



10

Biomarkers and Use in Precision Medicine

PETER LIBBY AND PAUL M RIDKER

OVERVIEW OF BIOMARKERS, 102

Clinical Applications of Cardiovascular Biomarkers, 103
Novel Technologies in Biomarker Identification, 105

CLINICAL MEASURES OF BIOMARKER PERFORMANCE, 105

Sensitivity, Specificity, and Positive and Negative Predictive Value, 105
Discrimination, C-Statistics, and Receiver Operating Characteristic Curve, 106
Accuracy and Calibration, 106

Risk Reclassification, 106

External Validation and Impact Studies, 107
Practical Example: High-Sensitivity C-Reactive Protein, Lipids, and Reynolds Risk Score, 107

CONCLUSION, 107

REFERENCES, 107

Clinicians use biomarkers daily in the practice of cardiovascular medicine. Moreover, the use of biomarkers can continue to improve physicians' ability to provide clinically effective and cost-effective cardiovascular medicine in the years ahead. Appropriate risk stratification and targeting of therapies should not only help improve patient outcomes but also assist in responding to the urgent need to "bend the cost curve" of medical care. In particular, excessive use of imaging biomarkers increases the cost of medical care and can jeopardize patient outcomes (e.g., from radiation exposure or complications of administering contrast material or investigating incidental findings). Inappropriate use or interpretation of blood biomarkers (e.g., cardiac troponin levels) can lead to unnecessary hospitalization or procedures as well.

Despite the current usefulness of biomarkers, their future promise, and the critical need to use them appropriately, much misunderstanding still surrounds their current clinical application.¹ In addition, contemporary technologies can greatly expand the gamut of biomarkers relevant to cardiovascular practice. Emerging genetic, proteomic, metabolomic, and molecular imaging strategies will surely transform the landscape of cardiovascular biomarkers (see also Chapters 7, 8, and 25).

This chapter provides a primer on cardiovascular biomarkers by defining terms and discussing how the application of biomarkers can assist in clinical care. The literature abounds with descriptions of biomarkers offered to apply to various clinical situations. Advances in cardiovascular biology and the application of novel technologies have identified a plethora of novel cardiovascular biomarkers of potential clinical usefulness—begging the question of whether a novel biomarker adds value to existing and often better-validated biomarkers. Thus clinicians need tools to evaluate these emerging biomarkers, to discern which may actually elevate clinical practice and improve patient outcomes. To help the reader in this regard, we also provide a guide to the rigorous evaluation of the clinical utility of biomarkers. Chapter 8 explicates the application of proteomics and metabolomics to discover novel biomarkers.

OVERVIEW OF BIOMARKERS

For regulatory purposes, the U.S. Food and Drug Administration (FDA) first defined a *biomarker* in 1992 as "a laboratory measure or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful end point that is a direct measure of how a patient feels, functions, or survives and is expected to predict the effect of the therapy." At that time the FDA considered a *surrogate endpoint* as "reasonably likely, based on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit." The National Institutes of

Health (NIH) convened a working group in 1998 that offered some parallel operating definitions to guide the biomarker field (Table 10.1).² NIH defined a *biologic marker*, or biomarker, as "a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention." Thus the NIH definition includes not only soluble biomarkers in circulating blood but also "bedside biomarkers," such as anthropomorphic variables obtainable with a blood pressure cuff or a tape measure at the point of care.

This broad definition encompasses measurements of biomarkers in blood (Fig. 10.1A) as well as measurements from imaging studies (Fig. 10.1B). Imaging biomarkers can include those derived from classic anatomic approaches. Imaging modalities now offer functional information, such as estimates of ventricular function and myocardial perfusion. Molecular imaging has the potential to target specific molecular processes. A functional classification of biomarkers helps sort through the plethora encountered by the clinician, in that biomarkers can reflect a variety of biologic processes or organs of origin. For example, as shown in Figure 10.1B, to a first approximation, cardiac troponin reflects myocardial injury, brain natriuretic peptide reflects cardiac chamber stretch, C-reactive protein reflects inflammation, and cystatin C and the estimated glomerular filtration rate reflect kidney function.

The NIH working group defined a *surrogate endpoint* as "a biomarker intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm) or lack of benefit (or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence." (Note that the NIH definitions do not include the commonly used term *surrogate marker*.) Thus a surrogate endpoint is a biomarker that has been "elevated" to surrogate status. This distinction has particular importance in the regulatory aspects of cardiovascular medicine. For example, the FDA previously accepted a certain degree of reduction in hemoglobin A_{1c} (HbA_{1c}) as a criterion for registration of a novel oral hypoglycemic agent; thus HbA_{1c} was considered a biomarker accepted as a surrogate endpoint. Current FDA guidance now requires a cardiovascular safety study for the registration of new medications that target diabetes. This policy indicates regulatory doubts about the fidelity of a decrease in HbA_{1c} as a surrogate endpoint for reduced cardiovascular risk, despite its value as a biomarker of glycemia.

The NIH working group defined a *clinical endpoint* as "a characteristic or variable that reflects how a patient feels, functions, or survives." Pivotal or phase III cardiovascular trials aspire to use clinical endpoints so defined. The distinction among biomarkers, surrogate endpoints, and clinical endpoints has crucial implications as practitioners, regulators, and payers increasingly demand evidence of improvements in actual clinical outcomes rather than mere manipulation of biomarkers as a criterion for adoption of a treatment in clinical practice.



In an effort to dispel persistent confusion regarding definitions in the biomarker arena, in 2015 a joint effort of the FDA and NIH developed an online resource denoted BEST (Biomarkers, EndpointS, and other Tools).³ They formulated a living online resource that furnishes an extended glossary of terms to facilitate standardization (Table 10.2). The BEST definitions overlap with those of the NIH Workshop, but provide more detail of particular relevance to those interested in the regulatory aspects of biomarkers (Table 10.3).

Clinical Applications of Cardiovascular Biomarkers (see also Chapter 5)

Much of the prevailing confusion regarding biomarkers involves framing the question that the clinician wants to answer with the use of a biomarker (Fig. 10.1C). We can classify the goals of application of cardiovascular biomarkers into the following rubrics:

1. **Diagnosis.** Daily medical practice uses many biomarkers for cardiovascular diagnosis. The current universal definition of myocardial infarction, for example, requires elevation of a biomarker of myocyte injury, such as cardiac-specific isoforms of troponin.
2. **Risk stratification.** Familiar examples of biomarkers used in risk stratification in cardiovascular medicine include systolic blood pressure (SBP) and low-density lipoprotein cholesterol (LDL-C). These biomarkers reliably predict future risk for cardiovascular events on a population basis.
3. **Goals for therapy.** Contemporary guidelines often specify cutoff points for targets of treatment, for example, a specific level of a biomarker (e.g., SBP, LDL-C) in a particular group of individuals. Practitioners of cardiovascular medicine typically use the biomarker

international normalized ratio (INR) to titrate the dosage of warfarin administered to an individual patient. Abundant data support the clinical benefit of maintaining the INR within a certain range in various patient groups, an example of a widely used biomarker that has proven clinical usefulness as a goal for therapy.

4. **Targeting of therapy.** In clinical practice, using biomarkers to target therapy has great usefulness and promise as we move toward a more comprehensive “personalized medicine” approach to practice. Examples of biomarkers used to target therapy include troponin measurements to triage patients with acute coronary syndromes for early invasive management and measurement of high-sensitivity C-reactive protein (hsCRP) to allocate statin treatment to individuals without elevated LDL-C.
5. **Drug development, evaluation, and registration.** Biomarkers have critical importance in the development of new pharmacologic agents. Biomarkers can provide early signals of efficacy that will help prioritize agents more likely to provide benefit on clinical endpoints in large-scale trials. Clinical trials not infrequently fail because of inappropriate dose selection. Judicious use of biomarkers can help in selecting an appropriate dose of an agent to study in a large endpoint trial. Biomarkers accepted as surrogate endpoints also prove useful to regulatory agencies in granting approval for novel therapies.

Clinical use of cardiovascular biomarkers requires a clear understanding of *how* they should be used. Many biomarkers provide

TABLE 10.1 Biomarker Definitions (NIH Working Group)

Biologic marker (biomarker) A characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

Surrogate endpoint A biomarker intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm) or lack of benefit (or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence.

Clinical endpoint A characteristic or variable that reflects how a patient feels, functions, or survives.

From National Institutes of Health (NIH) Biomarkers Definition Working Group, 1998.

TABLE 10.2 Biomarker Categories According to the FDA-NIH BEST Resource

Diagnostic biomarker
Monitoring biomarker
Pharmacodynamic/response biomarker
Predictive biomarker
Prognostic biomarker
Reasonably likely surrogate endpoint
Safety biomarker
Susceptibility/risk biomarker
Understanding prognostic versus predictive biomarkers
Validated surrogate endpoint

From the BEST (Biomarkers, EndpointS, and other Tools) Online Resource <https://pubmed.ncbi.nlm.nih.gov/27010052/>

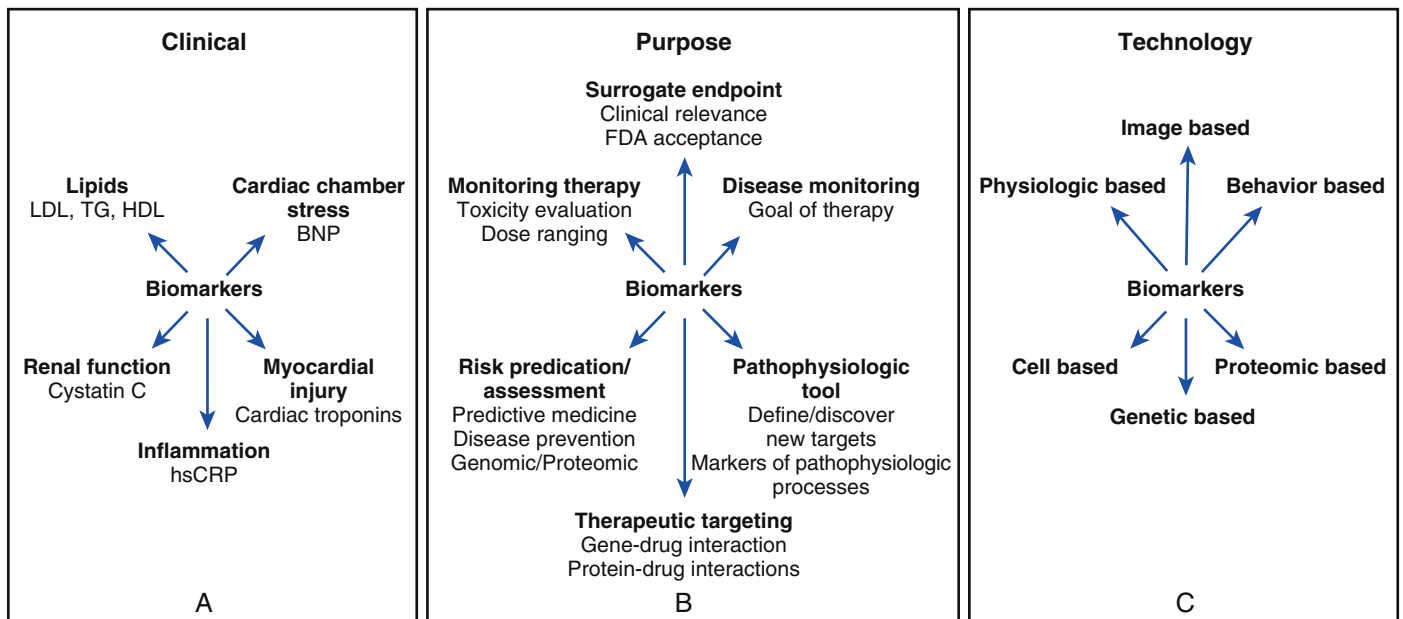


FIGURE 10.1 Examples of commonly used clinical biomarkers for cardiovascular disease (A), as well as research-oriented biomarkers categorized according to purpose (B) and technology (C). *BNP*, Brain natriuretic peptide; *hsCRP*, high-sensitivity C-reactive protein; *TG*, triglyceride.

TABLE 10.3 Selected Biomarker Definitions from the BEST Glossary

Biologic marker (biomarker): A defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions. Molecular, histologic, radiographic, or physiologic characteristics are types of biomarkers. A biomarker is not an assessment of how an individual feels, functions, or survives.

Surrogate endpoint: An endpoint that is used in clinical trials as a substitute for a direct measure of how a patient feels, functions, or survives. A surrogate endpoint does not measure the clinical benefit of primary interest in and of itself, but rather is expected to predict that clinical benefit or harm based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence. From a U.S. regulatory standpoint, surrogate endpoints and potential surrogate endpoints can be characterized by the level of clinical validation: validated surrogate endpoint, reasonably likely surrogate endpoint, candidate surrogate endpoint.

From the BEST (Biomarkers, Endpoints, and other Tools) Online Resource <https://pubmed.ncbi.nlm.nih.gov/27010052/>

clinically useful information when measured once at “baseline.” A baseline measurement of high-density lipoprotein cholesterol (HDL-C), for example, correlates inversely with future risk for cardiovascular events. However, serial measurement of biomarkers to document a change does not always guarantee a clinical benefit. In the case of HDL-C, recent large-scale trials that have measured clinical endpoints have cast doubt on the fidelity of a rise in HDL-C as a predictor of clinical benefit (see Chapter 27). A single measurement of coronary artery calcium score (CAC) ably predicts future events in statin naïve individuals. Yet, serial measurements of CAC may prove misleading because statin therapy increases calcification, but decreases coronary events.

Biomarkers require rigorous validation before adoption into clinical practice. In cardiovascular medicine, LDL-C has high reliability as a biomarker; it satisfies the modified Koch postulates. LDL levels prospectively predict cardiovascular risk, and decreases in LDL generally correlate with improved outcomes. Not all biomarkers, however, have proved as faithful in predicting clinical events. In the 1960s and 1970s, for example, most of the cardiovascular community considered ventricular premature depolarizations on the electrocardiogram (ECG) as important biomarkers for lethal arrhythmias. Numerous strategies have been aimed at suppressing ventricular ectopy. The Cardiac Arrhythmia Suppression Trial (CAST), however, showed that drugs capable of suppressing ventricular premature depolarizations actually worsened survival. The short-term improvements in indices of cardiac contractility produced by inotropic agents similarly led to worsened clinical outcomes, including increased mortality. These examples illustrate the necessity of rigorous validation of biomarkers before adoption into clinical practice.

Another important consideration in the use of cardiovascular biomarkers involves the question of *causality*. LDL-C exemplifies a *causal biomarker*, one that clearly participates in the pathogenesis of atherosclerosis.⁴ Its levels prospectively correlate with risk for cardiovascular events and the development of atherosclerotic lesions identified by a variety of imaging modalities. A variety of independent manipulations of LDL-C levels correlate with clinical outcomes. In addition, very strong genetic evidence based on mendelian disorders (e.g., familial hypercholesterolemia) and unbiased genome-wide association scans, as well as mendelian randomization analyses, have established LDL-C as a causal risk factor in atherosclerotic cardiovascular disease and as a generally valid surrogate endpoint offering great value in clinical practice (see Chapter 27). Even a well-validated causal biomarker such as LDL-C, however, may mislead under some circumstances. For example, lowering of LDL-C with certain cholesteryl ester transfer protein inhibitors does not appear to lead to clinical benefit.^{5,6} Other lipid measures such as plasma triglycerides and lipoprotein(a) predict risk, but currently lack definitive evidence that intervention reduces that risk. Ongoing trials of triglyceride reduction⁷ and of lipoprotein(a) reduction⁸ may move these relationships from casual to causal.

Other biomarkers, although clearly clinically useful, do not participate in the causal pathway for disease. For example, fever has served

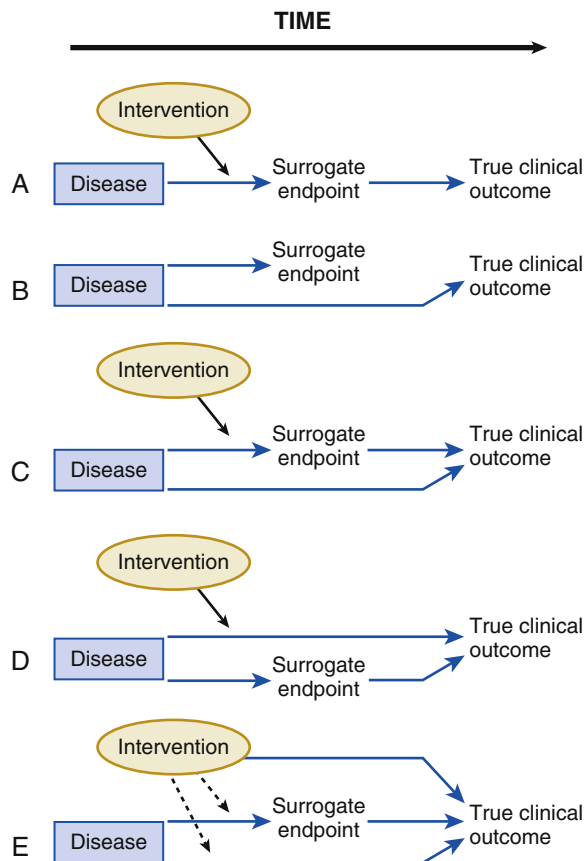


FIGURE 10.2 Biomarkers as surrogate endpoints in clinical research. **A**, The setting that provides the greatest potential for the surrogate endpoint to be valid. **B**, The surrogate is not in the causal pathway of the disease process. **C**, Of several causal pathways of disease, the intervention affects only the pathway mediated through the surrogate. **D**, The surrogate is not in the pathway of the intervention's effect or is insensitive to its effect. **E**, The intervention has mechanisms of action independent of the disease process. Dotted lines represent possible mechanisms of action. (Modified from Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med.* 1996;125:605.)

since antiquity as an important biomarker of infection. Resolution of fever correlates with successful resolution of infectious processes. However, fever does not participate causally in the pathogenesis of infection but merely serves as a biomarker of the host defenses against the infectious process. Sometimes a non-causal downstream biomarker can serve as an effective clinical surrogate for an upstream causal biomarker. For example, the use of hsCRP measurements improves the prediction of cardiovascular risk, and reductions in CRP correlate with clinical benefit in many cases. However, mendelian randomization studies do not support a causal role for CRP itself in the pathogenesis of cardiovascular disease. By contrast, intervention trials demonstrate that upstream drivers of CRP, in particular IL-1b, are indeed in the causal pathway leading to myocardial infarction, stroke, and cardiovascular events.⁹

These examples illustrate how a biomarker does not need to reside in the causal pathway of a disease to have clinical usefulness. A clear and early exposition of the uses and pitfalls in the application of biomarkers emerged from the landmark schema of Fleming and DeMets (Fig. 10.2). Biomarkers have the greatest potential for validity when there is one causal pathway and when the effect of intervention on true clinical outcomes is mediated directly through the biomarker surrogate (Fig. 10.2A). However, biomarker development can fail when the biomarker is found not to be in the causal pathway, when the biomarker is insensitive to the specific intervention's effect, or when the intervention of interest has a mechanism of action (or a toxicity) that does not involve the pathway described by the biomarker (Fig. 10.2B-E). These examples do not mean that biomarkers lack value; few if any novel biologic fields could develop without biomarker discovery and validation. Still, surrogate endpoints probably will not replace large-scale

randomized trials that address whether interventions reduce actual event rates.

Novel Technologies in Biomarker Identification

The limitations of currently available biomarkers for screening or prognostic use underscore the importance of identifying “uncorrelated” or “orthogonal” biomarkers associated with novel disease pathways (see Fig. 10.1A). Most current biomarkers have been developed as an extension of targeted physiologic studies investigating known pathways such as tissue injury, inflammation, or hemostasis. By contrast, emerging technologies now enable the systematic, unbiased characterization of variation in proteins and metabolites associated with disease conditions (see Chapter 8). The rapid development of polygenic risk scores for cardiovascular disease promises to provide biomarkers that may permit the targeting of therapies particularly in primordial and primary prevention before pathological processes have progressed to the point of altering disease biomarkers (see Chapter 7.)^{10–13} The applications of machine learning and artificial intelligence (see Chapter 11) will doubtless add to the development of novel candidate biomarkers that will require rigorous evaluation for clinical utility as outlined below. The burgeoning field of wearables will provide new inputs into biomarker science as well (see Chapter 12). The development of point-of-care technologies will likewise facilitate the clinical application of biomarkers by rendering their use more practical in the field and in urgent situations.¹⁴ Digital technologies, for example, smart phone apps and implanted sensors, will also provide “real world” biomarker input outside of the confines of the traditional medical enterprise.

CLINICAL MEASURES OF BIOMARKER PERFORMANCE

When considering any biomarker in a clinical setting for risk prediction, physicians should ask two interrelated questions:

- Is there clear evidence that the biomarker of interest predicts future cardiovascular events independent of other already measured biomarkers?
- Is there clear evidence that patients identified by the biomarker of interest will benefit from a therapy that they otherwise would not have received?

Unless the answer to both these questions is a clear “yes,” measurement of the biomarker will not likely have sufficient usefulness to justify its cost or unintended consequences. Such judgments require clinical expertise and will vary on a case-by-case basis.

Biomarker evaluation also typically involves repeated testing in multiple settings that include varied patient populations and that use different epidemiologic designs. *Prospective* cohort studies (in which the biomarker or exposure of interest is measured at baseline, when individuals are healthy, and then related to the future development of disease) provide a much stronger form of epidemiologic evidence

than do data from *retrospective* case-control studies (in which the biomarker of interest is measured after the disease is present in the case participants).

After discovery by the technologies described earlier or identification by a candidate approach, a novel biomarker typically requires development in a translational laboratory for refinement of its assay to address issues of interassay and intra-assay variation before any clinical testing begins. Focused studies in specific patient populations typically follow and eventually broaden to encompass the population of greatest clinical interest. Beyond simple reproducibility, biomarkers under development for diagnostic, screening, or predictive purposes require further evaluation with a standard set of performance measures that include sensitivity, specificity, positive and negative predictive value (NPV), discrimination, calibration, reclassification, and tests for external validity.

Sensitivity, Specificity, and Positive and Negative Predictive Value

The validity of a screening or diagnostic test (or one used for prediction) is initially measured by its ability to categorize individuals who have preclinical disease correctly as “test positive” and those without preclinical disease as “test negative.” A simple two-by-two table is typically used to summarize the results of a screening test by dividing those screened into four distinct groups (Table 10.4). In this context, sensitivity and specificity provide fundamental measures of the test’s clinical validity. *Sensitivity* is the probability of testing positive when the disease is truly present and is defined mathematically as $a/(a + c)$. As sensitivity increases, the number of individuals with disease who are missed by the test decreases, so a test with perfect sensitivity will detect all individuals with disease correctly. In practice, tests with ever-higher sensitivity tend to also classify as “diseased” many individuals who are not actually affected (false positives). Thus the *specificity* of a test is the probability of screening negative if the disease is truly absent and is defined mathematically as $d/(b + d)$. A test with high specificity will rarely be positive when disease is absent and will therefore lead to a lower proportion of individuals without disease being incorrectly classified as test positive (false positives). A simple way to remember these differences is that sensitivity is “positive in disease,” whereas specificity is “negative in health.”

A perfect test has both very high sensitivity and specificity and thus low false-positive and false-negative classifications. Such test characteristics are rare, however, because there is a trade-off between sensitivity and specificity for almost every screening biomarker, diagnostic, or predictive test in common clinical use. For example, although high LDL-C levels usually serve as a biomarker for atherosclerotic risk, up to half of all incident cardiovascular events occur in those with LDL-C levels well within the normal range, and many events occur even when levels are low. If the diagnostic cutoff criterion for LDL-C is reduced so that more people who actually have high risk for disease will test positive (i.e., increased sensitivity), an immediate consequence of this

TABLE 10.4 Summarizing the Results of Screening, Diagnostic, or Predictive Tests

	DISEASE PRESENT	DISEASE ABSENT	
Test positive	a	b	a + b
Test negative	c	d	c + d
Total	a + c	b + d	
Sensitivity = $a/(a + c)$			
Specificity = $d/(b + d)$			
Positive predictive value = $a/(a + b)$			
Negative predictive value = $d/(c + d)$			

a = Number of individuals for whom the screening test is positive and the individual actually has the disease (true positives).

b = Number of individuals for whom the test is positive but the individual does not have the disease (false positives).

c = Number of individuals for whom the test is negative but the individual actually has the disease (false negatives).

d = Number of individuals for whom the test is negative and the individual does not have the disease (true negatives).

Modified from Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther.* 2001;69:89–95.

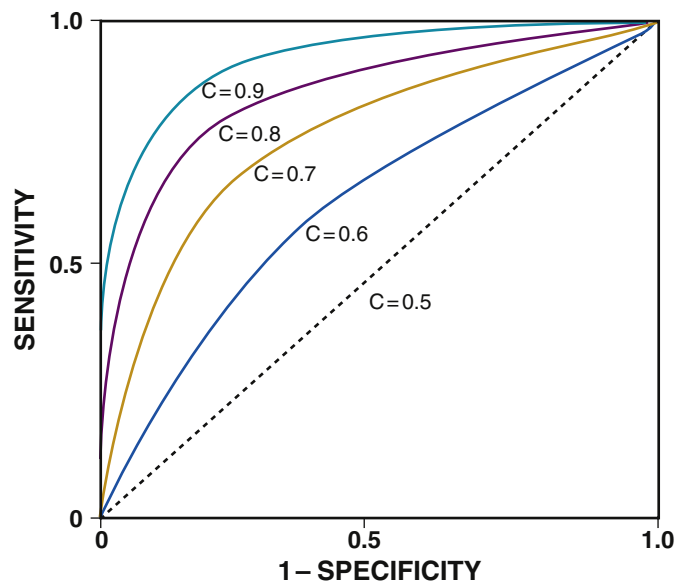


FIGURE 10.3 Receiver operating characteristic (ROC) curves for a series of biomarkers or risk prediction models with incremental improvement. The diagonal line corresponds to a random effect (C-statistic = 0.5), whereas the increasing C-statistic corresponds to improving model discrimination.

change will be an increase in the number of people without disease in whom the diagnosis is made incorrectly (i.e., reduced specificity). Conversely, if the criterion for diagnosis or prediction is made more stringent, a greater proportion of those who test negative will actually not have the disease (i.e., improved specificity), but a larger proportion of true cases will be missed (i.e., reduced sensitivity).

In addition to sensitivity and specificity, the performance or yield of a screening, diagnostic, or predictive test also varies depending on the characteristics of the population being evaluated. Positive and NPVs are terms used in epidemiology that refer to measurement of whether an individual actually has (or does not have) a disease, contingent on the result of the screening test itself.

The positive predictive value (PPV) is the probability that a person has the disease of interest, given that the individual tests positive, and is mathematically calculated as $PPV = a / (a + b)$. High PPV can be anticipated when the disease is common in the population being tested. Conversely, the NPV is the probability that an individual is truly disease free, provided that the test has a negative result, and is mathematically calculated as $NPV = d / (c + d)$. High NPV can be anticipated when the disease is rare in the population being tested. Although sensitivity and specificity are largely performance characteristics of the test itself (and thus tend to be fixed values), PPV and NPV depend in part on the population being tested (and thus tend to vary).

Discrimination, C-Statistics, and Receiver Operating Characteristic Curve

Discrimination is the ability of a test (or prognostic model) to separate those with disease or at high risk for disease (cases) from those without disease or at low risk for disease (controls). The most common method used to measure discrimination has been the *area under the receiver operating characteristic (ROC) curve* (AUC), which relates sensitivity (on the y axis) to (1 – specificity) (on the x axis) across a full range of cutoff values for the test or screening algorithm of interest (Fig. 10.3).

Given a population of individuals being evaluated, the area under the ROC curve, also called the *C-statistic*, equals the probability of correctly ranking risk for individuals by using the test or model under evaluation. A random test with no clinical usefulness would have a C-statistic (ROC AUC) of 0.5, which corresponds to the diagonal line in Figure 10.3. A perfect test that completely discriminates individuals with disease from those without disease would have a C-statistic that approaches 1.0. As the C-statistic increases from 0.5 to 1.0, model fit (or test accuracy) improves; thus the change in the C-statistic has served

historically to judge whether a new biomarker can “add” significantly to those already in use. This approach permits direct comparison of the relative efficiency of multimarker panels. For example, using comparative C-statistic analyses, investigators in the Emerging Risk Factors Collaboration recently found that the incremental clinical usefulness of CRP has similar magnitude as that of total and HDL-C. Thus when change in the C-statistic can be demonstrated and the overall power to do so is adequate, this test can assist in understanding the impact of novel pathways and novel risk biomarkers on prediction and prevention.

Unfortunately, the traditional C-statistic approach is limited in that biomarkers with large associations may have minimal effect on ROC AUC. For example, a predictor (or set of predictors) would need an odds ratio (OR) as high as 16 (>2 standard deviations [SD]) to lead to a substantial increase in the C-statistic. Almost no test in common use for risk prediction or prognostication in cardiovascular medicine has an OR in this range; high cholesterol, smoking, high blood pressure, and diabetes yield an OR of less than 2 and thus have negligible individual impact on ROC AUC. Consequently, sole reliance on the C-statistic as a method for developing and evaluating new biomarkers, at least in the setting of risk prediction, is insufficient.

Accuracy and Calibration

Discrimination is only one measure of model accuracy. The other important measure is *calibration*, or the ability of a predictive model to assign risk estimates accurately compared with the actual observed risk in the population being tested. Unlike discrimination, which is based solely on relative rankings of risk, calibration compares the risk predicted from a model or test with that actually observed.

For *binary* outcomes (e.g., disease or no disease), calibration is often evaluated with the Hosmer-Lemeshow test, which places individuals within categories of estimated risk by using the test biomarker or multivariable model and compares these estimates with the proportions actually observed. These “predicted” and “observed” probabilities can be compared with standard goodness-of-fit tests across categories of risk (e.g., across estimated quintiles or estimated deciles of risk). Calibration becomes particularly important when addressing a biomarker in different populations from the population in whom it was originally developed. A biomarker may calibrate well in men but not in women, or among whites but not among blacks. This consideration also applies to multimarker panels, such as the Framingham Risk Score, which calibrates well in whites but less well in other population groups. Newer risk models such as the Reynolds Risk Score show improved calibration, as well as discrimination, compared with the traditional Framingham model.

Risk Reclassification

To address the shortcoming of biomarker validation via the C-statistic alone, contemporary biomarker development programs for risk prediction now use a series of “reclassification statistics,” as initially developed by Cook and Ridker and refined by Pencina and Xanthakis and associates. Rather than addressing whether a new biomarker of interest adds to ROC AUC, reclassification addresses whether the biomarker can shift overall risk estimates upward or downward in a clinically meaningful way. Specifically, reclassification methods compare risk strata formed from prediction models with and without the new biomarker of interest and then determine which model leads to the most accurate classification of risk. Risk reclassification is particularly useful when actionable and clinically relevant risk categories already exist. For example, in primary cardiovascular prevention, 10-year estimated risk is often categorized as being less than 5%, 5% to 10%, 10% to 20%, or greater than 20%, and those above or below these cut points are frequently targeted for interventions such as aspirin and statin therapy. Thus a biomarker that reclassifies a proportion of individuals upward (or downward) might well be highly effective for targeting (or avoiding) drug therapy, even if the overall effect on discrimination is modest.

Mere reclassification of an individual by a given biomarker does not provide sufficient evidence to support clinical use. Rather, an effective

biomarker should correctly reclassify risk higher or lower and thus lead to more accurate overall risk assessment. The *reclassification calibration* (RC) statistic is a tool that tests how well the average predicted risk within a given cell agrees with the observed risk of individuals who actually experience the event. Accordingly, the RC statistic addresses whether the predicted risk estimates after reclassification (using the new biomarker) are more accurate than before reclassification (without the new biomarker). Superior reclassification occurs when the new prediction model places case individuals into higher-risk categories and places control individuals into lower-risk categories, and when the net shift in these two effects is in the overall correct direction. This characteristic can be addressed by using the *net reclassification index* (NRI), analogous to a test of discrimination (the ability to separate cases from controls) in the context of a reclassification table. Broadly, the NRI does not depend as much on the actual predicted probabilities as on movement across a categorical risk border that is the result of the new probabilities predicted. When reclassification is not addressed across categories, an alternative measure is used, called the *integrated discrimination improvement* (IDI), based on the Yates slope, or the difference in predicted probabilities among case and control individuals. Despite their relatively recent introduction, reclassification statistics have rapidly become the standard for clinical evaluation of emerging biomarkers and alternative multi-biomarker prediction panels.

External Validation and Impact Studies

A final but important test for any biomarker or biomarker panel when used for prognostication, *external validation* refers to the ability of the panel to function with clinically acceptable levels of sensitivity, specificity, discrimination, and calibration in external populations, distinct from the population used for generation of the panel. As Moons and coworkers note, prognosis research and prognostic biomarkers differ from those used in diagnosis and screening.

Prognostic research involves three distinct phases in the development of multivariable prediction models. The first phase is identification of relevant predictors, assignment of weights to the model, estimation of predictive performance, and optimization of fit. The second phase involves validation or formal testing of calibration and discrimination in new patient groups, which can be similar to those used in the development stage or purposely different. The third phase involves impact studies to quantify directly whether use of a prognostic model in daily practice actually changes physician behavior and decision making, and whether this occurs in a net positive manner and is cost-effective. *Prognostic impact studies* also focus on the incremental usefulness of a given biomarker beyond simple clinical and non-clinical characteristics. Such studies tend to be less biologically driven than biomarker discovery work and recognize that prediction does not necessarily involve a causal pathway.

Practical Example: High-Sensitivity C-Reactive Protein, Lipids, and Reynolds Risk Score

The use of hsCRP in clinical practice is an example of how biomarker development programs can move from pathophysiologic principles to clinical use and onward to multinational trials evaluating novel targets for vascular risk reduction.¹⁵ A prospective cohort of initially healthy individuals showed that hsCRP predicted future risk for a heart attack and stroke in men, an observation externally validated and quickly extended to women. Multiple commercial hsCRP assays—reproducible, internally calibrated, and externally validated to improve assay precision—then became clinically available. Multiple studies have shown that statins reduce hsCRP in a manner largely independent of reduction of LDL-C, thus suggesting that statins have both lipid-lowering and anti-inflammatory effects. The addition of hsCRP to the family history and HbA_{1c} was formally incorporated into the Reynolds Risk Score in 2008.¹⁶ This score was subsequently externally validated and shown to have superior calibration, discrimination, and reclassification over the more traditional Framingham Risk Score. Using hsCRP to define a high-risk population in need of treatment, JUPITER (Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin)

reported in 2008 that statin therapy (vs. placebo) in those with elevated hsCRP but low levels of LDL-C resulted in a 50% reduction in myocardial infarction and stroke and a 20% reduction in all-cause mortality. By 2010, more than 50 prospective cohort studies evaluating hsCRP were subjected to meta-analysis, which affirmed that the magnitude of vascular risk associated with a change of 1 SD in hsCRP was at least as large as that of a comparable change in cholesterol or blood pressure. An updated 2012 meta-analysis of clinical usefulness and risk prediction found that the change in C-statistic associated with hsCRP was similar to that associated with the use of total and HDL cholesterol. On this basis, several national guidelines incorporated hsCRP screening in primary and secondary prevention, and the FDA approved a labeling claim for the use of statin therapy in those with elevated hsCRP levels.

CRP itself, however, probably does not cause atherothrombosis, but rather serves as a biomarker for the underlying inflammatory process. Thus validation of the inflammatory hypothesis of atherosclerosis required measurement of clinical outcomes, not solely biomarkers. CANTOS (the Canakinumab Anti-inflammatory Thrombosis Outcomes Study), evaluated interleukin-1beta inhibition in individuals who were in a stable phase postmyocardial infarction, and were well treated with standard secondary prevention measures, but had an hsCRP greater than 2 mg/L. CANTOS demonstrated that an anti-inflammatory intervention (as gauged by lowering of hsCRP and other inflammatory biomarkers such as IL-6) that did not affect atherogenic lipoprotein levels could reduce recurrent major adverse cardiovascular events.^{9,17} Trials with colchicine have validated this concept as discussed in [Chapter 25](#),^{18,19} and new trials addressing the potential role of IL-6 inhibition for atheroprotection are underway.²⁰ The use of hsCRP as a biomarker laid the groundwork for CANTOS, illustrating the value of biomarkers in advancing both pathophysiologic understanding and therapeutics.

Congestive heart failure is another arena where multiple biomarkers ranging for troponin to ST2 to IL-6 to GDF-15 have been proposed to have unique prognostic utility. However, other than data suggesting that interventions targeting IL-1b such as canakinumab²¹ and anakinra^{22,23} may reduce hospitalization for heart failure, little confirmatory evidence is available.

CONCLUSION

We use biomarkers in our daily clinical practice, and cardiovascular journals contain numerous reports regarding biomarkers, new and old, that purport to show how they may aid clinical practice. Moreover, many cardiovascular trials use biomarkers—thus the current practice of cardiovascular medicine requires a firm foundation in understanding and evaluating biomarkers. The road map to the field of biomarkers provided in this chapter, including their use, development, and methods for evaluating their usefulness for various specific applications, should give practitioners tools to sort out the various uses of biomarkers encountered in practice and in the cardiovascular literature. Informed use of biomarkers can aid in decision making in daily patient care. Biomarkers should provide a key for personalized management by directing the right therapy to the right patient at the right time. They can also shed mechanistic insight on human pathophysiology that is difficult to obtain in other ways. Rigorous and careful use of biomarkers can aid in the development of novel therapies to address the residual burden of cardiovascular risk.

REFERENCES

For citations to the older literature, see previous editions of this textbook.

Biomarkers: General Considerations

- Libby P, King K. Biomarkers: a challenging conundrum in cardiovascular disease. *Arterioscler Thromb Vasc Biol.* 2015;35:2491–2495.
- Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther.* 2001;69:89–95.
- Robb MA, McInnes PM, Califf RM. Biomarkers and surrogate endpoints: developing common terminology and definitions. *J Am Med Assoc.* 2016;315:1107–1108.
- Borén J, Chapman MJ, Krauss RM, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease: pathophysiological, genetic, and therapeutic insights: a consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur Heart J.* 2020;41:2313–2330.

Biomarkers Versus Clinical Endpoints

- Group HTRC. Effects of anacetrapib in patients with atherosclerotic vascular disease. *New Engl J Med.* 2017.



6. Lincoff AM, Nicholls SJ, Riesmeyer JS, et al. Evacetrapib and cardiovascular outcomes in high-risk vascular disease. *New Engl J Med.* 2017;376:1933–1942.
7. Pradhan AD, Paynter NP, Everett BM, et al. Rationale and design of the Pemafibrate to Reduce Cardiovascular Outcomes by Reducing Triglycerides in Patients with Diabetes (PROMINENT) study. *Am Heart J.* 2018;206:80–93.
8. Tsimikas S, Karwatowska-Prokopczuk E, Gouni-Berthold I, et al. Lipoprotein(a) reduction in persons with cardiovascular disease. *New Engl J Med.* 2020;382:244–255.
9. Ridker PM, Everett BM, Thuren T, et al. Antiinflammatory therapy with canakinumab for atherosclerotic disease. *New Engl J Med.* 2017;377:1119–1131.

Evolving Biomarkers

10. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet.* 2018;50:1219–1224.
11. Inouye M, Abraham G, Nelson CP, et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. *J Am Coll Cardiol.* 2018;72:1883–1893.
12. Mosley JD, Gupta DK, Tan J, et al. Predictive accuracy of a polygenic risk score compared with a clinical risk score for incident coronary heart disease. *J Am Med Assoc.* 2020:323.
13. Aragam Krishna G, Natarajan P. Polygenic scores to assess atherosclerotic cardiovascular disease risk. *Circ Res.* 2020;126:1159–1177.
14. King KR, Grazette LP, Paltoo DN, et al. Point-of-Care technologies for precision cardiovascular care and clinical research. *JACC (J Am Coll Cardiol): Basic to Transl Sc.* 2016;1:73–86.

A Practical Application of a Biomarker

15. Ridker PM. A test in context: high-sensitivity C-reactive protein. *J Am Coll Cardiol.* 2016;67:712–723.
16. Cook NR, Paynter NP, Eaton CB, et al. Comparison of the Framingham and Reynolds Risk scores for global cardiovascular risk prediction in the multiethnic Women's Health Initiative. *Circulation.* 2012;125:1748–1756. S1–11.
17. Ridker PM, Libby P, MacFadyen JG, et al. Modulation of the interleukin-6 signalling pathway and incidence rates of atherosclerotic events and all-cause mortality: analyses from the Canakinumab Anti-Inflammatory Thrombosis Outcomes Study (CANTOS). *Eur Heart J.* 2018;39:3499–3507.
18. Tardif JC, Kouz S, Waters DD, et al. Efficacy and safety of low-dose colchicine after myocardial infarction. *New Engl J Med.* 2019;381:2497–2505.
19. Nidorf SM, Fiolet ATL, Mosterd A, et al. Colchicine in patients with chronic coronary disease. *New Engl J Med.* 2020;383:1838–1847.
20. Ridker PM, Devalaraja M, Baeres FMM et al. Effects of interleukin-6 inhibition with Ziltivekimab on biomarkers of inflammation among patients at high risk for atherosclerotic events. In press.
21. Everett BM, Cornel JH, Lainscak M, et al. Anti-inflammatory therapy with canakinumab for the prevention of hospitalization for heart failure. *Circulation.* 2019;139:1289–1299.
22. Buckley LF, Abbate A. Interleukin-1 blockade in cardiovascular diseases: from bench to bedside. *BioDrugs.* 2018;32(2):111–118.
23. Abbate A, Toldo S, Marchetti C, et al. Interleukin-1 and the inflammasome as therapeutic targets in cardiovascular disease. *Circ Res.* 2020;126:1260–1280.