Predicting Clinical Outcomes Using Molecular Biomarkers

Harry B. Burke

Professor of Medicine, Department of Medicine, F. Edward Hebert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, USA.

ABSTRACT: Over the past 20 years, there has been an exponential increase in the number of biomarkers. At the last count, there were 768,259 papers indexed in PubMed.gov directly related to biomarkers. Although many of these papers claim to report clinically useful molecular biomarkers, embarrassingly few are currently in clinical use. It is suggested that a failure to properly understand, clinically assess, and utilize molecular biomarkers has prevented their widespread adoption in treatment, in comparative benefit analyses, and their integration into individualized patient outcome predictions for clinical decision-making and therapy. A straightforward, general approach to understanding how to predict clinical outcomes using risk, diagnostic, and prognostic molecular biomarkers is presented. In the future, molecular biomarkers will drive advances in risk, diagnosis, and prognosis, they will be the targets of powerful molecular therapies, and they will individualize and optimize therapy. Furthermore, clinical predictions based on molecular biomarkers will be displayed on the clinician's screen during the physician–patient interaction, they will be an integral part of physician–patient-shared decision-making, and they will improve clinical care and patient outcomes.

KEYWORDS: cancer, prediction, molecular, biomarker, outcome, translation, surrogate outcome, clinical outcome, treatment

Introduction

He will manage the cure best who has foreseen what is to happen from the present state of matters (Hippocrates, The Book of Prognostics, 400 B.C.E.).

For thousands of years, prediction has been central to the practice of medicine. Even before there were effective therapies, physicians sat at the bedside, observed their patients and, based on their observations, predicted their patients' outcome. Predictions integrate the complex of facts that constitute a disease and its management in a way that guides clinical care.[1,2] In fact, the ability to make accurate disease-related predictions is the hallmark of medical science. Furthermore, predictive medicine is a critical component of precision medicine, and biomarkers, biological predictive factors, are central to predictive medicine.

Medical progress requires that we discover and clinically use molecular biomarkers that (i) are necessary components of the disease process, (ii) accurately predict clinical disease outcomes, and (iii) give rise to biomarker-related interventions that retard or halt the disease process.

Over the past 20 years, there has been an exponential increase in the number of biomarkers. At the last count, there were 768,259 papers indexed in PubMed.gov directly related to biomarkers. Although many of these papers claim to report clinically useful molecular biomarkers,

embarrassingly few molecular biomarkers are currently in clinical use.[3–6] One reason why this situation exists is that researchers may not fully appreciate the complexity inherent in the discovery, translation, and use of molecular biomarkers to clinical medicine.[4–6] The result is studies replete with misinformation and a literature that contains incorrect, and many times even contradictory, results.[4–6] In other words, many molecular biomarkers have been called but, so far, few have been chosen.

I suggest that a failure to properly understand molecular biomarkers has prevented their widespread adoption in treatment, comparative benefit analyses, and their integration into individualized patient outcome predictions for clinical decision-making.[4–7] This study presents a straightforward, general approach to understanding how to predict clinical outcomes using molecular biomarkers.

Medical Prediction

A prediction is an inference about an unknown present or future state or event based on known information. Although predictions may be of any character, here we are interested in quantitative predictions of medical outcomes that are based on empirical clinical information. In other words, we want to use current empirical information to predict the occurrence (or nonoccurrence) of the true outcome, where the true outcome can be a patient state or event.

Quantitative predictions are usually generated by entering a patient's predictive factors into a trained statistical model, the output of which is a patient's probability of the occurrence of the outcome.

All predictions must be time denominated because the meaning of a prediction depends on the time interval; for example, the meaning of the probability of an outcome occurring within one year is different from it occurring within 10 years. In other words, a medical prediction that is not time denominated rarely has a clinical meaning. Furthermore, the prediction of an individual's lifetime probability of death is usually not clinically useful because it: (1) provides information about a population average, which is not necessarily relevant to an individual patient, (2) is strongly affected by infant mortality, and (3) does not take into account how long the patient has lived to the time of the prediction (ie, conditional survival).

Until recently, medicine had not provided highly accurate predictions because: (1) it did not possess powerful predictive factors, (2) statistical methods seemed far removed from the practice of medicine and, as a consequence, predictions were not routinely integrated into the practice of medicine, and (3) there was no interest in, or mechanism for, providing patients with medical outcome predictions that were specific to individual patients.[8] The rise of molecular biology, the use of computers in medicine, and the discovery of molecular biomarkers that are directly involved in the disease process (and are, therefore, powerful predictors of disease outcomes), for example, estrogen receptors in breast cancer, have been major advances[1,2,8–10] in medical prediction.

Temporal and Biological Determinism

Prior to the advent of molecular biology, the dominant prognostic factors in cancer were anatomic because until relatively recently the only treatment for cancer was surgery and anatomic factors could be collected during the surgical procedure. These anatomic factors were the location of cancer and its spread (surgical exploration) and the tumor's size and grade (surgical pathology). Location was categorized as the solitary tumor (local), the presence of involved lymph nodes (regional), and evidence of distant metastasis (distant).[8] In other words, for many years, all predictions were based on patients undergoing surgery, on surgery being their only therapy, and on surgical-related predictive factors.[8]

In 1953, the French surgeon Pierre Denoix proposed to the Union Internationale Centre le Cancer that these anatomic factors be standardized, integrated into stages, and used across many solid tumor sites.[9] The variables tumor size (T), lymph node status (N), and distant metastasis (M) were used to create a uniform and easy-to-use prognostic system, namely, the TNM staging system. The TNM staging system grouped patients into four stages (I–IV), where higher stages meant worse survival. The patient's prognosis was predicted to be the average survival of all the patients in that patient's stage.

As medicine progressed, the disadvantages of this system became more obvious and troublesome. Because the TNM staging system was based on patients only receiving surgery, it did not take other treatments into account in its predictions. The system did not do so because if it had, it would have had to include predictors for each therapy and each combination of therapies, and this would have created a bin system that was too complex to be practical.[11] In addition, as we learned more about the TNM systems, its fundamental assumption was called into question. While it was true that many patients with large tumors and distant disease died more quickly than those with small tumors and no distant disease, it was also true that some patients with small tumors and no distant disease died more quickly as those patients with large tumors and no distant disease.

All disease has both a temporal and a biological dimension. Although these dimensions are clearly distinguishable, they are not completely independent. The temporal dimension indexes how long the disease has existed in the body. The biological dimension indexes the disease process itself. If the biology of the disease did not change over time, then one could observe the temporal course of the disease by simply measuring a temporal factor repeatedly over time and using that information to extrapolate in linear time the patient's prognosis. There are at least four problems with this approach to prognosis. First, we cannot measure cancer over time because once it is detected it is removed, so we cannot plot its temporal trajectory. Second, we cannot assume that the biology of the disease does not change over time. Third, the primary tumor is, many times, not what determines the patient's prognosis, rather, metastases provide lethality. Fourth, an effective therapy changes the biology of the disease and, therefore, the patient's prognosis, which means that if the patient receives an effective therapy, the predicted progression of the disease no longer applies because the progression has become discontinuous.

Cancer has both a temporal and a biological dimension. Ideally, in order to know the patient's prognosis one would know both dimensions, namely, how long the tumor has been growing and its aggressiveness. We cannot know how long it has been growing, but we can learn about its aggressiveness by the assessment of molecular biomarkers from a tumor biopsy, or a resected tumor, or other types of samples. Our current use of anatomic factors tends to conflate the temporal and biological dimensions. When we see a large tumor, we assume that it has been growing for a long time because we know that it is generally associated with a shorter survival than a small tumor. Using anatomic factors in this way, as temporal indices, is called temporal determinism,[1] and the TNM staging system is basically a temporal system because it does not take into account the biology of the disease.

When anatomic factors are used as temporal indexes for temporal predictions, the predictions are contaminated

by the temporal assumption, ie, the factors index time. This phenomenon has been called lead-time bias, but it is not actually a bias, rather, it is simply a natural consequence of temporal determinism. This temporal effect occurs because earlier detection increases the length of time we follow the patients, and thus, they appear to live longer with cancer. In other words, the earlier the tumor is discovered, the better the patient's prognosis appears to be. The reason for this effect is that the temporal determinism assumption confounds the temporal prediction. So long as anatomic factors are assumed to index time, ie, the larger the tumor, the longer it has been growing and the worse the prognosis, this confounding will continue. Everything else being equal, if the anatomic factors were analyzed for biological aggressiveness, this confounding would disappear. Aggressiveness factors would confer a worse prognosis, regardless of when they were discovered.

Another issue with the TNM staging system is individual patient predictions. It predicts that a patient will have the same prognosis as the mean survival of a group of patients with the same TNM variables, ie, the same TNM stage. The problem is that if one examines the outcomes of all the patients in a TNM stage, one finds a wide variability in outcomes. In fact, the overlap in outcomes between the four stages is such that many patients in one stage would have an outcome similar to those in a stage they were not assigned to. This means that predictions based on the TNM staging system will have low accuracies. For example, the TNM receiver operating characteristic (ROC) scores were 0.69–0.72, 0.74, and 0.53 for breast cancer, colorectal cancer, and prostate cancer, respectively.[12] Although these predictions were poor, the situation has only gotten worse. With the advent of screening and early detection, there has been a stage migration away from stages III and IV to stages I and II. This has resulted in the TNM staging system being no more accurate at predicting breast cancer survival than flipping a coin.[13] Finally, the TNM does not recognize the biology of an individual patient's disease, which is the antithesis of precision medicine.

We want to know the tumor's biology, including its aggressiveness. In other words, we want to assess biological determinism.[1] We are interested in aggressiveness to the extent that aggressiveness is related to lethality. Molecular biomarkers can index the biology of the disease, they can provide information about the aggressiveness of the disease and, as a consequence, they can provide information regarding the individual patient's prognosis. Although we have been using the term aggressiveness to illustrate biological determinism, in reality, it is not that simple. It is important to realize that the growth rate of the primary tumor is related to, but not identical with, lethality. Lethality is a great deal more complicated than the primary tumor's growth rate because, for many cancers, it is not the primary tumor that kills the patient but rather the tumor's metastases. Aggressiveness and lethality

are determined by biological factors that are directly related to both the disease and the patient-host.

Medicine must move from predictions based on temporal determinism to predictions based on biological determinism.[1] Cancer should not be defined by the anatomic stage of the disease at the time of detection. Rather, it must be defined by the molecular characteristics of both the tumor and the host. Biological determinism takes the view that the characteristics of the disease at detection must be related to the biology of the disease rather than to our methods of tumor detection. Biological determinism is the systematic combination of the molecular biomarkers that index the biology of the disease and of the patient-host and the use of this information to make individual patient outcome predictions, including the selection of an effective therapy for the individual patient. In other words, treatment must be driven by the molecular biology of the tumor and host rather than by the method of disease discovery.

Clinically Predictive Biomarkers

It appears to me a most excellent thing for the physician to cultivate Prognosis; for by foreseeing and foretelling, in the presence of the sick, the present, the past, and the future, and explaining the omissions which patients have been guilty of, he will be the more readily believed to be acquainted with the circumstances of the sick; so that men will have confidence to entrust themselves to such a physician (Hippocrates, The Book of Prognostics, 400 B.C.E.).

In terms of clinical use, a predictive factor is any measureable attribute of an individual that can be used to infer a health-related outcome. Here, we are interested in disease-related attributes that predict a disease-related outcome. There are at least three levels of medically related predictive factors, namely, demographic (eg, race, gender, socioeconomic status, etc.), anatomic/cellular (eg, tumor, stroma, cellular attributes, etc.), and molecular (eg, proteins, genes, etc.).[8,9] A predictive biomarker is a biological predictive factor. A molecular biomarker is any measurable molecular attribute of an individual that can be used to predict a disease-related outcome by virtue of its relationship to the disease process over a specific time interval.[14] All predictions must be accompanied by a time interval. If the prediction is of a current outcome, then the interval is instantaneous; if the prediction is of a future outcome, then there must be for a specified duration of time.

There are three types of medical predictions, risk, diagnosis, and prognosis (Table 1).[8,15] They differ in their clinical uses, their outcomes, and their accuracy. *Risk*, as the term is commonly used, is ambiguous. It can refer to the risk of the occurrence of incident disease, or it can refer to the chance of occurrence of a medical outcome. Here, risk is used to refer to the risk of incident disease over a specified interval of time. In place of the word *risk* in the phrases

Table 1. Clinical type (risk, diagnosis, and prognosis) and clinical use (natural history, prevention/therapy-specific, and post-prevention/therapy-specific) of predictive factors, their patient predictions, and clinical rationale.

CLINICAL TYPE AND USE	INDIVIDUAL PATIENT PREDICTION	CLINICAL RATIONALE
RISK	Predicts that the patient will in the future exhibit incident disease, over a specified time interval, the probability is much less than 100%	To identify patients who have a high likelihood of disease, the goal is to prevent or retard the occurrence of incident disease
Natural history risk	Probability of incident disease if the patient does not receive a prevention intervention, over a specified time interval	To determine whether a prevention intervention is necessary
Prevention-specific risk	Probability that the patient will respond to a specific prevention intervention, over a specified time interval	To determine the optimal prevention intervention
Post-prevention risk	Probability that the patient responded to the prevention intervention, over a specified time interval	To determine whether the prevention intervention was effective without waiting for the occurrence of a clinical outcome
DIAGNOSTIC	Predicts that the patient currently has the disease at this instant in time, the probability is close to 100%	To diagnose the patient
PROGNOSTIC	Predicts a future disease-related outcome in a patient with the disease, over a specified time interval, the probability is variable	To identify patients who have a high likelihood of an adverse outcome, the goal is to retard or stop the progression of the disease in those patients
Natural history prognostic	Probability of a disease-related outcome if the patient does not receive any therapy, over a specified time interval	To determine whether therapy is necessary
Therapy-specific prognostic	Probability that the patient will respond to a specific therapy, over a specified time interval	To determine the optimal therapy
Post-therapy prognostic	Probability that the patient responded to the therapy, over a specified time interval	To determine whether the therapy was effective without waiting for the occurrence of a clinical outcome

risk of recurrence and *risk of death*, the word *probability* will be substituted, as in the probability of recurrence and probability of death.[15]

A risk biomarker predicts that the patient who has not been diagnosed with the disease will exhibit incident disease over a specified time interval.[15] The risk biomarker, either alone, or in combination with other risk biomarkers, is always less than 100% accurate in predicting incident disease over a specified time interval. If a risk biomarker's prediction is close to 100% accurate, it is a diagnostic biomarker. There are three clinical uses of risk biomarkers. A *natural history* risk biomarker predicts the probability that the patient will exhibit incident disease if the patient does not receive a prevention intervention over a specified period of time. The goal of a natural history risk biomarker is to determine whether a prevention intervention is necessary, ie, does the patient have a sufficiently high likelihood of incident disease that a prevention intervention should be considered. A *prevention-specific* risk biomarker predicts the probability that the patient will respond to a prevention intervention over an interval of time. The goal of a prevention-specific risk biomarker is to determine the optimal prevention intervention. A *post-prevention* risk biomarker predicts that the patient responded to a preventive intervention over a specified interval of time. The goal of a post-prevention risk biomarker is to determine if the intervention was effective without waiting for the occurrence of a clinical outcome. The risk biomarkers that are the most powerful predictors are usually directly connected to the development of the disease and they are the best targets for prevention.[16]

In other words, risk has three subtypes of predictions: (1) the probability that a patient who does not currently have the disease will exhibit the disease over a specified time interval; (2) the prediction that a prevention intervention will reduce the probability of incident disease over a specified time interval; and (3) the prediction of whether the administered prevention intervention reduced the probability of incident disease over a specified time interval.

The reason that most risk biomarkers possess poor predictive accuracy is because no matter how carefully the at-risk population is selected it will almost always be heterogeneous for the occurrence of incident disease by the end of the time interval. Even strong risk biomarkers exhibit poor predictive accuracy when assessed in a heterogeneous population. For example, tobacco smoking is a very strong risk factor for lung cancer, but most smokers will not be diagnosed with lung cancer. The better the at-risk population is defined, for example, using lung cancer susceptibility genes in smokers to define a risk group, the more homogeneous it is.

A diagnostic biomarker predicts that the patient who was not known to have the disease currently has the disease at this instant in time.[15] The diagnostic biomarker, either alone, or in combination with other diagnostic biomarkers, must be close to 100% accurate in predicting incident disease at that moment in time. The goal is to diagnose the disease in the patient. A biopsy that demonstrates invasive cancer is close to 100% predictive of incident disease at the moment the biopsy is taken.

A prognostic biomarker predicts a future disease-related patient outcome in a patient with the disease.[8,15]

The prognostic biomarker, either alone, or in combination with other prognostic biomarkers, predicts the probability of a disease-related outcome over a specified time interval. Although prognostic biomarkers are usually stronger predictors than risk biomarkers because all the patients in the population have the disease, when the disease process is complex, as it is in cancer, it is rarely the case that a lone biomarker will accurately predict a disease outcome.

There are three clinical uses of prognostic biomarkers. A *natural history* prognostic biomarker predicts the probability of a disease-related outcome if the patient does not receive any therapy over a specific period of time. The goal of a natural history prognostic biomarker is to determine if a therapy is necessary, ie, does the patient have a sufficiently high likelihood of a poor outcome that a therapy should be considered. A *therapy-specific* prognostic biomarker predicts that the patient will respond to a specific therapy over a specific period of time. The goal of a therapy-specific prognostic biomarker is to determine the optimal therapy. A *post-therapy* prognostic biomarker predicts that the patient responded to a therapy over a specific period of time.[8,15] The goal of a post-therapy prognostic biomarker is to determine whether the therapy was effective without waiting for the occurrence of a clinical outcome. The prognostic biomarkers that are the most powerful predictors are usually those directly connected to the disease process and they are the best targets for therapy.[16]

It should be noted that there is no necessary reason why a biomarker that is predictive at the time of the biomarker is acquired and measured will continue to be predictive throughout the disease process. In other words, one cannot assume that a biomarker, once discovered, will continue to be an accurate predictor as the disease progresses. In fact, it is more likely that an early prognostic biomarker will not participate in later disease and, as a result, will lose its predictive power and that a late prognostic biomarker will not be in evidence early in the disease process.

It is important to understand that an effective therapy always changes the patient's prognosis by improving the patient's outcome. To the extent that the patient's prognosis has been changed by the therapy, the predictive power of the natural history biomarker must commensurately change (because the patient's outcome has changed). This change is almost always a reduction in the biomarker's prognostic power. Finally, natural history molecular biomarkers can be used to measure disease aggressiveness and this, in turn, can be used to stratify patients into biologically meaningful severity of illness groups.

A therapy-specific prognostic biomarker is, as its name implies, specific to a particular treatment. It predicts that a patient will respond to that therapy over a specified time interval.[8,15] For example, estrogen receptor status in breast cancer patients predicts a patient's response to antihormone therapy and, to the extent that she responded to the therapy,

her post-therapy outcome will no longer match her natural history outcome. This highlights the importance of taking therapy into account in every prediction model. Every effective therapy the patient receives must be modeled in order to accurately predict his or her outcome.

A post-therapy prognostic biomarker measures whether the therapy changed the patient's outcome without having to wait for the occurrence of a clinical symptom.[8,15] A therapy-specific prognostic biomarker can be assessed for a change in its value, and the direction of the change, over a specified time interval before and after therapy. A change in the biomarker may predict a change in the patient's outcome. The nature of, and relationship between, therapy-specific and post-therapy-specific biomarkers is quite complex due to the biological mechanisms that drive cancer.

Our ability to target a molecular biomarker and observe a change related to therapy, a change that indicates a change in the patient's outcome, is limited by: (1) the fact that there can be multiple alternative pathways such that if one pathway is blocked, another can come to the fore, (2) the role of the biomarker in the disease process can change over time, and (3) how effective the therapy is at affecting the biomarker, eg, whether the biomarker was the target of the therapy. These become critical issues when one attempts to use a post-therapy biomarker as a surrogate outcome.

In other words, prognosis has three subtypes of predictions: (1) the prediction of the probability of a disease-related outcome occurring without the patient receiving any treatment over a specified time interval; (2) the prediction that a therapy will reduce the probability of a disease-related outcome over a specified time interval; and (3) the prediction that the administered therapy reduced the probability of a disease-related outcome over a specified time interval.

There has been some confusion regarding the meaning of the words *prognostic* and *predictive*. Gasparini et al[17] suggested ad hoc definitions for *prognostic indicator* and *predictive factor*. He said that *prognostic* is an indicator at the time of diagnosis that provides information regarding clinical outcome and that *predictive* is a factor that selects patients that are likely to respond to a therapy. There are several problems with Gasparini's nomenclature. Prognosis is a prediction, which means that prognosis must be a subset of prediction. Furthermore, since prognosis is a subset of prediction, predictive factors and prognostic factors cannot be completely different. In addition, risk is a prediction and therefore it, like prognosis, must also be a subset of prediction. Finally, if as Gasparini suggests, a predictive factor must always be a factor that provides information regarding treatment in patients with disease, then risk factors cannot be a subset of predictive factors. These contradictions suggest that Gasparini's nomenclature is incorrect. What Gasparini was trying to do but failed to accomplish, was to distinguish between natural history prognostic factors and therapy-specific prognostic factors.

Three of the most important aspects of a predictive bio-marker are: (1) how strongly is it connected to the disease, (2) its relationship to other disease-specific prognostic biomarkers, and (3) its clinical use.[16] The disease-related predictive power of a biomarker is determined, in part, by how strongly it is connected to the disease process. The less connected the biomarker is to the disease, the less predictive it is.[14] There are three types of connectedness relationships between a biomarker and its disease. The strongest connection occurs when there is a direct relationship, where the molecular biomarker is a part of the causal disease process. This means that the molecular biomarker is an integral (ie, necessary and/or sufficient) part of the biology of the disease.[14] A weaker connection occurs when there is an indirect relationship, where the molecular biomarker is involved in the disease process but is neither necessary nor sufficient for the activity of the disease. An indirect connection means that it is not an integral part of the disease process, but it is related to it in some way, for example, being an occasional component of the disease process.[14] The weakest connection occurs when there is an epiphenomenon relationship, for example, pus to an infection. The biomarker is a byproduct of the disease process but does not participate in the disease process.

Direct biomarkers are the most predictively powerful because they directly participate in the disease process, indirect biomarkers are less powerful, and epiphenomenon biomarkers are the least powerful. In addition, direct biomarkers are the best targets for therapeutic intervention. In fact, one can use the strength of association between the biomarker and the true outcome, as measured by its accuracy, to determine which molecular biomarkers to target for therapeutic interventions.[16]

The clinically related predictive power of the molecular biomarker depends on the clinical question being addressed, ie, the biomarker–outcome relationship being assessed.[14,18] For some questions, the biomarker–outcome relationship will be clinically weak. For example, a biomarker taking part in the initiation of the disease process may become less associated with the disease as the disease changes during its progression, whereas, it may be a strong risk factor for the early detection of the disease. Furthermore, if there is no effective therapy, then only the natural history biomarkers will have clinical utility.

Generally, determining whether a molecular biomarker is a predictive biomarker requires that: (1) the biomarker is measured in a defined population, (2) the population is followed until a sufficient number of outcomes have occurred (eg, deaths), and (3) the relationship between the biomarker and the outcome is determined.[14,15] If the biomarker predicts the outcome with *sufficient* accuracy (where sufficient depends on the clinical question being addressed),[15,18] it is called a predictive biomarker.

Prognostic Outcomes

There are several types of prognostic outcomes including a recurrence after a response to therapy, disease-specific mortality, and all-cause mortality. Prognostic biomarkers obtain their prognostic power by virtue of their relationship to the disease; thus, it is critical that outcomes be related to the disease (and the host) if we are to observe the biomarker's power in predicting the clinical outcome. The problem with all-cause mortality is that if many of the patients die of causes other than the disease then the biomarker, which cannot predict the nondisease-related outcomes, will appear to be a weaker predictor than it actually is.

Surrogate Outcomes

Determining whether a molecular factor is a predictive biomarker can take a long time since one must wait for the prediction time interval to have elapsed. One way to shorten the initial investigation of the relationship between a putative predictive biomarker and the true outcome is through the use of large, long-term, comprehensive, annotated specimen banks.[19] Additional guidance regarding the acquisition and use of archived specimens for evaluating prognostic biomarkers has been proposed.[20,21] Another approach is to use biomarkers as surrogate outcomes.

A surrogate outcome is the use of a disease-related predictive biomarker as if it was the true outcome.[15] When a predictive biomarker takes the place of the true outcome it is called a surrogate outcome since it is acting as a surrogate for the true outcome.[15,22] The key concept is that a surrogate outcome is only useful to the extent that it is directly related to the true outcome. If a predictive biomarker is not related to the true outcome then it is useless as a surrogate outcome. On the other hand, if it is perfectly related to the true outcome, then it can always take the place of the true outcome. This means that if a biomarker changes its value after a preventive intervention or treatment and, if the biomarker is perfectly associated with the true outcome, then by observing the change in the biomarker we are observing a change in the true outcome. For example, if a circulating serum biomarker that is perfectly connected to the disease process disappears after treatment then, without any direct knowledge of the true outcome, we would say that the patient's prognosis has improved. Surrogate outcomes (in clinical trials, they can be called surrogate end points or intermediate end points) are usually employed in an attempt to shorten the duration of, or reduce the number of patients in, a prospective risk or prognosis trial. We must be careful with the word associated, for there are many ways by which a biomarker and an outcome can be associated. Only some of these associations will lend themselves to the biomarker being a surrogate outcome.

Several criteria for a surrogate outcome have been proposed. Prentice[23] proposed that the surrogate outcome must be associated with the final outcome, that the treatment must affect the outcome, and that the full effect of the treatment on the final outcome is fully explained by the surrogate outcome. Prasad et al[24] recently proposed three levels of evidence, namely, a biological rational, the

correlation of the surrogate and the final outcome, and a high correlation of the treatment effect on the surrogate outcome and the final outcome.

In order to assess the strength of a biomarker as a surrogate outcome, one must have the following information: (1) a necessary and sufficient definition of the biomarker as the surrogate outcome including a description of how to detect and assess it (but just because one can measure it does not mean one should measure it)[25] and a method for dealing with error in the biomarker detection and assessment process, (2) a necessary and sufficient definition of the true outcome including a description of how to detect and assess the true outcome and a method for dealing with the error in the true outcome detection and assessment process, and (3) a quantitative understanding of the strength and direction of the relationship between the surrogate outcome and the true outcome over a specified time interval.[15]

Prasad et al[24] reviewed the evidence related to the strength of the association between selected surrogate outcomes and their true outcomes. They defined three levels of correlation between the two, namely, less than or equal to 0.70, which they termed low strength, greater than 0.70 but less than 0.85, which they termed medium strength, and greater than or equal to 0.85, which they termed high strength. They found that most surrogate end points in oncology have a low correlation with overall survival. They conclude, "Our findings call into question the widespread use of surrogate end points in oncology as the basis for treatment decisions" (p. 1392). It is instructive that Prasad et al[24] used overall survival as their true outcome. In terms of prediction, proper disease outcomes are disease-specific, they are not all-cause. If one does not want to show an association between a disease-related biomarker and a true disease outcome, one uses all-cause mortality because it contains many causes of death that are not related to the disease. Furthermore, Prasad et al used linear correlation to assess the connection between the surrogate outcome and the true outcome, but the association between a disease-related biomarker and a disease-related outcome is, many times, not linear. This is because complex systems, and cancer is a complex system, are usually nonlinear and interactional, neither of which are taken into account by a linear model.[12]

It is relatively straightforward to use molecular biomarkers as surrogate outcomes when they are direct biomarkers, when they affect a single prevention or therapy-mediated biological pathway, and when the targeted pathway is the only pathway driving the disease. But this almost never occurs. One reason is because a disease usually has multiple possible pathways and a therapy may only affect one of its pathways.[22,26] One possible solution to the multiple pathway problem is to obtain a set of relevant pathways and observe the effect of prevention or treatment on each of the pathways.[23,27]

An important issue is how long must one must wait for a change in the surrogate outcome. An optimal surrogate outcome changes its value shortly after the treatment.

This means that if there is no change after the treatment, another treatment should be instituted without delay.

Explanatory Levels of Analysis

Molecular biomarkers do not exist in isolation from other predictive factors, every factor is imbedded in a hierarchical matrix of interconnected factors and a factor's relationships with other factors affect the factor's power.[8,9,16] As discussed previously, prognostic factors can be viewed in terms of at least three explanatory levels of analysis. *Demographic* factors exist across patients and include social, cultural, and environmental factors, *anatomic/cellular* biomarkers exist within an individual as tissue and include the tumor cells, surrounding stroma, and lymphatics, and *molecular* biomarkers are biochemical entities and include genes and proteins.[8,9] These levels are important because of their hierarchical relationship. Molecular biomarkers are the origin of anatomic factors in a reductionist manner. Thus, anatomic factors are a realization of the molecular factors and the molecular factors' rules of organization. Anatomic factors can be thought of as molecularly *compound* factors. Because anatomic factors are composed of numerous lower level (molecular) factors, and because different lower level factors can give rise to the same higher level factor, there need not be a one-to-one correspondence between what is observed at the intersection of the anatomic and molecular levels. In addition, the number and complexity of the factors increases as one moves from the anatomic to the molecular level.[8,9,16]

Because anatomic factors are a level away from the molecular disease process, they tend to be less powerful predictors than the molecular factors from which they are constituted. Although the movement to a lower level of analysis usually increases predictive power, it also results in the proliferation of factors, which are more difficult to operationalize both statistically and clinically. Furthermore, there can be analytic complications, for example, if a compound factor from a higher level and one of its components from a lower level are placed in the same statistical model, the model will exhibit colinearity and the predictive power of the two factors will be diminished. In other words, a factor's predictive power is related to how it affects, and is affected by, other factors both at its level and at other levels. In addition, because the disease process changes over time, the relationship between a factor and the disease changes over time.[8]

Predictive Accuracy

Predictive accuracy measures the relationship between the prediction of the occurrence of the outcome and the actual occurrence of the outcome. The outcome can be a patient state or event. The closer a prediction is to the outcome, its target, the greater its accuracy. For a specific disease, for its relevant outcome, and for a specified time interval, the predictive accuracy of a biomarker depends on: (1) how intimately it is connected to the disease process (its power),

(2) its orthogonality with relation to other known biomarkers (degree of predictive overlap), and (3) how precisely it can be measured.[14] Clearly, connectedness is related to power. To the extent that two biomarkers are related to the disease process in the same way, when they are placed in the same predictive model, they will dilute their individual predictive power but not the power of the model. If two biomarkers are connected to the disease process in different ways (orthogonal), then when they are placed in the same predictive model, they will be additive in the model.

There are quite a few ways to measure accuracy. Percent correct is a commonly used measure of accuracy. For a binary outcome with an event rate close to 50%, for example, for survival where half the patients are dead and half are alive at the end of the specified time interval, percent correct is an acceptable method for assessing the predictive accuracy of survival predictions. But in the context of statistical models that learn from the data, percent correct can be influenced by the frequency with which the event appears in the population because the model can learn to predict the most frequent event. For example, if 99% of patients are alive in five years, a model can learn to predict the most frequent event, namely, predicting that all the patients will be alive and, in this situation, it will have an almost unbeatable 99% correct rate. It should be noted that in situations where there is a high event rate, it will always be difficult for any predictive model to do better than betting the frequency.

A well-developed method for assessing the accuracy of a predictive model in terms of its discrimination, and for comparing models, is the receiver operating characteristic (ROC).[28] It was discovered by Somer,[29] and it can be directly calculated by the Somer's D formula. In addition, it can be approximated by a trapezoidal area calculation.[30] The ROC is a nonparametric measure of discrimination. It is independent of both the prior probability of each outcome and the threshold cutoff for categorization. Its computation requires only that the prediction method produces an ordinal-scaled relative predictive score. In terms of mortality, the ROC estimates the probability that the prediction model will assign a higher mortality score to the patient who died than to the patient who lived. The ROC varies from zero to one. When the predictions are unrelated to survival, the ROC is 0.5, indicating no predictive accuracy (flipping a coin). The farther the ROC is from 0.5, the better the accuracy. The c index[31] is equivalent to the ROC, and it is useful in situations in which there is censoring, for example, in assessing the results of a proportional hazards model.

It should be noted that ROC values are nonlinear. This means that it is easier to achieve an ROC of 0.50 than 0.60, and it is easier to achieve an ROC of 0.60 than 0.70. In other words, it becomes progressively more difficult to achieve higher ROC values. The reason for this is that prediction is easiest at the extremes and harder as one moves toward the indeterminate middle. For example, it is relatively easy to predict who will die of breast cancer within one year and who will live at least 10 years, but it is much harder to predict when patients will die in the time interval between two and 10 years. Initially, the model predicts the extremes and it achieves a certain ROC score, but as it moves more toward the indeterminate outcomes it becomes harder and harder to predict correctly, resulting in more errors and less improvement in the ROC score.

There is an intimate relationship between the ROC, sensitivity, and specificity. The ROC can be thought of as all possible sensitivity/specificity pairs. The ROC can be used to assess individual variables, individual models, and to compare models. Significant differences in the ROCs between two models can be tested by following Hanley and McNeil,[32] or by calculating their asymptotic variances, or by calculating the empirical variance using the bootstrap method.[33]

Discrimination is how well the predictions are ordered in terms of the true outcomes and calibration is how close the predicted values are to the true outcome values. Although both discrimination and calibration should be used to assess model accuracy,[34,35] one should not initially assess calibration because a well-discriminating model can always be calibrated (postprocessor calibration), but the accuracy of a well-calibrated but poorly discriminating model can rarely be improved.[36]

Accuracy allows one to assess the predictive power of individual biomarkers in a multivariate model. Here, we are interested in a biomarker's predictive power and not its statistical significance. One way to assess the predictive power of individual biomarkers in a model is to perform the serial removal with replacement of the biomarkers. The method to accomplish this is as follows. All the biomarkers are placed in a statistical model and the model's ROC is determined. A biomarker is removed, the model is run, the model ROC is determined, the biomarker is replaced, and another biomarker is removed and the model ROC is determained. This process is repeated until all the biomarkers have been assessed in terms of their contribution to the overall accuracy of the model. If the removal of the biomarker has little or no effect on the model ROC, then it does not contribute to the accuracy of the predictive model. The reason it does not contribute may be because it has no predictive power or because it is correlated with other biomarkers. Furthermore, it can be the case that biomarkers that are not significant in the model contribute to predictive power, so a biomarker should not be removed from a model just because it lacks statistical significance. Finally, to determine whether the observed model accuracy is due to chance, the significance of a model can be assessed.[32,33]

Although I have been discussing the accuracy of individual biomarkers, it is important to realize that we are usually not interested in a single biomarker because cancer is a complex system. The biomarker components of a complex system do not act alone—they act in concert. Biomarkers are interacting, interdependent, multipurpose parts of a biological system and by modeling them we index the cancer. The model's accuracy

measures how well the model captures the clinical behavior of the cancer. Furthermore, there is not one unique model, there are many models of a complex system like cancer.

Significance and Accuracy

Some investigators believe that if a biomarker or a group of biomarkers can stratify patients into two statistically different outcome groups, then the biomarker or a group of biomarkers is an accurate outcome predictor. Furthermore, they believe that the greater the statistical significance of this stratification, the higher the predictive accuracy of the biomarker. This belief is incorrect. The ability of a biomarker to separate patients into two groups at a nonchance level tells us very little about how good the biomarker is at making individual patient predictions. In other words, significance tells us how probable the observed perimeter estimates are if the null hypthesis is true, whereas accuracy tells us how good we are at predicting individual patient outcomes. It is rarely the case that there is an invariant relationship between significance and accuracy, for example, that a greater significance always means a higher accuracy.[14,16] One reason for the discrepancy between significance and accuracy is that significance is usually enhanced by middle values because they affirm the parameter estimate and minimize its variance and it is usually degraded by extreme values. On the other hand, accuracy is enhanced by extremes, because extreme outcomes are the easiest to predict, and accuracy is degraded by middle values because they are usually the hardest to predict.

Another issue related to significance is the truth of the assessment. By increasing the sample size, one may improve significance, but it should not be the case that truth depends on the sample size. Generally, accuracy only depends on achieving a minimum event (outcome) rate, ie, a rate sufficient to define the relationship between the predictor variable and the true outcome. This rate is usually 10–20 events per independent variable.[37] The event rate for a binary variable is the outcome that has the lowest frequency. If the population has a sufficient event rate, then adding patients to the population does not affect variable or model accuracy.

Validation

A molecular biomarker must go through three stages of assessment. The first stage is its discovery and initial characterization of the biomarker. The second stage is to train and test a model. The third stage is its validation on an independent dataset. The assessment of molecular biomarkers has been a concern since the earliest days of molecular research. Over the last 20 years, significant problems have been noted, and recommendations regarding solving these problems have been made, but few of these proposals have been adopted. Pepe et al[38] proposed a model for clinical validation of biomarkers for the early detection for disease, yet subsequent publications on early detection suggest that the confusion did not recede after this publication.[39–42]

A word about the initial dataset and its use. One should always randomly split one's data into two subsets, namely, train (two-thirds to three-fourths of the data) and test (the balance of the data). An investigator can do anything with the train dataset, including looking at the data, optimizing thresholds, and developing models. But the results on the training dataset cannot be reported because the investigator looked at the data and optimized the analysis. The investigator applies the optimized model derived from the training dataset on the test dataset once. The test dataset results are reportable. Note that because the train and test datasets are essentially the same dataset, validity has not been established. For validity, one must take apply the optimized model to an independent dataset.

Clinical Acceptability

Whether a molecular biomarker is clinically acceptable depends on its accuracy, independence, and clinical utility.[8,9,16] Accurate means that the biomarker, at its lowest accuracy, is a powerful predictor for a specified group of patients. A helpful heuristic is that a validated ROC of 0.60 is usually required to surmount a predictive model's variance, a validated ROC of 0.70 is the lower bound of a clinically useful model, and a validated ROC of 0.75 and above indicates excellent clinical accuracy.[2] It is interesting to observe that most ROCs in medicine fall below 0.80. Achieving ROCs above 0.80 can occur due to low task difficulty, ie, the outcomes are easy to predict in a particular dataset, or to reporting on training data, or to special methods of analysis.

Independent means that the biomarker retains its predictive value when it is placed in a multivariate model that contains other relevant predictive biomarkers. Clinical utility means that it addresses a clinically important problem, that is, it improves patient care and outcomes.[14,19] Clinical utility does not require that there be an effective therapy. When there is no effective therapy, a biomarker's utility is its ability to inform patients of their outcome, ie, to predict the natural history of the disease, so that patients can properly prepare for their fate. The importance of a biomarker in providing information to patients regarding their outcomes, even when the outcome cannot be changed, should not be underestimated.

Biological function and clinical utility are distinguishable. Although it has been suggested that biological plausibility is a criterion in the evaluation of a predictive biomarker,[43] it is not necessary to understand the function of the biomarker in order to use it to make clinical predictions.[16] It is certainly the case that a biomarker's predictive value rests on its function in the disease process, but it is not necessary to know its function in order to use it predicatively.

Reporting Guidelines

A number of reporting guidelines have been proposed,[44] including REMARK[6] and TRIPOD.[45,46] We proposed reporting criteria,[8,16] several of which have been incorporated

into reporting guidelines. In summary, the report of the clinical use of a molecular biomarker should contain, at a minimum, the following information: (1) The disease, the method of patient identification and inclusion, the number of patients, a detailed description of all the relevant demographic, anatomic, and molecular variables, and the therapies received by patients in this population. (2) The biomarker, the method used to assess it and, if it is not treated as a continuous variable, its prespecified threshold. (3) The type, ie, risk, diagnostic, or prognostic, and the clinical use, ie, natural history, therapy-specific, post-therapy, of the biomarker. (4) The clinical outcome, the specific time interval, and the event rate. (5) The type of statistical method used to create the model that makes the predictions and the justification of any important assumptions. If all the relevant variables were not included in the multivariate model, then there must be a justification for why they were not included. (6) The method of predictive accuracy assessment, the accuracy value of each variable and of the statistical model, including confidence intervals, the method of assessing the significance of the accuracy values of each variable and of the model, and their significance values. Ideally, this information should be used to design, as well as report, prediction studies.

Limitations

Specific cancer sites and their molecular biomarkers are not discussed for two reasons. First, such a discussion would have entailed a substantial increase in the size and complexity of the manuscript. Second, there are many excellent reviews of molecular biomarkers by cancer site.[47–50] The goal of this exposition is to promote an understanding of how to clinically use molecular biomarkers.

Conclusion

Vast sums of money have been spent on the discovery of almost a million biomarkers. Unfortunately, very few molecular biomarkers are in clinical use. One reason for this profoundly embarrassing situation is that many people believe that the hard part of molecular biomarkers is their discovery and the easy part is their use. They believe that the translational process involves simply integrating a biomarker into a clinical study and publishing the study results. The reason for this mistake is that they do not understand the complexity inherent molecular biomarkers. This failure to properly understand, and clinically assess and utilize, molecular biomarkers has prevented their widespread adoption in treatment, in comparative benefit analyses, and their integration into individualized patient outcomes predictions for clinical decision-making.

A straightforward, general approach to understanding how to predict clinical outcomes using risk, diagnostic, and prognostic molecular biomarkers has been presented. It must be acknowledged that this exposition is only a small first step in the process of understanding how to discover and use powerful, accurate, and clinically acceptable molecular biomarkers into the practice of medicine to improve clinical care and patient outcomes.

In the future, molecular biomarkers will drive advances in risk, diagnosis, and prognosis, they will be the targets of powerful molecular therapies, and they will individualize and optimize therapy. Furthermore, clinical predictions based on molecular biomarkers will be displayed on the clinician's screen during the physician–patient interaction, they will be an integral part of physician–patient shared decision-making, and they will improve clinical care and patient outcomes.

Disclaimer

This study does not represent the views of the U.S. Federal Government, the Department of Defense, or the Uniformed Services University of the Health Sciences.

Author Contributions

Conceived the concepts: HB. Analyzed the data: HB. Wrote the first draft of the manuscript: HB. Developed the structure and arguments for the paper: HB. Made critical revisions: HB. Author reviewed and approved of the final manuscript.

REFERENCES

1. Burke HB. Outcome prediction and the future of the TNM staging system. *J Natl Cancer Inst*. 2004;96:1408–1409.
2. Burke HB. The power of prediction. *Cancer*. 2008;113:890–892.
3. Khleif SN, Doroshow JH, Hait WN. AACR-FDA-NCI Cancer Biomarkers Collaborative. AACR-FDA-NCI Cancer Biomarkers Collaborative consensus report: advancing the use of biomarkers in cancer drug development. *Clin Cancer Res*. 2010;16:3299–3318.
4. McShane LM, Altman DG, Sauerbrei W. Identification of clinically useful cancer prognostic factors: what are we missing? *J Natl Cancer Inst*. 2005;97:1023–1025.
5. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Statistics Subcommittee of the NCI-EORTC Working Group on cancer diagnostics. *Exp Oncol*. 2006;28:99–105.
6. Mallott S, Timmer A, Sauerbrei W, Altman DG. Reporting of prognostic studies of tumor markers: a review of published articles in relation to REMARK guidelines. *Br J Cancer*. 2010;102:173–180.
7. Twomby R. Identity crisis: finding, defining, and integrating biomarkers still a challenge. *J Natl Cancer Inst*. 2006;98:11–12.
8. Burke HB, Henson DE. Criteria for prognostic factors and for an enhanced prognostic system. *Cancer*. 1993;72:3131–3135.
9. Burke HB, Hutter RVP, Henson DE. Breast carcinoma. In: Hermanek P, Gospadoriwicz MK, Henson DE, Hutter RVP, Sobin LH, eds. *UICC Prognostic Factors in Cancer*. Berlin: Springer-Verlag; 1995:165–176.
10. Ransohoff DF. Cancer. Developing molecular biomarkers for cancer. *Science*. 2003;299:1679–1680.
11. Burke HB, Hoang A, Iglehart JD, Marks JR. Predicting response to adjuvant and radiation therapy in early stage breast cancer. *Cancer*. 1998;82:874–877.
12. Burke HB. Statistical analysis of complex systems in biomedicine. In: Fisher D, Lenz H, eds. *Learning from Data: Artificial Intelligence and Statistics V*. New York: Springer-Verlag; 1996:251–258.
13. Burke HB, Henson DE. Histologic grade as a prognostic factor in breast carcinoma. *Cancer*. 1997;80:1703–1705.
14. Burke HB. Chapter 1 Integrating multiple clinical tests to increase predictive accuracy. In: Hanausek M, Walaszek Z, eds. *Methods in Molecular Biology, Vol. XX: Tumor Marker Protocols*. Tonowa, NJ: Humana Press; 1998:3–10.
15. Burke HB. Increasing the power of surrogate endpoint biomarkers: the aggregation of predictive factors. *J Cell Biochem*. 1994;19S:278–282.
16. Burke HB, Henson DE. Evaluating prognostic factors. *CME J Gyn Onc*. 1999;4:244–252. (Ch. 13, Prognostic Factors in Epithelial Ovarian Carcinoma, Péter Bõsze, ed.).
17. Gasparini G, Pozza F, Harris AL. Evaluating the potential usefulness of new prognostic and predictive indicators in node-negative breast cancer patients. *J Natl Cancer Inst*. 1993;85:1206–1219.

18. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:1487–1490.
19. Burke HB, Henson DE. Specimen banks for prognostic factor research. *Arch Pathol Lab Med*. 1998;122:871–874.
20. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst*. 2009;101:1446–1452.
21. Bell WC, Sexton KC, Grizzle WE. Organizational issues in providing high-quality human tissues and clinical information for the support of biomedical research. *Methods Mol Biol*. 2010;576:1–30.
22. Prentice RL. Surrogate and mediating endpoints: current status and future directions. *J Natl Cancer Inst*. 2010;101:216–217.
23. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med*. 1989;8(4):431–440.
24. Prasad V, Kim C, Burotto M, Vandross A. The strength of association between surrogate end points and survival in oncology: a systematic review of trial-level meta-analyses. *JAMA Intern Med*. 2015;175(8):1389–1398.
25. Ciani O, Davis S, Tappenden P, et al. Validation of surrogate endpoints in advanced solid tumors: systematic review of statistical methods, results, and implications for policy makers. *Int J Technol Assess Health Care*. 2014;30(3):312–324.
26. Fleming TR, DeMets DL. Surrogate end points in clinical trials; are we being mislead? *Ann Intern Med*. 1996;125:605–613.
27. Baker SG. Surrogate endpoints: wishful thinking or reality? *J Natl Cancer Inst*. 2006;98:502–503.
28. Swets JA. *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers*. Mahwah, NJ: Lawrence Erlbaum Associates; 1996.
29. Somer RH. A new asymmetric measure of association for ordinal variables. *Am Sociol Rev*. 1962;27:799–811.
30. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating graph. *J Math Psychol*. 1975;12:387–415.
31. Harrell FE Jr, Lee KL, Califf RM, Pryor DB, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247:2543–2546.
32. Hanley JA, McNeil BJ. The meaning and use of the area under the receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29–36.
33. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall; 1993.
34. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:1432–1435.
35. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ*. 2009;338:1373–1377.
36. Rosen DB, Burke HB, Goodman PH. Improving prediction accuracy using a calibration postprocessor. In: Proceedings of the World Congress on Neural Networks. Hillsdale, NJ: Lawrence Erlbaum Assoc. Inc; 1996:1215–1220.
37. Harrell FE Jr, Lee KL, Matchar DB, Reichert TA. Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep*. 1985;69:1071–1077.
38. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst*. 2008;100:1432–1438.
39. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev*. 2004;4:309–314.
40. Andre F, McShane LM, Michiels S, et al. Biomarker studies: a call for a comprehensive biomarker study registry. Biomarker studies: a call for a comprehensive biomarker study registry. *Nat Rev Clin Oncol*. 2011;8:171–176.
41. Pepe MS, Feng Z. Improving biomarker identification with better designs and reporting. *Clin Chem*. 2011;57:1093–1095.
42. Simon R. Clinical trials for predictive medicine. *Stat Med*. 2012;31(25):3031–3040.
43. McGuire WL. Breast cancer prognostic factors: evaluation guidelines. *J Natl Cancer Inst*. 1991;83:154–155.
44. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer*. 1994;69:979–985.
45. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). *Ann Intern Med*. 2015;162(10):735–736.
46. Burke HB. Transparent reporting of a multivariate prediction model for individual prognosis or diagnosis. *Ann Intern Med*. 2015;162(10):735.
47. Yiu AJ, Yiu CY. Biomarkers in colorectal cancer. *Anticancer Res*. 2016;36(3):1093–1102.
48. Inoue K, Fry EA. Novel molecular markers for breast cancer. *Biomark Cancer*. 2016;8:25–42.
49. Attard G, Parker C, Eeles RA, et al. Prostate cancer. *Lancet*. 2016;387(10013):70–82.
50. Guo H, Zhou X, Lu Y, et al. Translational progress on tumor biomarkers. *Thorac Cancer*. 2015;6(6):665–671.