarkers are currently used in clinical practice. The reasons for this disjunction are manifold and reflect the long and difficult pathway from candidate biomarker discovery to clinical assay, and the lack of coherent and comprehensive processes (pipelines) for biomarker development. Conceptually, the development of new biomarkers should be a process that is similar to therapeutic drug evaluation - a highly regulated process with carefully regulated phases from discovery to human applications. In a further effort to address the pervasive problem of inadequacies in the design, analysis, and reporting of biomarker prognostic studies, a set of reporting recommendations are discussed. For example, biomarkers should provide unique information that adds to known clinical and pathologic information. Conventional multivariable analyses are not sufficient to demonstrate improved prediction of outcomes. Predictive models, including or excluding any new putative biomarker, needs to show clinically significant improvement of performance in order to claim any real benefit. Towards this end, proper model building, avoidance of overfitting, and external validation are crucial. In addition, it is important to choose appropriate performance measures dependent on outcome and prediction type and to avoid use of cut-points. Biomarkers providing a continuous score provide potentially more useful information than cut-points since risk fits a continuum model. Combination of complementary and independent biomarkers is likely to better capture the biologic potential of a tumor than any single biomarker. Finally, methods that

Correspondence and reprint requests: Shahrokh F. Shariat, M.D., Ph.D. Division of Urology/Department of Surgery Memorial Sloan-Kettering Cancer Center 1275 York Avenue, Box 27 New York, NY 10065 sfshariat@gmail.com Telephone: 469 363 8500.

incorporate clinical consequences such as decision curve analysis are crucial to the evaluation of biomarkers.

Conclusions—Attention to sound design and statistical practice should be delivered as early as possible and will help maximize the promise of biomarkers for patient care. Studies should include a measure of predictive accuracy and clinical decision-analysis. External validation using data from an independent cohort provides the strongest evidence that a model is valid. There is a need for adequately assessed clinical biomarkers in bladder cancer.

Keywords

biomarker; diagnosis; prognosis; treatment; nomogram; decision-analysis; bladder cancer; statistics; statistical analysis

> While the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any one man will be up to, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician.

~Arthur Conan Doyle

Introduction

Biomarker research is an all encompassing term for investigation of biologic processes and their potential use in early detection, diagnosis, monitoring of disease and treatment decision. Most biomarkers represent changes in proteins or genes associated with disease that can be measured. Since there are innumerable alterations associated with disease states, it is necessary to develop statistical tools to identify which changes are significant and which can be clinically useful.

The recent mapping of the human genome together with advances in high-throughput genetic and proteomic technology have led to an explosion of biologic information and the identification of a plethora of candidate molecular biomarkers and therapeutic targets. Furthermore, work in proteomics is also improving our understanding of systems biology. As the proteome changes constantly with the state of the organism, biological variations that occur over time can easily be addressed. Proteomics therefore has the potential expected to discover new oncologic biomarkers that could help diagnose cancer at an early stage or establish tumor specific profiles that could predict tumor aggressiveness[1]. High-throughput platforms for analysis of protein expression and functionality in clinical samples and protein microarrays have been developed. These array platforms provide a quantitative or semi-quantitative means for measuring protein expression and also provide the ability to identify post-translational modifications, such as phosphorylation. Their increasing use in health research will significantly help the rapid validation and therefore translation of the massive amounts of data generated by modern genomic and proteomic technologies, thereby fostering a new synergy between bench and bedside. This explosion of knowledge about the basic biological processes and the genetics of cancer has led to increasing optimism that this knowledge can be put to practical clinical use in the near future for individualized medicine[2,3]. Indeed, important examples of translational approaches can already be found in the areas of drug discovery and development, disease diagnosis and classification, selection of therapeutic regimens for individual patients, and designing clinical trials. These are important developments but, as with any new approach, there is a danger of unwarranted enthusiasm and premature clinical application of laboratory results based on insufficient evidence. To carry out the translation of knowledge into practice with maximal efficiency

and effectiveness, it is essential to conduct studies with appropriate designs and analyses based on sound statistical principles.

There is considerable evidence that contemporary biomarker research falls far short of using appropriate research design and statistical methods. For example, Vickers et al. reviewed a `snapshot' sample of 129 studies to determine whether the statistical analyses used would allow conclusions to be drawn regarding the clinical value of the biomarkers studied. The authors found that the majority of articles regarding biomarkers for cancer focused on testing the null hypothesis of no association between the biomarker and cancer outcome with only a small minority (11%) reporting predictive accuracy. Not a single paper used a statistical method or study design that addressed whether use of the marker in practice would improve clinical outcome. Other authors have reported on different statistical problems in marker research, such as the correct application of statistical tests, multiple hypothesis testing, data-dependent choice of cut-points, power, and missing data.[4–6] Finally, deficiencies in study design, representative at-risk population, definition of endpoints, standard prognostic factor data, or laboratory methods may also limit the conclusions of a study.[7] Given these pervasive problems with study design and data analysis and reporting, this article attempts to address some of the statistical considerations needed for appropriate biomarker discovery and evaluation with focus on bladder cancer.

What is exactly a biomarker? Refining the definition

According to the NIH, a biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmaceutical responses to a therapeutic intervention[8,9]. The term biomarker is now typically shorthand for a molecular biomarker. There is a wide range of variation in the complexity of biomarkers going from the simplest (hair color, blood pressure or cholesterol levels) to more complicated examples (mRNA profiles of tumors and more recently combinations of proteins). Cancer biomarkers are either produced by the tumor or by the body in response to the tumor. Six different types of biomarker can be differentiated in bladder cancer (table 1):

1) Early detection: this biomarker is used for evaluating patients with either risk factors for or symptoms of cancer. Urine is an ideal source specimen for bladder cancer screening and early detection as it can be obtained in a non-invasive, easy and relatively cheap manner.[10] Several such biomarkers that have been proposed and are currently used such as nuclear matrix protein 22 and UroVysion.[11–14] Despite all the biomarkers that may be applied as screening tools, bladder cancer screening may never be cost-effective.[10]

2) Diagnostic: this biomarker can help classical histo-pathologic characteristics in assessing presence or absence of cancer. Uroplakin[15] is such a biomarker for bladder cancer.

3) Prognostic: this biomarker is used to predict the outcome of patients based on different risk of recurrence, progression, or death, thereby allowing individualized/tailored management. This type of biomarker provides information about the likely clinical course of a disease hence, guiding therapeutic decisions[16]. Classical prognostic markers are hormone receptors in breast cancer and PSA in prostate cancer[17]. Regarding bladder cancer, a panel of tissue biomarkers was recently found to be associated with a worse prognosis in patients with localized bladder cancer[18–20] and a current phase III adjuvant clinical trial of systemic chemotherapy in pT1-3N0 patients with high-risk biomarker status is ongoing.

4) <u>Predictive</u>: this biomarker is used to predict whether the treatment (drug or other therapy) will be effective, and/or to monitor the effectiveness of the treatment. It can help identify the best treatment modality. In bladder cancer, to our knowledge, there is no predictive biomarker in use or in clinical trial.

5) <u>Therapeutic target</u>: this biomarker can help identify the patients who will benefit from a particular treatment regimen. It identifies the molecular targets of novel therapies and is affected by therapy. For example, breast tumors expressing estrogen receptors are best treated with an antiestrogen such as tamoxifen or an aromatase inhibitor[21]. In bladder cancer, to our knowledge, there is no therapeutic target biomarker in use or in clinical trial.

6) <u>Surrogate endpoint</u>: this biomarker is used to substitute for a clinical endpoint and/or to measure clinical benefit, harm or lack of benefit or harm. Surrogates could replace traditional endpoints, such as mortality due to disease or the recurrence or relapse of disease. Biomarkers can reduce time factors and costs for Phase I and II clinical trials by replacing clinical endpoints. An example for a potential surrogate endpoint is post-therapy biomarker level to evaluate response to drugs in a clinical trial setting or in clinical practice. In bladder cancer, to our knowledge, there is no surrogate endpoint biomarker in use or in clinical trial.

For the purposes of the current review: a biomarker is the end result of a bioassay, for processing biological material from humans, expressed quantitatively or categorically. Imaging results could also be included as a biomarker if we interpret "bioassay" broadly enough. Biomarker measurement can be either binary (e.g., positive or negative), categorical (e.g., high, medium, low), quantitative (e.g., the level of some urine/serum protein), or multidimensional (e.g., a genomic signature or "metagene").

Biomarker conundrum in bladder cancer

A PubMed Search on "bladder cancer" AND ("biomarker" OR "molecular marker") in English language yielded 3434 hits (accessed 12/25/09; figure 1). The number of articles published on bladder cancer biomarkers have increased steadily since 1980s reaching close to 300 articles per year in 2009. Despite this plethora of biomarkers reported to be clinically "promising", only one biomarker is routinely used by urologists- cytology.[13] Why are bladder cancer biomarkers not living up to their promise?[22] For one, there are remarkable analytical and regulatory barriers to the application of biomarkers in cancer care (table 2).[23] These include but are not limited to the status of intellectual property protection, availability of standard reference materials for the assay, complexity of assay format, implementation of quality control to assure reproducibility and accuracy, sufficient market testing size to assess methods of commercialization, lack of clear guidelines for good manufacturing/laboratory practice and quality control requirements for all phases of biomarker development, cost and effort required to accumulate clinical data under appropriately designed, Institutional Review Board-approved prospective trials, and the interval required for resolution of patent issues, assay standardization, validation, testing, and regulatory approval.

Besides analytical and regulatory barriers, the lack of bladder cancer biomarker use in daily clinical practice is also a result of poor application of statistics and study design. Indeed, typical biomarker studies have involved convenience samples from poorly defined populations, non-standardized assays, and small numbers of patients subject to missing data. Biomarker research is usually done in context of usual clinical care, not clinical trials and has been largely guided by intuition and experience rather that well-structured analyses. This has led to the fact that most biomarkers findings are not reproducible.[24–26] Indeed, most biomarkers that appear biomedically and statistically significant at one center are not

confirmed by other centers.[24–29] These observations have prompted the development of guidelines (discussed below) intended to ensure that biomarker studies conform to some basic standards of design and reporting.

The discovery-validation-implementation paradigm: a schema for biomarker development and evaluation

Conceptually, the development of new biomarkers should be a process that is similar to therapeutic drug evaluation. Drug development is a highly regulated process with carefully regulated phases from discovery to human applications. In 2002, the National Cancer Institute's Early Detection Research Network developed a five phase approach to systematic discovery and evaluation of biomarkers.[30,31] The phases of research are generally ordered according to the strength of evidence that each provides in favor of the biomarker. The results from earlier phases are generally necessary to design later phases. Similar to other research groups[32] and previous efforts of this group[33] we have modified these phases to classify studies into a sequence of phases, from discovery to validation and assessment of benefit (figure 2):[23,34]

- *Preclinical testing*: biomarkers are developed *in vitro* or in animal models. This phase is essentially a hypothesis-generating step.

- *Phase 0 (Assay development):* development of preliminary assays on patient samples to assess test reproducibility and robustness. It may also be during this phase that the adequate source for testing (i.e. urine, serum, pre-treated extracts from tissue,..) is evaluated and the prevalence of the alteration to be used as determinant is present and relevant. This phase does not take into account the potential benefit of the biomarker.

- *Phase I (Feasibility and clinical prevalence):* the biomarker is tested on a small group of patients (retrospective, small studies) to determine its ability to discriminate between health or disease (or similar appropriate question). This stage includes discovery and marker optimization processes: define the marker, and determine the assay cut-off points in a well-defined population. At this point, the assay may still require refinement.

- *Phase II (Validation and standardization for clinical utility):* independent validation of the accuracy of the assay. The performance of the presumed biomarker is externally validated in larger sample of independent cases. The reproducibility and robustness of the assay are appraised and reference ranges are determined. The assay should demonstrate an adequate dynamic range and show that the biomarker has a relationship to drug exposure and/or the grade or stage of the biological process. Once proven to be analytically robust, standards for external use may be determined and the assay must then be measured in a relevant, representative target population to determine normal or background variation between individuals and within individuals over time. The thorough analyses of the outcome measures of this phase are a prerequisite for either entering the next phase or modifying the assay (i.e. including specific preparatory measures to ensure the robustness).

- *Phase III (Independent confirmation studies)*: the efficacy is determined in large patient populations other than the discovery population (typically multi-institutional collaboration). The objective is to determine the sensitivity (ability to detect true positives) and specificity (ability to avoid false positives) of the biomarker. Ideally this phase would consist of randomized trial to show that the use of assay results generates better clinical-decision making than current standards.

Discrimination, calibration, and decision analysis (discussed below) play a major role in this phase.

- *Phase IV (Impact assessment):* post-approval reporting and testing for other disease processes or disease stages. Ideally, once an assay has passed all previous phases and entered the clinical practice, registries of large scale continue to collect outcome measures. It is not rare that in this phase the interest has dropped and lack of additional data hampers the discovery of misleading assumptions for decades.

This schema is not only an intellectual process but also provides a clear scale by which researchers, patients, reviewers and investors can evaluate the status of the biomarker in the development process. The expected failure rate of biomarkers in development can be expected to be similar to the one of drugs. Biomarker development must not be considered to be any easier than drug development. It is crucial that researchers be aware of the complexity and poor success rate of potential biomarkers trying to enter the clinical arena.

Reporting of biomarker data

In an effort to address the pervasive problem of inadequacies in the design, analysis, and reporting of biomarker prognostic studies, a set of reporting recommendations such as the REMARK has been developed and adopted by many prominent journals (figure 3). The goal of these guidelines is to encourage transparent and complete reporting and to help readers judge the data and understand the context in which the conclusions apply. Indeed, it provides a detailed description as to the minimum amount of information that should be given in the reporting of results from biomarker studies. The REMARK lists 20 items that investigators should attempt to report in any biomarker study.

Another tool aimed at standardizing the quality of biomarker research is the Tumor Marker Utility Grading System (figure 4). The TMUGS is a scale of levels of evidence, designed to help place biomarker studies into a context of validity. Overall this framework and suggestions by others[35,36] have led to an increasing understanding that biomarker studies should be designed, conducted, and analyzed and reported with the same rigor as classic laboratory and clinical therapeutic studies.

Statistical consideration during early phases of biomarker discovery and validation

Most people use statistics like a drunk uses a streetlight, for support instead of illumination.

Andrew Lang

An issue that has received less attention is the degree to which research on biomarkers has made sufficient use of clinically relevant statistics, such as the assessment of predictive accuracy, decision analysis, and/or experimental methodology. The fundamental idea behind the concept of personalized medicine is that it is possible to identify patterns of demographic, clinical, genomic/proteomic, and other types of biological data that can be used together to benefit individual patients. Most biomarkers do not provide sufficient information to be used independent of other information. The optimal use of biomarkers lies in incorporating it in a model that includes standard clinical data as well as relevant additional data with regards assay performance (i.e. medication or nutrition that may interfere with the assay).[17,34,37–39] The model would then be used to provide individual patient care for diagnosis or treatment.

There is vast statistical literature on how to build predictive models. Multivariable models allowing simultaneous association of biomarkers and predictors with clinical outcome, such as logistic regression for presence/absence of the event of interest or Cox regression with survival, are common building blocks of biomarker-based risk prediction tools. More algorithmic statistical models, such as recursive partitioning and support vector machines or neural networks, can also be used.

To determine the value of a new biomarker, it is not sufficient to show that it is significantly related to the outcome, statistically significant in a multivariable model including the standard clinical and pathologic factors, or more significant than the standard clinical and pathologic factors. A variable that is statistically significant in a multivariable model might not substantively improve the model's predictive accuracy. P-values and odds or hazard ratios do not meaningfully describe a biomarkers' ability to classify patients. For a biomarker to be potentially clinically useful, it is necessary to show that adding the biomarker to an existing model based on the most important clinical and pathologic factors substantively improves the predictive accuracy (discrimination and calibration) of the model. In other words, we are asking whether the biomarker adds information to the current risk group classification.

Given that biomarkers will often be integrated with clinical variables into statistical models, a central concern of biomarker development is that of model development. The key issue for statistical models is that of appropriate validation. Models are generated using a particular data set, and there are numerous statistical reasons why models generally make better predictions for this "training" data set than for a new patient to whom the model will be applied to. There are two methods of correcting for this effect: internal and external validation. Internal validation techniques are based on a single data set but use various techniques to ensure that models are tested and evaluated on different patients. In cross-validation, for example, the study cohort is split into ten subgroups. A model is then built on the first nine groups and used to predict outcome for the tenth group. This process is repeated sequentially 10 times, so that a prediction is made for every patient in the data set based on a model developed excluding that patient. The predictions are then compared with true outcomes to give measures of discrimination and calibration.[40] Another method of internal validation involves bootstrap techniques.[40–45]

Although internal validation circumvents problems of statistical overfit, a model may fail to make good predictions on a new cohort because the patients in that cohort are unlike those in the original training data set. For example, a model developed on bladder cancer patients undergoing cystoscopy for surveillance of non-muscle invasive bladder cancer at a tertiary care center may not be appropriate for patients presenting in primary care with hematuria. Hence it is important that models be externally validated on a cohort entirely separate to that of the training set. Validation on heterogeneous external data sets allows for evaluation of the generalizability of the risk prediction tool to wider populations than originally reported.

An example of this process is shown in the research on the prognostic value of Ki-67, an established marker of cell proliferation, in bladder cancer. Several groups have shown the prognostic value of Ki-67 in bladder cancer. For example, investigators at UT Southwestern, Dallas, Texas, have shown in a cohort of 226 consecutive patients treated with radical cystectomy at a single academic center, that high Ki-67 overexpression was independently associated with both disease recurrence and cancer-specific mortality using traditional multivariable Cox regression analyses that adjusted for the effects of standard clinicopathologic features.[46] The group than externally validated these findings in a large external multi-institutional cohort of 713 patients treated with radical cystectomy and tested whether Ki-67 expression improves the accuracy of predictive models that include standard

clinico-pathologic features for prediction of disease recurrence and cancer-specific survival.[47] They first calculated the predictive accuracy of the standard multivariable model, which included tumor stage, grade, nodal status, and lymphovascular invasion. They then repeated their analysis after including Ki-67 in the model along with the other predictors. The addition of Ki-67 improved accuracy by 2.9% for disease recurrence and 2.4% for bladder cancer–specific death. The new model was well calibrated. They concluded that assessment of Ki-67 expression provides the clinician with information not captured by established predictors. However, it is unclear whether this increase in accuracy and risk assessment translates into improved individualized, evidence-based recommendations with eventual superior patient outcomes. Indeed, an improvement in the predictive accuracy, although necessary, is not sufficient to assess whether using the marker in practice would actually benefit patients.

Establishing clinical relevance of a biomarker test for guiding therapy decisions requires demonstrating that it can classify patients into distinct subgroups with different recommended management. An experimental assessment, such as the randomization of patients to clinical management that does or does not incorporate information from the biomarker, provides the most reliable data on the clinical value of a new biomarker. The different designs for randomized clinical trials of biomarker-guided therapy are discussed in another article (i.e., *Considerations in Implementing Diagnostic Markers into Decision Making in Bladder Cancer*) in this issue. In keeping with the example on Ki-67, as part of a phased, systematic evaluation and validation of biomarkers, the group at UT Southwestern designed such a trial that randomizes patients with pT1-3N0M0 after radical cystectomy to adjuvant chemotherapy based the expression of a panel of five biomarkers (cyclin E1, p53, p21, p27 and Ki-67).[48] The substantial human and financial resources necessary for randomized trials make such an evaluation impractical for most biomarkers. Decision analytic techniques can often be used instead of trials. The key point of decision analysis is that the consequences of clinical decisions are incorporated in analyses. Decision analysis allows one to weight the relative value of the benefits (true positives) to the harms (false positives) and thus incorporate the consequences of a clinical decision. There exist very simple decision analytic tools, requiring only basic math, which incorporate considerations such as it being more harmful to delay a diagnosis of cancer than to cystoscope a patient unnecessarily. One such tool is the decision curve analysis which is based on the principle that the relative harms of false positives (e.g., unnecessary cystoscopy) and false negatives (e.g. missed cancer) can be expressed in terms of a probability threshold.[49–52] For example, if a patient would opt for cystoscopy if he was told that his risk of bladder cancer was 10% or more, but not if his risk was less than 10%, it can be shown that he considers that harms associated with a delayed diagnosis of cancer to be nine times greater than the harms associated with an unnecessary biopsy (i.e. the ratio of harms is the odds at the probability threshold). This threshold probability can therefore be used both to determine whether an individual patient's test result should be defined as positive or negative and to weight the clinical consequences of true and false. These tools provide a clear indication whether a biomarker is worth using at all and whether an additional predictor is worth measuring. The decision analytic evaluation should be performed during later stages of research and before clinical implementation of the tool. Wider adoption of decision analytic methods will provide better insights as to whether any of the plethora of new biomarkers improve health.

Cut-points for biomarker classification

Many biomarkers are measured on a continuous scale such as for example nuclear matrix protein 22 (NMP22), which varies from 0.1 ng/mL to over 100 ng/mL. A common inclination with continuous markers is to chose a cut-point to distinguish positive from negative results, such as with 10 ng/mL for NMP22.[53]

There are numerous reasons to be suspicious of this practice. First, use of cut-points makes no biologic sense, on the grounds that it is implausible that there exists a fixed threshold of a marker at which risk abruptly changes. There is no reason to believe, for example, that a man's risk of bladder cancer does not change between 8.0 and 9.9 ng/mL, but then increases dramatically from 9.9 to 10.0 ng/mL.[11] As a corollary, dichotomizing markers before adding them to statistical models generally leads to lower accuracy than keeping them as continuous. Second, choice of cut-points can add statistical noise. It is quite possible that a biomarker that has no association with an outcome can be found to have a p value <0.05 if a large number of cut-points are investigated. Third, cut-points are clinically problematic. Again using the example of cystoscopy, let us imagine that a patient had an elevated NMP22, but was poorly tolerant of invasive procedures and wanted to know whether he really had to go through with the cystoscopy. A urologist might be willing to let things go if NMP22 was 10 ng/mL but not if NMP22 was 22 ng/mL. Yet how much higher than 10 ng/mL would be "high enough" to recommend cystoscopy in such a patient?

This last consideration suggests the natural alternative to cut-points, which is the use of predicted probabilities. Biomarker levels, with or without other clinical data, can be converted to probabilities by straightforward statistical methods. These probabilities can then be used in decision making, with patient or physicians choosing their own cut-points depending on personal preferences. Take the example of a marker for bladder cancer recurrence after cystectomy. An older patient with a low tolerance for drug toxicities may require a higher risk of recurrence to justify adjuvant therapy than a younger patient raising a family.

Predictive tools

Traditionally, physician judgment has formed the basis for risk estimation, patient counseling, and decision making. However, clinicians' estimates are often biased due to both subjective and objective confounders[54–57]. To obviate this problem and to obtain more accurate predictions, researchers have developed predictive tools that are based on statistical techniques[58]. Recently, predictive tools have been introduced to predict the outcome of interest for the individual patient. Predictive tools have been shown to perform better than clinical judgment when predicting probabilities of prostate cancer outcome[57,59,60]. Whether this also applies to bladder cancer remains to be studied. That said, physician input is obviously essential in medical decision-making, both for the measurement of variables that are used in the prediction process and for the interpretation and application of tool-derived outcome predictions in clinical practice.

There are a number of types of prediction tools such as Kattan-type nomograms[11,61–66], risk groupings, artificial neural networks (ANNs), probability tables, classification and regression tree (CART) analyses, probability formulas, look-up and propensity scoring tables, and risk-class stratification prediction tools reviewed in[65]). Prediction tools can be compared based on several criteria: discrimination, calibration (correlation between predicted and observed risk throughout the entire range of predictions), generalizability, level of complexity, adjustment for the effect of competing risks, use of conditional probabilities, and application of decision-analysis. The most important of these considerations are discrimination, calibration, and decision-analysis.[4]

A good predictive tool is able to discriminate between patients with or without the outcome of interest.[64] Discrimination is quantified using the area under the curve (AUC) for binary outcomes (e.g., presence or absence of cancer), the c-index for censored data (e.g. recurrence after radical cystectomy), and the Brier score[67]. For both AUC and c-index, 0.5

represents no discriminating ability (coin flip), whereas a value of 1.0 represents perfect discrimination.

While discrimination quantifies the ability of a prediction tool to distinguish between patients, calibration quantifies the accuracy of a prediction for an individual patient. Calibration plots graphically illustrate the relationship between predicted and observed rates of the outcome of interest. Ideally, a well calibrated prediction tool exhibits a 1:1 relationship between predicted and observed rates, resulting in 45 degree slope.

Methods that incorporate clinical consequences are crucial to the evaluation of biomarkers. This type of analysis allows insight into the consequences of using a biomarker in the clinic.[68,69] Available methods include decision curve analysis which is a method that combines simplicity with efficient computations.[49–52] In brief, the method is based on the principle that the relative harms of false positives (e.g. unnecessary cystoscopy) and false negatives (e.g. missed cancer) can be expressed in terms of a probability threshold. This threshold probability can be used both to determine whether an individual patient's test result should be defined as positive or negative and to weight the clinical consequences of true and false. Recently, Vickers et al.[70] used decision-analytical methods to prove that using a nomogram[71] to determine referral to adjuvant chemotherapy would lead to better decisions than using stage and nodal status. The decision analytic evaluation should be performed during later stages of research before clinical implementation of the biomarker.

Perspectives

Statistics can be made to prove anything - even the truth.

~Author Unknown

Biomarkers have the potential to be used clinically to screen for, diagnose or monitor the activity of diseases and to guide molecular targeted therapy or assess therapeutic response (table 1). However, discovery experiments have often overemphasized the significance of novel biomarkers, and efforts to further credential such candidates have been rare. Open any journal today and you will find multiple articles on the "next" bladder cancer biomarker. However, very few biomarkers have reached commercial availability and with the exception of cytology none has been widely utilized. This has been due in large part to a lack of effective and efficient strategies to determine which biomarker candidates justify the great investment of time and money required for assay development, optimization and demonstration of analytical robustness. In brief, many of biomarkers lack clinical utility. Demonstration of clinical utility and compliance with regulatory requirements remain formidable, uncertain and costly steps toward the commercialization of novel biomarkers. For a new biomarker to be clinically useful, it has to answer a clinically relevant question and provide information that is not available in a more simple and cost-effective way. Any new biomarker needs to provide a benefit over standard criteria or at least improve their accuracy. Before a biomarker assay can be implemented in the community setting, it needs to address four concepts: "*better, easier, faster and cheaper*".[23] In this review we have discussed the statistical tools needed for "better". Too often, the newly described biomarker does not have any repercussion in clinical practice despite scientific value and/or validity. For example, many biomarkers are correlated with tumor stage or grade such that measuring the marker does not provide physicians with information that they do not already have. While an association between a marker and tumor stage may help understand molecular pathways of the disease, it does not influence clinical management. A large concerted effort is required to advance the field of biomarker through systematic discovery, verification, and validation- each step coupled with adequate statistical analysis. Biomarker discovery and development have to shift toward a more organized and industrialized setting similar to that

of drug development framework. When these changes occur, we can expect improvements in both screening and development of biomarker, although we should stay aware that very few molecules will make it to the routine clinical practice. In fact, it is most unlikely that a single biomarker will have the single decision as to a diagnosis and/or a prognosis of a particular pathology. It may rather be that a constellation of biomarkers will have more predictive power.[17–19] There is no doubt that progress will continue based on the collaboration of basic researchers, clinicians and biomedical firms.

References

1. Egawa S, Kuruma H. Search for biomarkers of aggressiveness in bladder cancer. Eur Urol. 2006; 50(1):20–2. [PubMed: 16530926]

2. Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, Dowsett M, Fitzgibbons PL, Hanna WM, Langer A, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. J Clin Oncol. 2007; 25(1):118–45. [PubMed: 17159189]

3. van Gils MP, Stenman UH, Schalken JA, Schroder FH, Luider TM, Lilja H, Bjartell A, Hamdy FC, Pettersson KS, Bischoff R, et al. Innovations in serum and urine markers in prostate cancer current European research in the P-Mark project. Eur Urol. 2005; 48(6):1031–41. [PubMed: 16054748]

4. Shariat SF, Kattan MW, Vickers AJ, Karakiewicz PI, Scardino PT. Critical review of prostate cancer predictive tools. Future Oncol. 2009; 5(10):1555–84. [PubMed: 20001796]

5. Vickers AJ, Lilja H. Cutpoints in clinical chemistry: time for fundamental reassessment. Clin Chem. 2009; 55(1):15–7. [PubMed: 19028819]

6. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 21(1):128–38. [PubMed: 20010215]

7. Schmitz-Drager BJ, Fradet Y, Grossman HB. Bladder cancer markers in patient management: the current perspective. World J Urol. 2008; 26(1):1–3. [PubMed: 18060549]

8. Ilyin SE, Belkowski SM, Plata-Salaman CR. Biomarker discovery and validation: technologies and integrative approaches. Trends Biotechnol. 2004; 22(8):411–6. [PubMed: 15283986]

9. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001; 69(3):89–95. [PubMed: 11240971]

10. Svatek RS, Lotan Y, Karakiewizc PI, Shariat SF. Screening for bladder cancer using urine-based tumor markers. Minerva Urol Nefrol. 2008; 60(4):247–53. [PubMed: 18923361]

11. Shariat SF, Zippe C, Ludecke G, Boman H, Sanchez-Carbayo M, Casella R, Mian C, Friedrich MG, Eissa S, Akaza H, et al. Nomograms including nuclear matrix protein 22 for prediction of disease recurrence and progression in patients with Ta, T1 or CIS transitional cell carcinoma of the bladder. J Urol. 2005; 173(5):1518–25. [PubMed: 15821471]

12. Grossman HB, Messing E, Soloway M, Tomera K, Katz G, Berger Y, Shen Y. Detection of bladder cancer using a point-of-care proteomic assay. JAMA. 2005; 293(7):810–6. [PubMed: 15713770]

13. Herman MP, Svatek RS, Lotan Y, Karakiewizc PI, Shariat SF. Urine-based biomarkers for the early detection and surveillance of non-muscle invasive bladder cancer. Minerva Urol Nefrol. 2008; 60(4):217–35. [PubMed: 18923359]

14. Goebell PJ, Groshen S, Schmitz-Drager BJ, Sylvester R, Kogevinas M, Malats N, Sauter G, Barton Grossman H, Waldman F, Cote RJ. The International Bladder Cancer Bank: proposal for a new study concept. Urol Oncol. 2004; 22(4):277–84. [PubMed: 15283883]

15. Huang HY, Shariat SF, Sun TT, Lepor H, Shapiro E, Hsieh JT, Ashfaq R, Lotan Y, Wu XR. Persistent uroplakin expression in advanced urothelial carcinomas: implications in urothelial tumor progression and clinical outcome. Hum Pathol. 2007; 38(11):1703–13. [PubMed: 17707461]

16. Schnitt SJ. Traditional and newer pathologic factors. J Natl Cancer Inst Monogr. 2001; 30:22–6. [PubMed: 11773288]

17. Shariat SF, Karam JA, Walz J, Roehrborn CG, Montorsi F, Margulis V, Saad F, Slawin KM, Karakiewicz PI. Improved prediction of disease relapse after radical prostatectomy through a panel

of preoperative blood-based biomarkers. Clin Cancer Res. 2008; 14(12):3785–91. [PubMed: 18559597]

18. Shariat SF, Tokunaga H, Zhou J, Kim J, Ayala GE, Benedict WF, Lerner SP. p53, p21, pRB, and p16 expression predict clinical outcome in cystectomy with bladder cancer. J Clin Oncol. 2004; 22(6):1014–24. [PubMed: 14981102]

19. Shariat SF, Karakiewicz PI, Ashfaq R, Lerner SP, Palapattu GS, Cote RJ, Sagalowsky AI, Lotan Y. Multiple biomarkers improve prediction of bladder cancer recurrence and mortality in patients undergoing cystectomy. Cancer. 2008; 112(2):315–25. [PubMed: 18008359]

20. Karam JA, Lotan Y, Karakiewicz PI, Ashfaq R, Sagalowsky AI, Roehrborn CG, Shariat SF. Use of combined apoptosis biomarkers for prediction of bladder cancer recurrence and mortality after radical cystectomy. Lancet Oncol. 2007; 8(2):128–36. [PubMed: 17267327]

21. Andre F, Pusztai L. Molecular classification of breast cancer: implications for selection of adjuvant chemotherapy. Nat Clin Pract Oncol. 2006; 3(11):621–32. [PubMed: 17080180]

22. Grossman HB. Are biomarkers for bladder cancer beneficial? J Urol. 183(1):11–2. [PubMed: 19913828]

23. Bensalah K, Montorsi F, Shariat SF. Challenges of cancer biomarker profiling. Eur Urol. 2007; 52(6):1601–9. [PubMed: 17919807]

24. Malats N, Bustos A, Nascimento CM, Fernandez F, Rivas M, Puente D, Kogevinas M, Real FX. P53 as a prognostic marker for bladder cancer: a meta-analysis and review. Lancet Oncol. 2005; 6(9):678–86. [PubMed: 16129368]

25. Check E. Proteomics and cancer: running before we can walk? Nature. 2004; 429(6991):496–7. [PubMed: 15175721]

26. Diamandis EP. Proteomic patterns to identify ovarian cancer: 3 years on. Expert Rev Mol Diagn. 2004; 4(5):575–7. [PubMed: 15347249]

27. Dalbagni G, Parekh DJ, Ben-Porat L, Potenzoni M, Herr HW, Reuter VE. Prospective evaluation of p53 as a prognostic marker in T1 transitional cell carcinoma of the bladder. BJU Int. 2007; 99(2):281–5. [PubMed: 17155984]

28. Svatek RS, Jeldres C, Karakiewicz PI, Suardi N, Walz J, Roehrborn CG, Montorsi F, Slawin KM, Shariat SF. Pre-treatment biomarker levels improve the accuracy of post-prostatectomy nomogram for prediction of biochemical recurrence. Prostate. 2009; 69(8):886–94. [PubMed: 19229851]

29. Bensalah K, Lotan Y, Karam JA, Shariat SF. New circulating biomarkers for prostate cancer. Prostate Cancer Prostatic Dis. 2007

30. Verma M, Srivastava S. New cancer biomarkers deriving from NCI early detection research. Recent Results Cancer Res. 2003; 163:72–84. discussion 264–6. [PubMed: 12903844]

31. Winget MD, Baron JA, Spitz MR, Brenner DE, Warzel D, Kincaid H, Thornquist M, Feng Z. Development of common data elements: the experience of and recommendations from the early detection research network. Int J Med Inform. 2003; 70(1):41–8. [PubMed: 12706181]

32. Lokeshwar VB, Habuchi T, Grossman HB, Murphy WM, Hautmann SH, Hemstreet GP 3rd, Bono AV. Getzenberg RH, Goebell P, Schmitz-Drager BJ and others. Bladder tumor markers beyond cytology: International Consensus Panel on bladder tumor markers. Urology. 2005; 66(6 Suppl 1): 35–63. [PubMed: 16399415]

33. Goebell PJ, Groshen SL, Schmitz-Drager BJ. Guidelines for development of diagnostic markers in bladder cancer. World J Urol. 2008; 26(1):5–11. [PubMed: 18253738]

34. Shariat SF, Canto EI, Kattan MW, Slawin KM. Beyond prostate-specific antigen: new serologic biomarkers for improved diagnosis and management of prostate cancer. Rev Urol. 2004; 6(2):58–72. [PubMed: 16985579]

35. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. J Clin Oncol. 2009; 27(24):4027–34. [PubMed: 19597023]

36. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: one size does not fit all. J Biopharm Stat. 2009; 19(3):530–42. [PubMed: 19384694]

37. Shariat SF, Karakiewicz PI. Perspectives on prostate cancer biomarkers. Eur Urol. 2008; 54(1):8–10. [PubMed: 18262336]

38. Shariat SF, Karakiewicz PI, Roehrborn CG, Kattan MW. An updated catalog of prostate cancer predictive tools. Cancer. 2008; 113(11):3075–99. [PubMed: 18823041]

39. Shariat SF, Karam JA, Margulis V, Karakiewicz PI. New blood-based biomarkers for the diagnosis, staging and prognosis of prostate cancer. BJU Int. 2008; 101(6):675–83. [PubMed: 17941930]

40. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001; 54(8):774–81. [PubMed: 11470385]

41. Kattan MW. Nomograms. Introduction. Semin Urol Oncol. 2002; 20(2):79–81. [PubMed: 12030248]

42. Efron, B.; Tibshirani, RJ. An introduction to the bootstrap. Champman and Hall/CRC; Boca Raton, Florida: 1993. p. 275-281.

43. Kattan MW. Comparison of Cox regression with other methods for determining prediction models and nomograms. J Urol. 2003; 170(6 Pt 2):S6–9. discussion S10. [PubMed: 14610404]

44. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol. 2003; 56(5):441–7. [PubMed: 12812818]

45. Steyerberg EW, Harrell FE Jr, Goodman PH. Neural networks, logistic regression, and calibration. Med Decis Making. 1998; 18(3):349–50. [PubMed: 9680001]

46. Margulis V, Shariat SF, Ashfaq R, Sagalowsky AI, Lotan Y. Ki-67 is an independent predictor of bladder cancer outcome in patients treated with radical cystectomy for organ-confined disease. Clin Cancer Res. 2006; 12(24):7369–73. [PubMed: 17189409]

47. Margulis V, Lotan Y, Karakiewicz PI, Fradet Y, Ashfaq R, Capitanio U, Montorsi F, Bastian PJ, Nielsen ME, Muller SC, et al. Multi-institutional validation of the predictive value of Ki-67 labeling index in patients with urinary bladder cancer. J Natl Cancer Inst. 2009; 101(2):114–9. [PubMed: 19141773]

48. Lotan Y. Role of biomarkers to predict outcomes and response to therapy. Urologic Oncology: Seminars and Original Investigations. 2010; 28:97–101.

49. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006; 26(6):565–74. [PubMed: 17099194]

50. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak. 2008; 8:53. [PubMed: 19036144]

51. Vickers AJ. Decision analysis for the evaluation of diagnostic tests, prediction models and molecular markers. Am Stat. 2008; 62(4):314–320. [PubMed: 19132141]

52. Elkin EB, Vickers AJ, Kattan MW. Primer: using decision analysis to improve clinical decision making in urology. Nat Clin Pract Urol. 2006; 3(8):439–48. [PubMed: 16902520]

53. Shariat SF, Casella R, Wians FH Jr, Ashfaq R, Balko J, Sulser T, Gasser TC, Sagalowsky AI. Risk stratification for bladder tumor recurrence, stage and grade by urinary nuclear matrix protein 22 and cytology. Eur Urol. 2004; 45(3):304–13. author reply 313. [PubMed: 15036675]

54. Elstein AS. Heuristics and biases: selected errors in clinical reasoning. Acad Med. 1999; 74(7): 791–4. [PubMed: 10429587]

55. Vlaev I, Chater N. Game relativity: how context influences strategic decision making. J Exp Psychol Learn Mem Cogn. 2006; 32(1):131–49. [PubMed: 16478346]

56. Hogarth RM, Karelaia N. Heuristic and linear models of judgment: matching rules and environments. Psychol Rev. 2007; 114(3):733–58. [PubMed: 17638504]

57. Ross PL, Gerigk C, Gonen M, Yossepowitch O, Cagiannos I, Sogani PC, Scardino PT, Kattan MW. Comparisons of nomograms and urologists' predictions in prostate cancer. Semin Urol Oncol. 2002; 20(2):82–8. [PubMed: 12012293]

58. Ross PL, Scardino PT, Kattan MW. A catalog of prostate cancer nomograms. J Urol. 2001; 165(5): 1562–8. [PubMed: 11342918]

59. Specht MC, Kattan MW, Gonen M, Fey J, Van Zee KJ. Predicting nonsentinel node status after positive sentinel lymph biopsy for breast cancer: clinicians versus nomogram. Ann Surg Oncol. 2005; 12(8):654–9. [PubMed: 16021535]

60. Walz J, Gallina A, Perrotte P, Jeldres C, Trinh QD, Hutterer GC, Traumann M, Ramirez A, Shariat SF, McCormack M, et al. Clinicians are poor raters of life-expectancy before radical prostatectomy or definitive radiotherapy for localized prostate cancer. BJU Int. 2007; 100(6): 1254–8. [PubMed: 17979925]

61. Karakiewicz PI, Shariat SF, Palapattu GS, Gilad AE, Lotan Y, Rogers CG, Vazina A, Gupta A, Bastian PJ, Perrotte P, et al. Nomogram for predicting disease recurrence after radical cystectomy for transitional cell carcinoma of the bladder. J Urol. 2006; 176(4 Pt 1):1354–61. discussion 1361– 2. [PubMed: 16952631]

62. Karakiewicz PI, Shariat SF, Palapattu GS, Perrotte P, Lotan Y, Rogers CG, Amiel GE, Vazina A, Gupta A, Bastian PJ, et al. Precystectomy nomogram for prediction of advanced bladder cancer stage. Eur Urol. 2006; 50(6):1254–60. discussion 1261–2. [PubMed: 16831511]

63. Shariat SF, Karakiewicz PI, Palapattu GS, Amiel GE, Lotan Y, Rogers CG, Vazina A, Bastian PJ, Gupta A, Sagalowsky AI, et al. Nomograms provide improved accuracy for predicting survival after radical cystectomy. Clin Cancer Res. 2006; 12(22):6663–76. [PubMed: 17121885]

64. Shariat SF, Karakiewicz PI, Suardi N, Kattan MW. Comparison of nomograms with other methods for predicting outcomes in prostate cancer: a critical analysis of the literature. Clin Cancer Res. 2008; 14(14):4400–7. [PubMed: 18628454]

65. Shariat SF, Margulis V, Lotan Y, Montorsi F, Karakiewicz PI. Nomograms for bladder cancer. Eur Urol. 2008; 54(1):41–53. [PubMed: 18207314]

66. Lerner SP, Bochner B, Kibel AS. The use and abuse of data: nomograms and talking to patients about clinical medicine. Urol Oncol. 2007; 25(4):333–7. [PubMed: 17628303]

67. Begg CB, Cramer LD, Venkatraman ES, Rosai J. Comparing tumour staging and grading systems: a case study and a review of the issues, using thymoma as a model. Stat Med. 2000; 19(15):1997– 2014. [PubMed: 10900448]

68. Vickers AJ, Jang K, Sargent D, Lilja H, Kattan MW. Systematic review of statistical methods used in molecular marker studies in cancer. Cancer. 2008; 112(8):1862–8. [PubMed: 18320601]

69. Vickers AJ, Kattan MW, Daniel S. Method for evaluating prediction models that apply the results of randomized trials to individual patients. Trials. 2007; 8:14. [PubMed: 17550609]

70. Vickers AJ, Cronin AM, Kattan MW, Gonen M, Scardino PT, Milowsky MI, Dalbagni G, Bochner BH. Clinical benefits of a multivariate prediction model for bladder cancer: a decision analytic approach. Cancer. 2009; 115(23):5460–9. [PubMed: 19823979]

71. Bochner BH, Kattan MW, Vora KC. Postoperative nomogram predicting risk of recurrence after radical cystectomy for bladder cancer. J Clin Oncol. 2006; 24(24):3967–72. [PubMed: 16864855]

72. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst. 2005; 97(16):1180–4. [PubMed: 16106022]

73. Hayes DF, Bast RC, Desch CE, Fritsche H Jr, Kemeny NE, Jessup JM, Locker GY, Macdonald JS, Mennel RG, Norton L, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. J Natl Cancer Inst. 1996; 88(20):1456–66. [PubMed: 8841020]
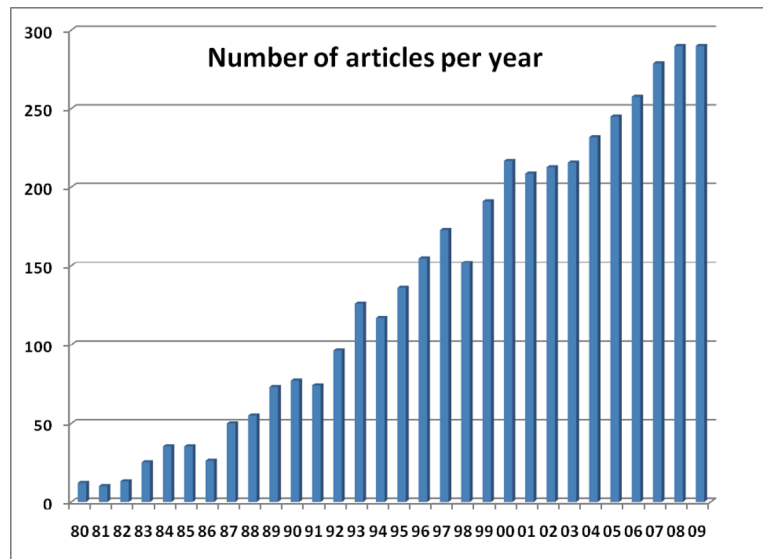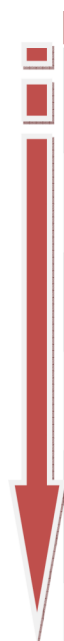
Figure 1.
Number of articles per year based on a PubMed Search using the terms "bladder cancer" AND ("biomarker" OR "molecular marker") published in English language (accessed 12/25/09).

Phase	Goals/aims	Experimentation	Sample details
Preclinical Testing	**Exploratory; nominate and rank candidate biomarker profiles**	Preclinical study for hypothesis generation	Possible bias: small size and convenience sampling
0	**Develop an assay with clinically reproducible results**	Reproducibility and robustness of assay; No assessment of benefit	
I	**Test on small sample to determine benefit**	Marker optimization, establish prediction rules, determine cut-offs	Sample population assay developed from candidate biomarker profile
II	**Determine operating characteristics & internal validation**	Retrospective design	Sample population should be the target population
III	**External validation**	Retrospective or prospective, Generaliziability, Impact on clinical decision-making	Multi-institutional, large study
IV	**Assess whether biomarker reduces the burden of disease**	Post-approval reporting and testing for other disease processes or disease stages	

Figure 2.
Modification of the structured phase-approach to the systematic discovery, evaluation, and validation of biomarkers[23,34]

Item
INTRODUCTION
1 State the marker examined, the study objectives, and any pre-specified hypotheses.
MATERIALS AND METHODS Patients
2 Describe the characteristics (e.g., disease stage or co-morbidities) of the study patients, including their source and inclusion and exclusion criteria.
3 Describe treatments received and how chosen (e.g., randomized or rule-based).
Specimen characteristics
4 Describe type of biological material used (including control samples) and methods of preservation and storage.
Assay methods
5 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
6a 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.
6b Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.
7 Precisely define all clinical endpoints examined.
8 List all candidate variables initially examined or considered for inclusion in models.
9 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
10 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.
11 Clarify how marker values were handled in the analyses; if relevant, describe methods used for cutpoint determination.
RESULTS Data
12 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.
13 Report distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumor marker, including numbers of missing values.
Analysis and presentation
14 Show the relation of the marker to standard prognostic variables.
15 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 analyzed. For the effect of a tumor marker on a time-to-event outcome, a Kaplan-Meier plot is recommended.
16 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.
17 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.
18 If done, report results of further investigations, such as checking assumptions, sensitivity analyses, and internal validation.
DISCUSSION
19 Interpret the results in the context of the pre-specified hypotheses and other relevant studies; include a discussion of limitations of the study.
20 Discuss implications for future research and clinical value.

Figure 3.
Reporting recommendations for tumor marker prognostic studies (REMARK) *NCI-EORTC (with permission from McShane et al.[72])*

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.

Figure 4.
Levels of evidence for grading clinical utility of biomarkers: Tumor Marker Utility Grading System (TMUGS) (reproduced with permission of *Hayes et al.*[73])

Table 1

Clinical uses of biomarkers in bladder cancer

• Clinical Trials
– Stratifying study populations
– Conducting interim analysis of efficacy/safety
– Applied toward regulatory approval
• Clinical practice
– Determine the risk of developing disease
– Early detection/screening
– Establish diagnosis
– Determine prognosis
– Predict response to therapy
– Therapeutic target
– Monitor disease/treatment response

Table 2

Challenges to and advances that may facilitate the development of clinically useful bladder cancer biomarkers

CHALLENGES

Biologic factors

- Progressive biologic heterogeneity with transient expression of certain features is a characteristic of tumor cells. Biological heterogeneity is present both among cells within the tumor at a given time and in cells during the development of the tumor from earlier to later points in time. In addition, biomarkers may be affected by therapy and as yet uncharacterized, host factors. Biological heterogeneity includes multiple pathways to the same endpoints and the variable metabolism of biomarkers, including post-translational trans-modifications.

- Age and the presence of other diseases also introduce variation in biomarker levels between individuals. Other physiologic or pathologic processes may generate biomarker profiles similar to those found in patients with tumor disease states.

- Exogenous substances that affect biomarker presence and concentration. Foods, drugs, and natural alternative therapies are well known interferences.

Clinical pathologic factors

- Requirement to define and standardize more precisely concepts of the biological events against which biomarkers are to be measured (e.g., normal variation, different disease states). New tools for accurate detection of pre-neoplastic neoplasia, micrometastatic spread, and states of early and/or aggressive cancer recurrence need to be developed.

Analytical sensitivity and detection limit

- Assay sensitivity needs to be sufficiently high to allow biomarker quantitation at concentrations that have biologic relevance. Clinical detection and measurement of biomarkers of this type, at worst, could lead to unnecessary investigation and therapy or, at best, unnecessary chronic anxiety for the patient.

- There is a lack of definition of standard procedures, standard reference materials, and quality control schemes necessary to assure accuracy and reproducibility.

- There is a lack of clear guidelines for good manufacturing/laboratory practice and quality control requirements for all phases of biomarker development.

Intellectual property

- Ownership of a biomarker is a key element of its commercialization. Considering the cost of developing and validating a biomarker, no company will invest in a biomarker for which they can not be assured a reasonable return on investment by means of protection of intellectual property. Only a very small fraction of candidate biomarkers demonstrate real clinical utility, such that the demand for return on investment is similar to the drug development models. These issues are well appreciated by bio-industry, and demonstrated by their lack of interest in commercialization of the plethora of candidate biomarkers in the literature.

Health service factor

- It is not sufficient for a tissue biomarker to detect a particular phase of neoplasia. To be successful, the biomarker must also fit within the profile of health service factors with respect to cost-effectiveness, cost benefit, and relative value of biomarker strategy for cancer burden reduction.

FACTORS TO SUPPORT ADVANCES

Defining the biology of bladder cancer and its processes with precision

- Enhanced interaction among investigators of different disciplines and institutions.

- Greater appreciation of the biokinetics of both cancer and its biomarkers permitting more dynamic views of how cancers evolve.

Defining host biology: pharmaco-genomics and pharmaco-proteinomics

- Biological profiling does have the prospect of individualizing therapy, maximizing efficacy, and minimizing toxicity. Ideal markers would reflect both cancer activity and individual sensitivity to therapy.

Defining biomarkers and surrogate endpoints

- Need for developing a consensus about definitions that are widely accepted and applied.

Creating guidelines for appropriate clinical employment of each biomarker

---

CHALLENGES

- Plan a series of national multidisciplinary initiatives aimed at (1) surveying the QC programs, (2) coordinating from a scientific point of view the activities in this area and producing guidelines for the clinical employment of cancer biomarkers, (3) standardizing the procedures, and (4) developing laboratory QC programs for the analysis of cancer biomarkers of validated clinical relevance in multicenter clinical protocols.

Standardization and stringency of analytical technology

- Further standardize pre-analytical, analytical, post-analytical methodology. However standardization of biomarker assay technology involves considerations beyond analytical sensitivity and specificity. For example, advancing toward standardized technology, the advantages of comparability between various studies must be weighted against the desire and need for innovation and conditions that require protocol flexibility.

High-quality specimen and clinical data repository

- Need for specimen and data repositories that address in a bioethical manner patient consent, confidentiality, specimen provenance, technical preparation, and storage.